Agile Approaches for Governing Emerging Technologies

Conference report

Conference jointly organised with the OECD Directorate for Science, Technology and Innovation, and the OECD Directorate for Public Governance

Paris Dauphine-PSL University, December 3, 2024

DEBATE

Dauphine | PSL
CHAIRE GOUVERNANCE ET RÉGULATION

Conference jointly organised with the OECD Directorate for Science, Technology and Innovation, and the OECD Directorate for Public Governance

December 3, 2024

Synthesis n°96
Paris Dauphine-PSL University

Agile Approaches for Governing Emerging Technologies

Speakers

Daniel Andler | Paris-Sorbonne University, France

Vladislava Bar-Katz | Frontier Economics, UK

Virginia Dignum | Umeå University, Sweden

Erik Fisher | School for the Future of Innovation in Society and the Consortium for Science, Policy and Outcomes, Arizona State University

Mar Gonzalez | OECD ENV Directorate

Tammy JihHsuan Lin | National Chengchi University, Taiwan

Lyric Jorgenson | National Institutes of Health, USA

Jane Kaye | University of Melbourne, Australia & University of Oxford, UK

Pierre Larouche | University of Montreal, Canada

Brice Laurent | Centre de Sociologie de l'Innovation Mines Paris, France

Peter Mills | PHG Foundation, UK

Béatrice Oeuvrard | META, France

Daria Onitiu | Oxford Internet Institute, UK

Roland Ortt | Erasmus University Rotterdam, Netherlands

Howard Shelanski | Georgetown Law Faculty, USA

Pawel Swieboda | International Center for Future Generations (ICFG), Belgium

Anike Te | Lucideon & University of Bristol, UK

Isabel Webb | Department for Science, Innovation and Technology, UK

Claudia Werker | Delft University of Technology, Netherlands

David Winickoff | Head of the Responsible Innovation Unit, OECD

Moderators

Miguel Amaral | OECD

Éric Brousseau | Governance and Regulation Chair

Becky King | OECD

Daniel Nadal | OECD

Douglas Robinson | OECD

Joëlle Toledano | Governance and Regulation Chair

Summary

Session 4: The role of ethics in guiding the development of emerging technologies

Moderation: **Daniel Nadal** | OECD

Session 5: Using soft law to bridge business and government objectives on emerging technologies

Moderation: **Joëlle Toledano** | Governance and Regulation Chair

Douglas Robinson | OECD
Miguel Amaral | OECD
Becky King | OECD
Daniel Nadal | OECD
Joëlle Toledano | Governance and Regulation Chair

Introduction

Éric Brousseau | Director of the Governance & Regulation Chair and of the Club of Regulators, Dauphine Paris-PSL University

David Winickoff | Head of the Responsible Innovation Unit, OECD

Éric Brousseau

I would like to welcome you all to this conference, which is organised with the OECD Directorate for Science, Technology and Innovation, the OECD Directorate for Public Governance, and the Governance and Regulation Chair at Paris Dauphine-PSL University. I would particularly like to thank our speakers, some of whom have travelled significant distances to be here, and the session moderators.

David Winickoff

Thank you very much for this valuable collaboration. In my role at the OECD, I lead on the secretariat side of the working party for bio, nano and converging technologies. This group, created in 2015, addresses emerging tech policy issues from a range of perspectives, including regulation, innovation and governance. We greatly value our collaboration with Dauphine, which is an academic leader in agile regulation technology, and with the OECD Public Governance Directorate, who are represented here today.

This conference brings together the regulatory community and the innovation science policy community with a view to breaking down silos and recognising the recurring issues that occur at the intersection of innovation policy and regulatory policy. These issues are highlighted in the OECD's new framework for anticipatory governance of emerging tech. Innovation processes are unpredictable, and for good reason.

Thank you all for being here today and to Éric and his team for organising this session.

Éric Brousseau

Innovation processes can be unpredictable. On the supply side, technology is designed by innovators and industrial firms to address supposed needs; on the demand side, users discover the potential of the technology, adapt it to their needs, and discover ways in which it can or should change. This ongoing process of design, adoption, adaptation and redesign triggers many challenges for governance and regulation, including a need for anticipation and foresight, observation and analytical capabilities, and agility to continuously adapt regulations and policies in line with the trial-and-error nature of the innovation process and assess the impact of alternative regulatory practices. It is necessary to try soft tools before moving to hard law, and to co-design solutions within multi-stakeholder forums while avoiding the risk of capture. Regulation must differentiate between fundamental advances in science and technology, including AI, and specific implementation in various contexts (e.g. healthcare, mobility, human resources management, etc.).

David Winickoff

Our first session will consider forward-looking approaches for emerging tech governance, particularly how different approaches can enhance capabilities and capacities to govern well. We will then consider governance along the innovation journey and policy cycle and the need to integrate policy and innovation development processes more robustly. The third session will explore the issue of technology convergence and its implications for governance and regulation before we turn to the role of ethics as an important resource for governance and technology that engages with approaches based on hard law. Finally, we will look at how soft law can be used to bridge business and government objectives on emerging tech.

Session 1
Forward-looking approaches for emerging technology governance

Moderation: Douglas Robinson | OECD

Douglas Robinson

Emerging and converging technologies bring high uncertainty and complexity and require agile and responsible governance. Decision makers in government and other organisations are challenged to make choices in spite of inherent uncertainties, complexities and trade-offs.

We must consider how evidence or intelligence about these technologies is produced, as well as their potential impacts on society and the economy. Some of this work has already been done in the field of responsible research and innovation. The UK, for example, requires an anticipatory, reflexive approach that takes an inclusive and responsive attitude to multiple possibilities over the near-, mid- and long-term, is explicit about the frames and biases used in analytical processes, and is responsive and timely relative to the overall decision-making process.

The role of public expertise in regulatory prediction and foresight

Brice Laurent | Centre de Sociologie de l'Innovation Mines Paris, France

Science studies explores foresight and perspective, primarily in terms of how people and the instruments and device they use shape the future. Successful people, instruments and methodologies are those that manage to imagine a future that comes to pass. When foresight and prediction are viewed in these terms, it becomes clear that this is a political process. We must be aware of who creates this future, what happens in it, and how it can be made responsible, fair, inclusive and democratic.

It is interesting to consider the role of public expertise in prediction and foresight. I lead the social sciences department at ANSES, the French public health agency in charge of risk assessment. This department was established to integrate socioeconomic analysis into risk assessments, including the analysis of policy measures and their implications. This might involve assessing the risk of a chemical and the potential implications of regulating, labelling or even banning it.

The example of new genomic techniques (NGTs) and their applications provides a good sense of what is involved and how public expertise can contribute to debates about the future. Rather than inserting a gene from one plant into the genome of a target plant to obtain a desired characteristic, NGTs directly edit the genome of the target plant itself. The regulation of NGTs, specifically the legal question of whether or not they are genetically modified organisms (GMOs), is currently being discussed in Europe. The position of the European Court of Justice is that NGTs are GMOs and are subject to specific requirements around labelling and risk assessment. Other parties argue that, to promote innovation, NGTs should not be regulated as GMOs.

ANSES was asked to consider whether adaptations could or should be made to the current regulatory requirements for NGTs, and to document and analyse the socioeconomic implications and issues and implications associated with NGTs. It released a report in 2024 that contains expert opinions that are particularly relevant to our discussion today because they do not deal with socioeconomic issues by trying to predict future economic values, future benefits or future costs. Instead, the report discusses concerns raised by different actors and explores how they might play out in different scenarios. For example, a scenario where NGTs are not treated as GMOs might create issues with organic farming or traceability.

The report is intended to provide technical expertise and risk assessment only. It does highlight, however, that these extremely technical, extremely regulatory policy decisions ultimately determine the kind of society in which we live. The experts lay out the issues and the concerns and map their broader relationship to society as a whole with a view to ensuring that the right questions are asked during the public debate, that the right issues are identified, and that stakeholders can identify different options, their potential impact on different parties, and what is at stake more broadly.

This example is interesting because it highlights the role of public expertise in situations that are uncertain or where future developments must be anticipated. It shows that, rather than providing definitive answers and solutions, experts can play an important role in bringing issues to the fore, identifying what matters and for whom, and exploring the roots of disagreement and opposition. Decision-makers might want public experts to provide definitive answers, but it is perhaps more relevant for them to recognise that some issues go beyond purely technical or regulatory matters and are worthy of public debate. There are methodological and institutional challenges involved in developing methods that balance the need for openness and debate with the range of different concerns and possible futures, but this challenge merits attention if we want to use scenario making and foresight in fruitful and productive ways.

Policy development in the United States: objectives and examples

Lyric Jorgenson | National Institutes of Health, USA

Innovation is unpredictable and innovation policy and regulation involves many complex inputs and factors. In general, we aim to adopt proactive measures that cover the basics before using horizon scanning to identify likely future developments, understand how they might relate to the existing policy landscape and determine how our policies might need to be adapted. It is important to assess the landscape and identify frameworks that will promote and protect new technologies, allowing them to flourish without burdening them without undue regulation. Nascent technologies require continuous monitoring until the point when they are actually mature; technology maturation and readiness are an important part of the conversation. Two case studies will illustrate this.

AI and computational models are entering the mainstream and are increasingly in the hands of everyday practitioners. The US administration has issued an executive order with a view to ensuring that AI advances safely and responsibly and understanding what that means for the US and global frameworks. In the US, the AI policy landscape focuses on benefit and risk rather than technology regulation. While the policy framework might not address AI specifically, it does cover relevant areas including research participants, data protection, management and data sharing, health data privacy, intellectual property, biosecurity and biosafety. Horizon scanning and landscape assessment aims to identify where existing policies are sufficient and where there are gaps that need to be filled with 'promote and protect' frameworks.

The priority is to allow the technology to flourish in ways that are safe and secure for the US and the global population. This involves thinking about how AI can be used for good and bad purposes, dual-use research, synthetic biology, and the potential for AI to be exploited by nefarious agents. The government funds research in this area and, with the private sector, has developed a framework for nucleic acid synthesis screening to mitigate the potential misuse of AI-enabled biotechnologies. Working across government and the private sector supports a comprehensive approach and the continuous monitoring and assessment of the new and unpredictable ways that AI is being used.

We are also thinking in depth about the future of regulation. Our biosecurity board is working to develop a roadmap for in silico research, to enable our researchers to work with in silico models and develop and identify new targets and new agents to achieve positive goals and mitigate the risk of dual-use developments. This involves consideration of strategies for mitigating risks and publishing findings with a view to maintaining transparency and trust and protecting against misuse. There is a need to protect investments but also to promote research.

The COVID pandemic informed our thinking about policy as regards technology maturation and readiness. From the day the SARS-CoV-2 virus was sequenced, it took one year to deliver a tested vaccine. This record achievement was rightly recognised with a Nobel Prize, but vaccine hesitancy was a significant challenge around the globe. This shows that the ability to deliver science and technology must be met by an interest in adoption within society. We must ensure that technological readiness is matched by social readiness. I would argue that this is one of the most critical steps of an agile governance framework. With the COVID vaccine, we developed programmes to reach hard-hit communities with a view to understanding their hesitancy and building trust. On average, these programmes led to a 6% increase in vaccine uptake.

A governance framework that truly promotes research and technology must also protect it and bring the public along for the journey. We cannot predict the future or the technologies of the future, but we must prepare the public to engage with it when it arrives. Trust is earned in drops and lost in buckets. It is vital that we work in collaborative transparency in order to build and maintain this trust.

Is more better? The risks of innovation without consideration

Virginia Dignum | Umeå University, Sweden

AI governance and innovation is the subject of extensive discussions at all levels of society. When we talk about innovation in AI, we are talking about using technology that exists, not about genuine innovation that involves developing a new technology or advancing an existing technology in exciting ways. The discussion always assumes that there is some God-given right to use any technology we want in any way we want and the focus is always on the idea that we need more; that without more power, more data, more algorithms, AI will not work. Although the focus is increasingly on the old AI approach reaching a plateau and weaknesses in gen-AI, these points remain at the heart of many discussions around governance, policy and regulation of AI.

But what happens when we think in depth about the idea that more is better? When we get more data, we place it the hands of 12 institutions: one in Europe, one in China, and 10 in North America. This is what more is better means. We have learned to build algorithms that are some kind of logical regional system. An AI algorithm is a system that optimises for a set of preferences with more or less complexity. This is different from the way that people behave, reason and take decisions. It is usually difficult to understand the context in which AI systems make those optimisations. Many of the principles and expectations that we have of AI systems, such as rationality, agency and autonomy, are more aligned with the expectations of the global north than the global majority. We are trying to solve very complex social problems with systems that are basically chess on steroids, optimising for a set of rules. More is bigger in computation, but the result is increasing competition between us and systems for the use of resources that, in many cases, are limited. We do not yet know how to balance this competition.

One important narrative that hampers and directs discussions around the governance and regulation of AI is the idea that AI happens to us: we cannot do much about it; it is very complex; all we can do is deal with the consequences. But AI does not happen to us: it is a product of our choices. Regulation needs to address the question of who makes these choices, how these choices are made, and how different choices will be prioritised. AI does not exist in a vacuum. Regulating AI is not about trying to build a technology that solves all of our problems for us. It is not about trying to make the technology right. It is about regulating and enforcing the ecosystem in which this technology is developed and used.

We do need technology, not because a lack of technology will hamper innovation but because technology can support innovation through the use of existing technology and by enabling us to think about alternatives. Regulation and governance is about much more than just legislation, opportunities and scope. We need to consider possible alternatives and explore voluntary standards that enable due diligence, limit liability, and support integration between services, products, innovations and systems. We need to explore different processes, organisational structures and opportunities to address the design process for technology, and find ways to monitor and assess these technologies. It is easy to develop guidelines and principles that say that fairness and human rights are important, but it is very hard to assess if the operation of a piece of code meets these ethical standards. What does it even mean for a piece of code to be fair or aligned with human rights?

There are many concrete steps that could complement and extend the discussion around regulation if they were backed by a desire and a commitment to decide what the right thing is and to do it. It does not require laws to be in place or legislation to identify the direction of travel. Responsibility with AI is not a choice. It is not about limiting innovation but is a step towards innovation. We need to move away from the idea that innovation is simply using existing technology. Innovation governance can provide the directions in which to invest our efforts in innovation. Without it , we will not have opportunities to build trust, drive continuous acceptance and innovation in the field, and enable a transformation in directions that are not only responsible, but also necessary for genuine innovation in the future.

Regulation of emerging technologies in the United Kingdom

Isabel Webb | Department for Science, Innovation and Technology, UK

Emerging technologies, including AI, bring huge opportunities to add value and transform areas of the economy but also come with very significant risks. Good governance is vital if we are to find ways to capitalise on these opportunities without exposing ourselves to unacceptable risks.

One of the first things that the new UK government did when it was elected in July 2024 was to create a new body, the Regulatory Innovation Office, to address this specific challenge and step up its activities in this area. This body is intended to support regulators to update regulations, speed up approvals and ensure different regulatory bodies work together smoothly. It will inform the government of regulatory barriers to innovation, set priorities for regulators to align with broader ambitions and support regulators in the capability they need to meet and grow the economy.

The Regulatory Innovation Office is set up around three main pillars: a knowledge pillar to ensure that we understand these technologies and their relationship to the regulatory landscape and have a solid foundation and appropriate metrics to support strategic decision-making over time; a strategic pillar to establish priorities, particularly industrial priorities, and develop an agile, responsive regulatory system that can deliver the activities, regulatory changes and legislation required for these new technologies; and a capability pillar to address the fact that some regulatory frameworks are not ready to deal with emerging technologies and build the regulatory skills required to identify and respond to the significant economic and societal changes that these technologies will bring.

The Regulatory Innovation Office has developed four pilots to iterate, develop toolkits, and create a broader offering for the entire regulatory system. Engineering biology incorporates areas with huge potential and significant risk. Expert groups on semiconductors, responsible biological innovation, robotics and quantum technologies meet several times a year; their agendas are often set by ministers who require scientific advice and expertise on policy challenges. The UK is fortunate to have a two-way exchange between the scientific community and policy teams within government, which ensures that rapid technology assessments can be provided when policy decisions require input from the worlds of academia and industry and vice versa. Recent deep dives have included digital twins, metamaterials, different types of computer, novel semiconductors, and wireless and 6G. The research analysis includes opinions on the current state and likely trajectory of a given technology.

The public are an important stakeholder in any discussion of technology regulation. This year, we surveyed the opinions of 3,000 people on engineering biology. Although around 95% of respondents thought that it should be regulated, the strength of feeling varied significantly by age: only 50% of 16- to 25-year-olds held extremely strong opinions. Most of them did not know what this technology is and could not answer our questions until we had upskilled them. We need to keep talking to the public to educate them and to understand their views. Capability is about more than just regulations and innovation.

The Regulatory Horizons Council is a group of appointed experts who perform deep dives into the potential regulation of future technologies and submit recommendations to the government. The government is obliged to respond to the Council and state openly whether or not they intend to accept its recommendations. There are clear examples of the work of the Council affecting government decisions. For example, their report on genetically edited technologies directly influenced UK legislation on genetic editing and precision breeding. They are due to report shortly on engineering biology.

The threat from synthetic nucleic acids has interesting implications for strategy and practice in engineering and synthetic biology. Based on two different rapid technology assessments about the emergence and increased use of nucleic acid and synthetic genetic technologies, we convened an expert group who advised us that the potential misuse of DNA synthesis should be our focus and was a policy gap. This spurred us to develop our own guidelines as a framework for the producers and users of synthetic DNA and RNA. We are now monitoring and evaluating how this guidance is being applied, how it is affecting business, academia and research funding, and whether we might need to go further. This is a good example of how our team moves from recognising a problem to developing expertise that feeds future policy.

Regulatory sandboxes provide spaces for regulators themselves to innovate. We have a one that is exploring the regulation of drones that fly beyond the visual line of sight (BVLOS) and another on novel foods, particularly the cell-cultivated products known as lab-grown meat. This sandbox is helping UK food standards agencies to improve their understanding of what the regulatory pipeline should look like and help pioneering companies to road map and test that regulatory approvals process.

It is important to remember that our decisions must be based on evidence and that experimentation, capability-building and continuous evaluation are necessary to stay on top of these rapidly changing circumstances. As policy makers and regulatory partners, we cannot stay ahead of the curve alone: we must keep listening and continue to share information.

Discussion

Douglas Robinson

Our panel have discussed evidence and agile governance, new processes of gathering and mobilising intelligence, and the fact that evidence can be performative, political, inclusive and exclusive. We have considered the management of uncertainty and new capacities, and the need to deal both with new types of knowledge and with the absence of knowledge.

Virginia Dignum

One common theme today is the idea that regulation is an innovative practice. There is often a perception that regulation is a fixed object that cannot be changed and should be pushed away. It is important to emphasise that innovative practices, including regulation, require experimentation if they are to grow.

Isabel Webb

Technology does not respect borders. It is important that we continue to talk and share values in these forums and consider how we can learn from each other, particularly with regard to AI.

Lyric Jorgenson

I agree about the importance of policy and regulation being experimental and the need to gather evidence about what does and does not work. At the same time, regulated parties do not appreciate innovation in regulation as it can be disruptive. To allow technology to flourish, it is important to update and strengthen policies on a schedule that reflects the need for certainty and continuity in the system.

Luis O. Silva

I'm the Portuguese delegate of the BNCT and the chair, and a professor of physics at University of Lisbon at Instituto Superior Técnico. Regarding the question of public acceptance, I would argue that the issue is that the public in general is not ready for change: new technologies are coming so fast and people are becoming more critical. They do not ask where new technologies come from or how they work; they look on it as magic. This is even true of intelligent, educated people with a science background. We need to address, as quickly as possible, the fact that people generally do not think about where technology comes from.

Typically, trying to assess a new technology is like trying to catch a train that is already leaving the station. Sometimes we manage to catch it. But anticipating a new technology is like trying to catch a train when we do not know which station it is leaving from or when it is leaving. That is extremely difficult.

Virginia Dignum

The idea that technology is magic has not arisen by chance. This narrative is cultivated and enforced by people who have an interest in making it sound like magic because, if it is magic, no one is responsible and no one is accountable. At the moment, we are on a plane that they are flying at the same time that they are trying to fix it. We are allowing them to do that through a lack governance and a lack of regulation that means that technology is being dumped into the world. The tech CEOs, particularly the CEO of OpenAI, are open about the fact that their tools and technology are often imperfect, but they put it out into the world anyway and wait for people to help them fix it. We are flying in a plane that is being fixed and we are being expected to put out a fire in the wings. This narrative is part of the push against regulation: if it is magic, it cannot really be regulated. Not only are we failing to discuss the problem of a lack of regulation, we are faced with policymakers and governments who have bought into this narrative.

Isabel Webb

In my experience, civil servants are open about the fact that they do not know what the next innovation is likely to be, while academics say that they can identify the general direction of travel by looking at what labs are trying to achieve but have no idea of when or where the innovation will appear. This is why talking to experts is extremely important and thinking ahead is a significant challenge.

Increasing public education is key. It is important to find entry-level parts of the economy where people are more relaxed about technology. Trousers made of a smart material that can be discarded are easier to adopt than a food that will be ingested or a building that might collapse. Most people were happy to take a vaccine that had been developed extremely quickly given the alternative. Infiltrating areas where the stakes are low can help people to gain trust. I think that this is a useful way to go.

Lyric Jorgenson

I would like to push back against the idea that the public is not ready for technology. Having spent time touring communities in the US and discussing technology, we have found that people are usually excited to have these conversations. The issue is that everything is moving so quickly that scientists are not ready to talk about new developments in time to prepare the public for their arrival. It is not possible to have conversations about the risks and opportunities of every single technology. Instead, we can explain the process of science and innovation, the use of personal data and so on, so that the public have an understanding of key principles and processes when new technologies appear. Scientists cannot do everything, but we need to support communication and engagement in the community as standard.

Claudia Werker

I am happy to hear regulation described as an innovative exercise based on trial and error because this is not usually accepted by the general public or my professional community. The fact that technology is developing and being implemented more quickly than it can be regulated can be seen with the impact of social media on, for example, the attention span of young people. There is a huge difference between the generation that grew up with smart phones and previous generations but, for the most part, we just accept it. How do you deal with that? Implementation often occurs before regulation can even be considered.

How do you deal with issues like the digital divide? This is a prominent issue with AI use.

Erik Fisher

We have heard a lot about talking to experts and talking to the public. Where are the convergences and divergences in terms of how these two groups think about risks and regulations?

David Winickoff

If we seek agile, adaptive regulation, does that undermine the legitimacy of regulation if it is implemented at agency level? By moving quickly, do we risk moving away from the original mandate of the legislation? Can joint knowledge-producing activities to produce more adaptivity fill in that gap?

Douglas Robinson

It was mentioned that we have to take on evidence where there is no definitive answer, and even when there is no clarity about accountability and willingness to experiment.

Pierre Larouche

I do not like the analogy of a train leaving the station because it assumes that innovation is exogenous: it simply lands on our plates and we have to figure out what to do with it. Innovation is not that exogenous. We have a localised problem with AI coming out of a digital sector that we voluntarily and willingly let loose 25 years ago. The issue now is the need to rein in the industry and oblige it to behave like other industries by paying attention to public policy issues associated with their activities.

From the floor

There can be differences in the way that a government regulates its own adoption of emerging technologies and how it acts as a regulator of the private sector's adoption of technologies. Citizens expect governments and civil administration to provide better services through technology. There are various policy approaches built into that and government has to acquire the necessary capabilities to understand the technology and its societal and economic impacts. Some jurisdictions oblige government agencies and companies that process large amounts of data to appoint a data protection officer. The creation of technology ethics officers could be a way to address issues around knowledge base and technological capabilities.

Virginia Dignum

We are not late regulating, because we are already trying to regulate and, in any case, the important thing is the direction in which we are moving. The challenge is not just about legislating. We need to approach legislation and governance from many different directions and combine them.

I fully agree that public agencies need to use technologies, but they often forget to ask why they are using technology. They rush to use AI, but do not consider why and to what end. We need public agencies to ask why they are using technology and whether it is the best way to achieve their goals in a given situation.

Isabel Webb

The UK mandates a data officer for any organisation that handles data. I like the idea of extending this model. I do not know whether the UK government's chief technology officer has ethics in their portfolio. Every department of government has a chief scientific officer who is responsible for ensuring that the department keeps pace with scientific progress; perhaps this could be extended to include technology. The public and the experts have a point in common, which is a very long memory. In particular, the experts remember the public's long memory and can be reluctant to be caught in a mistake. The GMO debate in the UK is often cited as an example that has held us back for twenty years.

Lyric Jorgenson

The fast pace of technological development is beautiful and challenging at the same time. We need at least a degree of horizon scanning to see what is coming. We tend to deploy a series of policy tools, starting with monitoring and stakeholder engagement to establish collective norms. Instead of publishing standards, the US government usually works to understand what is required before issuing a harder regulation. This can mean that we are behind the curve but it helps us to balance promotion and protection.

I find that the public are often experts in how they use their technology. There is a tendency to listen to experts for guidance, but to talk to the public to gain their trust. Listening to and understanding both parties would help to build mutual understanding. The introduction of ethics officers is a good idea, although it would be an very challenging role that would vary significantly between people, countries and regions.

Session 2
Informing more effective governance along the innovation journey and policy cycle

Moderation: Miguel Amaral | OECD

Miguel Amaral

This panel will focus on ways to bridge the gap between technology and policy development to design more effective governance mechanisms. This is a critical gap, and significant efforts are required to bring these agendas and these communities together. The OECD framework is a keystone of the agenda going forward.

The OECD framework for the anticipatory governance of emerging technologies

David Winickoff | Head of the Responsible Innovation Unit, OECD

The work I will present today was developed in partnership with Becky King and Laura Kreiling, among others, and builds on earlier work by Miguel Amaral and on the framework for anticipatory governance of emerging technologies, which was welcomed by the Science and Technology Ministers of the OECD in April.

The framework uses five pillars to guide governments and innovation actors around anticipatory governance. The first pillar relates to the importance of debate around the values that ground governance in emerging technologies. The second relates to strategic intelligence, notably the range and portfolio of techniques and strategies that can be used to perform foresight, horizon scanning and technology assessment. The third pillar is stakeholder engagement, including public engagement and public-private interaction. The fourth and fifth pillars relate to agile regulation and international cooperation.

The paper argues for the need to analyse the innovation and policy processes together in order to identify points where they intersect and how that could be used to inform and influence upstream work. It is built around case studies on policymaking and practice, maps Technological Readiness Levels (TRL) and explores how policy cycles move through different TRL and can be repeat in everything from defining objectives to assessing design and implementation. The agile mechanisms are developed in a taxonomy that draws on framework categories: strategic intelligence, regulatory experimentation, outcomes-based regulation, and non-binding approaches. Each category highlights selected mechanisms, such as horizon scanning, policy prototyping, regulatory sandboxes and innovation testbeds, and non-binding approaches involving ethics.

Case studies are used throughout the paper. For instance, the EU's open innovation testbeds created economies of scale through shared testing and validation and made it possible to investigate and anticipate innovation barriers and regulatory barriers. By considering regulation as an integral element of innovation, it is possible to achieve better outcomes in terms of efficiency, efficacy and values. More upstream approaches to technology governance target TRL 3 and 4 and recognise that governance and values must be considered during the design phase of technology rather than as an add-on at the end of the pipeline. The paper aims to map agile mechanisms with TRL to provide a visual overview that summarises key findings and shows how different innovation phases at different TRLs could and should implicate different mechanisms from early stage to later stage. It maps the policy cycle across these mechanisms.

Finally, it provides six principles for greater agility and technology governance, the first three of which are iterate, learn continuously, and generate knowledge and data. Iteration is defined as revisiting and refining governance approaches as the innovation process develops. Continuous learning aggregates knowledge through feedback loops so it can be transferred into governance. High-quality, communicable knowledge is systematically generated through this process to align governance and technology designs. Additional case studies cover the evolution of geoengineering codes of conduct, the development of broader multi-stakeholder processes, and outcomes and information generated by Malaysia's national sandbox.

The fourth principle relates to the integration of governance considerations in design and development and throughout the innovation process. There can be a tendency to view 'by design' as a technological fix, whereby the insertion of a little code within the technology is enough to provide protection. However, as we are dealing with socio-technical systems, embedding certain mechanisms is not sufficient to achieve the desired values. Instead, by design technologies need to be understood within their social context.

The fifth principle is to activate and support science and technology development communities and the sixth is co-creation, specifically the use of different engagement strategies to incorporate diverse perspectives. The stem cell research case study provides an interesting example of a mechanism or modality that developed organically over time and has been fairly effective in governing the space.

Technology development and diffusion within the science technology market

Roland Ortt | Erasmus University Rotterdam, Netherlands

My research focuses on technology development and diffusion. I will explore this issue through five steps, with reference to the work of Van der Ven and with a specific interpretation of the science technology market process. The example of quantum technology shows how the shift from science to technology, and from technology to market, is more erratic than it may appear and its implications for strategy and policy.

The science technology market perspective is very iterative but it is not necessarily linear. Quantum technology involves the control and mastery of atomic-level phenomena in order to create functionalities at the level at which we live. For example, manipulating the orbit of an electron to another orbit results in the emission of a frequency with a very specific number that can be used to measure time.

There are three main branches in quantum technology: measuring and sensing, computing and simulation, and communication and encryption. It emerged as a family but it is not a single technology. Instead, it is a set of spaces and sub-spaces, all of which have different characteristics and, probably, require different regulation and strategies. The development of the first atomic clock in the late 1940s led to a realisation that its weakness, namely its sensitivity to gravity, electromagnetic fields, sounds and so on, could be exploited to measure the very things that disturbed it. The field of quantum sensing technologies was born. Some of these technologies are now well established and widely available: quantum technology is not a single area that can be regulated as such.

By diving down one level, we enter into specific quantum sub-technologies. The innovation diffusion paradigm in science states that, in an invention where the principle of the technology has been demonstrated, it is possible with careful management to make something that will diffuse gradually before being widely adopted. My exploration of 130 different breakthrough technologies shows something completely different: that between the first demonstration of a principle and its first introduction, there can be decades or even a century or more of failure, inaction, and projects developing in parallel and in chaos. There is huge variation, but even among the most successful technologies of the 20th century, around 80% did not start diffusing gradually on a large scale. Instead, they diffuse erratically in small niche applications for specific segments in an adaptation phase that usually lasts around 10 years. That phase has some exciting consequences, but companies, institutions and scientists tend to underestimate these initial phases and instead create the hype cycle and the disappointment that goes with that.

Niche applications are very important. Agility is required to make the step from science to technology and from technology to market. We have been very slow with AI: although the idea was coined in 1948 and the initial phases were unbelievably slow, we were surprised by the speed of large-scale diffusion a few years ago. It is difficult to be responsible in these situations because our legal institutions focus on the phase of large-scale diffusion but lack the resources to consider small segments and initiatives.

There is also a paradox between convergence and dispersion. Convergence increases links between technologies; dispersion occurs because these links enable the creation of even more technologies. As a consequence, corporate strategies are completely different in terms of expectations, goals and performance metrics depending on the model used. An organisation that targets diffusion will aim for a mainstream market, but if they try to make this progress during the niche phase, they will be disappointed and abort their activities. Instead, they should aim to experiment, build up knowledge, develop a niche market, and have realistic expectations and performance metrics. The same is true of policies.

We should share a realistic perspective on how technology is developed in diffusion. Without this, we will be agile but playing the wrong game. We must have a realistic perspective of how technology develops and diffuses and provide a system perspective that explains what happens, highlights the conditions that are required for the start of large code diffusion, and ask whether they can be used as a basis for strategy information. Instead of niche strategies, we need a system perspective and map that can serve as the basis for policies and institutional intervention for both companies and governments.

We anticipated some disruptions in our work on quantum technology, but we have been taken by surprise by the disruptions caused by the new global political situation and big tech influences. With hindsight, we can see that these were not sudden events but we failed to notice them until they were upon us. It would be useful to have instruments to identify these events in advance. Our scientists are committed to their projects, but to achieve our goals and missions and solve the issues we face, we must also consider alternative technologies that could compete with or complement them.

Miguel Amaral

The move from technology to science to technology to market is not linear but is highly complex, sometimes iterative, and sometimes erratic. There is definitely a need to obtain a precise understanding of this evolution and explore opportunities to apply it to the framework.

Regulating for safety and sustainability in emerging technologies: the challenge of time and data

Mar Gonzalez | OECD ENV Directorate

My work in the OECD Directorate of the Environment focuses on chemical safety, specifically the development of harmonised instruments to perform safety assessments on chemicals, nanomaterials, advanced materials and so on. We work extensively with agencies involved in occupational safety, human health and environmental safety, developing legally binding task guidelines and guidance documents to help them to respond to regulatory needs. Our work is mainly triggered by the need of regulatory compliance. We interact primarily with policymakers, industry and researchers with a view to reaching consensus. For the last two decades we have been working on manufactured nanomaterials to try to integrate them with chemicals and, in the last seven or so years, we have included advanced materials, namely materials that are more than 100 nanometres and are more complex than traditional nanomaterials (e.g. NanoCarriers, Mxenes).

We must consider whether we are prepared, from a regulatory standpoint, to deal with these materials when they come to market and understand whether we can identify and implement adaptations when required. There are no obligations to enact legislation, but there is a need to promote innovation and to ensure that it is safe and sustainable by design. We do not seek to pit regulation and innovation against each other, but rather to use regulation and standards as ways to build trust, develop transferable products, ease the path to market, and protect human health and the environment.

Time is a significant challenge. Research must have reached a reasonable level of maturity before it can be discussed, analysed and validated, so developing a standard can take anything from three to ten years. We cannot afford to spend this amount of time on new emerging materials that offer a plethora of applications possibilities and exists in different varieties, such as nanocarriers and graphenes. As a result, we have had to envisage new ways of working that can be carried out in parallel to our regular work and support a safer and more sustainable innovation approach. This involves a change of cultural mindset: we need innovators, regulators and industry to stop thinking about safety, sustainability and the market as discrete, sequential steps and instead to integrate questions about safety and sustainability from the earliest possible stage. The aim is to pivot away from risk management and towards risk prevention.

The safe and sustainable innovation approach has three main components. Regulatory preparedness aims to better anticipate and adapt governance to these new materials, for example by questioning whether new regulations are required or whether existing regulations can be adapted or expanded. Safety and sustainability should be included at the earliest possible stage in the process. When a trusted environment allows industry, stakeholders, innovators and governments to discuss confidential issues in security, it is possible to anticipate many considerations before complete data is available. For example, it may be possible to identify at an early stage that a proposal will fail because it does not answer the right questions. The aim is to identify products that will reach the market and, at that point, consider their functionality. This also requires a change in mindset and working descriptions of sustainability and Safe-and-Sustainable-by-Design (SSbD).

Our aim is very pragmatic: to extract the elements that we require to develop tools for the implementation of the Safe(r) and Sustainable Innovation Approach (SSIA). Once we had developed a description and key elements, we analysed and mapped different systems on safer design and compared them to our description, for example with regard to lifecycle, human health and environmental safety, economic aspects and social aspects. We aim to publish this as a simple tool in early 2025 and make it available to industry and regulators.

We are also working to identify the information that is needed to identify hotspots and understand readiness levels. Our aim is to have three components to identify functionalities, human and health implications, and sustainability aspects, and to screen materials accordingly. Trusted environments are a simple concept but a significant factor in bringing stakeholders together and helping us to refine our models. We have produced simple guidance about the elements that are required to enable a confidential dialogue and support the participation of small and medium enterprises. In the last year and a half, we have run workshops about very specific advanced materials, exchanged unpublished data, and identified hurdles that industry faces with the development of these materials.

With regard to new advanced materials entering the market, the regulator has to be able to anticipate key moments in the process, develop new testing guidance, review legislation and so on. With this in mind, we have developed the Early4AdMa, which is a risk governance tool based on a very simple set of screening questions around human health, hazard, shape, size and so on. It incorporates sustainability questions based on the SSIA as well as a graphical summary that highlights areas that might require more data or discussion and identifies areas where problems are less likely to occur. The idea is to help regulators anticipate and move forward as soon as they have a certain amount of data and knowledge.

It is important to avoid paralysis by analysis: we do not have all the data but we can extrapolate a lot of information based on experience. Our approach needs to be multi-disciplinary: we come from a safety community, but we need to learn more about sustainability. We are not proposing that everything must be safe and sustainable, but we must be aware of what we are putting into the market and how it is produced., and understand the trade-offs. For instance, there might be no point reducing water use if the water will be replaced by highly toxic chemicals. Our aim is to increase this kind of awareness.

Miguel Amaral

Regulatory preparedness is one of the core dimensions of the recommendation and the framework. We believe that it is absolutely key. The challenge is to use the outcome of this exercise to feed into policymaking and avoid the disconnect that can exist between the regulatory side and policy action.

Socio-technical integration: the role of scientists and engineers in governance

Erik Fisher | School for the Future of Innovation in Society and the Consortium for Science, Policy and Outcomes, Arizona State University

It is only in the last twenty years that policy makers have begun formally to explore the roles and functions of scientists and engineers within anticipatory and agile governance regimes. Two documents that form the basis of today's conference, namely the OECD documents on agile governance and anticipatory governance, refer both implicitly and explicitly to the roles, capacities, contributions and responsibilities of scientists and engineers. Both documents refer to the policy and innovation processes in their entirety, creating significant expectations that scientists and engineers will take more and more things into account as they develop science and technology. The agile governance document seeks to leverage the role that innovators can play in the governance of innovation, although the role and the leveraging process require definition, while the anticipatory governance document states that innovators should support technical aspects of social research and engineering while taking account of social, legal, ethical and policy aspects.

This radical idea that engineering should focus on social and ethical implications is known as socio-technical integration. Although it is now largely taken for granted, it is not always clear what it means, how to work with it, and how not to work with it. It can be thought of as a micro-foundation for responsible innovation and for the many other expanded roles that society has assigned to scientists and engineers. The implication is that this capacity will lead to strong science policy regimes, and that without it, the regime will only be as strong as the capacity of the scientists and engineers to think and act beyond their training.

This vast oversimplification can be explored by considering one version of the innovation journey, broken down into three iterative functions, and the research policy cycle. In this scenario, scientists and engineers play a dual role, performing R&D that advances private wealth generation while also serving public interests that have been authorised by policymakers. In this context, scientists and engineers operate in the midstream between these two cycles. However, policy studies show that the implementation phase is extremely difficult to engage due to the inevitable discretion that delegated agents have in the policy cycle. Although we rely on these agents to implement public and private policy, it is notoriously difficult to control and even to monitor these agents, whether they are police officers, judges, civil servants or experts. Monitoring is extremely important in agile governance but is extremely difficult. Scientists and engineers have a dual role as policy agents and strategic innovation agents. This is because the myriad individual choices that they make are regarded as relatively unimportant at the time, but aggregate in such a way that scientists and engineers not only develop technological trajectories but also help shape their direction and pace. In implementing public policy, they also shape its outcomes.

The question is how this concept of R&D discretion or 'lab-level bureaucracy' should be engaged with and governed. Possibly the most important aspect of socio-technical integration is the real and perceived cost of integration. As well as financial costs, there are costs associated with loss of time, loss of productivity and cognitive load, namely the additional burden of thinking beyond your area of expertise. Over the last twenty years, we have explored a number of possible interventions. Rule-based, prescriptive, compliance-based regimes are very specific about the integration of certain issues, considerations and values but tend not to be particularly effective because they are easily skirted, are overly general and hard to apply, and may be counter-productive. Process-based approaches seem to be more effective but, again, deliver mixed results. General prescriptions that set out broad aspirations also deliver mixed results. Interdisciplinary collaboration also has its own issues and, again, generates mixed results.

Instead of trying to implement integration, our programme in Socio-Technical Integration Research (STIR) uses semi-structured dialogues between social scientists, scientists and engineers to

understand the conditions under which it operates. Research has taken place on three continents across dozens of technology fields, from professors and CEOS to research students and technicians. We have learned that this elusive integrative capacity is routinely and productively exercised by scientists and engineers, but only under certain conditions that are temporal, experiential, praxeological and methodological: the timing, structure and content of the dialogues must be synergistically in relationship to the problem-solving activities to which the scientists and engineers are committed. Furthermore, their capacity to integrate broader considerations, values, questions into technical practices changes over time. At T1, a team of engineers might state that stakeholder considerations do not factor into their decision-making; at T2, following an exchange with a funder or a client, their position might change entirely.

This is a dynamic capacity that needs to be understood and cultivated as an endogenous ability that exists within the technical community. Unfortunately, exogenous expertise often enters as an overlay that is counterproductive and fails to recognise existing signals. It is important that any incentives, guidelines or interventions rely on voluntary contributions by scientists and engineers rather than coercive [expectations], and that these are synergistic with local problem-solving practices. Funding organisations should seek to identify, codify and incentivise existing and emerging best practices for collaborative interchange and experimentation. Policymakers and social scientists are often sceptical of giving autonomy to scientists and engineers on the basis that they will fail to consider the public interest. Instead, a balance needs to be struck between asking scientists and engineers to be more attentive to the public interest and allowing them to do what they do best without undue interference in the core processes that allow them to innovate.

Discussion

Roland Ortt

The trusted environment is an exciting concept that could be useful for bringing different actors together at an early stage. How do you see it working at later stages, after the product or innovation has been introduced to the market? Is it disbanded or does it evolve into something different?

Mar Gonzalez

A trusted environment can last for as long as it is needed, as long as there is flexibility about what it is for and how it is used. The aim might be to bring together innovators with regulators or other innovators, but you can decide who will participate and how their input will be relevant at different stages of the process. A regulator might offer information and insight at an early stage, but could also have responsibilities for product approval at the end. The trusted environment can be used at any stage in the process. For example, the 11 largest tyre companies operate a trusted environment to exchange information on specific topics; COVID obliged vaccine companies to develop a trusted environment for sharing information. Obviously there are issues around confidentiality, but as long as the approach is flexible and the purpose is clear, they can offer genuine benefits to participants and could facilitate communication in the future.

David Winickoff

All of us are framing the relationship between governance and technological development as a co-evolving process that changes over time. This is a descriptive position but also a model and foundation for identifying levers of policy at different points in those processes. Our paper, like others, tries to do this across a range of TRLs. It is not enough to say that they are co-evolving systems: we want to shape the world and shape technology towards values. We are identifying different lever points in these processes.

Erik Fischer

I agree. All four talks pointed towards the need for reflexive awareness of different models, stages and conditions. The policy cycle and science and technology practices are morphing and can be out of sync at times, but an awareness of where these cycles are and how they relate to each other can be useful for identifying levers. The discussion also emphasised the need to anticipate change and coalescence through reflexive awareness, and the need for a shift in cultural mindset from risk management to prevention. In all cases, there is a need to engage human agency to promote concerted action that is largely organic and processes information to determine what is signal and what is noise.

Mar Gonzalez

Even with agile approaches, we need time. We need time to digest information, be multidisciplinary, promote useful innovation and so on. You have developed a fantastic framework but we need to ensure that we have time to absorb it and develop practical tools. The presentations have been fascinating.

Miguel Amaral

Agile regulatory approaches often highlight trade-offs of agility versus time and agility versus predictability. When developing a business, it is important to have an understanding of the legal framework; it can be problematic if that will change every five years. We are not yet sufficiently agile to accommodate this.

Douglas Robinson

My question concerns institutional change and incentives for change. We have discussed training for lab-level bureaucrats on constructive technology assessment experiments. New risk assessment approaches, the fusion stage and adaptation have all been mentioned, as have different elements of our framework. What are the incentives and disincentives around change practices? How do you get those things working?

Lyric Jorgenson

The consequence of expecting scientists to be masters of all sciences as well as convergence, engineering, ethical issues, social engagement and policy is a huge cognitive load for people who are also supposed to be thinking about innovation and boldness. In your data and in your experiences, should we prioritise training people to be somewhat knowledgeable in all things, or structure teams to bring in that expertise in multiple ways? Both come with trade-offs in terms of time and funding.

Virginia Dignum

The concept of 'by design' was mentioned in several presentations. It usually leads to an interpretation or contextual interpretation of a target in a specific context, rather than to the implementation of the values themselves. How can we be explicit about this contextualisation? What kind of approaches do you see in terms of liberating this 'by design' approach from excessive contextualisation?

Erik Fischer

Based on my knowledge of scientists and engineers, a big incentive for change in practice is that you can do your job better. This is a massive claim for experts to make from the outside. There is a need to make this demonstrable and part of the experience. It presumes that the person's concept of their job expands. Key drivers are education, engagement with stakeholders and learning from failure.

Roland Ortt

Regarding cognitive load, it is interesting to consider that, in law, the roles of prosecutor and advocate are separate. This is not due to cognitive load but because some roles cannot be mixed. We must differentiate between things that cannot be mixed because they would dilute or be in conflict with each other.

Mar Gonzalez

Regarding incentives that promote change, we spend our time responding to flagrant needs, which distract us from other priorities. There is a need to deal with ideas like 'by design', which are not new but have gradually, through repeated appearances in papers, been transformed into policies. Once they become policies, there is political interest and we can start working on them. In my experience, scientists want to do a good job and develop products that make money and are as safe as possible. That is part of the incentive. The pace of change means that we need to anticipate more and have different scenarios. Regulation is the trigger and part of the solution. People are not interested in safe and sustainable by design because it is 'nice' but because it offers a return on investment or there is a political push in that direction. Even if something does not become a regulation, we have to work to prepare to go in that direction. The idea of 'by design' is important because it highlights the elements that are involved and ensures that users and regulators are aware of different aspects that must be taken into account when making decisions.

David Winickoff

Six words: frameworks, behavioural models, capacities, sticks-and-carrots, evaluation, and iteration.

From the floor

Do you envisage spontaneous regulation by engineers, or organisational self-regulatory efforts through engineers in their role as employees or partners of those organisations?

Erik Fischer

Primarily the latter. My idea relates to long-term change and capacity building based on problem-solving tools, techniques and rubrics that can be developed by individuals. Once engineers have the ability to optimise their procedures, they can link with tools and procedures that might develop organically. With funding and a lab environment to identify and promote the best, these could be codified and used to develop an organisational set of procedures that could then develop into more explicit policy frameworks.

From the floor

There is a tension between the people who develop a technology, that they know and care about, and the policymakers who see a technology as something to be used. The developers think the policymakers know nothing about the technology; the policymakers think the developers are unworldly geeks. If we think of technology as a tree, we can start from the trunk or the branches. What are the implications of these two visions for your concept of anticipatory governance?

In a past role as a particle physicist performing data acquisition, I had to be able to decide quickly when to retain data and when to discard it, and also to take quick decisions based on small amounts of data. The trick was to choose the right data with the right patterns. Perhaps this could form part of the framework for approaching situations where we lack data.

Roland Ortt

The question of scientists and policymakers looking from different directions can be combined with the question about incentives. A quantum scientist with five sensing technologies and 10 potential applications will want to explore all possible combinations in a comprehensive and systematic way. They will start with the first, their funding will dry up halfway through the list, and they will be annoyed with the policymakers. Instead, the policymakers need to ask them to start with option 43 because that has the best commercial potential and could generate the funding they need to complete their whole project. We need to develop a research agenda that allows both parties to meet in the middle.

Mar Gonzalez

I work with experts to develop standards. It can be very difficult to get subject experts to accept that they need to move on before they have collected all of the data that they think is necessary. We need to get usable documents and guidance for industry and regulators; they can be updated and completed over time as new data becomes available and understanding matures. As every invention brings new risks, there has to be a certain level of comfort and trust and weight of evidence in the development of standards. With regulation, it's important to be aware of what is good to know and what you really need to know.

Erik Fischer

We need to be aware of *kairos*, that is the right time to act. We can prepare with data and evidence and awareness of the right time to act, but inevitably we have to act, at which point acting becomes more important than information. Sometimes there is a need to make decisions in conditions of uncertainty, to rely on judgement and intuition, and to train, test and diversify our intuitions accordingly.

Joëlle Toledano

'By design' is wording that was used during the development of GDPR. Although the Brussels effect exists, GDPR is not the same all over the world. How is it possible to be responsible and sustainable 'by design'? It does not happen automatically.

Mar Gonzalez

This is a complicated question. 'By design' has its own complexity. In my work, we think about very specific aspects of the material, such as particle size and hazard, but we also increasingly think about functionality: how and where the invention will be used. We have many opportunities to respond to the sustainable goals, but if we do not think about functionality then we will continue to focus on the hazard and get stuck on that step. For me, 'by design' is a change in mindset that shifts from thinking only about risk profile and safety assessment to thinking about potential consequences.

Roland Ortt

The question implies that 'by design' is a slow and expensive luxury at a time when we need to move fast. This may be true, but in my experience, when companies have postponed the introduction of a product to spend more time on this phase, it has resulted in a shorter, less expensive introduction process overall. Taking more time in the short term can provide dividends over the long term.

Mar Gonzalez

One version of 'by design' is about failing quickly and failing cheaply, before investing in 20 years of research to develop an unusable product. It is about failure.

Session 3
Tackling the challenges of technological convergence for governance and regulation

Moderation: Becky King | OECD

Becky King

This panel will cover converging technologies with a view to understanding the emerging and evolving reality of these technologies, which have perforated into many different technology spaces this year. The fact that the Nobel Prize in Physics was won by two computer scientists further highlights the convergence in these areas. Technologies are also merging with our physical, social and biological selves. This offers enormous potential to develop and address global challenges but also raises questions around privacy, safety, security and ethical boundaries. We will consider how technological convergence is driving these complex challenges and their implications in terms of responsible innovation and governance approaches.

Reflections on the risks, benefits and regulation of converging technologies

Anike Te | Lucideon & University of Bristol, UK

I am the Chief Strategy Officer of Lucideon, a material sciences company working in areas such as aerospace, energy, construction, ceramics, advanced ceramics, and healthcare with a focus on inorganic materials as well as small molecules and medical devices. We provide consulting and testing, solve materials challenges, and optimise processes. We are now exploring synthetic and engineering biology to support the development of new materials and leverage our existing expertise in material science technology.

Converging technologies bring potential solutions and risks. They offer hope in the form of improved vaccines, therapeutics, sustainable development, climate change mitigation and new energy solutions. The risks associated with these technologies differ depending on the industry and innovation but could include regulatory failures, laboratory accidents and, importantly, biosafety and biosecurity.

We need to take a broad view and remember that we could be simultaneously innovating for multiple sectors. For example, by innovating for healthcare we could at the same time be innovating for food; by innovating for construction, we might also be innovating for aerospace or energy. While technology transfers between industries have always occurred, the potential for simultaneous multi-sectorial innovation is growing. Regulation must be an enabler rather than a restriction. We cannot stop curious minds. If we restrict, innovation will simply move to less restrictive regions.

Traditionally, regulators have focused on specific industries. As we develop new materials and technologies that have broader, cross-industry applications, we should consider the need for cross-disciplinary taskforces to develop enabling regulations across the entire ecosystem because all parties, from corporates to start-ups and SMEs, have something to offer. It might also be useful to think about a common language. Synthetic biologists, material scientists and AI experts have already found ways to communicate with each other, but there is a need for better communication around new technologies across a broader range of fields that often operate in silos, such as energy, aerospace, construction and healthcare. A common language could speed up innovation and collaboration, support regulatory frameworks and enable risk identification.

Finally, we should consider using tools to simulate more outcomes. For example, our expertise at simulating for the flu vaccine could be expanded to reduce reliance on animal testing or use scarce resources in different ways. Simulating outcomes of new biomaterials could help us to understand how they will perform within buildings or in other applications after 5, 10 or 15 years. Simulations should not be used in isolation as they come with their own risks, but they could be combined with other solutions and outcomes. As we are innovating simultaneously for multiple sectors, it is imperative that we take a broad view.

Regulation and governance of convergence: the example of neurotechnology

Pawel Swieboda | International Center for Future Generations (ICFG), Belgium

Technology convergence, specifically neurotech AI, looms large in neurotechnology. There is a continuum from interdisciplinarity to convergence; we are now approaching convergence. Interdisciplinarity focuses on people and aims to integrate knowledge, methods and perspectives to generate new understanding. Technology convergence aims to blend or integrate different technologies with the objective of creating new functionalities, applications and products. Often they merge into each other. Technology convergence builds on longstanding ideas about the combination and recombination of existing technologies.

The difference now is the scale of this exercise, which is becoming dominant and creeping into all technology areas. Convergence is now our base scenario. For example, multi-modal solutions in brain health integrate neurotech with AI and link to real data sets. Once this convergence starts, it is unlikely to stop.

Convergence addresses a number of technological challenges in neurotechnology. Wireless data and power transfer improvements are required to enable neurotechnology devices to work precisely in high resolution. Appropriate packaging is also important so that brain or spinal implants can operate reliably in the body over the long term. Neurotechnology devices stimulate the neural tissue, so improvements in microelectrodes are necessary to support more precise stimulation. New approaches, such as non-invasive photobiomodulation, are also important because, by reducing the invasiveness of neurotechnology, it should be possible to increase access to these techniques. Finally, closed-loop technologies are required to integrate brain recordings and computational methodologies with AI.

As well as technologies, we must consider outcomes. A spinal cord implant can enable a person with tetraplegia to walk by decoding motor intentions in the brain and translating them into leg movements. This has relatively low sensitivity from a governance perspective because it concerns motor functions, but consumer neurotechnology is an intermediate category that is growing very quickly and is also multimodal. The intentions of the developers creating devices for eye tracking and reading motor intentions, for example, are not always clear. Brain-computer interfaces (BCIs) that enable the decoding of thoughts require very close attention, particularly as both consumer and medical BCIs are being developed.

Examples tend to relate to the medical area but the need for safety, authenticity and mental privacy is very strongly magnified by the fact that there is convergence with AI. Someone with a condition requiring treatment with a BCI might gain a massive advantage by becoming able to communicate with the outside world, but the decoding process might also result in private internal monologues being shared with the outside world. This is an extremely sensitive topic.

The OECD framework includes a few references to convergence; I expect there will be more references in future iterations. Convergence results in a blurring of categories, making it more complex and harder to govern emerging technologies. The framework identifies six dimensions for assessing governance needs which could, I think, be usefully explored through the prism of convergence. The pace of technology emergence is likely to see similar dynamics, whether from single technologies or converging technologies; for the others, we are in a different world. The future is likely to see a higher degree of uncertainty, a greater need for risk analysis, greater public concern, greater strategic importance, and more governance gaps.

Convergence is likely to change the scope of guiding values and foundational values, such as respect for human rights, safety, security and privacy. The meaning of privacy is already changing through the convergence of technology. Technology-specific values will change and be rendered significantly more complex by technology convergence. This is already an issue in neurotechnology and AI because it deals with brain data, which is a special category of personal data due to its sensitive and intimate nature. Unique brainwave signatures, captured through EEG devices, can enable authentication and profiling. The convergence of AI tools and neurotechnology also makes it possible to draw inferences from brain data. We must consider not only the pure datasets, but also the potential to discern a person's intended behaviour. Virtual data, namely data that predictive models create on the basis of real-world data, are also being created by convergence. These data can be useful in medicine but are not currently protected.

Convergence has a clear impact on the deliberative processes envisaged by the OECD framework. There is a need for a dedicated process around technology convergence, possibly based on observatories and specific education for stakeholders about the implications of convergence. Convergence also has a very significant impact on most other dimensions of the OECD framework. Horizon scanning, deep dives and extended appraisals should take account of the likelihood of technology convergence, and tools for stakeholder engagement will also need to adopt an explicit focus on technology convergence. Regulation and international cooperation will require attention.

There is a genuine need to have very dynamic capabilities for governing technologies which are converging because they are likely to drive changes that go beyond anything that has come before. Dedicated entities and institutions are needed to assess technology convergence and strengthen stakeholder engagement. Convergence will change, very substantially, the practice and the substance of our governance approaches.

Convergence of immersive technologies and AI: a case study in education from Taiwan

Tammy JihHsuan Lin | National Chengchi University, Taiwan

My interests include communication technologies, including immersive technology, social media and digital games, and their effects on humans. I have collaborated with Meta and Apple to innovate around immersive technologies, which are also convergence technologies, and explore the VR-AR convergence with AI policy. I work with TikTok on AI implementation in social media. I will be sharing a case study from Taiwan to showcase how government, industry, experts and other stakeholders co-create the convergence of immersive technology with AI and to explore how we might view convergence regulations.

Taiwan manufactures AI chips and can create most of the technology it needs, but it faces significant challenges around governance and regulations due to the extremely rapid pace of change. Early regulations tend to be meaningless due to an insufficient understanding of the approach and core values. A co-creation approach is required to develop core value guidelines without snuffing out innovation. The lack of standards and standardised platforms, for example for VR headsets, also poses challenges and creates inefficiencies. It results in wasted effort and energy at the outset, but can encourage smaller companies to work with the government to aggregate their efforts and develop a shared business model.

Rapidly changing technologies also lack test beds. For example, although immersive technology is an excellent human technology interface for education, training and medicine, it is hard for private companies to develop without access to test beds and systems. With this in mind, Taiwan has built a core value system and core value governance and promotes test beds based on an innovation co-creation environment. This systematic sandbox enables industry and stakeholders to come together in an inclusive environment. Engaging all stakeholders at an early stage makes it easier to innovate and test. Digital trust is a key issue: it is important to build this trust between society and different stakeholders in a shared way.

The case study relates to the integration of VR-AR into our K-12 education system. The platform and standard were co-developed by the government, education experts and industry leaders, like HTC and Meta. Industry participants are free to use the platform and test their programmes in the K-12 system; content development is led by education experts. The impact of the system will be assessed by the public and NGOs at a later stage. The government has invested $20 billion in procuring the system and creating a 5G network for schools. Schools can apply to access the 5G network and obtain grants to purchase hardware, like VR headsets, AR iPads and so on. HTC and education experts have transformed all classroom education materials into VR and AR content and integrated them into a platform called Education Market. The government ensures that the platform meets core values, education experts design the content and assess the effects, the public and the children share feedback on their experience of a VR learning experience, and industry is able to accelerate innovation.

Over the last four years, 222 schools and classes from grade 1 to grade 12 have participated in this project. Small and remote schools are encouraged to connect to larger schools through 5G for distance learning and AI is now being used to develop VR content for global participation, for example through instant translation systems that enable students and teachers from other countries to communicate directly inside the platform. This is a unique case study for convergence. The next step is to develop a global platform for international collaboration.

Taiwan uses two kinds of sandbox to rapidly test ideas. The first is at government regulatory level, based on an annual competition to explore specific scenarios. The second is aimed at the public and the innovation phase and is used to test special projects and projects that require regulatory exemptions. The AI POLIS system is used to enable democratic discussion in public.

These sandbox techniques are intended to explore effects and regulation in parallel. Our core values are foresight based on specific participation and participatory governance and stakeholder inclusion at every stage in the innovation process.

Anike Te

Do you have any suggestions as to how the intention of devices could be governed?

Pawel Swieboda

This question gets to the heart of the grey zone between medical and consumer devices, which is not currently covered by governance frameworks. Consumer protection has its own framework and medical devices have a separate one. In practice, medical devices can evolve into consumer devices and vice versa. Some teams receive significant funding to perform medical research only for the knowledge to also be deployed in the consumer space. This merits our attention. Regulatory regimes need to consider these grey zones and assume that there will be attempts to cross from one field to the other. Rather than trying to create a separate framework for this grey zone, we need to anticipate situations where borderline cases emerge and address them within the existing frameworks.

Tammy JihHsuan Lin

Convergence might be the basic scenario, but many people are not aware of this. How can we communicate this more broadly, to ensure that stakeholders are aware of it? This is key to building trust.

Anike Te

We need to keep the public in mind at all times, because if they do not support our actions then neither will politicians or governments, at which point funding for research and start-ups will dry up. We need to find ways for the scientific community to engage with the public that are accessible and interesting.

Pawel Swieboda

We must deliberately insert convergence into the discussion. Sometimes people do not want to make the governance discussion more complex than it already is, but convergence is already widely present, and it needs to be demystified, as it is the reality of where technology is going. It should be included in all of our governance discussions and in the OECD framework, so that the issues that result from it can be addressed. The great majority of convergence issues are not particularly sensitive or controversial, but some aspects do need to be addressed for reasons of transparency, accountability and general public awareness.

Becky King

Similar comments have been made about the need to demystify AI, which also suffers from a lack of knowledge. There is a need for a narrative that will shape and build trust around these technologies.

Regulating converging technologies to promote competition and incentivise innovation

Howard Shelanski | Georgetown Law Faculty, USA

There is a need to align and harmonise different kinds of regulation. The discussion thus far has hinged on questions of governance around key technologies, such as AI, material sciences and neurotechnology. We must also consider how we enable innovation by emerging firms and technologies in marketplaces that are not hospitable to the entry of digital technologies.

While it is true that innovators will innovate, they are more likely to do that if they have a commercial path to earn returns on those innovations. Although large incumbents do not have an interest in encouraging new firms with disruptive technologies that might challenge their dominance, new entrants will seek any path they can find to give consumers, governments and businesses the kinds of products they want, even if those products are not safe or in the best interests of society.

Convergence requires us to consider how to create a regulatory system of governance that does not favour one line of innovation or one set of industries over another, is sufficiently aligned and converged to relate to a variety of different systems and products, enables competition and the emergence of new technologies in concentrated markets, and aligns with safety considerations. These different regulatory objectives have collided in the past. It would be useful if large visual platforms would share some of their facilities, training data and other assets with new and emerging technologies that might disrupt their dominance. This might tempt legislators to introduce regulation that requires interoperability, sharing, and the interconnection of certain kinds of capabilities, particularly data, but this would have consequences for privacy and data security. We must consider the extent to which we are willing to trade safety and governance regulations around privacy and data security against the ability for new technologies to enter and access markets. These changes are happening in real time: new players using new, generative, AI-enabled technologies are increasingly entering digital marketplaces, such as e-commerce and social media, that have operated, thus far, using more conventional algorithms. We must consider how new players can break in when the training data and user base they need is held by a large incumbent. The anti-trust trial in the US around Google's search monopoly and potential court-ordered remedies will bring these issues to a head.

It is interesting to consider which regulatory approaches will be most effective at generating new competitors and new competition, incentivising innovation, and introducing innovation to the market. There are separate regulatory considerations around the governance and safety of these technologies. We might not want lots more AI-generated technologies until we understand their potential use and social implications. The unquestionable promise of certain neuro-technologies for people who are tetraplegic does not diminish concerns about the potential invasiveness of thought decoding. As always, technologies that offer the promise of great benefits and great risks need regulation that enables it to emerge, compete, converge and exist alongside social considerations about safety and governance.

History tells us which regulations are more or less likely to be successful at enabling emerging start-ups to break into markets where there are strong incumbents. Structural separation, whereby the large companies are broken up, and line-of-business restrictions are often highlighted. Current examples might involve advertising being separated from search, or e-commerce platforms being allowed to makes sales or provide logistics and display but not both. I explore the history of divestitures and line-of-business restrictions in an upcoming paper and show why they can impose very high regulatory burdens without creating significant benefits for consumers or innovators. They may, in fact, slow innovation.

The history of telecommunications regulations highlights a more promising approach which drives cooperation and convergence between companies based on interconnection, interoperability and access regulation. To the extent that competition authorities and other regulators want new technologies to gain traction in the market, they must recognise that feedback effects and networks effects give a built-in and repeating advantage to large incumbents under fairly common economic circumstances. The only way to achieve faster deployment of innovations by new and emerging firms – rather than by the incumbents – is to require the incumbents to interconnect and interoperate with some of these new players. This allows the new players to overcome the network and feedback effects that advantage the incumbents. I expect that interoperability and interconnection remedies and regulatory frameworks will be increasingly present in the marketplace and that this will speed the deployment of AI into digital applications and support the rapid deployment and development of generative AI and large language models by non-incumbent players. It might result in incumbent platforms being displaced over time. There are separate questions about how concerns about competition and speed of deployment can be converged with concerns about the safety and governance of these new technologies.

Discussion

Becky King

These comments highlight the regulatory challenges that cut across many of the other areas discussed in this panel as well as the importance of regulatory design, desired outcomes and unforeseen consequences.

Daria Onitiu

It is often a deliberate choice not to say that the intended purpose of consumer wellness devices has a medical purpose because this would require them to comply with higher standards for safety and effectiveness. With the proliferation and increased importance of these devices in our homes, there will be an increased burden on users to verify that their information is correct or that the Large Language Model is not hallucinating. Making medical technologies more accessible shifts the burden and that involves a number of trade-offs.

Daniel Andler

What is the basis for saying that convergence will increase? Convergence has been a topic for 22 years but it has not made enormous strides in that time. If an institute was established to study converging technologies, I wonder how busy the generalists would be compared to the researchers who explore convergence within very specific fields.

Pierre Larouche

We are discussing convergence from a business or technological perspective. But from a legal perspective, it is not really convergence. It should be noted that consumer devices (in general) are not a regulatory category: these products are subject to default rules, meaning that there are no restrictions on bringing inventions to the market. Medical devices are an exception to that default rule and are subject to additional, more protective rules. In a dynamic of business or technological convergence, the area under the default rule puts pressure on the area under the exception. History tells us that the default rule generally wins. There is a lot of pressure to reduce regulation in the area under the exception in order to align it more with the default rule. This is the legal challenge that we face.

Pawel Swieboda

I believe that convergence is becoming increasingly central, but this may be more characteristic of neurotechnology. There is nothing wrong with individual technological development of each domain, but convergence is creeping in everywhere. AI plus other technologies is the most pronounced and visible version of this. Progress in neurotechnology is only possible through interdisciplinarity and convergence. I believe that convergence should be given more weight and reflected more broadly in governance discussions and frameworks. I agree with your reading of the default dynamic, with the caveat that convergence puts much greater pressure on questions that have a legal underpinning.

Regarding the question about the shift of the burden of proof to the consumer, user or patient, it is interesting to consider at what point we will conceptualise certain red lines in this area. This requires real thought. The convergence of neuroscience, neurotechnology and AI is enabling the decoding of intentions or thoughts. This is being done in a very rudimentary way at present but the direction of travel is clear and companies such as Meta are achieving promising early results. We will need to decide whether we want a society where such abilities are deployed at scale. The only question is whether we want to have that conversation now or when the technology has matured further.

Tammy JihHsuan Lin

In the next three years, Meta Ray-Ban glasses will be sold as AR glasses with integrated EMG and neurotechnology. How can we govern these products in an agile way and how can AI support this? The convergence of consumer and medical devices is resulting in products like Apple Vision Pro, which can help the visually impaired to see, and AirPods that can support people with impaired hearing. There is a blurred line between the governance of commercial devices with medical use. We must have an agile approach to cope with the many issues, including privacy, that this convergence raises.

From the floor

I struggle conceptually with the definition of convergence. How different do the things that come together have to be? When is a particular technology a convergent technology? We need to be specific about what is new in convergent situations, particularly as convergence can compound economic risks and safety risk. We also need to take account of the complexity of technology governance across domains and the need for radical interdisciplinarity to foster the convergence knowledge base.

It has been noted that with convergence comes dispersion. One aspect of convergence is the way in which a technology can be dispersed quickly into different industries. This is a serious consideration for IP: is there a risk that the convergence cuckoo will lay its egg in various different industries and applications and become a part owner of them all? Can the market work these issues out efficiently through licensing, or does the IP system itself need to adjust to convergence?

Howard Shelanski

We must aim to ensure that the diffusion and diversity of innovation does not get lost as we move into the deployment phase of these new technologies. It is possible that a technology that is powerfully enabling could be co-opted through IP, network effects or feedback effects, into the hands of one or a few firms. One of our regulatory challenges is preserving the dispersion and diffusion of innovation efforts through the deployment phase so that we continue to have a diversity of providers, follow-on innovators and technologies. This becomes critical with reference to technologies that are truly convergent to prevent someone with a chokehold over a particular input or application from gaining market power or control over an increasingly diverse array of technologies in which that input or application is embedded or deployed. This is an important problem that is closely related to the standard essential patent debate: how do we make sure that one contributor of an essential input into a standard essential patent does not gain control or a chokehold? How do we ensure that the developer of an important technological advance that converges into a variety of different things does not gain an increased chokehold or potential control over many of those technologies, especially as it may be ill-equipped to develop them properly?

Session 4
The role of ethics in guiding the development of emerging technologies

Moderation: Daniel Nadal | OECD

Daniel Nadal

Ethics and governance are foundational issues that will determine the direction in which emerging technologies develop and the extent to which they respond to the needs of society. The need for ethics in emerging technologies is best illustrated in situations where a lack of attention to ethics has had negative consequences. For example, the development of algorithms with inherent biases reinforces societal inequalities; data has been misused for election interference; and a lack of ethics in synthetic biology could lead to the development of bioweapons and bioterrorism.

A number of methods are used to embed ethics into technology and policy development, including codes of ethics, guidelines, and ethics by design. Systems for establishing these and identifying relevant values include participatory approaches, expert opinions, governing boards, and advisory bodies. The wide range of tools and approaches available makes it necessary to identify and apply the most relevant options for each context.

This panel will consider why and how ethics affects the development of emerging technologies throughout a Technology's Readiness Level (TRL), and how it shapes the policy cycle and its governance approaches. The OECD Framework for Anticipatory Governance of Emerging Technologies – which aims to equip governments, innovators and societies to get ahead of governance challenges, has two pillars relevant to this discussion. First, the framework identifies 12 values that are shared by liberal democracies and that the OECD believes should permeate innovation and innovation policies. Six are foundational values, such as human rights, equity and democracy, while the other six are more specific to technology and include trustworthiness, transparency and responsiveness. The Framework recommends that these values are debated in diverse, multi-stakeholder communities with a view to identifying opportunities for application and operationalisation in different contexts, and embedded in the innovation cycle to ensure that technology develops in a direction that is acceptable to society. A second pillar concerns stakeholder engagement, which is widely agreed to enrich our understanding of issues, help us to understand and address public hesitance, and promote good communication and broad education across society. This involves standard approaches like capacity building, interdisciplinary education and consultations, as well as more innovative methods like citizen juries, science cafes and co-creation strategies.

Challenges and benefits of inclusive research and innovation

Claudia Werker | Delft University of Technology, Netherlands

Recently, a Belgian TV presenter was very seriously injured in a car accident. News coverage focused on the fact that he was driving a classic Porsche that did not have modern safety systems. Most of us do not drive old Porsches, but we are not equally protected by modern car safety systems. In fact, even though they have fewer car accidents, women drivers who have car accidents are more likely to die or be seriously injured. The reason is simple: crash test dummies are built like men who are 1.8 metres tall and weigh 80kg.

While discussions about values are important, we must also consider the level of compliance with those values that we seek to achieve. Society would probably agree that women should not be at a higher risk of death or injury on the roads, yet a higher risk is accepted in practice. This elevated risk also applies to men who are very tall, very short, very fat or very thin. This issue is not unique to cars. Healthcare, for example, is rife with solutions that are not inclusive. There is some understanding that cancer treatments need to be personalised, but the treatment of heart disease is much less so: women present very differently with heart failure and, as a result, they are underdiagnosed.

Inclusive research and innovation is a strategic priority of Delft University of Technology. In particular, we seek to convince engineers, physicists, economists and other scientists look at the diversity of human being, i.e. not only at wealthy, white men, when constructing their solutions. Responsible research and innovation also promotes anticipation, flexibility, inclusiveness and responsiveness. I will focus today on inclusivity in engineering.

In 2014, the Europe Commission launched an important initiative through Horizon 2020 to promote inclusive research and innovation and encourage STEM researchers to integrate sex and gender as drivers of scientific discovery. Intermediate findings from 2017 showed that uptake was disappointing. Research teams were able to point to diversity within their own teams but struggled to show how their research contributed to inclusive solutions. Even when the will is there, this goal is difficult to implement.

Inclusive research and innovation in STEM involves innovative agents exploring and exploiting the potential of human diversity in all its facets to drive scientific discovery and innovation. A group of European universities of technologies are working to drive this goal and conceptualise it through a systems approach that encompasses industry, government, the academic sector, entrepreneurs and civil society. The question is how economics and ethics can be brought together and at what level they should meet. The aim is not to have a few EU projects studying inclusive crash-test dummies that never reach the market. To that end, an EU-level study is exploring product standards with reference to human diversity, which is intended to feed into an EU directive to ensure that this happens. A macro intervention of this type could impose top-down rules or allow a bottom-up approach to flourish. In practice, rules tend to emerge at the messo level, i.e. on the level of innovative ecosystems where stakeholders meet and negotiate rules.

In a study on 18 engineering projects funded by the EU ATTRACT 2 programme, we asked research teams whether they paid attention to human diversity. In every case, the answer was ‘no’. The project leaders suggested that these issues were a matter later-on in the process when their solutions were commercialised. Still, they found it interesting and obviously so important that eventually in the end-survey ten out of eighteen projects had considered diversity of human beings. Two even had constructed inclusive prototypes. This shows that engineers are not unwilling to engage with these issues. Yet they follow the often criticised linear model of innovation, where research and innovation and commercialisation are following each other and done by different agents. When encouraged engineers are happy to consider the diversity of human beings to come up with inclusive solutions.

To provide society with inclusive engineering solutions, researchers and innovators must consider the diversity of human beings from the outset, rather than designing for men and adapting when the product hits the market.

Legal, regulatory and ethical tensions in the deployment of emerging technologies

Daria Onitiu | Oxford Internet Institute, UK

The Trustworthiness for AI auditing project at the Oxford Internet Institute is a multi-disciplinary project that brings together researchers working in computer science and AI ethics, law and social psychology. Our work explores questions surrounding the social and institutional norms, legal mandates, ethical values and technical constraints guiding the development and governance of trustworthy AI systems. My role focuses on legal and regulatory tensions around the certifications of new systems and tools for medicine, specifically AI and medical imaging and assistive diagnosis, as well as current challenges in the design, evaluation, and deployment of large generative models.

Two aspects of the regulation and governance of emerging tech, which are relevant to the framework for anticipatory governance of emerging technologies, merit urgent attention. The first is the need to balance the risks and benefits of emerging technology. The second is the need to understand how a forward-looking approach could apply to potential risks and benefits of emerging technology scenarios.

The design and future deployment of scalable AI solutions, particularly in the field of diagnosis, is becoming available at a time when healthcare is at a crossroads. Healthcare systems are under increasing pressure from underfunding, an aging population and increasingly specialised treatments. AI seems to be a desirable and appropriate solution for some of these major problems. There is a market for AI-based medical imaging tools and wearable technology, particularly for diseases that require early detection and specialised treatment and follow-up. In order to regulate AI to respond to the magnitude of these problems and ensure optimisation through innovation, it is important to identify the ultimate beneficiaries of this technology, the long- and short-term implications, and the trade-offs involved in minimising potential risks.

With regard to the regulatory ecosystem and agile approaches to AI design, development and standard setting, it is important to be clear about the values that underpin the safety and benefits of AI and safety-critical systems. There is a trend towards risk framing the operationalisation of certain principles, such as transparency, explainability and human oversight, from a very narrow and specific safety angle. In the literature about AI, medicine and computer science, a plethora of publications emphasise the performance and accuracy of medical imaging tools tested under lab conditions. Once these systems are deployed in the wild, these claims may be heavily overstated. Trustworthy AI principles are incredibly important and a good starting point, as are efforts to interrogate the reasoning that underlies developers' efforts to articulate the intended use of a device. There is a need to deflate some of the hype and overstated claims about the safety and effectiveness of new innovations, such as the new wave of large generative models.

Risk framing can have significant implications for patient safety and effectiveness on the ground. In practice, issues will be resolved by weighing priorities and ethical principles and making hard value judgements. Trade-offs and value judgements can be formalised in light of data quality requirements and safety, but the real question is whether the risk management approach limits all risks to an acceptable level and whether issues of fairness mean that the intended use of the device or the device itself must be restricted. These are very delicate value judgements that require us to consider AI medical devices from a different angle.

Balancing the safety and benefit of technology innovations requires us to weigh the anticipated benefits of the product and the claims surrounding the product in ways that put stakeholders, patients and professionals at the forefront of device development. In this context, anticipatory, agile governance requires a dynamic approach that allows approaches to social, cultural and organisational factors to vary across the system life cycle. To translate this into practice, developers, AI companies, regulators and other stakeholders require common incentives and methodologies.

To implement this approach in practice will require a shared methodological feature with specific stakeholder involvement, collaboration between regulators, industry, civil society and product users, and increased international coordination and understanding around the balance between innovation and legal certainty, as well as capacity building and institutional mandates that specifically oversee and monitor technological developments and respond to incidents. Formalising safety needs to go beyond technical fixes towards a concept of responsible innovation that translates ethical principles into robust, deliberative mechanisms and ensures that intended use does not only reflect anticipated benefits but also the articulation of claims around different tasks, aims and users.

Daniel Nadal

The idea of common methodologies and shared understanding is particularly important to help drive collaboration between scientists and engineers who can often find themselves working in expertise bubbles.

Generative AI in a university environment: a case study from the University of Melbourne

Jane Kaye | University of Melbourne, Australia & University of Oxford, UK

In contrast to the relatively abstract and high level presentations thus far, my focus will be on practical policy responses at the institutional level, specifically the importance of guidelines, engagement and capacity building. I will present my activities as an academic convenor in AI and research at the University of Melbourne and discuss the university's response to ChatGPT.

ChatGPT became publicly available in November 2022 and was immediately adopted by millions of people around the world. Generative AI platforms create enormous challenges across society but are particularly challenging for universities, which are part of the knowledge-generation industry and need to comply with standards for academic integrity and manage issues around academic misconduct, copyright and knowledge commons. Many large platforms are also now looking to consume reliable sources of academic information. The pace, potential, scale and adoption of this innovation is unprecedented and merits careful reflection, planning and future scanning. This is difficult to achieve when a technology is being rolled out and adopted exceptionally quickly. It should be noted that the regulatory landscape in Australia is currently very uncertain, with the Privacy Act being redrafted and the government exploring the introduction of something akin to the European AI Act.

Our institution had a cohort of early adopters and a long tail of users. We needed to bring everyone along and ensure no-one was left behind. Initially, the university focused on developing guidance for supervisors, lecturers and students; more recently, attention has turned to research. Many universities have been scrambling to develop standards in light of the lack of sector standards. Our university held focus groups and workshops to understand what was happening, which led to a report that included recommendations that led to the creation of the academic convenorship in AI and research that is tasked with developing policies across the university. My focus is on two particular areas: the role of graduate students as future academic leaders and their need to experiment in a safe environment; and the need to shift from risk management to innovation by developing a learning environment that is conducive to experimentation. I also intend to develop a governance structure to support innovation and have established an AI working group to coordinate policy and activities across the university. My co-convenor has established an AI community of practice. Academic ambassadors have been appointed in each faculty to build capacity, provide informal advice, and promote the idea of graduate researchers as future leaders.

Existing policy across the university has been reviewed and updated. In the absence of a specific regulatory framework, the university adopted ethical principles to guide this work but policy is developed through a very formal procedure. To identify a more responsive and agile approach that would be suitable for AI, we ran all existing university policy that mentioned research through the university's own Gen AI platform, Spark, and asked it to identify gaps and deficiencies. This has shown us where policies need to be updated and how that can be done. New guidance is being developed for graduate students and supervisors. We are piloting an AI prize for graduate students to build capacity, recognise the use of AI methodologies and innovation across the university, and encourage people to explore ethical responses to these issues.

By appointing ambassadors, running focus groups and workshops, holding seminars and growing our community of practice, we are better able to understand our constituency, identify concerns and issues at an early stage, build resilience, and develop value-based approaches to new technologies. My position is intended to provide coordination and leadership around the introduction of new technologies; similar roles could be useful at national and international level. The new governance structures are designed to support innovation, build resilience, and act as early warning systems. Time will tell how successful they have been.

Our example highlights the importance of guidance when the legal framework is uncertain. The university introduced ethical policies but also flexible, responsive guidance that is able to adapt to change. Capacity building should help graduate students to become leaders in this field and support the university as a whole to shift from a risk-averse management position to an approach that truly supports innovation.

Uncertainty, ambiguity and values: ethical implications of emerging technologies

Peter Mills | PHG Foundation, UK

The PHG Foundation is a health policy think tank and a charity of the University of Cambridge. Its purpose is 'to make science work for health', which means that we explore the normative conditions that enable science to work for health and aim to answer three kinds of questions: What sorts of innovations should we promote? How should we control them when we use them? How do we know them to be good?

Our work at PHG promotes a future of preventative and personalised medicine, though we recognise that some of the relevant technologies have yet to demonstrate evidence of clinical utility beyond relatively limited circumstances. However, we need to go further and look beyond the utility of individual technologies, to the social value of the *personalisation paradigm* itself. Certain types of technology and technological convergence are driving a fairly profound change in the way in which we can deliver health care. As such, it is important to explore how these technologies relate to norms and limits governing things like privacy, safety, security and even democracy. When thinking about the adoption of technology, we must not only think of its comparative utility but also ask what kind of world we are making.

New technologies do not necessarily develop along a linear trajectory; they are better thought of as being assembled from a range of different knowledges, practices, products and applications. The challenge for regulation is to track, influence and circumscribe the ways in which this assembling takes place. Prospective technologies are often characterised by an uncertainty that takes us outside the scope of simple risk management and into something that is more difficult to manage because of its multiple dimensions. The narrowness of certain risk-benefit framings has already been mentioned. A second characteristic to which we should attend in addressing emerging technologies is ambiguity: the idea that different people place different value on different outcomes and features. Often, in the case of emerging technologies, these questions are initially ignored or left unresolved. A third characteristic concerns the transformative potential of emerging technologies to affect and restructure norms and social relations, for example through the insidious effect of high frequency, low impact outcomes. These risk being overlooked due to more obvious intentional, high impact outcomes. When we encounter uncertainties, we are obliged to take account of different kinds of knowledge that potentially reveal insights that are not about the contingent purposes of the technology. It is critical that values in bioethics are explored in ways that address questions of ambiguity and of the difference and distribution of meanings.

As an example, these are all in evidence in our recent consideration of AI-assisted pathology, where it can be shown that an AI could be more reliable than trained pathologists in analysing several thousand case slides. The point here is not assessing the AI's superior performance characteristics but the challenge that we are broaching when we consider the pressures to change, quite reasonably, the norms of diagnosis, potentially eschewing human input in the pursuit of the best achievable standard of care.

An interdisciplinary bioethics is a discourse of experts, but it is not an expert discourse – or should not be, in my view. It should be practical and problem-focused rather than knowledge-based, recognising that the questions being addressed are owned in common by people with shared interests. This is relevant to another question that has been raised about researchers shouldering additional cognitive load: it is impractical for an individual researcher to synthesise these epistemologically diverse reflections in their own person. As a result, the question becomes one of the relation of discourses, and our capacity to produce the conditions of their productive implication, always in a particular situation, faced with a defined problematic.

The common objection that emerging technologies lead us down a slippery slope arises from a failure to appreciate that norms respond to developments in the sociotechnical horizons that are present. To adapt Geoffrey Vickers' apothegm regarding public health: the moral history of technology is the history of successive redefinitions of the unacceptable.

Finally, I'd like to recognise a feature of the practice of bioethics are extremely important. Broad engagement has contributed enormously to informing ethical reflections on emerging technologies, both from a substantive point of view, by providing relevant but sometimes obscure insight, and also from a normative perspective, by ensuring that reflections on emerging technology respect epistemic justice. If we assume that all people are the same, we risk developing biased technologies that fail to account for very different forms of human embodiment and sources of human interest.

The innovation ecosystem should include all of those whose interests are engaged by it because no technology will be successful if it does not take account of those interests, and because the outcome will be improved if interests are allowed to inform, shape and produce the conditions of innovation. But we cannot have every innovation that is proposed. As such, from a societal perspective, we should value resistance to innovation, which potentially makes the innovations we have better.

Discussion

Daniel Nadal

Thank you to all the panellists for their remarks. It is indeed important to remember that the norms accompanying guiding values are malleable, and can evolve and be shaped over time.

Erik Fischer

My understanding is that standards for automotive crash testing are established and updated through an international or possibly transatlantic legal framework and that the US automotive industry is the primary force opposed to upgrading these standards. Is this correct?

Claudia Werker

I focus on EU legislation and am not familiar with how standards are set in the US. I understand that the US usually has lower safety and security standards, for example with regard to the safety of pedestrians.

From the floor

A number of speakers have mentioned the central importance of stakeholder engagement. How should we engage with stakeholders given that many or most of them do not have the skills and training required to understand the technologies in question?

From the floor

When developing applications that focus on differences to promote inclusivity, how can we ensure that this focus on differences will not result in tools being used to promote discrimination? We tend to think that our basic ethical principles are stable, but in practice the way that they are interpreted and enacted can change significantly over time and across geographies. How can we take account of this?

Claudia Werker

When considering human diversity, it is important not to fall into ableism or techno-ableism. This refers to the practice of assuming that the goal of people without an ability is to attain that ability, for example, that someone in a wheelchair must want to be able to stand and walk. It is important to respect people's wishes rather than assuming how they want a product to be used. It is important to think, talk and listen.

From the floor

There are also issues around differences in ethnicity and so on. At some point, people will exploit these issues. This needs to be integrated into our thinking.

Claudia Werker

Ethnicity can be relevant for health applications. For example, people with an Asian background metabolise medicine in a very different way from Caucasians. The aim is to respond to differences rather than simply identifying them, as long as that is what is wanted. If the difference is recognised in order to respect people's expressed desires, there is no discrimination.

Daria Onitiu

This is a narrow point about stakeholder involvement and product safety, but it should be noted that an important and longstanding requirement in medical device regulation is usability. This gains new importance with AI. It is extremely important to ensure that a system works with the intended user in the intended environment. Claims that a system works must be supported by usability studies in the target environment. This simple principle is easily overlooked, but it becomes an important piece of the puzzle when we consider that some AIs go beyond augmenting certain processes in order to assist decision making.

Peter Mills

It is possible that stratified or personalised medicine will disadvantage certain people, particularly because those suffering the greatest burden of ill health tend to be those who are least likely to be included in data sets and consultations. There are therefore both economic and moral reasons to ensure that we cover the whole population and argue for greater equity in those areas.

There is a risk of perpetrating epistemic injustice if stakeholders who might not have the capacity to grasp the full detail of a technology are excluded from the discussion. People do not need to fully understand the operation of a particular technology in order to discuss the values that should govern its use. In order to discuss public policy, we implicitly try to construct a sense of the public interest through engagement with the notional public, which includes people with different capacities and interests.

Jane Kaye

I agree with your comments on justice. In society, there are always people who know about things and people who do not. We organised a public participation and involvement panel as part of a project on AI and health that included seminars to increase participants' knowledge base, but we also tried to ensure that our exchanges were appropriate and proportionate and that we only sought advice in areas where their input could affect outcomes and public involvement in decision making was crucial. We must be aware of where and how we involve people, so it is not a burden to them or the broader engagement process.

Session 5
Using soft law to bridge business and government objectives on emerging technologies

Moderation: Joëlle Toledano | Governance and Regulation Chair

Joëlle Toledano

The governance of emerging technology is not a straightforward process. Soft law is a flexible approach to governance that is used for emerging technology and areas with limited information. It can support agile regulation by enabling public authorities to influence industrial practices and incentivise industry to engage in self-regulatory efforts and prevent risky practices. This panel will consider how effective soft law is for governing new technologies.

Soft law and innovation: examples from Meta

Béatrice Oeuvrard | META, France

At Meta, we believe that innovation is key to developing new technologies and that no technology has more potential to boost EU competitiveness than AI. Goldman Sachs estimated in 2023 that generative AI could boost global GDP by 7% over 10 years.

A major obstacle to economic development and innovation is fragmentation, which creates over-regulation and creates issues with regulatory consistency. For example, a French start-up aiming to scale up across the EU needs to navigate 27 different intellectual property laws, licensing rules, and data protection authority. Over the last 10 years, regulators have created around 100 tech-focused laws. There are over 270 regulators active in digital networks across all member states.

Meta recognises the importance and the positive impact of major European laws like the DSA, the DMA, the AI Act, and GDPR. Nevertheless, we identify two major issues. First, that national governments and local regulators try to create too many specific national laws. Second, that there is a lack of clarity in the way that different regulators understand, implement and interpret these laws. The Draghi report expresses similar sentiments about the impact of this institutional setup. It highlights regulatory uncertainty as a major issue, noting that the EU claims to put innovation at the centre of its activities but basically does everything it can to keep it at a low level. GDPR is estimated to have reduced the profit of small tech companies by 15%. Regulation balanced by innovation is key for economic development in any area.

Soft law is important as it provides more space for innovation. It can promote the deployment of innovation by establishing guidelines, encouraging transparency and accountability, fostering collaboration and knowledge sharing, providing flexibility and adaptability, encouraging industry cooperation and, ultimately, reducing regulatory burden.

I would like to share some examples of how Meta is deploying soft law. Meta published its first Facebook transparency report in 2013. This biannual report provides insight into government requests for user data, content restrictions based on local law, and other measurements that can affect user privacy and security. Over time, the scope of the report has grown; it now contains similar information to that required by the DSA. In 2014, we launched a user-facing AI tool that enables people to better understand and more easily control what they see on their newsfeed from friends or advertisers. This feature can be seen as a form of self-regulation, aligned with soft

law principles. By voluntarily providing transparency tools, we address users' concerns and social expectations about what people see on our platform. A series of system cards further help our users to understand how multi-modal generative AI systems work.

In February 2024, we committed to labelling AI-generated images on Facebook, Instagram and Threads. We have been working with industry partners to develop common technical standards for identifying AI content, including video and audio. We label images as AI-generated when possible, using a visible marker for users and an invisible watermark that allows other platforms to also label the content. This approach was developed with a group of industry players to improve the robustness of our systems and share better tools across the industry. We have also worked with a non-profit community of academics, civil society, industry, and media organisations to help AI to improve transparency across society. Our approach to watermarking exists within a synthetic media framework that provides specific recommendations on the responsible development, creation and sharing of content that has been generated or modified with AI.

Meta is a founding member of ML Commons, an AI engineering consortium built on a philosophy of open collaboration. Its mission is to create AI innovation and increase AI's positive impact on society by developing open industry-standard benchmarks and diverse data sets for model evaluation. These initiatives reflect Meta's open-source philosophy and belief that open source can result in better, safer products due to increased scrutiny. Vulnerabilities in open-source models are openly identified and discussed by the community, helping to create a repository of knowledge that the community can leverage to develop and fix any bugs. It speeds up innovation and maximises adoption by enabling researchers and academics to build on each other's achievements and reduce barriers to entry for start-ups and small businesses. Finally, open source can increase competition. This is key in the context of AI and emerging technologies. The LLaMA model, based on open source, has been downloaded 350 million times.

Soft law can be an effective tool in bridging gaps that exist between business and government objectives and emerging technology. The challenges associated with soft law can be addressed by establishing clear guidelines, encouraging transparency and accountability, and fostering collaboration and sharing. Soft law can also promote flexibility, adaptability and cooperation and reduce the regulatory burden. Meta believes that it is essential to continue to explore the potential of soft law in government and to develop innovative approaches that balance regulatory needs with industry interests.

Soft law and the regulation of AI: why transparency is key

Daniel Andler | Paris-Sorbonne University, France

AI governance must rest for the most part on soft law, but not on soft law as it exists. In my view, the relationship between ethics and soft law is simple. As the late philosopher Joseph Raz explained, ethics is the endeavour of giving substance to the abstract category of the good. I see it as an exploratory process, whereby experience, that is induction from cases, confronts common sense convictions, values, governing preferences and behaviour. This exploration process is accelerated in phases of novelty brought about by cultural, social, economic transitions or cataclysms such as war and technological innovation.

To provide guidance to society or relevant stakeholders, the state of play must be summarised as a set of principles and norms of conduct. These can be implemented through coercive force in the form of hard law, or applied through the persuasive force of soft law. Hard law is adapted to cases where the objective situation is stable and well known, there is a consensus on the normative situation, and there are effective ways to enforce the norms. Soft law is better suited when one or more of these three conditions is not satisfied. As fewer conditions are satisfied, the case for soft law becomes stronger.

AI appears to be a paradigmatic case where soft law is the better candidate for governance because the objective situation, namely the state of play in AI, is neither stable nor well understood. The technology is ill-defined, there is uncertainty about its nature and perimeter, and it is far from stable: it has evolved at an accelerated pace, is likely to continue to evolve rapidly, and its medium and long-term effects are disputed.

There is also no consensus on the normative situation, namely the values that should be defended, the priorities that should be enforced, and the risks that require the most urgent interventions. It is not clear how to weigh innovation against regulation or whether this is even a valid concern. Furthermore, between the three major blocks, namely Europe, China and the US, and between the different stakeholders within each block, and between the people and communities within each society, there is increasing complexity and diversity regarding perceptions of respect, fairness, sustainability, justice and so on. Enforcing the rules is particularly difficult due to the possibility of corrupting all-purpose AI systems and the difficulty of identifying malignant agents.

These factors mean that it would be very hard to apply hard law to AI. In addition, there are positive arguments in favour of soft law. It sets up provisional guardrails that can prevent or at least limit some of the more obvious damages, its agility means that it can be adjusted as the technology evolves and its applications expand, and it is a testing ground that makes it possible to explore and deliberate on ethics as AI generates new situations and gives rise to unprecedented situations that require ethical innovation.

Nevertheless, the mere fact that soft law is a better candidate than hard law does not necessarily make it a good candidate. There are problems with AI soft law as it stands. It exists mostly as sets of principles laid out as recommendations, guidelines, charts and codes. This approach does not improve with age. A cumulative process would allow questions raised by a proposed code to be answered from the field, resulting the development of new principles and the posing of new questions. We would move from one chart to a more focused, more streamlined chart. Except that, in practice, this does not happen. In 2023, an inventory of AI soft law standards identified more than 200 charts. A more recent survey identified 634 programmes involving up to 107 variables, catalogued according to 15 social themes and 78 sub-themes.

This proliferation of principles makes soft law less readable and less persuasive for real stakeholders, namely those who are in the business of either producing or deploying AI systems. This is a major problem given that soft law operates by way of persuasion. One reason why stakeholders do not play close attention to the charts is that all-purpose AI is, by vocation, hard to pin down. Stakeholders rightly feel that the real issues arise when AI systems are deployed in specific areas of professional, public or private life, close to their respective concerns.

The remedy for this is to go regional, as is done with nanotechnologies. General nanotechnology ethics mean relatively little, because they simply concern the scale of what is being done. With AI, ethical exploration requires experience and judgments that are only available within a given field of practice. In each of these fields, existing professional ethics for that field pre-empt general AI ethics, which are adjusted accordingly. The same applies to general use systems such as ChatGPT that are deployed in the private and public field: the norms of public and private life set the rule and priorities.

An objection to the regional solution is that there are commonalities, such as privacy, that could be more economically dealt with through broad principles. This sounds plausible, but on closer examination these demands apply not only to AI but to all providers of goods and services. AI-driven interventions mean quite different things in different areas of deployment: music algorithms, consumer goods, robotic surgery and lethal autonomous weapons do not require the same levels of transparency and responsibility.

In some cases, these questions make immediate sense; in others, they require conceptual clarification and extensive experience of the system in question. There is, however, one exception. The one feature of AI that is intrinsic, in the sense that it is independent of any particular intended application, and that requires across-the-board regimentation is that it mimics the human person. AI systems can pass off as genuine persons and they should not be allowed to hide their true nature. That is the transparency requirement in the sense of the AI Act. This transparency requirement should be part of hard law.

Soft law and hard law in the context of emerging technologies and the AI Act

Pierre Larouche | University of Montreal, Canada

Lawyers are often mesmerised by the fact that soft law has no binding effect or, depending on the text you read, no State enforcement to back it up. These two points are not equivalent: the absence of sanctions in the traditional sense does not mean there is no binding effect. The real question is what that binding effect means in the absence of State enforcement. Many lawyers are puzzled by this and argue that the lack of State enforcement negates the legal nature of soft law. However, legal theory long ago abandoned the idea that State enforcement is the hallmark of law. Unfortunately, this formalistic view of law lives on in other disciplines: engineers, scientists and others often express the idea that, if there is no State waiting to punish breaches, there is no law.

A more sophisticated view would argue that law is about the normative realm: the conduct of actors is influenced by a normative account of what they should do. My analysis is based on a scheme developed by Professor Lessig at Harvard 30 years ago. He argued that four forces influence the behaviour of actors: the law, market forces, social norms, and technological architecture. From this perspective, soft law is a hybrid: it contains norms of conduct, but their binding force comes from one of these four forces (including other legal instruments). This view raises very interesting questions about soft law.

The first question relates to the origin of the binding force. This could be *market pressure*, for instance in the form of rules on disclosure and transparency or "voluntary" international standards like 5G. It could be *social norms*, as with "comply or explain" systems. Furthermore, standardisation, even when voluntary, creates constraints by affecting the *technological architecture* that surrounds us. *Hard law* provides additional persuasive force in some other cases. US merger guidelines, for example, are effective at changing behaviour because, without compliance with them, there is the threat of enforcement of the actual merger control legislation. This raises further issues around the effectiveness and operation of soft law instruments.

Secondly, there is the question of who issues the soft law instrument. It can emanate from a State entity but it can also derive from private actors, as occurs with transnational private regulation. Although the State entity may appear to be more the legitimate actor, procedural considerations can call this into question.

The final questions concern the relationship between soft law and hard law, best expressed (by Senden) as a distinction between the pre-, post- and para-law functions. A soft law instrument can be a precursor to hard law, an interpretative instrument applied after the fact, or an alternative to law. This has an incidence on the legality of soft law: a precursor conditions what comes later and is thus relatively immune to legality challenges, but a post-legal instrument cannot detract or stray from the hard law instrument it is supposed to interpret.

By way of illustration of the above, the EU AI Act leaves significant space for soft law instruments to be enacted but also provides for a number of hard law instruments. I will explore three major categories: guidelines, standards, and codes of practice.

Soft law guidelines can be issued on any topic. The Commission has a general power to issue guidelines under the AI Act, and in addition the AI Act lists specific topics where guidelines can be expected. Legitimacy issues arise if there are attempts to use guidelines to circumvent political difficulties. Legal issues arise if guidelines stray from the Act or take it in a different direction. Tensions arise due the fact that the Commission is typically bound by its guidelines but national authorities are not and, at best, have a "comply or explain" obligation. The judiciary is not bound by guidelines at all. The Commission has already stated that it will issue guidelines on the definition of high-risk AI systems. It is likely that these will be exploratory guidelines intended to develop some issues and gather practice, possibly with a view to developing hard law at a later stage.

Harmonised standards are a central element of the architecture of the AI Act; compliance with these standards is presumed to be compliance with the Act. This raises issues of legality. The legitimacy of these standards is also questioned on the basis that the issues at stake with AI standardization are not purely technical, e.g. the shape of a power plug, but serious issues of public policy, privacy and so on. Standards organisations are definitely capable of handling public policy arguments, but the Commission will bear the onus of deciding if the standard conforms to the public policy requirements set out in the AI Act itself. The AI Act is intended to influence the standard-setting procedure and bring it closer to a democratic ideal. This is nice in theory but less realistic in practice. External organisations are likely to produce global standards, which will oblige the Commission to decide whether these standards meet European ideals or whether the EU needs to develop its own, separate standards.

The AI Act Code of Practice for general purpose AI (GPAI) models is being drafted and should be in place by May. There are already questions about whether the codes go beyond what is provided for by the Act and, more importantly, whether they will be able to improve upon the legitimacy of the standard-setting process. One thousand stakeholders have been registered as participants in this process, but the action appears to be taking place behind the scenes, which compromises legitimacy. The extremely short deadline may affect their effectiveness. There is a danger that we will seek refuge in a more formal compliance model which could seriously undermine the effectiveness of the Code of Practice and lull us into a false sense of security without effectively addressing security risks.

Soft law and innovative technologies: an economist's view on risk management and policy

Vladislava Bar-Katz | Frontier Economics, UK

Rather than asking whether hard law or soft law is better, it is interesting to ask whether there is any space for policymakers to support firms to adopt soft law and which incentives could persuade companies to adopt soft law. This assumes that soft law is beneficial and works effectively alongside hard law, and that firms seek first to maximise profits. Under these conditions, firms will adopt non-binding measures like assurance, standards and guidance if profits are lower without them.

There are probably three ways in which innovative technology can have a negative impact on the profits of a firm. The first is production impacts, whereby a firm adopts a technology with a view to obtaining productivity gains and experiences a direct impact on its profits if the expected gains are not achieved. This situation can be managed by outcomes testing, which tests whether systems, including AI systems, perform in the intended way so that companies can assess whether or not they will beneficial.

The second potential impact is more interesting and relates to how innovative technologies might affect or reduce sales. A tool that does not perform as expected, such as a hospital AI imaging tool that fails to diagnose patients correctly, could cause reputational damage. The firm adopting the tool only be affected if the harm is visible to consumers. Social harms, such as poor hiring decisions caused by a biased AI recruitment tool, can be more difficult to identify and to trace back to the firm. Even if tools were made available to prevent these kinds of damage, companies might not adopt them because the uncertainty about the potential exposure to harms and subsequently risk. Sales reduction can also be driven by a societal demand for assurances or non-binding measures to be adopted. This requires significant social awareness of the risks associated with those technologies and of their preferred approach to AI assurance and transparency.

The third potential impact is investment reduction. Many businesses rely on external investors to finance the adoption of new technologies. In order for investors to be incentivised to demand soft law or non-binding measures, they need to observe or anticipate the two potential impacts previously mentioned and be aware of the associated risks.

Even assuming that non-binding measures are good and create good outcomes, the adoption of these measures faces significant market barriers. In particular, companies are unlikely to adopt them unless there will be an observable, or clearly anticipated, impact on their profits. As such, there may be a role for government to support the adoption of non-binding measures. In this scenario, it can be assumed that companies would use different tools to assess the production impact of the AI or new technology and ensure it performs as intended. Multi-disciplinary teams would need to create those measurement tools and develop different models, assurance tools, or even non-binding measures.

It is unlikely that companies fully recognise potential risks around sales reduction at present. This may be due to a lack of understanding among senior management, or because there is a lack of tools to assess them in economic terms, for example by attaching a monetary value to the failure to consider those risks. It is important to ensure that consumers are educated about potential risks and non-binding measures. The example of GDPR shows that a highly educated subset of the population can drive the social education of the whole society. The situation with investors is similar.

Policymakers need to find ways to convince senior management teams that prevention is more cost effective than treatment after the fact. The risk of unknown future costs and harms is already preventing many people from adopting AI and innovative technologies; other people have unfounded confidence that things will work out fine in the end. As a rule, thinking about things in advance is a better approach than trying to change systems after the harm is done.

Discussion

Miguel Amaral

The idea that soft law and hard law work in opposition to each other can be misleading. A lot of value can be created by achieving the right combination of policy options. Outcome-based regulation, combined with soft law and policy guidance on how to comply with outcome-based regulation, could be extremely valuable for governments and for businesses. It will be important to choose the right mix of policy options.

Pierre Larouche

There is too much emphasis on opposition between hard and soft law, and not enough on how these two realms complement each other. For instance, in EU competition law, we have a model where the law is fairly general and clear ("you cannot abuse your dominant position and you cannot form cartels") but the elaboration of the details around these general statements is more flexible and is covered by soft law. This sort of model could emerge in AI regulation. The AI Act could turn out to be a statement of what we expect of AI as a society, but the implementation of these principles could be left to soft-law instruments. At that point, we would see the emergence of different interpretations in different sectors, with the AI Act becoming the common point of reference. In that case, there would be merit in keeping the legislation relatively short, to make an impactful statement of what we want as a society, without attempting to pin down the details and the trade-offs at the legislative level.

Béatrice Oeuvrard

This is not only about regulation but also about governments. There are already difficulties establishing the legal basis for AI within GDPR: is it legitimate interest or something else? The issue is not GDPR but the fact that the regulator is giving an interpretation without clear guidance. The consequence is that we need to repeatedly postpone the launch of a product in Europe because we do not have the right type of reactivity. There is hard law, but there are also questions about how hard law and governance are working.

Unfortunately, hard law is still seen as the main goal of legislators, regulators and politicians. Soft law has more industry involvement, requires more consultation, takes more time and requires more iteration. But soft law offers good value and offers many benefits. We should educate stakeholders more about the value and economic impact that it can provide.

Erik Fischer

I believe that 15 to 20 years ago, one of the US agencies had a soft law on the self-reporting of nanotechnology materials; response rates were extremely low. More recently, President Biden issued an executive order on AI; as far as I can tell, it has had no effect, even in symbolic or psychological terms, on anyone who was not already on board. These seem like failures. Can you think of any soft law success stories around emerging technologies?

Daniel Andler

Bioethics is often presented as a soft law success story. For example, the conversation around CRISPR-Cas9 is progressive and gives the impression of a community converging towards a set of values that are sanctioned by soft law. It is not entirely coincidental that a well-known effort to develop AI ethics is based on bioethics. This may also be intended to give AI ethics an air of respectability that otherwise evades it.

Vladislava Bar-Katz

If a society wants non-binding measures, it needs to understand the incentives and the blockers with regard to uptake, and the risks and expectations involved in acting or not acting. If there is not sufficient pushback from society to reduce sales, or if the probability (or the perception of probability) of something going wrong is too low, the measures will not work. There will be examples where the measures do work, to a certain extent, but these elements must be borne in mind when considering the motivation for action, the purpose behind it and the barriers that companies face. With AI, we do not yet fully know the future risks or costs of not abiding by these standards.

Éric Brousseau

There has been no mention of implementation and enforcement bodies: is this an elephant in the room? It is possible that the debate should not be organised around the distinction between soft law and hard law as these two styles of law are complementary. A legal standard or legal norm operates through a process of interpretation. This is key, as it establishes the norms and allow to assess whether a behaviour is in line with the norm. There are many potential vectors to interpret and guard norms, from judges and professional communities to industrial unions, citizens' assemblies, civic activists and so on. These actors play a strong role in setting behavioural norms, but they are often entirely ignored. The problem arises with the implementation of legislation, the clarity of norms, and the legitimacy of the bodies and actors that translate generic principles into behavioural norms.

Béatrice Oeuvrard

It is important to consider the way that authorities are working with the global ecosystem on this vital topic. This is a significant topic: it is the legal basis of AI and it has an impact on all types of companies and economies. We were shocked by the lack of public consultation on such an important topic. We leveraged the discussion to ensure that all stakeholders were aware that something was coming that would change the model for many companies. There has been a very dogmatic approach whereby the political aspect does not feature, economic aspects have a minor impact, and we are stuck in a position of trying to engage in discussion. It is interesting that the GDPR enforcement regulation is coming at EU level at the moment. I think that it is a missed opportunity.

Joëlle Toledano

Why do you use the term 'soft law' to refer to what, in my view, is self-regulation?

Béatrice Oeuvrard

We call it soft law because it is developed with other industry players across the ecosystem and because it can involve academic researchers, civil society, and bodies external to Meta. Self-regulation would apply if we were simply providing our own regulation. The transparency report that we launched includes an audit to ensure we can review the metrics and methodology that we choose to apply in that type of situation.

Pierre Larouche

Some elements of soft law are really self-regulation. For example, transnational private regulation sometimes delivers standards out of nowhere because they are wanted by the industry. But in many cases, as with the AI Act, public authorities are involved and provide direction about what they want to see. This can be described as co-regulation or co-construction.

Joëlle Toledano

My point was about the example of the transparency report. In my view, even if four companies decide to prepare the same report, this remains an example of self-regulation rather than soft law.

Pierre Larouche

It can be seen as self-regulation, but transparency is a legal requirement and the reporting is undertaken with a view to avoiding more direct intervention by the public authorities. There are times where an industry is simply doing its best. There are also times when action is taken with a view to avoiding legislation or regulation. It could be described as self-regulation in the shadow of the law.

Private standardisation bodies have their own mechanisms to interpret rules and generate further precisions as required. Usually that is good enough. If things become contentious, you leave the realm of persuasion and enter the realm of coercion. Lawyers will then enter the picture and try to litigate to pin down the interpretation to the last detail. However, as long as there not too much litigation, all of these instruments will have their own features for interpretation. The non-litigious elements of implementation work can be done within an organisation. I could describe some success stories if that is of interest.

T***ammy JihHsuan Lin***

Regarding the distinction between self-regulation and soft law, is there a need for third-party evaluation to assess the public impact of, for example, a transparency report? It is being suggested that this kind of report becomes soft law at the point where it goes beyond responding to government requirements and starts to affect an entire ecosystem. How do we assess the impact in a way that differentiates between soft law and self-regulation? Meta has done a good job on these reports, but are there any evaluation methods or organisations in France or Europe that evaluate digital trust projects of this type? In Taiwan, a third-party alliance formed of industry and academics provides annual evaluations that have a binding effect. What is the situation in Europe?

Pierre Larouche

We had a discussion in the Code of Practice forum. It was argued that, while it is good to have third-party evaluation, it is necessary to identify appropriate expertise that is capable of conducting that evaluation and is external to the main providers.

Béatrice Oeuvrard

We had relatively few options. We could have waited for legislation to be developed but we also wanted to respond to user requests for increased transparency. We rarely act entirely independently and we are supported by expert academic researchers to ensure that we are making the right choices.

Joëlle Toledano

There are big debates in Europe around transparency reports.

David Winickoff

One of the key activities of the OECD is to establish soft law through recommendations and guidelines.

Soft law is a promising approach for international governance: it is not binding, it is flexible, and it helps to establish expectations, guidelines and norms. However, as in other kinds of international law, there is competition between standard-setting bodies, forum shopping, a proliferation of instruments and issues with institutions competing for influence. It is productive but also difficult to collaborate with other international institutions and ensure that instruments are in sync.

There appears to be a narrative in technology that predicts a virtuous evolution from soft law towards hard law. Does that always hold, or are there certain situations where we want to provide strong or pre-emptive signals that some things are off limits or subject to controls? Do we want an evolutionary regime for bioweapons or human germline engineering for hereditary purposes? Are there cases where we want a pre-emptive approach that introduces strict, hard laws that can be softened through regulation?

Daniel Andler

I believe that we should have stated, at an early point, that it is not acceptable to make systems that could be taken to be real persons. If we had had that law, about half of the applications that are poisoning our world right now would not exist. A world without deep fakes would be a better world. The AI community, whether because it lacked perception, lacked foresight or prioritised technical fun and profits, chose to go ahead. That is my opinion. Other people might argue that making fake people is not a significant issue and does not merit a hard law. This subject can generate deep disagreements.

Closing session

Éric Brousseau

Thank you very much to all of speakers. The chairs of each panel will now share their concluding comments.

Douglas Robinson

We argued that anticipation, reflexivity, inclusivity and responsiveness are key concepts. The panellists presented work on new genomic technologies and risk assessment answers, described case studies around AI and COVID, discussed the challenges of implementing diverse governance approaches, and showcased new agencies, tools and approaches. They emphasised that evidence, knowledge and intelligence are performative, representative and shape what we do and presented a number of new processes, methods and forms of knowledge that must be absorbed into existing institutions. Accountability in decision making was discussed, as were questions about assessing evidence of intelligence in an evolving, moving target. Can knowledge ever be robust? Can processes of moving targets in terms of knowledge be robust? And how do we handle that if we are going to arm and feed our agile mechanisms?

Miguel Amaral

The panel highlighted the potential to go deeper into the understanding of technological developments: grasping the complexity of innovation processes seems absolutely key to further refining our approach to the governance of emerging technologies and innovation. They also pointed out the need for a change in mindset and highlighted drivers of change including the incentive regime, trusted environments, capacity training, and data acquisition. All of these points require further consideration going forward to ensure that the framework reflects reality.

The need for regulatory foresight and anticipation came through strongly for two main reasons. First, to avoid badly timed and unsuitable regulatory actions, and second, because anticipation is the best way to embed governance and regulatory approaches in the innovation process. International coordination will be critical for the implementation of agile approaches to global challenges, especially as regulating agile approaches in isolation is not likely to be successful. The OECD has been doing a lot of work in this area to issue practical guidance and drive this agenda forward. The paper you are developing is also an excellent step to improve practical guidance, collect more case studies, showcase good practices and derive lessons that could inspire other countries and the private sector.

In my view, two elements are missing: practical guidance on the right timing of different regulatory approaches; and clarification on the right mix of agile approaches to regulation. We have some ideas but we are not in a position to offer precise and tailored recommendations for governance. Finally, this agenda on bringing together technology policy communities and public governance communities raises a need for data to ensure that the right people are in touch and that decisions are based on the right information. Building evidence-based bridges between communities could be a significant challenge going forward.

Becky King

We ended our session on converging technologies by highlighting 3Ds: diversity, dispersion and diffusion. The discussion of diversity raised questions about the scope of technologies and the complexity of different combinations of technologies. Because innovation is multi-sectorial, we need to adopt a broader view of managing and regulating these issues and engage more and different stakeholders.

In terms of dispersion, we highlighted that different technologies do not develop in the same way, particularly as they spread out. The examples of neurotechnology and AR-VR showed that these technologies can end up having very different risk and governance profiles in different applications. Many governments are facing the challenge of tackling that dispersion from a governance perspective. The conversation on diffusion showed that, while convergence is not new, the current scale and pace of convergence and diffusion is causing major headaches for governance. The example of AR-VR highlighted that a lack of standards at the start may tie into aspects of soft law initially but could become an impediment as the technologies evolve further. Speakers alluded to the fact that governance is not simple and that companies are able to skip or circumvent systems that do not align with their objectives.

This brings me to a fourth concept: dynamism. Our technology and our values are inter-related but many emerging technologies are driven by large organisations. How will this shape our ethics and values? An individual has certain ethics and principles, but these can change if they are expressed through an organisational structure. The choice of what to regulate and when should be informed, in part, by the shape that we want our future values and ethics to take. The panel highlighted the need to demystify emerging technologies, build trust, promote stakeholder engagement and develop common languages for discussing issues and technologies. We need to learn from history about what regulations have done to promote competitiveness and fairness in the past. Finally, we emphasised the need for cooperation and coordination in governance.

Daniel Nadal

Three main themes emerged from our panel. The first concerns the broad role of ethics in determining the innovations that we should be promoting. Emerging technologies have the potential to reorder societal structures and norms and redefine the acceptable. Ethics are not set in stone and the norms and limitations carried by guiding values can be operationalised in different ways. The drivers for obtaining and applying ethical intelligence can come from top-down and bottom-up approaches, but the rules often emerge at the mezzo level in the ecosystem of actors.

The second key theme is inclusivity as a tool for anticipation. Considering inclusivity from the outset may prevent issues arising further downstream. While we should be making evidence-based decisions, it is key to question the robustness and inclusiveness of the evidence, and obtain ethical evidence in an inclusive manner through broad stakeholder engagement. At the same time, it is important to be selective about when and how we engage people to avoid overburdening them and to ensure that their input can have an effect.

The third key theme was multidisciplinarity. This could help researchers to deal with additional cognitive load of implementing ethical governance approaches, but we should not be satisfied with just having diverse teams - they need to translate into inclusive research outcomes. There is a need to be explicit about risks and benefits, perhaps via shared methodologies, to avoid narrow framing that could ignore effectiveness and overblow claims. This is intricately linked to the need for capacity building and international collaboration.

Joëlle Toledano

We started with simple definition of soft law, no binding effects, to more sophisticated, influencing behaviour of actors, and noted that a lack of clarity around enforcement mechanisms for soft law could raise issues of effectiveness, legitimacy and legality, as has been seen with aspects of the AI Act. Hard law and soft law are often complementary but not always. Soft law is thought to give more space for innovation due to its focus on interpretation. Hard law imposes regulatory burdens and increases the risk of national fragmentation. While enforcing rules could be difficult, soft law might in practice not be alone the right answer. For AI, usage and professional ethics could be the baseline for an ethical framework, with hard law preventing «mimic person".

Daniel Andler

We have failed to consider end users in our discussions. They also have an ethics and should also be subject to soft law and maybe hard law. Carmakers have certain obligations but so do drivers. There is an interplay between the different things that end-users and firms are subject to or tempted to violate.

David Winickoff

This was an excellent day of exchanges that probed many issues and revealed some interesting questions along with the outlines of possible answers. It bears repeating that learning from regulatory communities needs to be combined with expertise in science and technology. This meeting is a step in the right direction. I am very grateful to Eric and his team at Dauphine for hosting us today.

Éric Brousseau

Thank you very much. It was a pleasure to co-organise this event with you. Holding a two-day event would have allowed us to discuss the global governance system, which is collapsing before of our eyes. In that context and the increasing acrimony between the three main blocks – the US, Europe and China – it is difficult to see how we can implement many of these principles, for example around AI ethics. I hope that we will have opportunities to hold other conferences together to deal with these essential issues. Working at the level of the OECD could help to converge the approaches of the US and Europe. China also knows that it is in its interests to participate in a global governance system that, at the least, guarantees peace and trade. We must remain hopeful. Thank you all again for your participation.

Governance and Regulation Chair
Paris Dauphine-PSL Fondation
Place du Maréchal de Lattre de Tassigny - 75016 Paris (France)
https://chairgovreg.fondation-dauphine.fr/