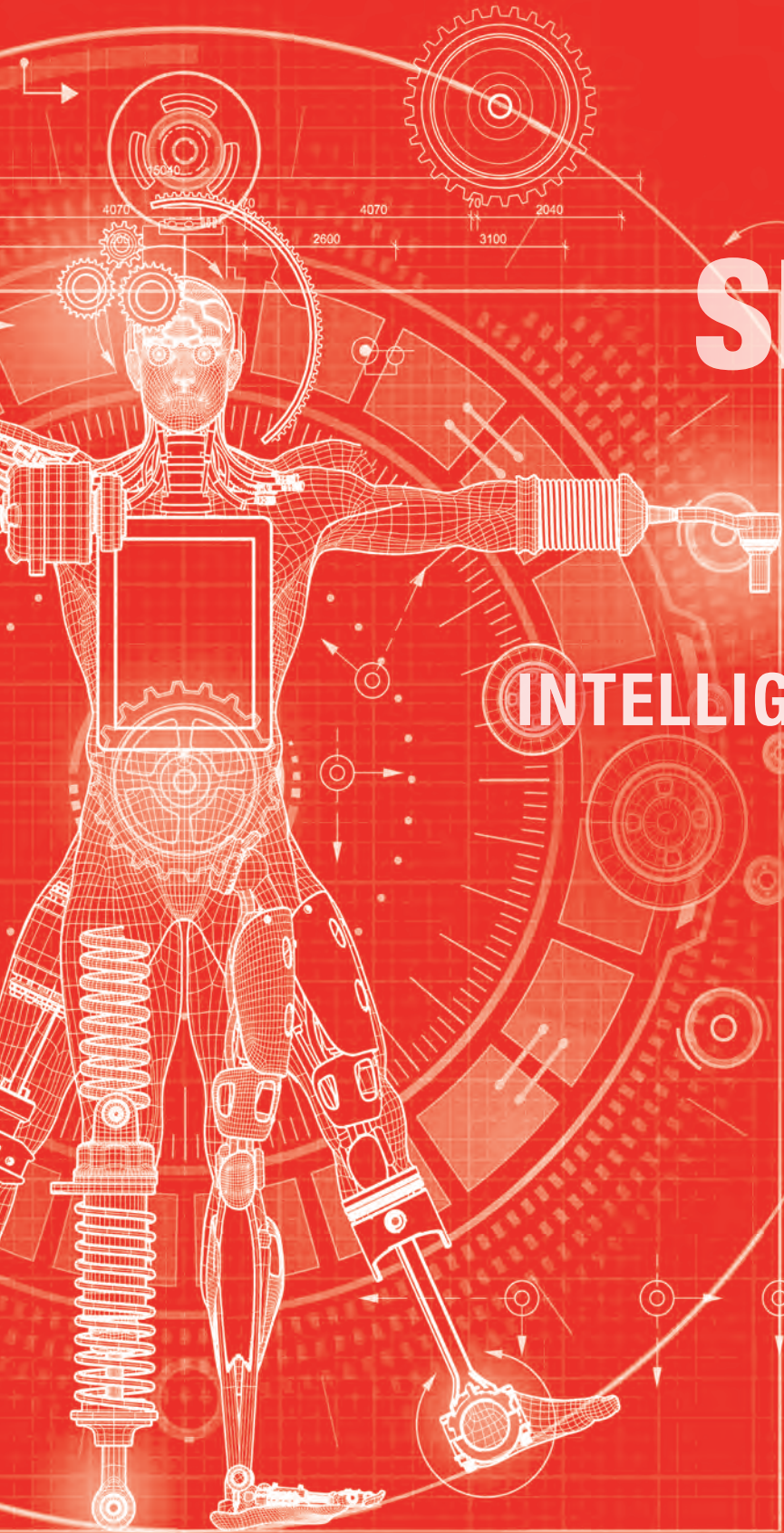




Vodafone Institute
for Society and
Communications



digitising
europe



SHAPING A NEW ERA

HOW ARTIFICIAL INTELLIGENCE WILL IMPACT THE NEXT DECADE

A collection of essays
and interviews by
Alexander Görlach with:

Sebastian Thrun
Edmond Awad
Alena Buyx
Niall Ferguson
Stephanie Hare
Viktor Mayer-Schönberger
Audrey Tang
Andreas Weigend
Sami Haddadin

2020



Vodafone Institute
for Society and
Communications



SHAPING A NEW ERA

HOW ARTIFICIAL INTELLIGENCE WILL IMPACT THE NEXT DECADE

A collection of essays
and interviews by
Alexander Görlach with:

Sebastian Thrun
Edmond Awad
Alena Buyx
Niall Ferguson
Stephanie Hare
Viktor Mayer-Schönberger
Audrey Tang
Andreas Weigend
Sami Haddadin

Table of Contents

3

Preface

Joakim Reiter,
Group Director of External
Affairs at Vodafone and Chair-
man of the Vodafone Institute

6

The big picture

Alexander Görlach,
Harvard University,
Carnegie Council for Ethics
in International Affairs

The Interviews

10

This will be the decade of artificial intelligence

Sebastian Thrun, CEO of Kitty Hawk and Founder of Udacity

14

The idea that everyone should be equally safe is contestable

Edmond Awad, Department of Economics at the University of Exeter Business School

18

Medicine is not an area where you can be disruptive

Alena Buyx, Institute of History and Ethics of Medicine at the Technical University Munich

24

It's not necessarily truth and beauty that holds people's attention

Niall Ferguson, Hoover Institution at Stanford University

28

I want a moratorium on facial recognition technology

Stephanie Hare, researcher exploring the nexus between technology, politics and history

33

We love the excess of availability

Viktor Mayer-Schönberger, Oxford Internet Institute at the University of Oxford

39

Democracy needs to evolve into a real-time system

Audrey Tang, Taiwanese Digital Minister

46

I am sceptical of people who say that technology hasn't changed anything

Andreas Weigend, former Chief Scientist for Amazon

50

How to drive technology in the direction we want as a society in general

Sami Haddadin, Munich School of Robotics and Machine Intelligence at the Technical University of Munich

54

Looking ahead – Shaping the vision for a sustainable Europe

Inger Paus, Managing Director Vodafone Institute

55

Imprint

Preface

Just two years ago, the Vodafone Institute published a collection of interviews and essays from leading scientists about the opportunities and challenges presented by artificial intelligence (AI). We called it "Entering a new era", and placed a particular emphasis on philosophical, ethical and historical considerations.

The influence of AI is rapidly increasing. This development is perhaps most visibly illustrated by the number of AI-based smart speakers in circulation. While around 32 million of these devices were sold worldwide in 2017, that number spiked to over 80 million in 2018. And in 2019, the total is expected to be around 150 million.

But alongside consumer considerations, AI will affect many facets of our lives in less visible, yet no less important, ways. AI can be applied to each of the Sustainable Development Goals of the United Nations, whether it be conserving resources, predicting disasters, enhancing food production, combating diseases or tackling climate change. The use of AI will play an essential role in the transformation towards a sustainable digital economy and society.

The many societal and environmental challenges facing our world require us to step up and fully deploy the tools available to us to protect our planet and improve the lives of all citizens. So, too, must companies such as Vodafone and we are ready to do our part. Vodafone technologies are already enabling other companies to reduce their CO₂ emissions significantly, and by 2025, if not before, 100 percent of the energy purchased by Vodafone will come from renewable sources.

Governments, too, must play their part, ideally hand in hand with industry – particularly when it comes to ethical issues and steering the use of AI in a manner aligned with our European values and way of life. Vodafone launched its own AI Framework in 2019, setting out how we intend to deploy this technology across our business in an ethical manner that respects fundamental rights. The framework addresses key ethical considerations to ensure responsible use of AI: human rights and fairness, transparency and accountability, privacy and security as well as diversity and inclusivity.

At this critical junction, with the recent launch of the EU's new Digital Strategy and to advance the dialogue on AI, together with our writer, Prof. Dr. Alexander Görlach, we have once again sought the views of renowned experts, this time focusing primarily on specific fields of application. What is being done with AI already? Where is it leading society? How do we want to shape our world with AI?

May we find inspiring answers to these questions in the lively discourse that follows.

Enjoy reading!

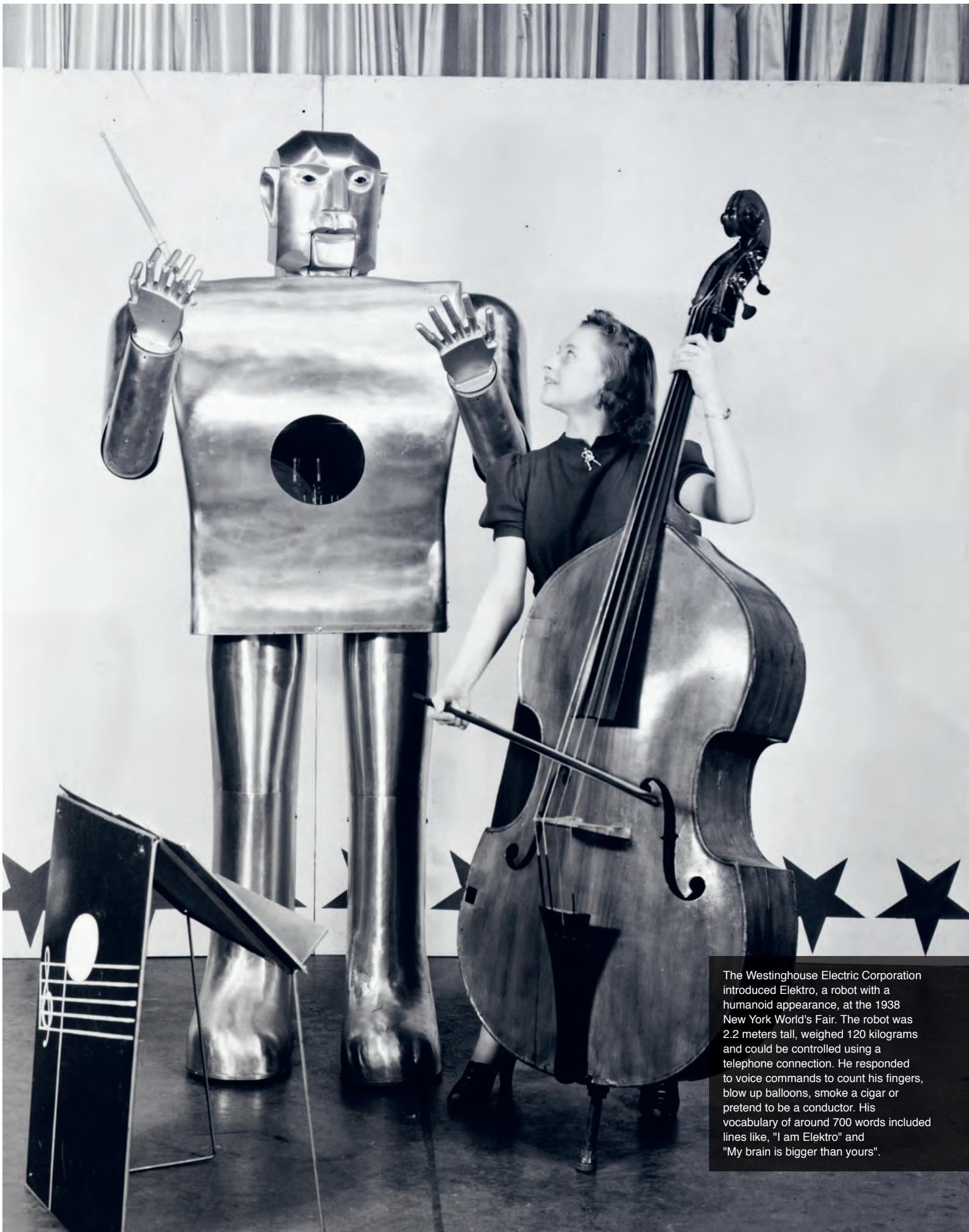
Joakim Reiter

“From entering to shaping a new era”



JOAKIM REITER

is Group Director of External Affairs at Vodafone and Chairman of the Vodafone Institute. Before joining Vodafone in April 2017, Joakim was the Assistant Secretary-General of the United Nations and the Deputy Secretary-General of the United Nations Conference on Trade and Development (UNCTAD). Prior to that, he spent more than 15 years in the foreign service of Sweden, including as Deputy Director-General of the Ministry of Foreign Affairs, Ambassador to the World Trade Organization and at the Permanent Representation to the European Union. He also served as an EU negotiator with DG TRADE at the European Commission. A Swedish national, Joakim holds a Master's in Economics from the London School of Economics and a Master's in Political Science from Lund University.



The Westinghouse Electric Corporation introduced Elektro, a robot with a humanoid appearance, at the 1938 New York World's Fair. The robot was 2.2 meters tall, weighed 120 kilograms and could be controlled using a telephone connection. He responded to voice commands to count his fingers, blow up balloons, smoke a cigar or pretend to be a conductor. His vocabulary of around 700 words included lines like, "I am Elektro" and "My brain is bigger than yours".

After losing both an arm and leg in an accident, Londoner James Young has worked with a Japanese gaming giant to personally design and develop his own advanced bionic arm – earning himself the nickname Metal Gear Man. The arm features a 3D-printed bionic hand that enables him to perform a number of gestures, all controlled by tensing his shoulder muscles. It has the capacity for USB-powered attachments and can charge his phone. The carbon-fibre limb is part art project, part engineering marvel.



The big picture

An Essay by Alexander Görlach

As technological innovation ploughs ahead, long-term consequences and ethical questions are often forgotten. But German academic and writer Alexander Görlach argues that a new way of thinking has already begun to take root. It involves engaging society as a whole to figure out how to make technology work for us – and not the other way around.

The German language is known for creating long words like *Technologiefolgenabschätzung*. The term refers to the process of assessing the long-term impact new technologies will have on society. In its own dispassionate way, this word confirms the primacy of technology above all other elements of society. Philosophers, sociologists and political scientists, not to mention lawyers and theologians, have rarely knocked the world off its axis. And lawmakers are often busy stuffing what has long since become reality into a regulatory corset based on antiquated standards.

Technological achievements modernise societies and force others to react. After 15 years of constant digital-platform development and living in the new economy this has created – one that has permanently changed the way we do business, communicate and engage in politics – we are now in a phase in which the medium- and long-term implications of this development are crystallising. It has become clear to social media companies, for example, how vulnerable their platforms are to manipulation. The automotive industry is asking itself what criteria should determine when self-driving vehicles

must brake for other road users. In medicine, algorithms are helping us identify tumours accurately and detect pandemics before their impact is widely felt. But this also raises the question of how best to meet the challenge posed by the fact that more and more people are staying healthy and living longer. Questions have been raised and are waiting to be answered.

Technologiefolgenabschätzung does not require us to demonise disruption and its consequences, but it means their impacts, both good and bad, require serious examination. And it entails thinking through scenarios that are not immediately apparent to an engineer tinkering with an invention or to a programmer on the brink of a fresh breakthrough. A new way of thinking that is willing to consider such ethical questions is already beginning to take root among some of the central players of our era. Data, a product of this new age, is useful in this endeavour. It has never been possible to investigate complex issues as quickly and comprehensively as we can today.

The Moral Machine developed at MIT, for example, collected and analysed data with the goal of developing ethical standards that decide when self-driving vehicles should apply their brakes. The answers varied depending on cultural presuppositions. In cultures that value the elderly, fewer people were willing to run over an older person in a worst-case scenario. It also revealed that certain ethical standards apply across cultures. People everywhere want to keep the number of potential victims as low as possible, and drivers around the world want to survive, regardless whether it costs one or more people their lives. The desire for self-preservation takes precedence over survival of the species.

The Cambridge Analytica scandal has made clear that political content is sensitive and cannot be handled like ads. We have seen how it can drive a wedge through societies if the same metric that befits the sale of mustard is applied to the distribution of political content. It is therefore logical that the future of democracy is now seen as related to “fake news” and “alternative facts”. It is becoming abundantly clear that societies cannot have an algorithm determine what rules they should live by or what constitutes good and evil. This is a task for all of society.


ALEXANDER GÖRLACH

hosts the event series “AI&I” organised by the Vodafone Institute for Society and Communications. He serves as a Senior Fellow at the Carnegie Council for Ethics in International Affairs and as an Honorary Professor of Ethics and Theology at the University of Lüneburg, among other duties. Before his current engagements, Alex was a Visiting Scholar and Fellow to both the Harvard Center for European Studies and the Harvard Divinity School. Alex holds doctorates in linguistics and comparative religion. His work focuses on narratives of identity and democracy in the digital age.

The internet ethicist Luciano Floridi, who teaches at Oxford, warns that people might forget how to think, evaluate and judge if we let algorithms relieve us of every decision, however small. Floridi points out that free and democratic societies are characterised by the decision-making processes of their citizens, who enjoy equal rights and access to education, including decisions about what restaurant to patronise, where to go on holiday or which new car to buy. A society only benefits its people when it negotiates a *bonum commune* (literally: a common good) and decides how it wants people to be able to live.

With so many marvellous opportunities currently at hand, answers are all the more urgently needed. In medicine, experts are negotiating questions about the beginning and end of life, about reproductive technologies and genetic modification. Climate protection is also a central issue. And questions about fair living conditions and the justness of our economic system must be optimistically addressed if we really want to extend our social contract.

The question of how we want to live is thus far from trivial, and involves more than lifestyle choices. When it comes to civil and social rights in today's liberal democracies, human dignity is becoming a practical matter. The right to vote, after all, is useless if you have nothing to eat. By this logic, a fair and just social order must include universal access to education and healthcare. Only if we agree on this as a society will we be able to use the data we have to maximally benefit the greatest possible number of people.

Today, new actors oppose this line of thinking. In the People's Republic of China, human rights are only understood socially. As long as the Communist Party delivers prosperity to the people, citizens gratefully reward it with political obedience. In the United States, a concept of humanity based on archaic natural law – according to which sexual self-determination, gender identity and homosexuality are aberrations – is once again gaining traction. Both are attempts to deny human rights their political dimension: in China in the name of maintaining the party's power, and in the US in accordance with an evangelical Christianity based on a literal reading of the Bible. When it comes to these questions,

“Today, as in the past, technology can be used for good or evil. With new technologies, democracy can be renewed or entire cities turned into prisons.”

Europe has everything it needs to present an alternative and to renew and further develop humanism, human dignity and the rule of law.

Social polarisation based on divisive traits is already measurable and powerful today, and not only in China and the US. Today, as in the past, technology can be used for good or evil. With new technologies, democracy can be renewed or entire cities turned into prisons. What, then, is the goal we are aiming for with this technological disruption: the digital citizen? Or the person under total surveillance?

Free societies have the privilege of being able to reach agreement through discourse over what their common goal – their *bonum commune* – should be. This question cannot be answered by the experts, the technologists and the programmers alone: Because the self-driving car and the skin cancer-detecting algorithm touch on the *humanum*, the essence of humanity. They are equally entwined with the question: “What is the human being?”

Just as this question reaches beyond the technological, so does its answer. People are not identical to that which

they create. At the same time, human beings do not exist without the implements they derive from the world. In the polarised debate now being held in many places, some supposedly natural human condition – an idealised moment in history – is invoked in order to immunise against innovations of all kinds, or, broadly speaking, against “liberal society”. As if the definition of humankind has been established by nature and is thus not subject to change.

This claim is false: There is no “humankind” without society and without culture. In this case, “nature” and “nurture” are rivals pitted against each other in the debate. However, from the domestication of the horse to the invention of the self-driving car, people have always “enhanced” themselves with the aid of technology.

But people can only fulfil the promise of enhancement they have made to themselves if they remain committed to their own humanity: to empathetic listening, to the desire to learn and understand and to evaluation and judgment. This holistic overall accomplishment and overview of the “big picture” is intrinsic to humankind. In this respect, humanity is still far from being surpassed by algorithms. Algorithms are already many times faster than people when it comes to completing a single task, but they are not able to connect different contexts or complex relationships.

In this sense, this collection of interviews is committed to the “big picture”, a holistic overall view. Because only in a conversation engaging society as a whole can we probe technologies for the good or ill that they may bring and improve them over time. In this context, ethics are not somehow appallingly aloof, but rather everyday and relevant. It is up to us to put new technologies into the service of human beings. Indeed, they exist for our benefit – not the other way around.

“This will be the decade of artificial intelligence”

Interview with Sebastian Thrun

As the influence of artificial intelligence on our daily lives continues to grow, it is important to demystify the technology, argues Sebastian Thrun, a former Vice President at Google. AI is merely a tool, he says, and it can have enormous benefits if we use it correctly.

ALEXANDER GÖRLACH: We hear so much talk about how AI is going to change healthcare, automobiles and so many other industries. What do you see emerging in the new decade?

SEBASTIAN THRUN: This will be the decade of artificial intelligence (AI). I like to compare AI to the agricultural revolution, when machines began doing a lot of physical work. Farming used to involve hard physical labour. Hundreds of years ago, almost all Europeans worked in agriculture, but today, that figure is

below 2 percent. I think of AI as being the same for menial work, for people who work in offices. They do extremely repetitive work, day in and day out. In the future, they can hand off some of the work to the machines.

Those changes, though, are also fuelling fears about job losses and other forms of displacement. Do you see parallels between the current debate over AI and the concerns during the Industrial Revolution that the changes would cause people to lose their livelihoods?

It's a great debate, and a necessary one because it will affect all of us. We need to use AI very responsibly and very carefully. If we use it well, then we will become much smarter as a human race. In addition to freeing ourselves from repetitive work, we can also improve our jobs. Doctors will be better in diagnosing and treating diseases like cancer – and who doesn't want to have a better doctor? Our lawyers will be better too, and so will our pilots.

It feels like the debate about the ethical aspects of AI – potential inherent biases and the disadvantages it can cause – is taking place even before the technology itself has been introduced.

There is uncertainty every time a new technology is invented. And that gives rise to questions like: What does it mean for me? What does it mean for my neighbours, for my community, for my country? And because every new technology affects humanity in some way, it's important to have this debate and to think about its risks and pitfalls in order to minimise the chances that something could go wrong. AI is no exception.

Is it our individual or collective responsibility to address the potential pitfalls of AI?

AI is a technology. It's a tool, in the same way that a kitchen knife is a tool. I can use a knife to cut my vegetables or I can use it to harm a person. The person decides whether to act ethically. And I want to demystify AI a bit: When we talk about AI, we really mean machine learning, which is a subfield of AI. Machine learning is a technique that allows computers to extract rules and patterns from data. If you feed a computer a lot of repetitive data, the computer is able to determine rules based on that data. If you feed a computer enough images of skin cancer, for example, it will eventually be able to detect if a person has skin cancer and provide a diagnosis. If you provide a computer with examples of emails from a criminal and the computer sees enough of them, it can eventually detect patterns and conduct the same task that a person can do. That's it. So, it's a tool that allows a computer to acquire the skills of highly repetitive work from people. The question is: How do we use this tool? These decisions need to be made by everybody and not just computer scientists.

Bias is one of the greatest challenges emerging in AI right now. Experience has shown that technologies like facial recognition can lead to discriminatory practices.

If we as a society believe, for example, that biometric facial recognition works equally well in every part of the population, then there must

“Because every new technology affects humanity in some way, it's important to have this debate.”

be techniques for using it that are appropriate. But if a facial recognition system is trained on one race, we should not be surprised that it performs better on the one it has been trained on. The tools only extract patterns, and it is us – the designers and we as a society – who have to use the tools responsibly, in accordance with our values. If we want facial recognition to work equally well for women and men, for Hispanics and African Americans and white people, then it's our responsibility to train the system appropriately.

AI is already being used in healthcare and in vehicles, both autonomous and otherwise. We have, in other words, gathered some experience on ethical issues. Do you see a potential rift developing – between, say, the Chinese, the Americans and the Europeans – when it comes to the values that inform our different approaches to AI?

It goes back to my kitchen knife example. AI is a tool. There are countries where kitchen knives are being used to feed people, and there might be countries where kitchen knives are being used to stab puppies. It's up to society to think about how to use a tool responsibly. I am convinced that the way AI develops will hinge on tomorrow's values. No artificial network will, by itself, say: “I hate white people and I love African-American people.” That's not the case. The mathematics underlying this model have nothing to do with any racial statement. It's the way in which we use the tool that predicates the outcome.

My conversations with programmers and scientists have made it sound as though a consensus is emerging that we should apply our constitutional values when determining which biases we need to prevent – that algorithms should be programmed to avoid discrimination based on gender, race, sexual orientation and other things. Do you think such a consensus in the tech industry is realistic?

When I apply AI in my own work, I use machine learning to make cars drive themselves and to make them drive more safely than people would be able to drive them. I have also used AI to diagnose deadly diseases like skin cancer, and we have saved many lives using this technology by assisting doctors in finding dangerous melanomas. The issue of ethics has not come up in any technical areas where we rely on machines to do work that humans also do well. There is no question that if we can make cars safer, we should indeed make them safer. I see no moral issue in making the diagnosis of potentially deadly diseases more accurate. We never even considered whether a cancer diagnostic works better with one group of people than another. I don't want to whitewash these important moral questions, but it's how we use the tools that makes all the difference.

Still, some of these technologies create other issues. Technologies that are great for security, for example, can also be used for surveillance, like in China, where the government has installed these technologies to control large numbers of people.

Germany itself is no stranger to mass surveillance. If you think about East Germany and about my generation, you know what that means. What makes me optimistic is that over 1,000 years, Europe grew from a state of almost continuous war into the current era of peace, hand in hand with an era of democracy. Nobody forced the Europeans into democracy – the people wanted it and they won. The people of East Germany prevailed over the East German government. West Germany and a now reunited Germany have worked hard to put the horrors of the Nazi regime behind them. And we have created values that have made lives a lot better than they were 300 or 400 years ago.

Take a look on Wikipedia, where you can see that England and France went to war against each other 26 times over the past millennium. Today, you couldn't even imagine a war between the two. As part of humanity, we have a social contract to use our technologies responsibly. And we do want this, because at the end of the day, we want to have freedom, we want to have liberty, we want to have opportunities, we want to have safety, and we

“If I can see 20 patients a day as a doctor now, it might be possible to see 40 a day in the future.”

want our children to have a better life than we had ourselves. This applies universally to every country.

As someone with years of experience living in Germany, Europe and the United States, do you think it's possible that the EU and the US could adopt a common approach to AI in the same way that the Europeans have worked together to address privacy concerns through the General Data Protection Regulation (GDPR), which regulates how people's private data can be used?

The one thing I would love to see more of, particularly in Germany, is for people to think more about the opportunities presented by AI and not just about the potential risks and negative consequences. We are at the beginning of an era that has the potential to transform people's lives more than any before it. If you take aspects of people's intelligence and put them into a computer, you will fundamentally transform every job. Every single job is repetitive. As you transform human expertise, you will also make humans much smarter.

We have technologies that can turn a nurse into a world-class doctor when it comes to diagnosing skin cancer, without the 10+ years of training necessary to become a certified dermatologist. We can enable a five-year-old to safely drive a car, which was impossible before AI. We need a debate on the possibilities so we can use this toolset to improve society and also ensure that Germany remains one of the world's best countries in the future. Germany is an exceptionally well-positioned country, with incredibly strong talent. On top of that, it is also an attractive destination for immigrants. Germany has the opportunity to become one of the absolute leaders in this nascent field.

You have repeatedly mentioned human-machine interaction and AI, which has been a major focus of yours. How do you think it will transform the workplace in the next decade?

Any worker could ask him or herself: What part of my work is highly repetitive and not super creative? Say, I'm a lawyer and I do repetitive work on drafting contracts. Or I'm a radiologist and I do repetitive work looking at X-rays. I would say that 50 to 90 percent of people's work is repetitive. If a machine watches you



SEBASTIAN THRUN

is Founder and CEO of the flying car company Kitty Hawk and Founder of the online university Udacity. Prior to that, he was a Vice President at Google, where he founded Google X, which is developing self-driving cars and other technologies. A native of Germany, Thrun spent many years as a Professor of Computer Science at Stanford University before joining Google.

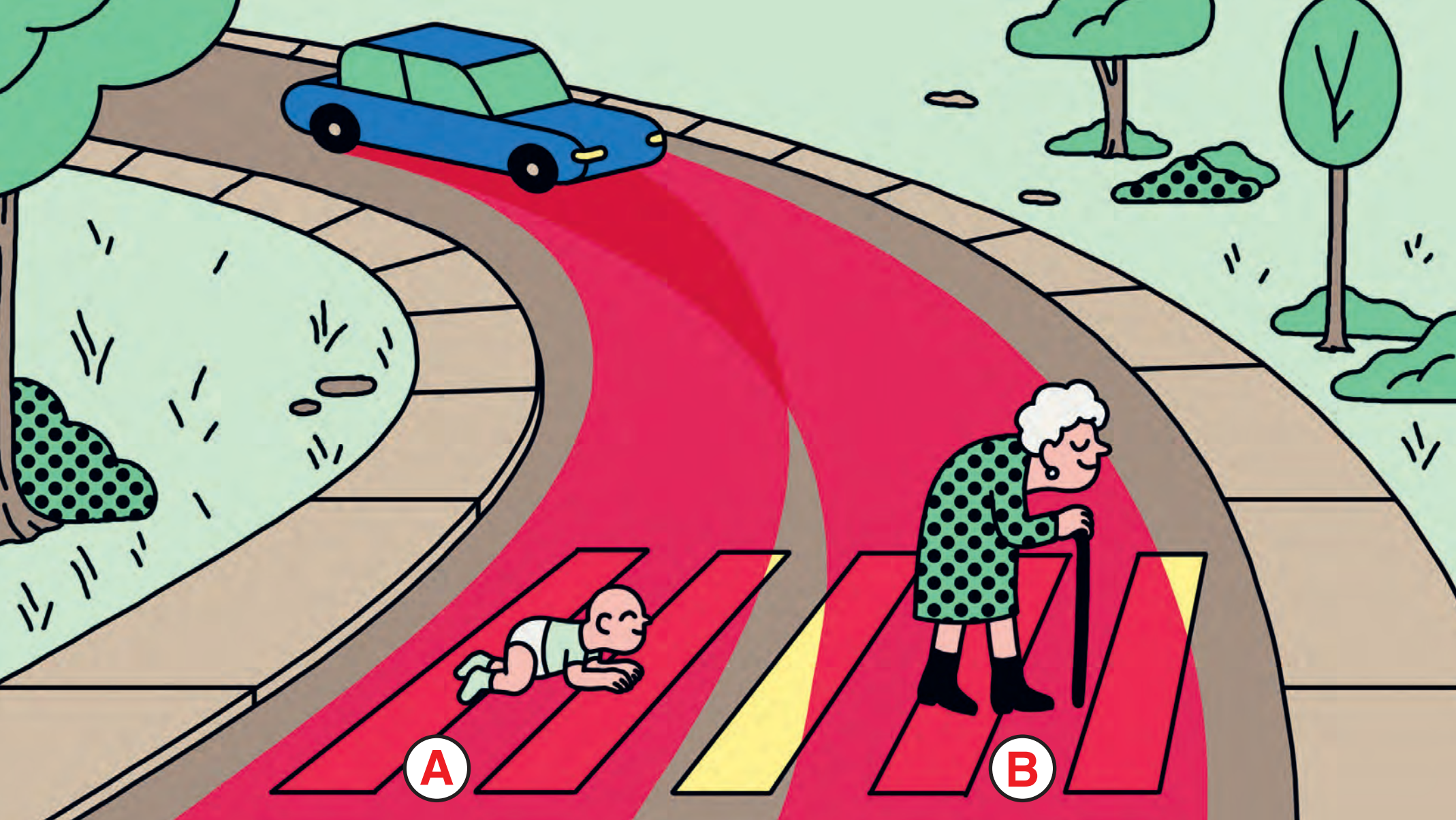
doing this work – and not just you, but every person in the world doing similar work – it will be able to detect patterns and gradually take over these tasks for you. I might need a lot of time to read and answer my emails now, but I might be able get it done in a few minutes in the future. If I can see 20 patients a day as a doctor now, it might be possible to see 40 a day in the future, while at the same time increasing the accuracy of my diagnoses. As a patient, this might mean that rather than paying \$1,000 a month for healthcare, as is common in the US today, I might only have to pay \$500 in the future.

It's not only menial tasks that could be automated in the future. Algorithms are growing so sophisticated that they can even write code themselves. What will we do with our time in an era of pervasive automation?

We have always found new ways of spending our time. Technical jobs like software engineer, TV moderator or pilot didn't exist 100

years ago. There are also non-technical jobs that didn't exist: There were no massage therapists 100 years ago, for example. And there are professions experiencing huge shortages in personnel, like teaching. We know that the best way to teach students is in groups of one to five, but most classes have 30 students. There is no shortage of work. As we free ourselves from the repetitive part of our work, we are free to get more creative jobs. This will increase the need for learning and training.

As a society, we have long believed that a single period of education is sufficient. We go to university and then never get any further formal training. We need to rethink this given that we now live so much longer, that many of us have several different jobs in our lifetimes and that society is changing so much faster. People will have to make formal training part of their entire life journey and not just a one-time event.



The human perspective on moral decisions made by machine intelligence: The self-driving car must choose between killing a toddler and killing an elderly person.

“The public interest surpassed our wildest expectations.”

So essentially, the Moral Machine was an effort to engage the public in the ethical discussion surrounding AVs. What were some of the challenges you faced?

The first challenge was to design scenarios that were close to real-life, while keeping the experiment manageable at the same time. We considered nine different factors that were combined to form realistic scenarios: type of intervention (stay-in-lane/swerve), relationship to AV (pedestrians/passengers), legality (lawful/unlawful), gender (male/female), age (younger/older), social status (higher/lower), fitness (fit/large), number of characters (more/fewer), species (humans/pets). We also considered 20 different characters, including male/female adults, elderly people, athletes, doctors, etc. Then, there were many other aspects that, had we considered them, would have made the scenarios more realistic. For example, we did not introduce uncertainty about the fates of the characters (life-or-death outcomes were certain), or about the classification of these characters (characters were recognised as adults or children and so on, with 100 percent certainty).

The second challenge was reach. The high number of factors and characters considered resulted in millions of distinct possible scenarios, and that needed to be matched with hundreds or thousands (if not millions) of participants.

How do ethics and morality play into the realm of pure technology?

We usually tend to evaluate the performance of technology in terms of whether it is doing what it is designed to do. But we don't usually stop there. We also take performance into consideration, in terms of factors like efficiency, safety and security. These factors are ethically relevant. The problem becomes more complex when we realise that these are not the only ethically relevant factors that we should care about. Recently, we started to realise that

“The idea that everyone should be equally safe is contestable”

Interview with Edmond Awad

The Moral Machine he helped design highlighted the ethical dilemmas of automated driving. But how should they be addressed? Computer scientist Edmond Awad discusses the risks of autonomous vehicles and where the ethical discourse is heading.

ALEXANDER GÖRLACH: What led you to undertake the Moral Machine experiment?

EDMOND AWAD: Our initial motivation for doing the experiment was twofold. First, we wanted to provide a simple way for the public to engage in an important societal discussion. Second, we wanted to collect data to identify which factors people think are important for autonomous vehicles (AVs) to consider in resolving ethical trade-offs. The public interest in the platform surpassed our wildest expectations, with the website going viral on multiple occasions in various countries. At some point, we realised that the sheer scale of the data enabled us to conduct a study that is much more ambitious than we had originally anticipated. In particular, the global coverage enabled us to make cross-cultural comparisons that have seldom, if ever, been possible with social psychology experiments.

other important, ethically relevant factors like privacy, fairness, transparency and agency were insufficiently considered in the design of technologies that we use every day. The work on the ethics of AI in the last few years has mainly focused on including such factors in technology at an earlier stage.

When it comes to the specific technology of AVs, where do ethical issues arise?

When driving on the road, AVs will make some decisions that we, as drivers, make without thinking much, such as how we laterally position the vehicle in the driving lane. When making such a positioning decision, AVs are probably optimising for something, such as efficiency, safety, liability or a combination of all those things. But even this mundane decision may have a societal impact in aggregate. Consider, for example, two different programmes, used by two different manufac-

“Programme A results in killing 15 passengers and five cyclists, and Programme B results in killing one passenger and 19 cyclists.”

turers, that differ in where they position AVs within their lanes. Suppose that after driving for thousands of miles, we notice that both have similar safety levels (resulting, say, in the death of 20 people in one month), but they differ in the proportions of subgroups being killed: Programme A results in killing 15 passengers and five cyclists, and Programme B results in killing one passenger and 19 cyclists. What would be the morally preferable programme? It’s easy to see how Programme B would be more popular among customers, which could motivate manufacturers to become more protective of their passengers, thus putting cyclists on the road at a disadvantage. This goes for other small decisions on the road as well, like the decision to brake or speed up when approaching a yellow light.

It seems likely that people from different cultures would make different choices in specific situations. Did that come through in the study?

We found that, while most countries agreed on the general direction of the preferences (such as sparing younger lives over older lives), the magnitude of these preferences were considerably different across borders. We also found that countries are broadly grouped into three main clusters: Western (including a majority of English-speaking, Catholic, Orthodox and Protestant countries), Eastern (including a majority of Muslim, Confucian and South Asian countries), and Southern (comprised of Latin America and former French colonies). Clusters

EDMOND AWAD

is a Lecturer in the Department of Economics at the University of Exeter Business School. Prior to that, he was a Postdoctoral Associate at the Media Lab at the Massachusetts Institute of Technology. In 2016, Edmond led the design and development of Moral Machine.



“There is a tension between the self-interest of individuals and the collective interest of society.”

largely differ in the weight they give to preferences. For example, the preference to spare younger lives over older lives is much less pronounced for countries in the Eastern cluster and much higher for countries in the Southern cluster. The study also found predicting factors when it came to country-level differences. One example is that the strength of rule of law in a country correlated with the preference to spare the lawful.

What lessons should we draw from these cross-country results?

The main lesson is that programming ethical decisions in AVs using a certain set of rules is likely to get different levels of pushback in different countries. For example, if AVs are programmed in a way that disadvantages jaywalkers, then such AVs may be more acceptable in some countries than in others. Or to put it another way, if one manufacturer figures out a software update that improves the safety of road users, but only at the expense of jaywalkers, should the software be implemented? Our findings predict that such an update may be more acceptable in countries where the rule of law is stronger.

An interesting aspect of the Moral Machine experiment was that those who participated through the website were confronted with a scenario in which an autonomous vehicle’s AI may choose to sacrifice the safety of an AV passenger to, for example, save the lives of a mother and child. Did the results indicate that humans are selfless in that regard and accept their fate in such a situation, or did they show that self-preservation is paramount in our ethical considerations?

When designing the Moral Machine experiment, we tried to focus on cases where website users were not direct participants in the scenario. They were just judging it from a bird’s-eye view. We made this choice because we anticipated that imagining direct involvement would strongly influence their answers

and would thus limit the scope of the experiment. A preceding study (published in 2016 by my co-authors, Bonnefon et al.) considered cases in which participants were asked to imagine being passengers inside the AV (alone or with a family member or co-worker). In dilemmas similar to the ethical thought experiment known as the “trolley problem”, where the AV would have to sacrifice one or two passengers to save 10 to 20 pedestrians, they found that participants preferred to buy an AV that would protect the passengers at the expense of the pedestrians. Interestingly, though, participants acknowledged that sacrificing the passengers in order to save up to 10 times more pedestrians is more morally acceptable. This indicates a tension between the self-interest of individuals and the collective interest of society, a situation known as the Social Dilemma.

Because of these and other dilemmas, some argue that AVs should make random decisions, choosing from various options that have emerged from surveys such as the Moral Machine. What is your view?

If everyone is equally safe or equally at risk in an AV-dominated environment, regardless of their physical features, such an approach does seem justifiable. But that doesn’t necessarily make it the solution. First, the idea that everyone should be equally safe is contestable. Some may argue that extra care should be given to keep vulnerable individuals safer, even at the expense of others. Others may argue that people who are more reckless, or those who participate in the generation of risk, should not enjoy the same safety as everyone else. Second, the idea that AVs would make random choices might have a direct effect on the use and adoption of AVs. An important step in all of this is the ability to quantify risk to individuals and to identify whether some groups enjoy (or will enjoy) more safety at the expense of others in a future, AV-dominated environment.



This still from a video taken by French artist Yves Gellie shows how interaction with a robot can brighten the day for some patients in hospitals and nursing homes. Here, a woman who usually does not speak, chats to the robot about birds and their shared dream of learning to fly.

“Medicine is not an area where you can be disruptive”

Interview with Alena Buyx

Many have high hopes that Big Data, AI and other technological developments will revolutionise healthcare. But with the advances that come with technologies like the genome editing tool CRISPR, we also need to weigh the potential harm, argues medical ethics expert Alena Buyx.

ALEXANDER GÖRLACH: How has AI changed the field of medical ethics?

ALENA BUYX: Not that much has changed, but there has been a clear progression. We have addressed Big Data approaches in genomics and personalised or precision medicine. Slowly, but quite intensely, deep learning and machine learning entered the field of data-rich medicine, and people discovered that if you

have medical, lifestyle, genomic data and so on, you can mine it quite successfully with machine-learning algorithms.

Now you have these machine learning-based algorithms that can make autonomous decisions in symptom diagnosis, making treatment suggestions, or responding in a therapeutic fashion to patients. That is the stage we are currently in.

Medical ethics has always looked at such developments and tried to understand what they mean for clinical practice, for doctor-patient relationships, and what kind of advantages and problems there might be.

Proponents of these developments say they will change the whole system, that we will be going from a treatment-focused medicine that looks at a sick patient to an antecedent health-propelling system.

Yes, with the predictive power that these algorithms bring, we have better chances of making strides towards prevention and understanding the many predispositions for illness. Of course, all the data-driven approaches are

very powerful, and I hate to be the party pooper, but I don't think we will see a radical transformation of the medical system. Hopefully, though, we will get to a place where we are much better at catching and treating things a lot earlier.

Yet for the first time in history, we have data that can make projections into the future and enable us to change our lifestyles in advance.

Some will be able to do that, but many people won't because it's very hard to change your lifestyle. Plus, illness is complex, multi-dimensional and multi-factorial. Often, it doesn't just have one cause that we can avoid or address, like a particular behaviour or a specific genetic component. With many illnesses, particularly those that kill most people, it's usually a confluence of a variety of factors and often happens over time. Even with new data-driven approaches, we can often only predict and address part of that complexity.

But even if we do have highly predictive power, which we *will* have down the road, it won't be available everywhere, and many won't be able to respond to it, or won't have the resources to do so. My hope is that we will make this technology work for the widest swath of society possible, but my worry is that it might widen health inequalities. It could go both ways. I strongly hope it will be very broadly distributed, but I'm not holding my breath.

“A big issue we have regarding the ageing population is loneliness. It's a ‘silent killer.’”

With people reaching more advanced ages than ever before, new diseases will emerge. Researching and treating them costs a lot of money. There's not much data available from the past, since people didn't live as long. Do you have high hopes for this particular niche of medicine?

Yes, it is a promising group, and there are some interesting assistive technologies. Robotics and AI-powered systems could really help people live at home for longer and be fitter and healthier. I expect it to be a very promising area but, again, ageing and illness are complex. Elderly people usually aren't digital natives, so there's a bit of a digital divide. Whether they will be able to quickly adopt these therapeutic and preventive measures is an open question.

But that's more in the realm of care than in medicine, isn't it?

Yes, and there are other questions. A big issue we have regarding the ageing population is loneliness. It's a “silent killer” and a huge factor for morbidity and early death. We know about it, but it's not really at the forefront of policy-making yet. Some say we will never be able to solve it with machines, and that we need to be better at integrating older people into society. Others say we might not always be able to do that with everyone and that if we can provide people with a social experience, it could help, even if it's not with another human. It's an interesting area for us to see if we can address some of these health determinants and do so in a responsible way.

ALENA BUYX

is the Director of the Institute of History and Ethics in Medicine at the Technical University of Munich and holds the Chair of Ethics in Medicine and Health Technologies. She is currently Chair of the German Ethics Council and member of the WHO Expert Advisory Committee on developing standards for Genome Editing.

Change in the last 10 years has been a mixture of hardware and software. In the health industry, it has also been multifaceted and there are multiple factors. It seems reasonable to assume that there will be other changes in society related to these algorithms.

Yes, you'll have the *unknown unknowns*. You'll have benefits that we can't even anticipate yet. You'll have synergies, you'll have sudden, unexpected interplay of certain interventions you didn't expect. And AI used in the workplace could even suddenly improve health for some reason. Of course, the same is true for risks and unanticipated harm. We'll have to be vigilant.

In what way?

We have to perform a careful analysis of the potential benefits and dangers with each application. That's something we have done for decades with new medical technology, so we know how to do it. We need to anticipate how the technology will help, who will benefit, and who will not. We have to anticipate potential dangers and try our best to predict the unintended consequences. With these algorithms, we know that there are areas where they can do harm, such as built-in bias, which would be very problematic in medicine. And there are, of course, a number of data ethics questions. So, we have to assess all this for the applications we want to develop and implement.

When will consumers begin to notice the effects of AI in the way they are treated or in the ways in which their health is monitored?

One thing I want to say, because I'm based in Germany, is that we are far away from anything like using this kind of data as part of a digital infrastructure in health. I'm not talking about consumer products. I hear a lot of, “But oh, we have smartphones, we have all of this technology, we can do anything.” Yes, in



principle we do have the technology. But in health, especially in countries like Germany, we don't yet have the infrastructure needed to actually use the data for clinical applications. One of the worries I often hear from developers is that this will all happen, and the tech will be great, but it will all be developed in China or the United States. That means we won't know if and how these applications will work with our German or Austrian populations.

Could you elaborate a bit more on what you mean by “unknown unknowns”.

Medicine as a business, but also as a profession, is always on the lookout for something new and, at times, we find something in unexpected corners. Medicine just takes whatever there is, even if it's a robot with an algorithm, and tries to see whether that can help patients. To me, that impulse is a wonderful thing.

“Editing embryos and implanting them in women is incredibly irresponsible.”

Physicians are usually very open to all kinds of innovations and technological developments. If you've spent at least a decade in medicine, you've seen some great advances, but also a lot of things that didn't work. Every once in a while, though, something comes along. Suddenly you can treat hepatitis C, people are surviving certain cancer treatments, and you have the first target molecule for amyotrophic lateral sclerosis (ALS), which nobody thought would ever be possible. That's cool!

With the algorithms, probably the biggest impact so far is that they allow the recognition of new patterns. In research, they help us study data in new ways that provide all kinds of ideas about which kind of molecules could work, or what kind of causal pathways we might not even have been aware of. It's a mix between a scientific and an entrepreneurial spirit, and it's very much geared towards innovation, which has probably been a part of medicine since the days of Hippocrates.

There has been a lot of news lately about CRISPR and genome editing.

As a tech-friendly ethicist, I'm so excited about a new technology like this. There's so much potential to do good. But it does have a dark side. Medicine is not an area where you can be “disruptive”. Nor should you be! You can't do innovation the way you do it in Silicon Valley, because you can't “break things and move fast”. Many are familiar with the Theranos case, the failed health tech company led

by Elizabeth Holmes, and its attempt to use this kind of rapid, transformative innovation framework to disrupt the entire laboratory medicine market. None of it worked – it was a huge fraud. It also highlighted that, at the end of the day, innovation in medicine is quite slow, and trying to jump ahead in a “disruptive way” can harm people. Whether you like it or not, it takes at least a decade to get from proof of principle to deployment in routine care – and for good reason.

What He Jiankui did in China with CRISPR and what Denis Rebrikov now intends to do in Russia – editing embryos and implanting them in women – is incredibly irresponsible. The technology is nowhere near the stage where we know if it's safe enough to try it in the clinic. This is unethical on so many levels and, again, an example of trying to be “disruptive”, trying to be first, and accepting a level of potential harm to those involved that should not be accepted. That's the dark side of this kind of innovation, and it is the reason we need to focus more on the ethics of innovation. We cannot just be medical innovators – we have to be medical ethicists at the same time, because if we are not, we will harm people.

As you pointed out, this isn't necessarily new. The novelty here is the scale. And the pace. Hardly a week goes by without a report coming out of Silicon Valley or elsewhere suggesting we can now predict how long we have to live – to the point that people are asking themselves what the future of medicine will look like. What can we really expect in the years to come? Will we have the ability to predict the exact date of our deaths?

No, not for quite some time. This is such a meta answer, but the marketing of new studies and major papers has gotten very good. That didn't used to happen. This is where the media, innovation and medicine converge.

There's a lot of news out there about medical advances, and some of it is grossly overhyped, driven by the desire to be recognised as the first, as the next innovator. We must be aware that we are not talking about just selling phones or cars, we are talking about living people. We should be a bit more responsible, also in communications.

Not too long ago, news broke that a way to reverse ageing had been discovered.

And yes, it was true to some degree. I'm not denying that there are steps being made, but what I am worried about is that with the overhyped reporting, people will get scared. They will think, “What the hell? This sounds terrible, I don't want to know when I'm going to die! And do I actually want to be able to reverse ageing?” That kind of fear also doesn't help build trust.

What are the cornerstones of the ethos you are advocating? And how can we ensure that this ethos is also heard in countries whose systems are based on human dignity?

That's what the theoretical part of my work is about. We have our frameworks and principles that we employ in medical ethics. We also have principles that we use in research, based on the declaration of Helsinki, and a mix of human rights-based frameworks, application-oriented or sector-oriented ethical frameworks, and of course regulations and laws. With regard to the overarching principles, we are fine. We have the main principles and we have the frameworks, and these also differ between places. To some degree, that's good. Ethical frameworks need to be built on universal principles, but they should be responsive to environments, resources and settings.

The task now is to understand what the principles mean in each context and how the frameworks apply.

“We must be aware that we are not talking about just selling phones or cars, we are talking about living people.”

So, when I ask: “You've written this algorithm. Can you make sure it respects patient autonomy?” – what does that mean concretely? Asking these questions and considering potential answers is something medical ethicists have been doing since the field has existed. It's a very exciting time for us, and there is going to be plenty to do.

There are not only lots of science and health stories every week, but also many ethics stories these days. That is new. Something has changed, and I think this has come from an erosion of trust. There is a higher interest in ethics. I wish people didn't have to be afraid and that this interest could come from a more positive place, but I'll take it.

In your opinion, will doctors develop a deeper sense of how their work affects ethics?

Yes, because ethics have been part of medicine for a long time and are part of the curriculum, even if only a small one. Every doctor is trained in medical ethics. I always tell my students that every doctor is also a practical ethicist, whether they know it or not. But, of course, it's not something that people are always conscious of.

Ethics has always been a part of the fabric of medicine. Maybe that wasn't always the case in the tech area. People have thought about ethics for decades there as well, but it wasn't such an accepted part of the field. This is changing now, and that's something I can only applaud.

“It’s not necessarily truth and beauty that holds people’s attention”

Interview with Niall Ferguson

Networks have been around for centuries and have proven to be vital when it comes to disseminating ideas, says historian Niall Ferguson. Internet-based social networks, though, are dangerous, he argues, because they prioritise outrage and extreme views.

ALEXANDER GÖRLACH: Your book, *The Square and the Tower*, traces the influence of non-hierarchical networks throughout history. Have such networks taken on a new life in the 21st century?

NIALL FERGUSON: My motivation for writing *The Square and the Tower* came partly from the ascendancy of Silicon Valley and, more specifically, the rise of the social network platforms Amazon, Google and Facebook. But I had been groping towards a book about networks for a while for different reasons. Right now, I am in the middle of writing a biography of Henry Kissinger; I’ve written the first of two volumes. The second volume, a work in progress, is partly the story of how he went from being a professor to being one of the most powerful men in the world. The book argues that, while it’s true that he’s brilliant – and that’s often, though not always, an advantage in the realm of power – it’s probably just as important that he’s a consummate network-

er who is extremely good at connecting, from the government to the media and Hollywood, from the United States to Europe and China. I wanted to explain his ascendancy after 1969 by showing that he instinctively builds networks.

What can his biography tell us about the way we network today?

Kissinger’s transition was part of the American metamorphosis, and indeed global shift, away from the hierarchical structures of the mid-20th century to something much looser. Hence, I was already thinking along those lines when I moved from Harvard to Stanford three years ago. Here we are, on the very periphery of Silicon Valley, where they think they’ve invented the world anew, and my overwhelming instinct is to say that they didn’t invent social networks. They already existed. They are how human beings organise themselves informally. What Silicon Valley did was to build bigger and faster networks than ever before, but it didn’t create social networks.

Both hierarchical and non-hierarchical networks tend to be values-driven. Do individual networks embody a specific set of values at a given time?

Historically, the best way of transmitting values or achieving a change in values is through a network. The success of both Christianity and Islam illustrates this point very well. You can’t really explain the extraordinary success of these two monotheistic faiths without realising that they were transmitted extremely rapidly, despite almost no technology, to vast numbers of people. And not only did it happen quite



The map Online Culture Wars is an overlay of hundreds of politicised memes, along with influential political figures and symbols. It offers a representation of online ideological and political frictions, integrated into the visual system of a Political Compass meme.

quickly, but the networks then proved remarkably resilient to persecution and conflict. From that experience, we know that networks are how ideas are transmitted, and history has shown that you can transmit any kind of idea, wicked or virtuous, by that means. There's a strange process that determines which ideas go viral, and it's not self-evident that the good ideas always win.

How exactly does technology impact the marketplace of ideas?

As you add technology, you change the ways human beings interact. First, we had the written word, then we moved to the printed word, and so forth, to the telegraph up to the present. You continuously make it easier and more comfortable for people to communicate, lowering the cost to zero and overcoming the distance problem. As a consequence, the potential exists for the extraordinarily rapid dissemination of ideas. The problem is that if you construct your networks on the basis of revenue from advertising, if that's how you actually monetise Facebook or Google searches, then you start to build perverse incentives via algorithms.

The most obvious is that you're selling people's attention. You want to hold their attention, and it's not necessarily truth and beauty that holds people's attention. It's fake news and extreme views. I think what's interesting is that all technologically facilitated networks, especially the big online networks, are quite dangerous. The values that are propagated and the ideas that get amplified are not the kind of ideas that John Perry Barlow thought would benefit from the internet back in the 1990s, when he wrote his Declaration of the Independence of Cyberspace.

Are these networks responsible for the polarisation between, say, the cosmopolitan camp and the nationalist camp that we are seeing today?



NIALL FERGUSON is the Milbank Family Senior Fellow at the Hoover Institution at Stanford University and a Fellow at the Center for European Studies at Harvard University. He is the author of 15 books, including most recently *The Square and the Tower: Networks and Power*, from the Freemasons to Facebook. He has won numerous prizes for his work, including the Ludwig Erhard Prize for Economic Journalism in 2013.

The ideas we encounter in the United States or Europe today are not new. It's almost as if there exists a library of ideas that people choose from based on which seem the most appealing at the time. I don't think anything original has been said about politics in decades, but because people don't really study history anymore, they are impressed by the novelty of old things, and very few people can recognise that Trump's populism is completely derivative of the late 19th century populists. And yet, so many people are running around acting as if something unprecedented is happening. By "unprecedented", they mean "I never read history".

What role do social networks play in this repetition of old ideas?

The connection is that the ideas are out there, and the network architecture determines which ones get the most traction. White

supremacy never went away as an idea. I remember that as a small boy, the first idea about politics that my parents ever taught me was that apartheid was wrong, indeed, I think it's the first thing I ever wrote down as a political idea. So there's nothing new about claiming some pre-eminence for people with light-coloured skin. But why would this dumb idea, which has no basis in science, make a comeback?

I think the answer is: If you set up social media so that extreme views get more traction than moderate views and outrage is the source of engagement, then anybody who is willing to say outrageous things will get way more views than me, with my boring observation that white supremacy is a stupid idea without scientific basis that's been invalidated by historical experience. So that's the problem at the moment. We've built these engines of polarisation and confirmation bias. It's not surprising, therefore, that outrageous ideas make a comeback when the whole machine is set up for that to happen.

What you are arguing is that the networks themselves play a significant role in the ideas that are disseminated. Has that always been the case?

One should never decouple ideas from the network structures that propagate them. Most historians of thought, whether it is political thought or economic thought, tend to talk about ideas as if they float through the air. In reality, ever since there's been a market for journalism, books and people giving speeches or lectures, a reading public has decided which ideas get traction. We have this kind of cognitive blockage in academic life, which is that there are always academics telling us which books were important in the past. But when you look at the bestseller list for the years in question, there's a total mismatch. I don't think those academic judgments are very relevant to a historian.

“White supremacy is a stupid idea without scientific basis that’s been invalidated by historical experience.”

What is interesting is which books sell a lot of copies and which ideas are being replicated. Richard Dawkins loves the concept of the meme, which is an idea that self-reproduces over time. This is very helpful here, because back in the late 19th century, ideas about race underwent a metamorphosis. They became more pseudo-scientific; they became biologically infused. People started digging in the United States; they start adding ideas about miscegenation. These ideas, which had a lot of resonance in the 19th-century United States, got exported across the Atlantic. Those ideas didn't fly over the sea. You have to ask the question, who is translating this stuff? Who is writing about this stuff in German in the late 19th century and giving currency to these new concepts of race – concepts that would subsequently be operationalised by the Third Reich?

I'm fascinated by the fact that there hasn't been a sustained effort to track and graph the networks that transmit ideas. What I say in *The Square and the Tower* is that an idea goes viral partly because of its inherent appeal and partly because of its network structure. We can't write the intellectual history of humanity as if ideas fly around of their own accord.

Is that interplay between appeal and network structure to blame for the dichotomy we are now seeing in the United States, Britain and elsewhere?

We have to be careful of thinking that it's just a dichotomy. That's an oversimplification of what's going on and probably attributes too much importance to the noisiest people. It's like giving all your attention to the Antifa and to ultra-right, neo-Nazi groups. I recently came across a lovely map of the culture war: On the left, things are quite deliberately fragmented because of intersectionality, while the right is deeply divided between what I'll call traditional conservatives and the populist right. It's a multi-front war. Maybe what the social network platforms do is give a false sense of the dualism of our time.

The anonymity scarf was designed by Sanne Weekers, a student at the University of the Arts Utrecht. By overloading them with information, facial recognition systems get confused, rendering the wearer invisible.

“I want a moratorium on facial recognition technology”

Interview with Stephanie Hare

We've gotten ahead of ourselves with some advanced technologies, particularly when it comes to biometrics and surveillance. It's now time to hit the pause button to give scientists time to address the ramifications of these innovations. Researcher Stephanie Hare argues that a moratorium on facial recognition is overdue.

ALEXANDER GÖRLACH: As the line between technology and humankind grows more blurred, some foresee a sort of hybrid existence in the not too distant future. How do you see this developing?

STEPHANIE HARE: In some ways, this has been the case since humans first began using tools. As I type this, I have a pen stuck behind my ear, an example of wearable tech! Anyone who carries around a notebook and pen, or a Swiss army knife, or a bag full of all the implements needed to support a baby or child when taking them out of the house, is living a hybrid existence with technology and tools. We've taken that to the next level in recent years with smartphones, which are minicomputers, cameras, voice recorders and communication (and surveillance) devices all in one. Some people use “wearable devices”, such as Fitbit or the Apple Watch, or they are even requesting to be microchipped. Samsung just announced smart contact lenses that can take photographs and record videos.

Technological advances are the catalyst for innovation in society, forcing lawmakers, regulators, academia, journalism, ethics and the arts to play catch up. How do you explain the delay between progress and society's reaction to it?

Although technology sometimes spurs innovation in society, it is also often a *response* to social, economic, political and even environmental changes. It's a two-way street, but it's even more than that: It is organic, multi-dimensional and holistic. Technological innovation acts on, and it is acted upon. I view it as a force, and an important one. But I don't worship it as *the* force. It is one of the many lenses through which I consider the world.

Nascent technologies like AI and Big Data can also be used to infringe on our rights. Scandals like Cambridge Analytica appear to be only the tip of the iceberg. What does the age of AI mean for people living in democracies?

“I would like to see technology play a role in improving the functioning of democracy.”

I am very concerned about the development of biometrics and surveillance technologies in liberal democracies, which is taking place without any real challenge or checks and balances from lawmakers or regulators. This, in turn, is forcing civil liberties groups, researchers and concerned citizens to raise awareness of the risks and mount legal challenges. Biometrics are some of our most powerful data – our DNA, fingerprints, face, voice, etc. And the use of our biometrics by law enforcement, governments and the private sector without a democratic debate and legal framework to enshrine our rights in law is one of the biggest threats I see to democracies and citizens today.

I have called on lawmakers to pass laws that protect our biometric rights, since the opportunities for abuse here are many and, in the extreme scenarios, terrifying. We also need to empower our regulators so they can investigate possible abuses and take action to hold people accountable and, ideally, incentivise them to protect people’s rights in the design and deployment of their technologies.

Do you think that companies and governments can be relied upon to ensure that some of the worst abuses don’t become reality? To police themselves?

The core concepts of transparency, accountability and responsibility do not feature in how a lot of companies and governments currently use AI and how they plan to use it. This is bad for everyone. Transparency means the ability to know how our data is being used, for what purposes, who it is shared with, how long it is kept, etc. Accountability means the ability to interrogate a decision that is made using our data. Why we were denied a mortgage, for instance, and what we can do to improve our chances of receiving one the next time we apply. Or why did an AI-powered recruiting tool that scans our curriculum vitae not select us to proceed to an interview? Or why did AI-powered

medical software upgrade our diagnosis from not serious to serious? And responsibility means taking care to ensure that using our data in any way, but especially with AI, does not violate our rights or have harmful consequences.

Companies try to hide behind the excuse of “intellectual property and proprietary software”, but that is unacceptable when we are talking about decision-making that affects people’s lives, that may infringe on their civil liberties, that may be riddled with bias and other forms of discrimination. It is particularly unacceptable when it is being used on taxpayers and *paid for* by taxpayers. If you don’t want to expose your algorithms and data gathering practices to audit and scrutiny, then you also cannot expect to get public sector contracts. This technology must be held to account.

How do our institutions need to change to adapt to the sweeping changes brought about by technological advances?

We are seeing a stress test of our democratic institutions, but this has always been the case. We have adapted them when we need to. I would like to see technology play a role in improving the functioning of democracy. For that, though, we need to be able to trust technology – and that is where we are currently so underserved. We have fake news. We have online radicalisation. We have voting machines that are hackable and many that are not backed up by paper, so there is every reason not to have confidence in them. We have lawmakers who are unable, for example, to hold Facebook CEO Mark Zuckerberg to account for his firm’s egregious abuses of millions of people’s data. The reason is that the majority of these lawmakers have not even learned about technology, which means they are in no position to craft laws that will protect us. And we have regulators, such as the US Federal Trade Commission (FTC), that fined Facebook only \$5 billion for those data abuses, while simultaneously granting the compa-

ny’s officers legal immunity for anything that happened before 12 June 2019. That’s a bargain for Facebook! And it sends a signal to the rest of society that the FTC is not prepared to punish Facebook.

In the United Kingdom, our parliament has been stuck in quicksand over Brexit. The result being that lawmakers have done nothing to pass laws on biometrics and surveillance technologies, even though the Science and Technology Committee in the House of Commons called for a moratorium and the Surveillance Camera Commissioner, the Biometrics Commissioner and the Information Commissioner’s Office, all three of the main data regulators, have urged for new legislation *repeatedly*. Yet parliamentarians do nothing, government does nothing and, meanwhile, the police and the private sector keep rolling out this technology. So, we have to wait for a verdict on a landmark legal action to see if this sorry state of affairs will continue.

In China, we are seeing how AI can infringe on people’s rights through the social credit system, which has seen people barred from boarding a plane or embarking on a train journey – and there is no appeals process. China has obtained all its citizens’ data to the point that it can easily be used for subtle manipulation or to discriminate against or incarcerate entire segments of the population. Is this the dystopia we were warned about?

One person’s dystopia may be another’s utopia, or at least another person’s neutral. It all depends on our values and the priorities we set as a result of those values. China is building a system that reflects the values and priorities of the Communist Party and possibly also many other stakeholders. Part of the problem with studying China is that it is not a free society, so people may not be able to express their opinions about whether they consider the system there to be a dystopia.



STEPHANIE HARE

is a Researcher exploring the nexus between technology, politics and history. She has worked as a Principal Director at Accenture Research, as a Strategist at Palantir, as a Senior Analyst at Oxford Analytica, as an Alistair Horne Visiting Fellow at St. Anthony’s College, Oxford, and as a Consultant at Accenture. She holds a PhD and Master’s of Science from the London School of Economics and a BA in Liberal Arts and Sciences (French) from the University of Illinois at Urbana-Champaign. Her book *Technology Ethics* is slated for publication later this year.

“Companies have already shown that they cannot be trusted to self-regulate.”

Certainly, for the Uighur Muslims – who are monitored down to their biometrics, including their DNA, face, voice and fingerprints, and over 1 million of whom are being kept against their will in concentration camps – the system in China is authoritarian and even totalitarian. But there may be other people in China who are fine with all this, and others who may not like it but don't feel they can do anything.

For liberal democracies, it's trickier: We have a tradition of human rights and civil liberties, and some of us even believe in it. So, the introduction of biometric and surveillance technologies really challenges us. The fact that we are currently doing it with no legal framework creates an even bigger challenge. We are seeing some pushback against companies and the government gathering and keeping so much data. Companies may change before governments do on this if they sense that their consumers value privacy and data security more than the trade-offs currently on offer, which are mainly convenience and so-called "free" services (that is, free in exchange for our data). Governments may take longer to realise that all this data collection does not actually make us safer – data collection is itself a security threat. It will be interesting to see what the tipping point will be for governments when they start viewing poor data security as a threat to the economy and national security. It's only a matter of time.

Would you recommend temporary prohibitions on certain new technologies until we

can determine their downsides and upsides and make sure that scientists give appropriate consideration to the potential influence their innovations will have on societies from a human rights and democracy standpoint?

I want a moratorium on facial recognition technology use by the police, other branches of the government and the private sector with immediate effect and lasting at least until lawmakers have passed laws that enshrine our rights about data relating to our bodies. That is the bare minimum. Companies have already shown that they cannot be trusted to self-regulate on this; on the contrary, they are building, testing and deploying these technologies enthusiastically.

As for how to make scientists and technologists think about the implications and consequences of their work, many are already doing this. There is also a wealth of scholarship out there already in the Science and Technology Studies (STS) discipline for any who are not. This discipline has existed since the inter-war period and really began to take off after the Second World War, and it is some of the most fascinating thinking I've ever come across, contextualising hard science and technology in ethics, law, culture, history, sociology, gender studies, etc. It's one of the most exciting and dynamic fields, but there is some truth to the criticism that it is not well known inside most of our technology companies and possibly not in companies dealing with other hard sciences (defence, chemical and biotech, for instance), nor has it received the attention it deserves in business, law, politics and journalism.

Your new book, *Technology Ethics*, is set for publication soon. What aspects of the topic are the focus of the book?

I hope to deliver actionable insights so that the book is as useful and provocative to a computer scientist or engineer as it is to a C-Suite executive, board member, lawmaker, journalist or even an ordinary citizen who wants to engage with my main question, namely: How do we build and use technologies so that they maximise benefits and minimise harm?



“We love the excess of availability”

Around three million products are waiting for dispatch on the shelves of the Amazon logistics center in Brieselang, Germany.

Interview with Viktor Mayer-Schönberger

The disruption brought by Big Data and technological change is reconfiguring society on a scale so massive that our political systems are ill-equipped to deal with it. The best answer, argues Viktor Mayer-Schönberger, is to look back at evolution.

ALEXANDER GÖRLACH: You've been studying the impact of Big Data on society for years. How would you summarise its effect?

VIKTOR MAYER-SCHÖNBERGER: We will witness a period in the coming years that resembles the phase of internal social conflict we experienced in the Enlightenment, a public and societal debate on whether we should become more rational in our views and our decision-making. At the time, the Enlightenment went against the grain of so many longheld beliefs, superstitions and dogmas; it led to large social upheaval but also to tremendous progress. We have continuously pursued the Enlightenment programme without ever returning to the original assumptions.

What few of us realised in the 1970s, 1980s and 1990s is that humans are quite biased and far from the rationalist descriptions of the Enlightenment. So, making ourselves behave more rationally has inherent limitations. That's where Big Data comes in: Used correctly, it can help us overcome some of our human biases and lead to a somewhat more diverse, or at least less emotionally biased, view of the world. Of course, that may be disconcerting to some because we humans also indulge in our biases.

Does this phenomenon also underlie the resurgence of populism in many democracies?

Absolutely. Our assumptions and prejudices give us comfort. We like to cling to the popu-

list fear even if the data analytics suggest otherwise. In addition, from the 1950s until the 1970s, many in the West were content with second-best, as long as it offered economic and social progress. Back then, there was no perceived finite end to resources and, accordingly, there were fewer emotional quarrels about them. That has changed.

It seems to have led to a kind of cognitive dissonance: We can now customise products for ever smaller groups in societies and yet we still expect a piece of the bigger pie.

That is true, and in a way, we are still like the proverbial kid in the candy store. We love the excess of availability. In the 1970s, my parents were content to go on a vacation with me to Italy and considered it a badge of honour to lie on the beach next to the Müllers, the Meyers and the Hubers. We all got the same beach. Today, that's not good enough anymore. We don't want to be part of the group that gets the same thing, we want difference. At the same time, a lot of people now realise that the pie is finite – and that exacerbates fears of being on the losing end of the distributive struggle.

On the one hand, we're speaking about consumerism, but on the other, about the political and societal system that holds us together. Certainly, both are connected, especially in democratic theory, but if we look at Big Data, it has the potential to change both in different ways. How has the

VIKTOR MAYER-SCHÖNBERGER

is Professor of Internet Governance and Regulation at the University of Oxford and has published a dozen books and over 100 articles and book chapters on issues relating to the data economy and digital transformation.



realisation that the pie is finite altered our preferences?

We've now grown up in a world of diverse preferences. We know that we don't have to follow in everyone else's footsteps. We've grown accustomed to a world in which, in theory at least, most of our preferences are fulfillable. And now we yearn for their fulfilment. When you look at the first wave of e-commerce, it was all about getting the best price. With all the new recommendation algorithms, we now want precisely the right good or service. With this comes a sense of entitlement.

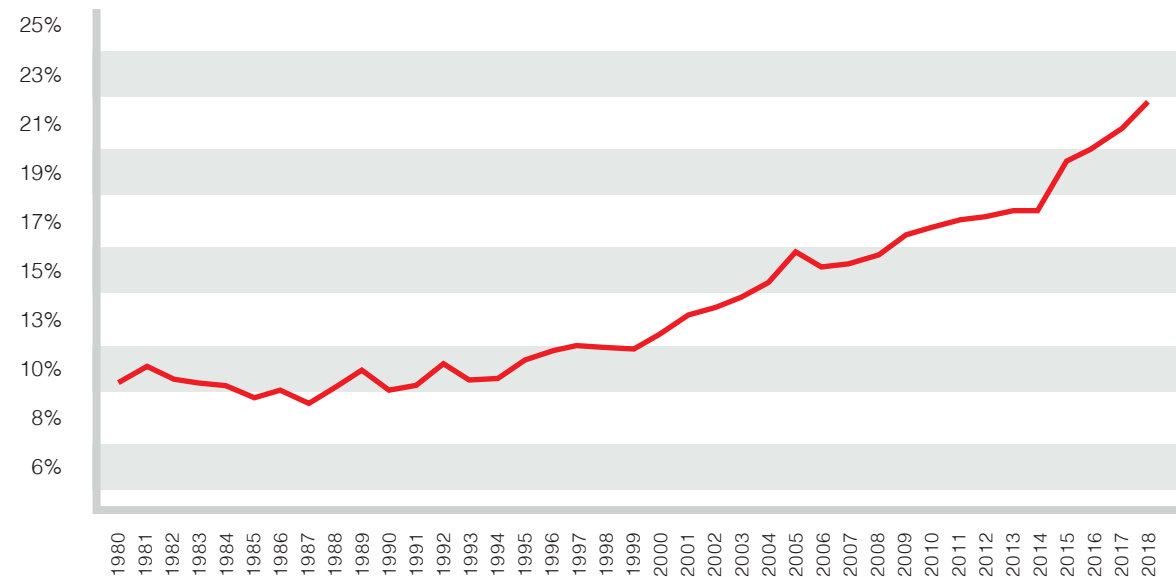
Is the perfectly customised good still utopian or is it already a reality?

I don't think it's utopian anymore. But as usual, we have high expectations that aren't

always met. Although it has never denied or confirmed the rumours, it is estimated that Amazon generates a third of its profits from its recommendation engine. It's not because Jeff Bezos hypnotises us to buy his wares; it's because we look at it and we say, "Yes, that's exactly what I want." This improved matching of supply and demand creates consumer surplus and value. The same is true in the travel industry. The fact that travel platforms enable us to find exactly the right flight and ride share is amazing. Thirty years ago, we mostly bought standard package holidays.

The mobility industry has already been shaken up by technological change to a considerable extent, and its giants are increasingly under pressure.

The rise of the populist parties in Europe



Source: Timbro Authoritarian Populism Index 2019

Average share of votes for populist and extremist parties in 33 European countries 1980-2018

Absolutely. And they completely misread the situation: As they continued to discuss the utility of the diesel engine, Uber was already preparing for the driverless age and pushing shared rides, which increase efficiency but reduce the demand for drivers and cars. If I were in the car industry, I'd be concerned not only that my sales are going to decrease but also that I will be selling to large fleet managers rather than individuals. These fleet managers will have much greater bargaining power than you and me. In crude terms, the ecosystem that the car manufacturers have built to milk society is falling apart. My concern is that one half of the industry is in denial and the other is too blind to see it.

Many car manufacturers would argue that those days are still far away. As long as China only has 25 cars per 100 citizens versus the approximately 70 in most advanced economies, there is still room for the market to grow.

That is a dangerous argument. China is increasingly implementing environmental policies that make many gas-guzzling cars obsolete, while at the same time promoting electric vehicles. By contrast, German car manufacturers haven't yet become champions of e-mobility. We see similar trends in the financial industry. There's a reason Deutsche Bank is laying off tens of thousands of its employees: The business model of being

intermediaries in the monetary chain is quickly growing obsolete and is no longer creating as much value.

At this point, should we be focusing on the micro level, such as business models, or on the macro level, in the form of systemic overhaul?

It depends who you talk to. Policymakers have to look at the macro level, whereas normal people are more concerned with the micro level. The real problem for policymakers is not a single issue, but rather the reconfiguration of society on a massive scale. Our political systems aren't very resilient to the tensions that result from this.

Earlier, you suggested that the core of the Enlightenment was the intellectual shift to rationalism. But wasn't it also about technological change?

I am rather careful about granting any agency to technology. There were societal changes that enabled technological ones and vice versa. This influence has always been going in both directions. What we see at the beginning of the 19th century, for example, is the rise of commerce. Along with it came better reporting and accounting in firms that in turn led to better decision-making. So, it's not a simple technical change that "did it". It's also about structural changes. Together, they create economic opportunities that have been driving change.

And it is hard to predict the consequences of technical innovations. For instance, Thomas

“The real problem for policymakers is not a single issue, but rather the reconfiguration of society on a massive scale.”

Edison invented the phonograph and thought it would be used to send messages, while Alexander Graham Bell believed that the telephone he invented would be used to listen to concerts. And, of course, once their inventions were released to the public, they were reimagined and appropriated, ultimately being used in the exact opposite way. Technology is far more plastic than we think and can be adapted to the needs of individuals.

Nonetheless, I would still argue that without the invention of the steam engine, the story would have been completely different.

That is true. However, the steam engine had also been around for a while before it took on such a central role. It was initially much less efficient than human work. It required the invention of the governor, for instance, to make it more controllable. Even this story is far more complicated than it seems on the surface.

Who should we be turning to for predictions about the changes the future will bring? Comparative historians?

Evolutionists. At the end of the day, that is what we have to come back to. Quite some time ago, we abandoned the notion that evolution is a simple, linear process. Instead, we now speak of punctuated equilibria, which are long phases of relative stability and short phases of radical change. And when there is radical change, evolution does not hunker down – it goes into overdrive. The solution in

times of unpredictability is to try everything, which increases the chances of discovering a solution that works.

If you take that together with what you said at the beginning about the crisis of democracy, is the idea that we need to see how we can combine civil liberties and economic prosperity and then adapt the solution in different democracies?

I started off as a lawyer. In Singapore, we applied comparative law by considering best practices in other regions of the world and adapting them. That may sound nice, but it was a rather silly idea in retrospect. Every legal system is wedded to its specific sociological, political and economic context.

So, I'm very hesitant to jump quickly to universal recipes or advocate "best practices". What may work in one society may not work in another. That's also true for the forces that undermine democracy. If you look at the rise of populist movements, from Alternative for Germany to Italy's Lega to National Rally in France, they have some similarities, but also a lot of differences, as can be seen from the difficulties they have had in forming a joint faction in European Parliament. Populism is a very loose category.

We all have friends scattered around the globe, which makes it easy to come to the realisation that we are living in relatively undogmatic times. At the same time, this produces dissenting voices in the populist camp.

That is not something incredibly novel. It happens from time to time. In Europe, the 1920s were an era in which many old dogmas were discarded. That was a time of liberation, but it also sowed the seeds of what was to come.

Similarly, Erasmus of Rotterdam taught about religious freedom a few years before religious wars broke out across Europe.

Yes. These periods of change and opening recur from time to time, and they always come with the risk of a backlash. European economic integration reached its first zenith in the 1890s and then it declined. It took until 1970 for Europe to become as economically integrated again as it was back then. I wrote my master's thesis in the late 1980s on the United States Supreme Court confirming a law criminalising homosexuality. That's almost unthinkable today. We have come a very long way in 30 years. But for the proponents of Enlightenment this means not solace but the need for vigilance: Nothing is irrevocable, as much as I would like it to be.

“Democracy needs to evolve into a real-time system”



Taiwan's government introduced a process that uses a combination of online and offline debate to find consensus among engaged citizens on specific issues of law and regulation. At its heart is an online platform called pol.is.

Interview with Audrey Tang

Before the internet, it wasn't possible to crowdsource civic life. That has changed, but Audrey Tang, Taiwan's Digital Minister, argues that recent technological advances have shown us that machine learning is no substitute for collective intelligence.

ALEXANDER GÖRLACH: How has democracy changed since the advent of digital technologies?

AUDREY TANG: It has meant two different things: First, digital technologies have made it much easier for people to listen to one another. Pre-internet technologies, such as radio or TV, made it simple for one person to speak to millions. Arguably, that's how World War II was started, and to some extent, World War I as well. It made communication hierarchical.

With the internet, everybody has more-or-less symmetrical connections, meaning they have an equal amount of bandwidth for receiving and sending. This makes "listening-at-scale," as I call it, possible – i.e., one person can listen to millions of people, but more importantly, millions of people can listen to one another. This is a fundamental configurational change. And for the first time, it has made horizontal organisation easier than hierarchical organisation.

This has had implications for parliamentary democracy as well.

Essentially, in the pre-internet era, representative democracy was limited by the reality that it was impossible to listen to more than 20 people at a time. Now, it is possible to listen to 200 or even 400 people at once. We do that every day on Twitter or other social media. It has become easier to coordinate

with what we call "weak links" via the internet instead of stepping into a physical town hall. Representativeness has given way, to a degree, to representation.

And what is the second significant shift?

Previously, when governments had to notify the public about planned legislation or regulations, they could only do so using snail mail or the phone. In some cases, it was a single person's job to handle all of the tens of thousands of incoming messages pertaining to controversial regulations. People were unaware that there were 5,000 other people in line before them, so the administrator essentially became a bottleneck.

Didn't some administrations find innovative approaches for dealing with this phenomenon?

President Barack Obama's White House hired volunteers and had a dedicated staff whose job it was to handle the massive amount of emails he received. They used a kind of rating mechanism – a qualitative algorithm that selected around 10 emails every day to be read aloud to Obama as a way of ensuring both a diversity of voices and a direct channel to the president.

Do you see that as an efficient way of communicating with voters?

It represented less than 0.1 percent of the total mail received. As I said, before internet technologies, it was impossible to crowdsource. It was impossible for those who wrote letters to interact with other letter writers as they now can on the internet. It was impossible for a consensus to emerge.

What does this form of collective knowledge do to leadership in politics? Do these technologies simply prolong the decision-making process?

If we open the agenda-setting power up to the crowd, we can reflect back to people

what they agree or disagree on. Public servants no longer have exclusive ownership of the agenda. But these technologies are still relatively new, so there isn't a clear strategy yet for using them in governance.

But they've certainly had a recognisable impact on public discourse, with Taiwan being one example.

Yes, they provide an accurate reflection of everybody's feelings on specific issues. That has certainly had a healing effect, because if you only look at the mainstream media or, indeed, some social media, you | will see at least five divisive issues constantly being repeated. This has led to the illusion that people are inherently tribal and against each other.

But this technological change is also anti-elitist, in a way.

It might be populism, because it allows everybody to escape the lies of the elite and share their feelings. But it's not tribalism, because it becomes clear to people that we are more or less part of the same tribe, no matter how different we may be superficially.

Would you say these forms of communication can help bridge the gaps that have emerged in many democracies around the world?

Certainly. It is not just an open space, it's a reflective space. And reflective spaces make people aware of the common good. Everybody can relate to a fellow citizen sharing their feelings. It creates the missing link between the data, which is objective and always there, and the notion that everybody can have different ideas, even though only a few might actually be acted upon. We have created a kind of reflective stage between facts and ideas, allowing people to see each other's feelings. The best ideas are the ones that take into account the most people's feelings.

It seems to me that the crisis of democracy stems from the decoupling of civic rights from social rights. Would you agree?

Very much so. It often feels like you have to join one tribe or the other to organise, find a voice and force public servants to take action and alleviate tension.



AUDREY TANG

is the Taiwanese Digital Minister, a position which she has held since October 2016. Considered one of Taiwan's brightest computer programmers, Tang dropped out of school in junior high and founded her first company at the age of 16. She retired from entrepreneurship at age 35 and has since been working on the development of open-source software and freeware.

That is the traditional form of social activism or mobilisation. But people don't organise like that anymore. Just look at #MeToo or #fridaysforfuture. It's basically just a meme that people identify with, take control of and apply however they want.

Is that also what is happening with right-wing parties across the globe?

Basically, it creates hope for people suffering from a lack of representation. This hope depends on outrage against the old, well-functioning democratic system. By using hashtags, people feel like they're part of each other's consciousness in real time, whereas with democracy, you only get to express yourself every four years. It creates an asymmetric hope, and that feeds outrage against the slow pace of democracy.

Democracy needs to evolve into a real-time system. People need to feel that, regardless of

Taiwan recently unveiled its first autonomous electric vehicle, a minibus dubbed WinBus. It is scheduled to undergo a "sandbox" trial before the end of the year and enter mass production before the end of 2021.



the issue, they have a say. There are many ways of doing this, like the Pol.is system for crowdsourcing legislation we use in Taiwan, but also participatory budgeting, e-petitions, etc.

How can we transition into such a real-time democracy?

The United Nations' High-Level Panel on Digital Cooperation recently coined the term COGOV, or collaborative governance. It is a rebranding of multistakeholderism, which has existed in internet governance for decades. The old model essentially asks two questions: First, given stakeholders' positions, are there common values that can be identified? Second, given the shared values, can anyone deliver innovations that represent a Pareto improvement – i.e., that are good for some without being bad for anyone? It's an innovation-focused democracy.

A key component, in other words, is reducing the time between idea and implementation?

In Taiwan, we're adopting "sandbox" laws, which means you get one year to try out your innovations in a heavily regulated area. You will not be fined for breaking the law or violating regulations. Indeed, you're encouraged to do so. But there are two conditions: First, you have to propose an alternate regulation for the one-year trial period. Second, all your data must be open: Your

innovation must be transparent for one year so everybody can see. After that, people may decide it's a bad idea. In that case, we thank you and your investors for paying the tuition. But if people think it's a good idea, then we adopt the entire innovation.

In any field?

Almost any. The Justice Ministry has said you cannot experiment with money laundering or the funding of terrorism, because we already know what would happen. [Laughs.] Otherwise, it's all fair game.

That's a fascinating model. It suggests people are willing to relax some of their standards if they feel like they are in a fair, balanced and discursive arena.

Exactly. I wouldn't say that we're deliberative. The informed decision part is utopian: It assumes that people bring their interests to the table honestly, which often is not the case, even in the very deliberative Swiss referendum model, for example. We usually use the word collaborative, which means we only identify some common values, and we're satisfied with that. It's not really a consensus. It's more like consent, which signals someone can live with something, not necessarily that they would sign their name to it.

China, meanwhile, is taking the opposite approach.

Using more or less the same technology that we are employing to make the state more transparent to its citizens – by publishing budget data, procurement data, implementation data, planning data, etc., and then asking people to participate and comment openly – the People's Republic of China is making its citizens radically exposed to the state. It is doing so with its social credit system and many other programs. Accountability is lacking because even though they are introducing the technology in the private sector, the private sector is de-facto owned and controlled by the state using non-market forces.

What is the critical difference?

There's no due process. It's the same technology we use to foster transparency, but it is used for exactly the opposite purpose. As I mentioned, the "sandbox" method essentially allows the public to judge private-sector regulations. The Chinese Communist Party, by contrast, insists on installing a party delegate in all large companies to ensure that they follow the official agenda. Again, this is exactly the opposite of our method, but applied to the private sector rather than the social sector. I think it's fascinating, like a mirror image of philosophies, using more or less the same technology.

Social progress has always taken place via technological progress, both good and bad, depending on the context.

Yes, it's an amplifier. Whatever your dominant philosophy is, it's bound to amplify it.

As someone from a country that exports technology, how would you address this ethical dilemma, given that innovation is unstoppable?

If there is a high chance of abuse, I would actually argue for export controls. Anti-proliferation is the best path, but it depends on the case. Some things are simple to rediscover based on first principles, in which case it's a lost cause and you might as well export those technologies. But if there are things that are heavily dependent on what we call an increasing return on research, then you can set your research agenda accordingly.

With technology constantly at the forefront of advancement, laws and regulations naturally lag behind.

“As a conservative anarchist, my mission is to foster social innovation.”

The recent Facebook example is a good one: The fine levied against Facebook is more about setting a norm than doing much harm to the company's bottom line, which is crucial in internet governance. It is primarily about setting a cybernorm. People will write code that conforms to incentives and policies and, finally, law. Law always lags behind – but for an obvious reason: If you started with the law, it would mean lawmakers know more about innovations than the innovators do. I wouldn't claim such a thing, and I wouldn't accuse any sane lawmaker of claiming such a thing, either. It's like having a law that dictates the value of π , and we all know how that went.

You prefer the social sector to be the driving force behind innovation?

As a conservative anarchist, my mission is to foster social innovation. This means that the social sector can take control of emerging technologies that benefit society. Just last night, I went to a meetup for RadicalxChange. I'm also on the board. I was joined by Danielle Allen and Vitalik Buterin, who invented Ethereum, the blockchain technology. Our vision is basically to use technologies like Ethereum, but for them to be owned by the social sector. Because Ethereum is open source, anyone who wants to can "fork," meaning, use it to pursue a different vision. This kind of social innovation legitimates governance, but in a way that is free of the legitimacy provided by representative elections. This way, people can see when social innovations work better.

Still, popular opinion can be influenced by foreign forces. Is mainland China trying to interfere with this year's election in Taiwan?

Our president recently coined the new phrase, "the Chinese continent." I like how she referred to Taiwan as "an island off the Chinese continent." In any case, there are undoubtedly Chinese continental forces. The main insight I would share from Taiwan is that we basically have three defences, as well as

“Old media forced people to consolidate their viewpoints to the degree that the human factor almost became indistinguishable.”

three proactive responses. The first of the three defences is that we clarify any trending rumours within one hour, because more people are likely to hear the clarification within that timeframe.

Here's a real example: There was a rumour going around that the administration would fine anyone who permed and dyed their hair on the same day. Within one hour, we had the premier, Su Tseng-chang, post a message that went viral on social media. It read: "There's a popular rumour going around claiming that perming your hair will subject you to a fine of 1 million. It's not true. Although I have no hair now," – with a picture of the prime minister when he was young – "I would not punish people like that. However, perming and dying within a week can damage hair, with serious cases ending up like me."

[Laughing]: That's a funny response.

More people saw this clarification before they saw the rumour itself. The clarifications serve as inoculating agents. This is the first defence. The second defence, of course, is collaborative verification. In Taiwan, we have members of the International Fact-Checking Network (IFCN) and the collaborative fact-checking community highlighting fake news they see on social media.

And what happens with the flagged material?

When something gets flagged, it is a designation that it doesn't belong in the public space. There are international organisations, like Spamhaus, that serve as clearinghouses for flagged material. That's how the spam wars are won. These organisations publish the signatures of people who match those emails, so that machine learning can be used to send mails with those signatures directly into the spam folder.

IFCN basically gathers the material people have flagged as disinformation and they publish

fact-checking reports. Once the Thai fact-checking centre rates something as false, for example, it goes back to the Facebook algorithm so that people stop seeing it on their news feed. As a result, it's shared less. It's not entirely censorship, because if you go to that friend's page, you'll still see the post. All in all, it reduces its virality to around one-fifth or less while increasing the virality of the clarification.

And what is the third cyber defence?

Finally, during elections, our campaign-donation law requires political donations to be designated. Anyone can get structured data, like Excel files, of individual donation records. During the last election we observed that people had been investing in precision-targeted, political advertisements. Because those were not accounted for, we're now changing the law so that they are subject to the same transparency requirements as campaign donations. That is to say, advertisers must disclose where their payments came from. If the purchaser at the end of the chain is not a Taiwanese citizen, everyone in the chain is subject to a fine of 50 million New Taiwan dollars.

Won't Facebook protest?

Facebook, just like the Japanese messaging service Line, which is popular in Taiwan, signed on to best-practice principles for stopping disinformation.

That's part of the lesson learned from Cambridge Analytica, I suppose.

Exactly. People trust Facebook less than tobacco and liquor companies. [Laughs.] It's interesting because there are similarities between the two: Facebook also sells an addictive product and creates externalities for society, so it's going to be regulated the same way. Once these donation lists are published, it won't matter what Facebook's agenda is, because it will be required to publish all of this data, including – crucially – precision-selected targets.

You also mentioned three proactive measures to prevent intrusion in the first place.

First, we have a team in each ministry responsible not only for clarification, but for inviting people with differing opinions to collaborative meetings. Anyone can use e-petitions. We have found that the most active users are

around the ages of 15 or 65. What they have in common is that they seem to be more interested in the public sphere than the private one. Just this month, for example, Taiwan banned the use of plastic straws at many types of restaurants. Everyone is using recyclable straws or glass straws now. This came about a year and a half ago, after an e-petition went viral and quickly amassed 5,000 signatures. The leader of the movement was only 15 years old.

How does that affect disinformation?

Empowerment will make disinformation less likely to spread. Secondly, the citizens can see that the government isn't refuting their agenda and is on their side – and that it is reacting in real time, not just every four years. The office hours set up by the government are also part of this. Furthermore, the conversations are not private. Anyone can go onto Google and find my position on this matter.

Finally, the private sector amplifies flagging and clarification. Both LINE TV and Mashup Television recognise that clarifications, like the one the premier released, can be useful news items to spark a conversation. It becomes good business – and making a business case out of this is also very important. It allows people to participate in the norm-setting. It is no longer just the private sector against the public sector: The private sector also sees that it can be cool to spread clarifications.

Your premier making fun of his own baldness is a good example of the human factor within the digital sphere. What is your take on the human factor in all this?

Old media forced people to consolidate their viewpoints to the degree that the human factor almost became indistinguishable. There was simply no bandwidth in newspapers to deliver a two-way interaction. What we are now seeing is not a transactional configuration of policymaking, but a reactional one. When people participate in e-petitions, they're not just doing it to support a 15-year-old. One also starts identifying with a common goal of improving life on this planet. I think this is essential.

That also requires empathy, which is a uniquely human ability. How can we

incorporate that into technology?

There are a few things. One is that divisiveness and a lack of empathy are direct results of a lack of imagination, so it feels like we're caught in a zero-sum game. It doesn't have to be like that. In many democracies, including Taiwan, the planning horizon is usually four to eight years. However, the hardest problems require structural solutions that take a decade or two. The first thing we do in our collaborative meetings is consider the stakeholders that haven't been born yet. That extends everybody's horizon.

We had a real case involving, on the one side, people who were very much supportive of marriage equality, and on the other, people who were very conservative about whether we should extend artificial insemination rights to people who are not in a marriage. We were able to make this discussion fruitful, instead of violent or divisive, by posing the "How might we ...?" question: For example, how might we ensure an inclusive and accepting society for children born into such a family?

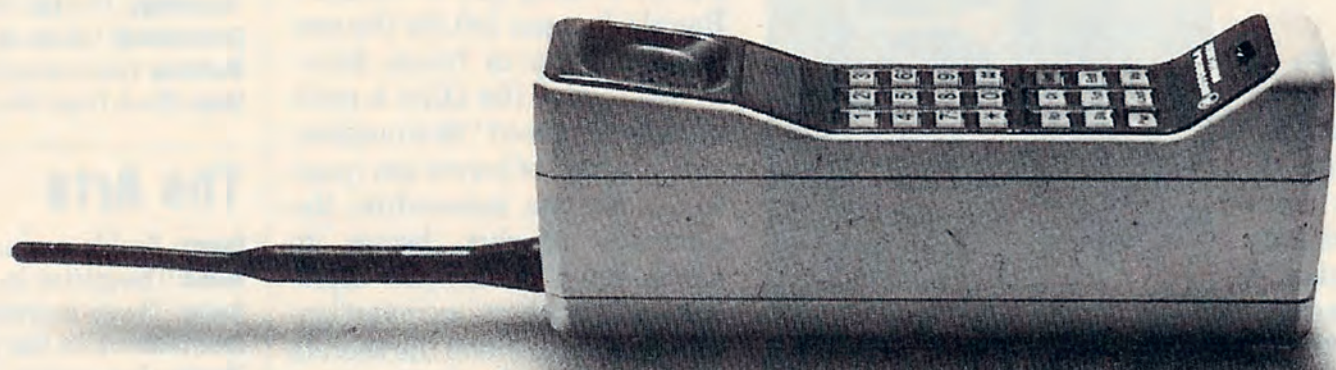
That rephrasing results in the question, "What kind of society do we want to create for people who haven't been born yet?" If you use the planning horizon of 10 or 20 years down the road for a conversation, people who might otherwise disagree can work very collaboratively.

Is that your approach to digital ethics as well, taking the long view?

Yes. We extend the horizon whenever possible. If it's next quarter, there's hardly even time for consent. If it's 20 years in the future, it's much easier to find consensus. It's deeply human to care about the next generation. AI or machine learning cares about the past because that's where the data comes from. The human capacity to imagine, even the poetic capability of envisioning alternate futures, is the key principle for making humanity and the digital coexist.

This is a Motorola cellular portable telephone.

In 1983, the Motorola DynaTAC 8000X came onto the market. The first mobile phone ready for serial production weighed 800 grams and cost almost \$4,000, equivalent to approximately \$10,000 in 2020.



First in the market and first in use. It's the best-selling hand-held cellular portable on earth.

© 1985 Motorola, Inc. Motorola and  are registered trademarks of Motorola, Inc.

“I am sceptical of people who say that technology hasn't changed anything”

Interview with Andreas Weigend

AI has changed the way we think and make decisions, and more changes are coming. Quickly. But there is a dark side to those changes and we must be vigilant, says Big Data expert Andreas Weigend.

ALEXANDER GÖRLACH: In your former position as Chief Scientist for Amazon, your focus was on exploring the possibilities of technology, on things that we now refer to as “algorithms” and “AI”. What has changed since then?

ANDREAS WEIGEND: Jeff Bezos always says: To only look at what is changing is not enough. You also need to look at what *isn't* changing. As a physicist, I'm always interested in the invariances, the things that don't change, and what hasn't changed is that the value of data derives from the influence it has on decision-making. That's how it was 50 years ago, and it will still be that way in 50 years. What has changed, of course, is the breadth of decisions that are affected by data. And the way we make decisions has also changed.

Indeed, I am sceptical of people who say that technology hasn't changed anything, or that it doesn't matter to them. Technology has fundamentally changed *how* we do things and has also fundamentally changed *what* we

do. Technology has had an impact on work, on the future of work, and on how we perform and define work. *Everything* is affected by technology.

When you look at AI, you have a huge amount of complex data but just one thing you want to simplify and learn. Do you think that, in the popular imagination, we attribute too much to AI, or do you agree that its capabilities can be quite frightening?

I don't think algorithms, the foundation of AI, have changed very much over the last 20 years. What has changed is the amount of data and computing power. According to Moore's Law, this doubles every one-and-a-half years, by a factor of 10 after five years, by a factor of 100 after 10 years, and so forth.

The scary thing is that the learning cycles for AI take only one year to grow by a factor of 10. It's almost impossible to imagine, but that is the current speed of the AI cycle. This means

that not only are computers getting faster, but huge amounts of resources are being poured into AI. There is no question that it will have an effect, and it will happen at a speed five times faster than we have seen with mobile phones and other technology that follows Moore's Law. You can slice and dice Moore's Law in many ways. If someone wants to invest money into storage, that means that every five years, you will need a factor of 10 more capacity. We can kind of grasp that. But contrast that with AI, where growth of that magnitude doesn't take five years, it only takes one. It is almost incomprehensible.

What are some potential benefits of this massive growth? How can it make our lives better?

This speed of change makes forecasting difficult. Just take speech recognition: Five years ago, if you talked to a computer and expected text to come out, it would have been nearly impossible. Now, though, a small team of 30 engineers from Silicon Valley can create an app that produces real-time transcriptions. The default has shifted. Every day, in the 15 or so hours I am awake, I produce text. An app called Otter pulls out text that it thinks was important in my day and highlights it. This is no longer the future, this is the present. What is important to me in this conversation is how the default settings in life are changing. My book has a first section where everything is recorded and, more than that, also understood, transcribed, and made searchable. That's a huge effect that AI has had on life. It is a paradigm shift. We don't have to rely on our memories anymore. How will it change how people interact or think about themselves if they have their lives indexed?

In linguistics, we talk about intertextuality, which means that texts are somehow connected through our cognitive ability. Now we have all this data, but what will we do with it? Can we do some good with this



massive amount of data, or will things automatically go downhill?

No, not automatically downhill. There are always good and bad sides to any technology. Certainly, whatever technology can be used here to detect our emotions or to detect terrorists in a crowd can also be used to find dissidents in China. It comes down to what you do with it.

Is regulation futile? We can regulate a company, but the data behind everything is still there.

People are trying to learn. But to mention one example, German Finance Minister Olaf Scholz was sitting next to me recently at dinner and said, "We don't need Uber, we have something

ANDREAS WEIGEND

is the former Chief Scientist for Amazon and author of *Data for the People*, published in 2017. He is a member of Germany's Digital Council and also teaches courses on data science, AI and the social data revolution at Stanford and UC Berkeley and at universities in China, Singapore and Thailand.

called Easy Taxi." He is a smart man, no question. But I was sitting there thinking: How do I tell someone who has a bulletproof car with a driver parked outside, someone who has not stood outside in the Hamburg rain looking for a taxi, how do I explain to him that a platform like Uber is different than Mercedes-Benz writing an app for cab drivers?

I really think that the best thing politicians could do is spend an hour a month with a teenager and ask them how they are leading their lives. People in power have no idea how people under 50 form their opinions. Commissioning a study probably doesn't help either. For me, my students help fill that role. Every now and then, they explain to me the cognitive dissonance and how they see the world versus how I see the world.

But when it comes to transparency, you can force corporations to be transparent. There can be policy pressure and pressure from things like the Panama Papers, for example. Or take the very thoughtful open letter from Amazon workers to Jeff Bezos, urging him to take a more active stance on climate change and transform Amazon into a carbon neutral company.

But then look at Amazon Web Services. Facial recognition at the border is built on AWS. People want to know what the technology is really being used for. The standard example is the role of IBM punchcards in the Holocaust. IBM actively helped by sending engineers to optimise their punch cards for the logistics necessary to get Jews to the gas chambers. Now, people are saying we do not want Amazon to be talked about in 50 years or whenever with regards to a genocide. What can we do now? What could IBM have done in World War II? What could their employees have done to prevent that technology from facilitating the Holocaust?

All fields of innovation are now under scrutiny. Are we entering a more ethically aware period of time? Or are we just being dreamers?

“There is a fundamental power shift taking place from the person as an employee to the person as an individual.”

I think there is a fundamental power shift taking place from the person as an employee to the person as an individual. If a person doesn't like what Company A is doing, he or she can quit and go to Company B. It's also surprising to me how many people give me their Gmail address as opposed to their company email address. I think it's great to see. It's the power of the individual over that of the company.

That's an interesting perspective. You are saying that we have more agency even as many are saying that AI is actually taking agency away from us by giving us recommendations and making us unlearn decision-making.

In the past, we had very limited inputs to, say, find a restaurant. Today, we have many more options. Even in the town where I was born, my brother and I will use an app to find dinner. It truly has the potential to be democratising. I have always been a rebel and I love that these traditional institutions can be blown up.

Your positive outlook on individualism and empowerment is refreshing. In your most optimistic scenario, where do you think we are headed?

This is a moment of maximum uncertainty. I don't know if it's more likely that we'll end up in a surveillance society much worse than we ever imagined, or if it will empower people to express themselves and make better decisions based on the data they create. It is neither one or the other, of course, but I fear the overall distribution has shifted for the worse. I was more optimistic a few years ago than I am now, because I have learned about the Facebooks of the world. They have all the data about you and they aren't trying to help you, but only themselves. I am a big fighter for data literacy. But isn't it amazing that when the news hit that Facebook might be levied a \$5 billion fine that their stock went up?

“How to drive technology in the direction we want as a society in general”

Interview with Sami Haddadin

Europe is taking a more deliberate approach to the tech revolution than powerhouses like the US and China. But Sami Haddadin, an AI adviser to the EU, argues that exactly this is the continent's greatest asset. The focus of technological development, he says, should be on people.

Many have criticised Europe for thinking and worrying too much, and for acting too slowly. As a member of the European Union's High-Level Expert Group on AI, what is your view?

SAMI HADDADIN: Critical discourse is extremely important, so I'm not an advocate for moving forward without discussion. This technology shift we are now seeing represents a fundamental transformation and it is moving at a very rapid pace. I don't necessarily mean the basic scientific progress, which is the product of decades of hard work. But lots of the things that have been developed in the labs and technology centres around the globe are now reaching the real world. And that is really something that must be discussed in a modern society. We're a democratic ecosystem of European countries and we must collectively lead this discussion in the right direction.

Are we there yet? Have we settled on the right direction?

My view is that the discussion shouldn't be focused on what we don't want, but on what we do want. And how we can get there, hopefully at the forefront of technology and science. I truly believe that our wealth is based on scientific and technological discoveries, on our curiosity and on us having been drivers of progress over the last centuries.

At the same time, we have to be responsible. I think the label of trustworthiness is a good one. So the question is: How can we be forward-looking in terms of technology and science while being responsible and respectful of an ecosystem that includes industry, policy and society? As such, the discourse is vital, because if we didn't have it, it wouldn't be Europe.

InMoov is a humanoid robot, developed for artistic purposes by French sculptor Gaël Langevin. The robot is unique in that it can be reproduced in a smaller format using a 3D printer and its files stored under a Creative Commons license.

“The next generation will be about having humans and machines in the same workspaces and interacting in everyday life situations.”

The debate in Europe consistently focuses on ethical standards and values. Is that true at your university as well?

Absolutely. I think they're actually part of the university's genetic makeup, if you will. We drive science and technology forward, but we ensure that the human is at the centre of everything.

What does that look like in practice?

There are a number of questions we seek to address. The first is: What is it that we want to change? But the very next question is: How does it affect society? We engage with the social sciences and with philosophers to really understand how to drive technology in the direction we want as a society in general. Finally, we are also focused on educating the leaders of tomorrow on how to be mindful that they are not only cutting-edge technologists and leaders in the field, but that they should also be aware of what this technology can do and be responsible with their decisions.

More broadly, what should be the focus of a European AI strategy?

I would argue that we should build on the strengths we have. It is well known that science and technology are really strong in Europe, as is the next generation of startups. This is the product of tremendous investments on the European and national levels, investments that obviously have to continue. But now we need to face the transition into the real world. And we need to embrace that transition, because it is a chance for Europe as a continent, and for Germany as a country, to become established on the sustainable side of technology. It is an opportunity to get to the next level of what is being referred to these days as AI – which basically means relying on deep machine-learning techniques to perform large-scale data analytics. Doing so in a variety of applications is just the first phase of AI. The next generation will be about having humans and machines in the same workspaces

and interacting in everyday life situations. That is where the potential and the opportunity for Europe lies, because that is where we are strong. We are effective in uniting engineering and computer science to create something new. The competition has just begun and there is still a future to be determined. But our focus should be on building on the strengths we already possess.

Last year's Tech Divide study from the Vodafone Institute revealed that Europeans are particularly concerned about AI. Yet you argue the focus should be on interactions between humans and machines. Have researchers and academics underestimated the challenge of building trust with people?

I wouldn't say it has been underestimated. It's a concern for the entire community. Our goal has always been to establish a relationship and show that the science and technology we're pushing forward is for the benefit of humanity. And this is the fundamental core of European science, a precious thing we should value highly. I don't think that has necessarily been the focus of science and technology in other parts of the world. But that approach informs the discussion we are having now, a discussion that I truly believe is the right way forward. So I would say that the building of trust is being taken very seriously.

What part of the human-machine relationship are people actually afraid of?

I think what people really fear is being governed or controlled by technology and machines. We have to show them the direct benefits they offer to humanity and to individuals – not by having debates about science fiction, but through real-life technology. We also have to consciously choose which problems we want to solve – such as the future of work and making inhumane work obsolete in Europe, or working toward providing everybody the right medical treatment.

It's about democratisation that we can finally embed. Technology is the solution to the problems that we have as a society, not a hurdle to those solutions. But it's up to humans, to us as a society to determine exactly what we need to do to get to where we want to go. And this is a process, it's how democracy works. So we shouldn't fight it. We have to get out there, talk to people, help them understand how the technology works, get their feedback and co-evolve together!

You seem to be extremely optimistic about Europe's future in tech. What do you hope to see in the coming years?

I really hope the right decisions and the right investments are made. We have to ensure that we have the right people, the right mindset, the right markets and the right scientific background. We need all these ingredients and we have to be strategic about that. More than anything, I would like to see Europe develop a greater awareness of its diverse strengths and a recognition that diversity is our greatest strength. I grew up in a divided Germany with a multi-national background, and seeing how united and strong the entire continent is today makes me think that we need an extremely positive vision of the future because Europe is really something worth fighting for, and science is at the core of it. We are paving the way for the future and educating the next generation of leaders now. So we need to be positive and proactive to make this optimistic vision of the future a reality – a future that is rewarding for our entire society.



SAMI HADDADIN

is Director of the Munich School of Robotics and Machine Intelligence at the Technical University of Munich (TUM) and holds the Chair of Robotics Science and Systems Intelligence. He previously served as Professor and Director of the Institute of Automatic Control at the University of Hannover. He also worked in various capacities as a research associate at the German Aerospace Center.

Interview: Alice Deißner

Looking ahead

Shaping the vision for a sustainable Europe

by Inger Paus



INGER PAUS

is Managing Director of the Vodafone Institute for Society and Communications and is responsible for Vodafone Germany's corporate responsibility strategy. Prior to joining Vodafone, she held numerous positions in corporate affairs and corporate communications at Microsoft, developing campaigns and initiatives on issues ranging from digital education and industry 4.0 to the future of work. She holds a degree in strategic communications and planning from the Universität der Künste in Berlin.

As a new decade begins, the challenges we face could hardly be greater: climate change, epidemics, migration, recession fears and more. The rise of populism in many parts of Europe is yet another indication of just how fundamental these issues have become for democratic societies. But what role does digitisation play in meeting those challenges? Will it disrupt our traditional business model and destroy social cohesion? Or is the reverse true? Will we only be able to overcome these massive problems with the help of technology?

Most Europeans tend to share the latter view, as our most recent Digitising Europe Pulse report (Vodafone Institute/Ipsos, 2020) shows. Accordingly, 79 percent of respondents are convinced that digitisation is a key to mitigating climate change. Measures taken by the European Union in recent years also play a role. In addition to the European Green Deal, the EU is now planning multi-billion-euro support for the development of AI.

For the German economy alone, AI's potential impact between now and 2025 has been estimated at around 500 billion euros (see: Artificial Intelligence (AI): Its Potential and the Lasting Transformation of the German Economy; Arthur D. Little; Vodafone Institute; 2019). Indeed, it is difficult to disagree with Sebastian Thrun's prediction that AI will be the most decisive force in the coming decade.

Yet despite their positive outlook, technology optimists are – consciously or unconsciously –

triggering a considerable amount of economic and ecological controversy. Because what may seem today like a necessary measure to promote long-term sustainability can present a considerable jolt to the current social order in the short term. People in the workforce are worried about losing their jobs, while politicians must prevent the collapse of social welfare systems. Past waves of automation have frequently only benefited ensuing generations. But as the third decade of the 21st century gets underway, we don't have that much time anymore. Our trans-continental trilogy of studies called The Tech Divide revealed the striking differences in attitudes toward automation between the "Old West" (Europe and the US) and the "New East" (Asia). One side is sceptical of progress and views the future with trepidation. The other is euphoric and full of optimism.

The natural tension between technological progress on the one hand and social responsibility on the other has been the focus of the Vodafone Institute's work from the very beginning. This publication represents a seamless continuation of that mission. It is an outgrowth of the AI&I series, which has already involved Alexander Görlach holding discussions with experts like Audrey Tang, Vinton G. Cerf, Sir Martin Rees, Nigel Shadbolt, Nuria Oliver and Luciano Floridi.

All these projects are part of the overarching leitmotif Digitising Europe, which serves as a platform for bringing together decision-makers

and digital pioneers to debate desirable futures. Two prime examples of that mission are the Digitising Europe Summits held in 2014 and 2019, both of which were opened by German Chancellor Angela Merkel. It is of particular importance to us that this discussion not be confined to just Berlin and Brussels. As such, we seek out exchange with internationally renowned partners like Access Now, bitkom, eco, the Data Pop Alliance, the European Institute for Technology, United Europe, the Mercator Institute for China Studies and the Oxford Internet Institute.

Despite the politically challenging environment in which it currently finds itself, Europe must find a way to avoid squandering the technological, economic and ethical achievements it has accumulated in past centuries and keep up with globally acting disruptors. As part of that, we are examining the contributions digitisation can make to mitigating climate change and to building more sustainable economies – the central challenges facing us in the 21st century.

In what kind of Europe do we want to live in the future? We are eager to contribute to a far-reaching societal debate on the search for inclusive and sustainable solutions.

Imprint

Published by

Vodafone Institute for Society and Communications
Behrenstraße 18
10117 Berlin
www.vodafone-institut.de

Executive Committee Members

Inger Paus, Thomas Holtmanns

Responsible for Content/Editorial Management

Friedrich Pohl (Head of Communications)

Publishing House

Tempus Corporate GmbH –
Ein Unternehmen des ZEIT Verlags
Alt-Moabit 94
10559 Berlin
www.tempuscorporate.zeitverlag.de

General Management

Jan Hawerkamp, Kai Wutte

Project Management

Dr Joachim Schüring

Editors

Charles Hawley, Daryl Lindsey

Art Direction

Christopher Delaney

Copyediting

Susanna Forreast

Production

Dirk Woschei

Print

Peschke Druck, Parsdorf

Date of Publication

March 2020

Pictures: Cover: Ikon Images/Alamy Stock Photo; p. 3: Vodafone GmbH; p. 4: Bettmann/Kontributor/Getty Images; p. 5: David Vintiner; p. 8: David Elmes/Harvard University; p. 13: CreativeShot Photography; p. 14: Simon Landrein; p. 16: privat; p. 18: Yves Gellie/filmstill from the movie "The Year of the Robot"; p. 21: Deutscher Ethikrat; p. 24: disnovation.org; p. 26: Dewald Aukema; p. 28: Sanne Weekers; p. 31: Mitzi de Margary; p. 33: Sean Gallup-Staff/Getty Images; p. 35: Peter van Heesen; p. 39: Digitale Vision/Getty Images; p. 41: public domain; p. 42: Ministry of Economic Affairs Taiwan; p. 46: Motorola; p. 48: Social Data Lab; p. 50: ROBOT LOVE/Gael Langevin/InMoov-Photo Yethy; p. 53: Astrid Eckert/TU München; p. 54: Vodafone GmbH



About us

The Institute is Vodafone's think tank. We explore the potentials and responsible use of digital technologies for sustainable innovation and economic growth as well as social inclusion. Through research and events, we provide thought-leadership and offer a platform for dialogue between business, academia and political players. We are committed to improving access to technology for all parts of society and thus develop and support on-the-ground projects for female empowerment. The wide-ranging expertise of the Advisory Board members reflects the Institute's intention to act as a cross-sectoral platform.

www.vodafone-institut.de



Vodafone Institute
for Society and
Communications

