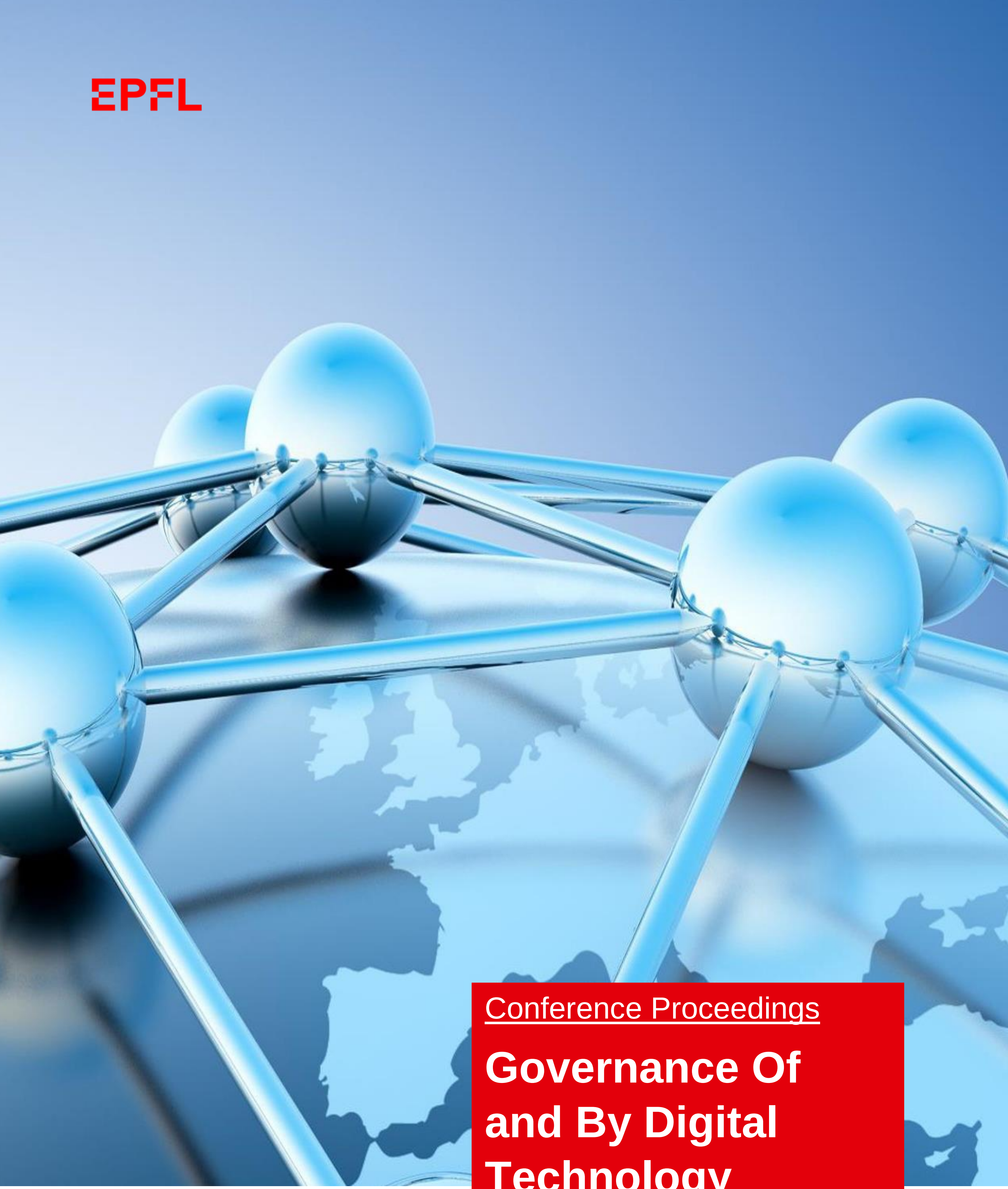

Conference Proceedings

Governance Of and By Digital Technology

18 November 2020

International Risk Governance Center

This paper should be cited as:
IRGC (2020). Governance Of and By Digital Technology, 18 November 2020
Conference Proceedings. Lausanne: EPFL International Risk Governance Center.

Available from irgc.epfl.ch

The **EPFL International Risk Governance Center** is a neutral platform for dialogue about opportunities and risks related to science and technologies, with the aim of providing recommendations for their governance.

Contents

Introduction

On 18 November 2020, the EPFL International Risk Governance Council (IRGC) organised the *Governance Of and By Digital Technology* conference to discuss the governance challenges raised by the rapid pace of technological change and the deepening reliance of societies on digital technologies. This conference convened a multidisciplinary group of expert researchers and policymakers to discuss how to balance the benefits of new technologies with the risks they pose.

The conference was organised under the auspices of the Horizon 2020 TRIGGER project, which focuses on Europe's role within the evolving global governance landscape. The project is led by CEPS (Centre for European Policy Studies), with IRGC heading the work dealing with digital technologies, focussing in particular on machine learning, data protection, distributed ledger technologies, open source software and open standards.

The purpose of this conference was to address the following questions:

- What kind of governance is needed to prevent new digital technologies from causing harm and to keep pace with the rapid pace of innovation?
- Can a technology like machine learning be deployed in ways that address issues of accuracy, bias and misuse?
- Is there a role for algorithmic decision-making in policy, or will it undermine important societal decision-making processes?
- What lessons can be learned from the rapid roll-out of Covid-19 contact tracing apps?
- How should trade-offs be made when technologies bring values into tension – for example, how should performance be balanced against fairness?

These questions were addressed in three sessions: the first looking specifically at trade-offs of privacy and efficacy faced by industry, public institutions or citizens, such as in COVID-19 tracing apps; the second focussing on governance of technology and the challenges of regulating machine learning; and the third looking at governance with and by technology and the potential benefits and risks of digitally-enabled policymaking. There were also two keynotes from world-leading experts including Stuart Russell, who wrote the standard textbook on AI, and Joanna Bryson, a leader in the field of AI ethics.

This report was written by the IRGC secretariat. The detailed programme, slide presentations and video recordings are available at https://GOBDT.ch.

Key takeaways

The conference provided a wide-ranging survey of the issues raised by the increasingly important governance role that digital technologies play, both in terms of the challenges of regulating these technologies, but also in terms of how digital technologies are being used in various governance processes. There were a number of direct and even provocative proposals made during the conference—such as a prohibition on the use of AI for public policy or a ban on impersonation using manipulated images, video and text—but the conference panellists' nuanced and informed approach emphasised the complexity of the challenges in this area. The paragraphs below highlight some of the key issues raised at the conference.

Covid-19 contact tracing apps

The development and use of tracing apps to help control the spread of Covid-19 illustrate the trade-offs that exist between the challenges and opportunities of digital technologies. Issues of privacy and data usage have been raised in most countries, and this has been a key consideration in determining whether apps are built in a centralised or in a decentralised manner. Privacy has been a critical concern for many citizens, as well as for developers of these apps.

Confidential computing

The privacy questions raised by Covid-19 apps are one example of a much wider challenge of finding ways to maintain people's privacy while also realising the societal benefits that data processing can offer. We need to see more and better implementation of various types of privacy-preserving computational techniques that allow for greater amounts of sensitive data to be processed without undermining privacy. These technologies present many opportunities, but there are shortcomings too, and so it is crucial to have governance structures and regulatory responses in place that reflect clear societal decisions about how these technologies should be used.

Regulating AI and other digital technologies

When it comes to the governance or regulation of digital technology—and, in particular, artificial intelligence and machine learning (AI/ML)—it is essential to tackle serious potential problems including malicious use, surveillance, and the deliberate manipulation of individual's choices and decisions. There are genuine risks of poor design as well as malfeasance. We need to take seriously the societal-scale risks of digital technologies. AI should be human-centred. Its use must be fair, transparent, trustworthy and safe. Organisations that use it should be made fully accountable for how they do so. The regulation of digital technology is an immediate problem with wide-ranging impacts, including on democracy and the global world order. The growing role of technology as a source of geopolitical and geoeconomic tensions will make these issues all the more difficult—and crucial—to resolve. There is some concern, particularly in Europe, that regulation will suppress innovation, but in the long run what people and corporations want is software that works on their behalf, that they can trust and that is secure. An important lesson for the regulation of AI/ML is to focus on domain-specific uses of the technology rather than on the technology per se. Regulators should look at particular uses of the technology with a view to clearly identifying and managing risks while trying not to prevent the beneficial use of the technology.

Digitally-enabled policymaking

There are many areas of public governance in which machine learning algorithms may be able to help to improve the accuracy, reliability, efficiency and overall performance with which decisions are made by public administrations and public services are delivered. Robust technology can help to underpin responsive governance systems. However, greater governance and more precise rules are needed. Particular concerns arise when the use of AI/ML tools is proposed in public systems that wield great power, such as policing or the court system. More generally, there are critical questions and risks around the extent to which algorithmic decision-making should ever be used in policy governing humans. Machine learning algorithms are trained on data, not on rules, so AI-based policymaking would be decided by algorithm instead of ruled by law. This means that decisions may be shaped by learned biases rather than legitimately and democratically established principles. A great deal of care is needed. Public debate should be encouraged about what the appropriate goals should be when AI/ML and other digital technologies are used by governments. Otherwise there is a substantial risk that increased efficiency will be the default goal, and this may lead to important adverse impacts being overlooked. This is an area where the trust and the legitimacy of the government are called on. Policymakers are advised to remember that when they make decisions about a particular technology, it is a decision about how to use a tool to accomplish what society wants.

The challenge of global governance

A uniform global system of governance for digital technologies may not be desirable, because diversity is valuable and redundancy is important. Existing national and regional principles already have significant global governance implications. For example, the EU’s data protection rules have explicit extraterritorial scope, and the whole world is affected by the rules of China’s firewall.

Welcomes and introductions

Marie-Valentine Florin, IRGC

IRGC Executive Director Marie-Valentine Florin began by welcoming everyone and explaining the role of the IRGC as a multidisciplinary centre that organises collaborative work on risk governance with the purpose of “connecting the dots between science and research on the one hand, and policymaking on the other hand.” Florin explained that the IRGC began working on this specific topic in 2018 looking at algorithmic decision-making, focusing on when there is no or little human supervision.

This conference was organised under the auspices of the EU Horizon 2020 TRIGGER project, which has as its goal to provide tools and guidance to the EU for global digital governance. The motivation for holding this conference as part of the TRIGGER project was to address the need to define clear rules and possible regulations to manage the risks caused by digital technologies, while also looking at the benefits brought by digital technologies. It is indeed possible to improve the performance of governance, broadly defined as ‘how decisions are taken in society’, by using such technologies. By ‘performance’, we mean: efficacy, cost-efficiency, accuracy, reliability. However, the counterpart of efficiency could be the loss of resilience, and the counterpart of accuracy could be discrimination, so trade-offs will have to be addressed, with a careful balancing of efficiency and resilience, and accuracy and discrimination, to protect fundamental rights. Therefore, we need to think about the ‘governance of technology’, and define clear rules and possible regulations, to manage risks caused by digital technologies. These rules must be based on important governance principles that are aligned with our value systems, including democracy, human rights, civil liberties, privacy, but also accountability, social justice and, overall, fairness. This will require a careful balancing of innovation and precaution/fundamental rights. The more we are governed by digital technology, the more we also need to think in terms of governance of that technology.

There has been lots of initiatives in the past two or three years to establish AI principles, which aim to be a reminder of the importance of values that support our societies: AI should be trustworthy, ethical, responsible, human-centric, beneficial, inclusive, sustainable, etc. Similarly, digital transformation should serve the goal of developing a society that is more ethical, responsible, inclusive, sustainable, etc. At the same time, there have been concerns raised about so-called ethics-washing, particularly in the private sector that may just put a normative/ethical wrapping around business as usual. There are some concerns that principles remain at the theoretical level, and that it is simply too late to take action to control AI in cases when outcome of its applications is not desirable. It is increasingly recognised that sector-specific governance or regulation is needed to embed these principles into the reality of domain-specific applications and software, whether the domains are regulated or not. However, it is important to be pragmatic, and to look at how individuals, corporations and public sector organisations behave in the practice, often not fully aligned with principles and regulation.

Andrea Renda, Centre for European Policy Studies (CEPS)

Andrea Renda, the coordinator of the TRIGGER Project and Senior Research Fellow and Head of Global Governance, Regulation, Innovation and the Digital Economy at CEPS, further explained the TRIGGER Project, that it is centred around the idea of governance and interrelations between EU governance choices and global governance. According to Renda, digital technology is a quintessential case of interaction between those governance forms.

He then offered a couple of examples of current work that speaks to both governance of and by technology. Looking at governance of technology, the European Commission's proposal for measures to boost data sharing and to support 'data spaces'[1], forthcoming at the time of the conference and published on 25 November, "could be a game changer in the way in which we look at data flows". The EC is also preparing the first-ever comprehensive proposal for regulating artificial intelligence. And regarding governance by technology, at the same time as this conference was taking place, the GAIA X initiative[2] was being presented in Brussels. This initiative to translate European legal rules into software code is an unprecedented attempt at governance by technology. A question that many of us are interested in is: will this enable Europe to gain weight in AI and other digital technologies, where the power of the US and China is striking and sometimes a source of concern when difference in value systems and culture may imply that Europe will lag behind?

James Larus, EPFL

James Larus is the Dean of EPFL's School of Computer and Communication Sciences along with the Academic Director of the IRGC. Looking back at this year, he said that if there were ever a year that justified having a risk governance centre at a university, that this was it. He also explained that while EPFL has many great scientists, researchers and engineers, that they do not often think about the risk issues associated with the technology they develop. In that way, the IRGC has brought a much broader perspective to the university.

[1] https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2102

[2] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

Session 1: Privacy, efficacy and the digital response to COVID-19

Larus moderated the first session, which explored trade-offs of privacy and efficacy, notably in the rapid roll-out of digital COVID-19 tracing apps used to notify people in case of exposure to the virus. The Swiss app uses the DP3T protocol, which was developed by a consortium of researchers led by EPFL and ETH Zurich. It is designed for use with both the Android and iOS (Apple) operating systems. Because the data are stored on users' phones rather than a central server, they are extremely difficult to hack[3].

Striking a balance between data privacy and effective machine intelligence for algorithm development Jeffrey Bohn, Swiss Re Institute

Jeffrey Bohn asserted that enterprise machine learning is still not reaching its potential and that valuable data remain unexploited, in part because we begin to see the limitations of AI and the difficulty for companies to fully embrace data science, and also due to lack of privacy-preserving calculation environments, which causes valuable data to remain unexploited. One example Bohn gave of this is that in the UK, only three percent of financial services companies are using machine learning to gain actionable insights from unstructured data. He dubs data as "the new oil" in its potential to transform society and said that multiple techniques are needed to address different privacy goals. Important in this is technology that prevents data from being changed, such as confidential computing and distributed ledger technology (supported by blockchain). In particular, we need to see more and better implementation of confidential computing, i.e. the various types of privacy-preserving computation techniques that exist for processing sensitive data (including differential privacy, secure multi-party computation, or fully homomorphic encryption, and trusted execution environment).

Looking specifically at COVID-19 and tracing apps, Bohn said that collaborative data-sharing presents complex choices for stakeholders, but is key to achieving much needed innovation and effective response. It is important to communicate with the public to fully explain the technology being used in the tracing apps. He asserted that data-driven apps offer the opportunity to better manage COVID-19 without necessarily having to go through a full shutdown as happened in spring 2020.

Data governance and privacy: lessons learned from the rapid-roll out of digital contact tracing Elettra Ronchi, OECD

Elettra Ronchi began by stating that data access and sharing have been critical to an effective response to COVID-19. Data can be used to understand and control the spread of the virus, improve the capacity of health care systems, evaluate the effectiveness of containment policies, organise vaccine distribution and evaluate effectiveness of the

[3] https://foph-coronavirus.ch/swisscovid-app/

vaccines. It is therefore important to evaluate whether data policies are fit for the purpose for which they are being implemented.

Two types of widely available data useful to monitor movements have been used for most digital tracking and tracing: geo-location data and close-proximity data. Different countries have been using a mix of these. In Germany for example, the main German telecommunications provider, *Deutsche Telekom*, has been providing anonymized “movement flows” data of its users to the Robert-Koch Institute, a research institute and government agency responsible for disease control and prevention.

Privacy has been a key concern for citizens and thus for developers of these apps, and countries[4] implementing them. Privacy enforcement authorities had a key role to play. The range of personal data that can be collected through a COVID tracing app is broad and many countries did not have the data governance frameworks in place for the kind of data collected by these apps.

For example, Switzerland had to introduce legal provisions in an amendment to the Epidemics Act in order to launch its SwissCovid app and Australia introduced a new law, which amends the Australian 1988 Privacy Act addressing some of the most pressing concerns about COVIDSafe and how its data is used. Under the new legal provisions, improperly disclosing data is punishable by up to five years in prison.

A list of privacy guidelines for different countries can be found through the Global Privacy Assembly[5].

Ronchi also brought up questions of efficacy – are contact tracing apps actually more efficient than manual contact tracing? Is there a certain threshold of uptake to make the digital apps effective? And would countries with low community spread have a reason to use an app at all? These questions have not yet been fully answered.

Ronchi concluded with some recommendations:

- That the digital response be necessary and proportionate (effectiveness and clear societal objectives).
- To limit data collection and data quality to what is necessary (fit for purpose).
- That there be a time limit on data retention.
- That there is transparency and accountability.
- That stakeholders are engaged in the process and users have control.
- That there is data protection and a privacy-by-design approach.

Privacy, infrastructure and the digital response to COVID-19 Michael Veale, University College London

Michael Veale began by exploring the allure of the precision of mobile interventions to avoid indiscriminate lockdowns, to make sure people observed quarantine and to measure flows of people. The most common use of mobile interventions, however, was tracing apps.

[4] https://public.flourish.studio/visualisation/2241702/

[5] https://globalprivacyassembly.org/covid19/covid19-resources/

Veale briefly outlined other potential forms of mobile data and explained why they weren't used. For example, GPS is imprecise and doesn't work indoors; cell phone data has surveillance issues; ultrasound data uses the microphone and also has privacy issues. In the end, Bluetooth low-energy was the best technology for tracing apps.

In order to create and deploy these apps, public-private partnerships are needed. Vertical integration in companies like Apple means that they have control over the hardware and software supported, as well as what kind of apps can be available in its app store.

Veale was involved in DP3T, the EPFL project to create the decentralised contact tracing apps such as SwissCovid. He explained that as part of the process to decide how the app would be built and function, they had to look at two main potential architectures for using Bluetooth: centralised and decentralised. With a centralised architecture, phones emit encrypted identity and only the central authority has the key database. Each phone records each encrypted identity it hears, and when a user tests positive, the phone sends the numbers received from others to the central authority to decrypt, identify and notify the other parties. With a decentralised architecture, phones emit random, constantly changing numbers and remembers all of those it emitted. Phones also listen for others' random numbers. When a user tests positive, the phone uploads all the random numbers it sent out to everyone else's phone, but never the ones it heard. The random "infected" numbers users receive are regularly compared to see which users were at risk. In this way, all the work is done between phones and no centralised authority is involved.

Centralised systems may cause privacy concerns because they allow central authorities to track people and could lead to runaway power. In contrast, platforms like Apple and Google only allow decentralised systems. This shows the trade-off between the power of platforms versus the power of central authorities, often controlled by governments. Platform power means that platforms determine the protocols that run on devices and have the power to analyse data privately. It also neuters the power of the general-purpose computer and has anti-competitive dimensions. State power turns citizens into sensors and actuators with great potential for abuse and loss of trust in the digital economy and society. There are also international dimensions – who sets the standards? And the age-old debate over allowing workarounds to end-to-end encryption.

Veale ended with some recommendations:

- There is a need for an international agreement for openness of computing systems.
- There is a need for a strong data and information regulator to ensure that this does not lead to misuse or insecurity.
- Governments need to support digital human rights in order to be trusted when they ask individuals to use their phones or other devices in concert with one another.
- Everyone loses when there is coercion from platforms or governments.

Session 1 Discussion

Larus moderated a discussion with Bohn, Ronchi and Renda. The discussion started with Covid-19 tracing apps, and the fact that some epidemiologists had been disappointed that the strong privacy protection of the DP3T system meant they were losing out on valuable epidemiological information. This raised the question of trade-offs between privacy and epidemiological goals, and how such trade-offs between two desirable goals should be dealt with. Participants stressed the need to educate the broader population on what technology can and can't do and get people more comfortable with privacy preservation. They noted that because we have been in an emergency situation, things have had to happen at scale and at pace out of necessity, because we did not have the time to wait and see. As a counter-argument, it was suggested the public has been very sensitive and responsive to everything going on with contact tracing technology, and that the suspicion of mission creep is justifiable, especially in more authoritarian countries.

As the discussion proceeded, it was suggested that Covid-19 contact tracing is a perfect example for those who want to consider the role of technology in society: technology is a means to an end and can be very powerful, but it is not always that best solution. Contact tracing apps present themselves as a potential panacea, but there is also the trade-off between the power of the technology and the privacy risks. Technology also has shortcomings and in isolation it is not sufficient; it needs a governance structure behind it to deal with the challenges that will inevitably arise: What happens when your phone tells you you've been exposed to a COVID phone? Do you go the hospital to get a test, or do you build a governance around it so there's a privileged testing lane for them? Is an exposure notice on a phone a good enough reason not to go to work? Technology is only a complementary means in a world of testing and tracing, which is not about technology.

In the Q&A, the panel was asked whether technology fundamentally changes the contact tracing process. In response, it was suggested that technology gives us a menu of choices, with the governance and regulatory response then helping us decide how to use them. The conversation then returned to privacy and educating the public, and the potential paradox of people saying they don't trust the privacy of Covid-tracing apps, while also having a lot of other apps on their phones with much weaker privacy protections, like social media or maps.

There was agreement among the panel that there's been a struggle with understanding privacy. It was suggested that to convince users of the efficacy and trustworthiness of tracing apps, it would have been helpful to already have a health data system in place, as in Finland, where there has been an ongoing conversation about sharing health data and about how that data is collected and stored. The discussion ended with a reference to Bohn's presentation, that in fact it is trust that is the new oil.

Keynotes

Governing AI: A Few Suggestions Stuart J. Russell, UC Berkeley

AI technology has come a long way since the early days and is moving quickly. Stuart Russell cited some examples of this point, such as the unexpected victory of the AlphaGo AI system[6] and the fact that we will likely have self-driving cars within the next decade. The benefits behind all this are pretty obvious – if you have human-level AI, you can have a better civilization. Russell cited the figure that there would be a 10-fold increase in world GDP with general-purpose AI – the net present value of that being $13.5 quadrillion.

But of course, there are downsides as well, such as issues of surveillance and control. Russell also highlighted autonomous weapons, making the distinction that while people often come to that issue through the lens of human rights (the inability to discern civilians from soldiers) or, from the speculative end, robots taking over the world, the real issue is the risk of deliberate mass slaughter. Autonomous weapons are scalable weapons of mass destruction (WMDs). We need a partial or total ban on these weapons.

Russell then discussed AI that impersonates humans through image and/or audio – deepfakes. The risk of impersonation is deception without overt fraud. And as humans, we owe different obligations to each other than we do to machines, and we should not be forced to fulfil those obligations to robots. Russell believes there should be a total ban on impersonation and that he has not heard a good argument against that.

Another risk of AI is the selection of social media content through algorithms. These algorithms have more power of propaganda than Hitler or Stalin ever did and influence people through most of their waking lives. The objective of these algorithms is to maximise clickthrough. Instead of learning what people like, these algorithms modify people to become the most predictable versions of themselves. Over time, this can lead to manipulation, the inevitable consequence of reinforcement learning with rewards that depend on human response. Russell suggested a possible ban on reinforcement learning in human interaction and that we need to have research trials with algorithms on human subjects.

AI offers enormous upsides, and we could make it work for individuals, to negotiate services, data exchange and organise privacy, if it can be adapted to individual preference, and help alleviate market failures. However, there are very real risks of poor design, and, as Alan Turing predicted in 1951, at some stage machines may take control unless a provably safe and beneficial foundation for AI can be constructed. Therefore, Russell concluded, we need to move quickly to establish basic rights to mental security[7] for humans and take seriously the societal-scale control risks.

Governing AI Made Easy Joanna Bryson, Hertie School

Joanna Bryson noted that Russell had discussed AI as a service whereas she would be discussing AI as a product, in such a way that the same ordinary obligations and legal concepts as we have for any product could be applied. Bryson also noted the importance of

[6] https://deepmind.com/research/case-studies/alphago-the-story-so-far

[7] the right to live in a largely true information environment

moving quickly on this topic, that it is an immediate problem with impacts on democracy and the global world order.

Bryson then defined some key terms:

- *Intelligence*: doing the right thing at the right time. It's a form of computation (not math) that transforms sensing into action. Requires time, space, and energy.
- *Agent*: any vector of change, e.g. chemical agents.
- *Moral agents*: considered responsible for their actions by a society.
- *Moral patients*: considered the responsibility of a society's agents.
- *Ethics*: the set of behaviours that creates and sustains a society, including by defining its identity. Ethics vary by society.
- *Artificial intelligence*: an artefact; built intentionally. Because AI is built intentionally, this intent equates to responsibility.
- *Responsibility*: a property that moral agents of a society assign to each other. Implies a peer relationship (as does trust).
- *Accountability*: a society's capacity to trace responsibility.
- *Transparency*: the means by which accountability is implemented. It is not an end in itself.
- *Trust*: is a relationship between peers where the trustee is not micromanaged and is allowed to defect, whether for pragmatic reasons or to allow innovation.

Her conclusion based on these definitions is that, "We should not trust large powerful agencies, but rather construct systems to hold them accountable."

Bryson then showed the OECD Principles of AI.[8] In summary, AI should be human-centred, fair, transparent, safe and accountable. Noting that 44 governments and G20 have already endorsed them, she suggested that "we can stop building stuff (new principles) now". The talk then moved to regulation, which Bryson defines as: "The means by which a complex entity perpetuates a recognisable version of itself into the future." She also noted that most regulation for companies is positive through tax cuts. She then defined governance as: "explicit, deliberate regulation".

When it comes to AI, what we regulate is not the micro details of how AI works, but how humans behave when they build, train, test deploy, and monitor it. It should be easier for digital systems to have transparency. Software companies in general are already adept at tracking their processes, but AI companies so far have not been nearly as good with their development and operations (devops).

Good (maintainable) systems engineering of software requires architecting the system, securing the system (including logs), documenting every change, and logging testing both before and during release.[9]

She then moved on to governments, asking the rhetorical question of "can we trust governments" and immediately answering "no". However, governments are a principle means to ensure that that everyone does their fair share. Rather than trust governments, we

[8] https://www.oecd.org/going-digital/ai/principles/

[9] cf Bryson, Diamantis & Grant 2017; Bryson & Theodorou 2019, Bryson OUP 2020

should be vigilant, demand they are transparent, and hold them accountable. She agreed with Russell’s assertion about weapons and mass genocide.

Bryson ended looking at different countries. Emphasising as false the recent narrative that China and the USA are locked in a binary cold war, Bryson showed that in terms of IP, the EU is ahead of China. Also, the rest of the world outside of the EU, China and the USA is ahead of China and the EU combined. And finally, the USA has more AI IP (and more market capitalisation in companies holding AI IP) than everywhere outside of the USA combined, so it certainly doesn't need protection from China by the EU. She emphasised that neither patents nor large market capitalisation show the strength of robust economic approaches like strong SMEs. She described the United States as “libertarian market logic, limited regulation”, China as “state capitalism, government directs actions of companies”, and Europe as “social market economy, multi-stakeholder models", with the GDPR as first step toward regulating Big Tech”.

From the keynotes, moderator Renda then took the programme into the next session, where Russell and Bryson, along with a group of new speakers, participated in a roundtable.

Session 2: Governance Of Technology - the challenges of regulating machine learning

Governing AI: Understanding the limits, possibility, and risks of AI in an era of intelligent tools and systems John Zysman, UC Berkeley

John Zysman began by referring back to the keynotes by Russell and Bryson, saying that he agreed largely with what they had said. He continued that AI applications are part of a suite of intelligent tools and the regulation of big platforms and data is part of the challenge of governing AI. Focussing on AI technology alone distorts the governance problem.

AI governance, at least for now, is more about managing the people who would create and deploy the technologies, than about managing super intelligent systems. One problem is that the application of AI tools for public systems, for example waste disposal, public transportation or policing, requires a great deal of care because of the substantial risk of confusing efficiency with a public debate about what the goal should be in the first place. The public values purported to be addressed by these AI systems are themselves a result of political and social conflict.

He also asserted that the economic implications of AI systems are easily exaggerated. This brings up the question of whether public funding should go to basic research or towards dissemination of the tools and training needed to implement the AI systems.

On global governance, he noted that as difficult as it will be to identify and implement the goals of AI systems in one community, it will be even more difficult to do so internationally. Any global agreement that goes beyond simple objectives and statements is unlikely to produce operational and implementable rules. We should rather pursue different interoperable systems instead of a single system.

Discussion

In a roundtable discussion following Zysman's remarks, a question was raised about whether there could truly be any international agreement on AI and if so, what its focus would be. The recommendation was to start with narrow, sector-specific goals and systems. For example, what do we want to do about retail? What do we want to do about policing? If a system recommends the wrong colour shirt, that's not so important, but if it hands out the wrong punishment for a parking violation, that is a much bigger problem. Framing the issue of how to manage AI on a sector by sector basis could be scalable globally.

Why the approach taken in the EU White Paper on AI is incorrectly described as 'risk-based' Karen Yeung, University of Birmingham

Karen Yeung's talk looked at the EU White Paper on AI. The White Paper proposes a risk-based approach to AI, but Yeung argues that the way it deals with risk is too simplistic. The principles underpinning a risk-based approach to regulation are well established. This approach means that the level of protection should be proportional to the level of risk. Red lines are possible in a risk-based approach, meaning certain risks that are seen as societally unacceptable, such as autonomous weapons, should be completely prohibited. And activities with very low levels of risk do not need to be regulated and worried about.

Yeung supports a risk-based approach in general, but argued that the White Paper’s approach cannot be appropriately described as such. Instead, it adopts a binary approach, meaning that AI applications identified as “high risk” are ex ante subject to scrutiny and everything else is considered to be no risk and subject to no scrutiny, is not sufficient. Moreover, it adopts an unacceptably narrow understanding of ‘risk’, ignoring collective risks altogether, and focusing almost exclusively on risks that have ‘legal or other similarly significant effects. And if individuals can ‘avoid’ these risks, then the White Paper regards them as the appropriate risk-bearers. This is seriously inadequate. She also said that the White Paper fails to grapple with another central cause of concern, namely people being at risk from the effects of AI without understanding what they’re being subjected to and what is happening. These are risks generated by automated systems more generally, but while they are not specific to AI, AI makes them more complicated.

Discussion

The discussion began with comments about risks being ignored and this led to the question of how to decide what is risky. This can be a subjective concept. The proportionality principle was mentioned, and it was suggested that if we can hold corporations liable for the amount of damage they do, then it would motivate corporations to regulate themselves and avoid risky behaviour. One panellist countered that while the proportionality principle makes sense in theory, the challenge is how to operationalise it in the real world.

The panel then discussed the five-level risk-based system of regulation that the German Data Ethics commission has called for, ranging from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones. This was noted as a more useful way to operationalise the precautionary principle compared to the binary approach that Yeung had described in the European Commission white paper.

It was argued that we need more ex-ante measures to deal with AI and that this raises problems with common law, which operates after the fact. However, there was pushback against the idea that there should be an AI testing centre where a developer would have to send their AI and some data to check that things are correct. One of the panellists described this sort of view of AI as technically, societally and commercially misguided.

Technical robustness and safety of AI based systems as a means for their governance Raja Chatila, Sorbonne University

Raja Chatila also questioned the binary system and pushed back against some examples from the White Paper, such as one that said that risk of flaws in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. He then quickly highlighted some of the many sectors where AI is used, such as healthcare, transportation, justice, public sector, security, insurance, finance, recruitment, management, personal services and assistance, and warfare, and reminded that adoption of a technology is founded on trust, and AI is no exception. Trust is founded on safety, transparency and governance. Governance requires technical as well as non-technical means. Looking at technical robustness and safety, Chatila brought up the idea of dependability, that the delivery of service that can justifiably be trusted.

Dependability of AI systems means that the system is available, reliable, safe, confidential, secure and demonstrates integrity. Justification is granted through verification, validation and explainability. Explainability is less clear and newer than verification and validation. With AI systems, the general belief is that it's a black box, and it isn't actually that nor should it be. Explainability depends on the target audience: what are we explaining, for whom are we explaining it, and why are we explaining it?[10]

Chatila concluded by stating that transparency is a means for governance. There must be traceability of the data and design, auditability of the AI system and design process, communication about and by the AI system, and certification. Governance requires a regulatory framework.

Discussion

Chatila's talk brought up a chicken/egg question: If you want to govern digital technology, you need to modernise the tools of governance by using those digital tools. But if you want to use digital technology in governance, you need to build a regulatory framework. So which comes first? One response was that we need a priori frameworks. We need to define a law to make that law applicable. Proven technologies need to be used to assess new technologies. When the governance is implemented well, then the technologies are there to assess. Governance is about predicting, not just reacting.

The discussion moved on to consider the international dimension of digital technology governance. It was suggested that we already have global governance in a way. For example, the entire world must respect the EU's GDPR, and likewise it must respect the rules of China's firewall. In any case, it was argued, we probably don't want a single system of governance because diversity is valuable and redundancy is important.

One participant stated that people worry that EU regulation will suppress EU innovation, but suggested that in the long run, what people and corporations want is software that works on their behalf, that they can trust and that is secure. This led to the suggestion that corporations in the EU should start building things people want, and then people around the world will want to buy-in too. Another participant argued that what matters are clear and stable principles, noting that the principle "let's not exploit people unjustly" would be a good place to start.

Returning to the relationship between regulation and innovation, it was suggested that regulation often sets conditions that can actually drive innovation, rather than stifle it. An example cited was the struggle over the ozone layer and the radical shift in technology that came out of the regulations put in place to protect it. This drew some agreement from the rest of the panel, but with the caveat that improperly designed regulation can stifle innovation.

[10] For more information on explainability, Chatila cites this paper:
Alejandro Barredo Arrieta, et al, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 2020. https://arxiv.org/abs/1910.10045

Governing AI Ecosystems Bernd Stahl, de Montfort University

Bernd Stahl leads the SHERPA (Shaping the ethical dimensions of information technologies – a European perspective) project[11], which develops ethical approaches to AI technologies. He briefly summarised the problem with governance of AI, saying that one key issue is the complexity of the discourse.

There are many different ways to regulate AI and many different stakeholders as well. This makes it clear that there will be no one-size fits all answer for getting the ethics of AI right. The SHERPA project looks at the ethics of AI discourse from the theoretical perspective of innovation ecosystems. This perspective allows to identify relevant characteristics of ecosystems and applying them to AI, for example the concept of boundaries, the variety of different members with different interactions and the fact that ecosystems change over time.

If we accept that AI is a set of overlapping and interlocking ecosystems, then the question should be what can we do in order to govern AI ecosystems in such a way that promotes human flourishing. There are certain requirements that must be met in intervening in these ecosystems to make them successful. The first has to do with the boundary limitations of the ecosystem: where are we aiming the intervention? These boundaries can be geographical, conceptual or technical. The second is that for these ecosystems to promote human flourishing, it must have some level of knowledge input. Then there must be adaptable and flexible governance structures. In addition, it is important that stakeholders are involved.

Discussion

During the discussion following Stahl's remarks, it was suggested that to create regulations that promote human flourishing, there must be a central node somewhere that brings the conversation together. And there must also be ways that these ideas can be taken up at the organisational level. This prompted some discussion about the kinds of institutional arrangements that would be feasible for AI, and the possibilities raised included a body analogous to the FDA, an ethics board for AI, and a potential role for the UN in organising international deliberation in this field.

Difficulties in regulating emerging and rapidly evolving digital technologies: Deepfakes Kelsey Farish, DAC Beachcroft

Kelsey Farish, a media and tech lawyer, went in-depth on one specific area of AI regulation: deepfakes, deliberate manipulation of images and audio that look and impersonate real individuals. She began with the example of Scarlett Johansson, a famous actress whose likeness has been used many times in deepfakes, including pornography. However, when Johansson was asked about taking legal action, she said it would be a useless endeavour.

This is because, as Farish explained, the laws are incomplete to tackle deepfakes. The law is slow to evolve and lawmakers and judges don't know tech well enough. While existing laws, such as against image-based sexual abuse, or protection of IP / copyrights and data / privacy (EU GDPR) could apply in some cases, there is still a lot of grey area, and problems of

[11] www.project-sherpa.eu

implementation and enforcement, along with technological and disruption loopholes that people can use to get around the rule. In addition, it is hard to define the emotional and reputational harm that can be caused by deepfakes, and personal rights need to be balanced against societal rights and freedoms.

For now, it is up to the industry mostly to self-regulate, leaving it to platforms such as Instagram and Facebook to request takedowns or ban deepfakes, but this is often in name only. In addition, these are complex and systemic issues that go beyond the internet ecosystem. For example, principles of expression and free speech matter, including for creating social cohesion, and must be balanced with the risk of dis- and misinformation. The many different stakeholders with competing interests, lobbying and industry influence, contribute different views and opinions, noting for example that the technology used for deepfakes can be used for lots of other legitimate activities such as in the movie industry.

Discussion

The discussion began by considering the idea that there should be a total ban on impersonation. Difficulties were highlighted in keeping such a ban from butting up against legitimate forms of impersonation like satire and political impersonation. One participant argued that while a total ban might be overly broad, that doesn't mean we should do nothing, because even doing nothing is doing something.

The tendency to anthropomorphise AI was raised, and was dismissed as fallacious and a cause of legal and personal errors. Another participant noted that all of these issues related to AI are contributing to trade and geostrategic conflicts that make it both more crucial and more difficult to resolve problems like these.

The discussion ended on the need for a stable governance framework. It was mentioned that there is a clear and successful principle in consumer law in Australia that misleading conduct is prohibited. This type of framework could be useful for distinguishing between legitimate and illegitimate personalisation. More generally, it was suggested that there is a need for future-proof principles for governing AI, and a mindset that is the antithesis of "move fast and break things".

Session 3: Governance With and By Technology – digitally enabled policymaking

The context of the final session is that, although a deep digital transformation is going on in many areas of life, the impact in the public sector has been relatively modest. After a session that discussed governing AI and digital technology applications, panellists discussed whether greater use could or should be made of digital technologies in developing and implementing public policy. They considered the extent to which ICT-enabled innovation can transform governance and policymaking, and under which condition is such transformation possible and desirable. Europe may wish to regulate the digital transformation but Europe may also wish to digitalise public services more than what is currently done. This could have a big potential for Europe, within Europe and outside of Europe, provided it is done *well*.

Digital Europe 2040: AI & Public Sector Innovation in a Data Driven Society
Gianluca Misuraca, Danube University Krems

Gianluca Misuraca's talk built upon findings from recent research he has been leading at the European Commission's Joint Research Centre on 'Exploring digital government transformation in the EU'[12] and AI Watch for the Public Sector[13].

He started his presentation by outlining possible future directions for digital transformation of governance at the horizon 2040, which revolve around the following scenarios:

1. Apathy and closed innovation – unregulated landscape and passive citizenry
2. Trust and open innovation – unregulated landscape and active citizenry
3. Fear and surveillance – regulated landscape and passive citizenry
4. Precaution and inclusion – regulated landscape and active citizenry.

Misuraca also highlighted some of the results of the AI Watch report on Mapping of AI use and impact in public services in the EU[14] and underlined that public administrations are generally not exploiting all the possibilities offered by AI. Technologies most used are chatbots, predictive analytics, computer vision and expert rules-based systems. Applications are mostly oriented to supporting service provision and engagement (38%), enforcement (20%) and internal management (20%). Important challenges remain in moving from data analysis to end-user adoption of AI for public services. Different European countries are currently in different phases of their work on national strategies for policy actions for AI in the public sector, with a high heterogeneity of adoption and use of AI.

He concluded by laying out some open issues and policy implications:

- There is a high level of heterogeneity of AI use across EU and the public value created is unclear.
- Governments are searching for a "holy grail" of best practices that can be replicated.

[12] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/exploring-digital-government-transformation-eu-understanding-public-sector-innovation-data

[13] https://knowledge4policy.ec.europa.eu/ai-watch/topic/ai-public-sector_en

[14] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/ai-watch-artificial-intelligence-public-services

- There is a need to ensure the path to institutionalise AI into mainstream services beyond always having pilots.
- There is little evidence of what works and what is actually threatening services quality.
- There are varying scope and depth of strategies to develop and adopt AI for the public sector.
- Innovative Public Procurement & GovTech are crucial for adopting AI solutions
- The public sector has a multi-pronged role in governance “with, of and by” AI.

In Misuraca’s opinion as a European citizen, success in the European Union would be to ensure that the services that are enabled by technology are human-centric, serve the purpose of the citizens, respect our globally-agreed values and at the same time support achieving the Sustainable Development Goals.

The EU approach to blockchain/DLTs Peteris Zilgalvis, European Commission

Peteris Zilgalvis began his talk by explaining that the EU is moving to implementation of regulatory policy, starting with the evidence as they should. This is through public consultation such as with the FinTech action plan[15] and the European Union Blockchain Observatory and Forum[16]. They then moved to the next stage, which was a ministerial declaration on blockchain signed by all 27 EU states plus Norway and Liechtenstein to build a European blockchain services infrastructure, which is working as a regulatory sandbox.

According to Zilgalvis, this initiative and technology will give control of identity back to the individual citizen, offer diploma certification across borders, authentication and publication of audit documents, regulatory reporting and work in the standardisation area. This is all in close cooperation with European and international stakeholders. In addition, after evidence gathering and reflection over the last seven years, they created the digital finance package for markets and crypto-assets.

It is also important to build advanced digital skills for European citizens and build skills among people who will utilise blockchain technologies. All of these actions are ways that the European Commission has brought into reality a concept of law and political economy of decentralised digital ecosystems to make sure these technologies can be adopted, adapted and produced in Europe. Blockchain is an ideal technology for multilevel governance systems such as the EU. It is not a technology that will solve everything, but when it comes to AI and the Internet of Things (IoT) and other technologies in the data economy, it can implement trust in data sharing where there is a lot of data being used.

[15] https://ec.europa.eu/info/publications/180308-action-plan-fintech_en

[16] https://www.eublockchainforum.eu/

When and why machine learning should never, ever be used in governance
Bryan Ford, EPFL

Bryan Ford's strong assertion is that AI has no role to play in policy governing humans.[17] He went on to explain the difference between mechanism and policy. Mechanism represents a toolbox of technologies and processes that are usable in many different ways, and that are generally oblivious to how they are used, for example identifying a house or a tree in a photo. Policy determines the right way to use the tools we have towards beneficial ends, for example for choosing which neighbourhoods need more police surveillance. According to Ford, AI policymaking is rule by opaque algorithm instead of rule by law, rule by learned biases instead of rule by principle.

There is sometimes an assumption that algorithms come from "good, pure" datasets, but these datasets come with bias from the past, meaning that instead of being ruled by what's right for the future, we are ruled by our past. AI guardians can never be human. As other speakers said earlier in the conference, the concern isn't about "robot uprisings" but rather more mundane opportunities we'll have to shoot ourselves in the foot.

Ford emphasised that we must confine the uses of AI to mechanism, not policy. "We should never allow our governments or their employees to abdicate their responsibilities to govern us as humans…Influence by AI on policy is just as pernicious as foreign influence on a sovereign nation's policymaking, and we should make it just as illegal. There are many powerful and justifiable uses of AI, but we must confine these uses to mechanism, not policy."

Session 3 Discussion

The discussion began with a reminder that when policymakers need to make decisions about a technology, it is a decision about choosing a tool for what we want to accomplish. We need to figure out how it can work for us and avoid the hype. Ford's presentation prompted a lively debate and among the responses of other panellists was the suggestion the potential risks of using AI could be offset by the opportunities it brings to governments and the benefits it would bring to citizens. The moderator brought up the example of predictive policing and asked whether it could ever be desirable, to which Ford responded that this is exactly the kind of thing we should not do with AI.

The discussion moved on to consider blockchain and AI, and it was stated that from a governance perspective, blockchain is a much safer technology built around transparency. Whereas blockchain starts from the point of transparency and then modifies it just enough for privacy, AI on the other hand comes from a place of non-transparency and then makes it a little more transparent.

On the potential uses of AI in the public sector, it was suggested that AI can help in areas such as social services or employment policy, but subject to ensuring the data are correct and well-structured. This is something where the trust and the legitimacy of the government

[17] This talk was taken from Ford's blogpost "AI for Governance Belongs in Mechanism, Not Policy", which can be found here: https://bford.info/post/2020-11-18-ai-governance/

is called on. This is an area with a real challenge and problem, because algorithms can produce unbiased outcomes only if they are trained on good datasets, which is very difficult. A possible place where blockchain can help is bringing together different sources of data not to one database but keeping them decentralised while still offering access.

The question of ownership arose, and in particular whether it matters if governments buy the technology, outsource to subcontractors, or develop their own solutions. In response, the example of a central bank digital currency was given: the central bank won’t sell data on its citizens but a private company likely would. It was noted transparency is what matters, and that both governments and companies can be transparent or not. Research was cited saying that developing the technology in-house can increase the uptake and legitimacy, but it depends on the sector.

Conclusion

Renda came back to congratulate everyone on the conference. He added that we have learned some of the intricacies of digital technologies, perhaps even beyond the “of” and “by” from the title, and in terms of what proposals could be made to draw boundaries such as what Ford mentioned. Do you delegate your decisions to this digital technology, use technology to help guide decisions, or eschew it entirely?

Florin concluded by relating the conference back to the TRIGGER project and picking out four takeaways, as four pillars of digital technology policy to be achieved: (1) Protect people, for example with the EU GDPR, (2) Support data sharing, for example with EU Data Strategy, (3) Build confidence, perhaps through a Digital Service Act and (4) Find a common goal, suggesting that at the moment, the digital transformation may be a tool to an end that has not been collectively defined yet.

Acknowledgments

IRGC thanks the following individuals who participated in the Governance Of and By Digital Technology conference, 18 November 2020: **Jeffrey Bohn,** Swiss Re; **Joanna Bryson,** Hertie School; **Raja Chatila,** Sorbonne University; **Kelsey Farish**, DAC Beachcroft; **Bryan Ford**, EPFL; **James Larus**, EPFL; **Gianluca Misuraca**, Danube University Krems; **Andrea Renda**, CEPS; **Elettra Ronchi,** OECD; **Stuart Russell**, UC Berkeley; **Bernd Stahl,** DMU; **Michael Veale,** UCL; **Karen Yeung,** University of Birmingham; **Peteris Zilgalvis;** European Commission; **John Zysman,** UC Berkeley.

Additional thanks to members of the TRIGGER project and to Hugo Hueber for technical support.

The views and recommendations contained in this report do not necessarily represent the views of individual workshop participants or their employers.

The writer of these proceedings was **Stephanie Parker** with contributions from **Marie-Valentine Florin** and **Aengus Collins**. Responsibility for the final content of these proceedings rests entirely with IRGC.

Appendix 1: List of Speakers

1. **Jeffrey Bohn**, Chief Research & Innovation Officer, Swiss Re Institute
2. **Joanna Bryson**, Professor of Ethics and Technology, Hertie School
3. **Raja Chatila**, Professor Emeritus of Robotics, Artificial Intelligence & Ethics, Sorbonne University
4. **Aengus Collins**, Deputy Director, IRGC
5. **Kelsey Farish**, Lawyer (Solicitor, England & Wales), DAC Beachcroft
6. **Marie-Valentine Florin**, Executive Director, IRGC
7. **Bryan Ford**, Associate Professor, Decentralized and Distributed Systems Lab EPFL
8. **James Larus**, Dean, EPFL School of Computer and Communication Sciences
9. **Gianluca Misuraca**, Research Fellow on eGovernance and Public Administration, Danube University Krems
10. **Andrea Renda**, Senior Research Fellow and Head of Global Governance, Regulation, Innovation & Digital Economy, Centre for European Policy Studies (CEPS)
11. **Elettra Ronchi**, Senior Policy Analyst, OECD
12. **Stuart Russell**, Professor of Electrical Engineering and Computer Sciences, UC Berkeley
13. **Bernd Stahl**, Director of the Centre for Computing and Social Responsibility, de Montfort University
14. **Michael Veale**, Lecturer in Digital Rights and Regulation, University College London
15. **Karen Yeung**, Interdisciplinary Professorial Fellow in Law, Ethics and Informatics, University of Birmingham
16. **Peteris Zilgalvis**, Head of Unit of Digital Innovation and Blockchain, European Commission
17. **John Zysman**, Professor Emeritus of Political Science, UC Berkeley

Appendix 2: Programme

14:00 **Welcome and Introductions**

- Marie-Valentine Florin, EPFL; Andrea Renda, CEPS; Jim Larus, EPFL

14:20 **Session 1: Privacy, efficacy and the digital response to Covid-19**

Data protection is at the heart of debates about the governance of digital technologies, but concerns have been raised about hampering innovation. This session will assess the relationship between privacy and technological efficacy, drawing on the recent development of Covid-19 contact tracing apps.

Moderation: Jim Larus

- Jeffrey Bohn, Swiss Re: *Striking a balance between data privacy and more effective machine intelligence for algorithm development*.
- Elettra Ronchi, OECD: *Are there lessons about governance of and by technology to be learned from the rapid roll-out of digital contact tracing apps as part of governments' urgent public health response to Covid-19*?
- Michael Veale, University College London: *Privacy, infrastructure and the digital response to COVID-19*

15:10 **Keynotes:**

- Stuart Russell**,** UC Berkeley: *Governing AI: A Few Suggestions.*
- Joanna Bryson, Hertie School: *Governing AI Made Easy.*

15:40 **Session 2: Governance of technology: the challenges of regulating machine learning**

The increasing role played by machine-learning algorithms in a growing range of decision-making processes raises legal, technical and ethical challenges. In this roundtable session, participants will discuss the priorities, constraints and trade-offs that policy-makers face in the regulation of machine learning.

A roundtable discussion with Andrea Renda (facilitator), Stuart Russell, Joanna Bryson and:

- John Zysman, UC Berkeley: *Governing AI: Understanding the Limits, Possibility, and Risks of AI in an Era of Intelligent Tools and Systems.*
- Karen Yeung, University of Birmingham: *Why the EU White Paper's approach is incorrectly described as 'risk-based'.*

- Raja Chatila, Sorbonne University: *Technical robustness and safety of AI based systems as a means for their governance.*

- Bernd Stahl, de Montfort University: *Governing AI ecosystems*.

- Kelsey Farish, DAC Beachcroft: *Difficulties in regulating emerging and rapidly evolving digital technologies – as illustrated by deepfakes.*

16:50 **Break**

17:00 **Session 3: Governance with and by technology: digitally enabled policymaking**

Given the transformation that new technologies have unleashed in many areas of life, the impact in the public sector has been relatively modest. This session will discuss whether greater use could or should be made of digital technologies in developing and implementing public policy.

Moderation: Marie-Valentine Florin, EPFL

- Gianluca Misuraca, Danube University Krems: *Digital Europe 2040: AI & Public Sector Innovation in a Data Driven Society.*

- Peteris Zilgalvis, European Commission: *Implementing the EU Blockchain Initiative*

- Bryan Ford, EPFL: *When and Why Machine Learning Should Never, Ever Be Used in Governance.*

EPFL International Risk
Governance Center

EPFL IRGC
Station 5 BAC (Bassenges)
1015 Lausanne
Switzerland

+41 21 693 82 90

irgc@epfl.ch