Brussels Privacy Symposium 2021

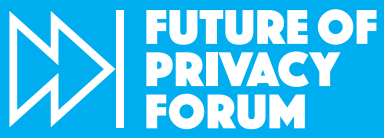
The Age of AI Regulation: Global Strategic Directions

Symposium Report

Authors: Sebastião Barros Vale, Katerina Demetzou, Lee Matheson

The authors thank Dr. Gabriela Zanfir-Fortuna, Rob van Eijk, Lina Jasmontaite, and Gianclaudio Malgieri for their input and feedback

16 February, 2022



The Future of Privacy Forum

The Future of Privacy Forum (FPF) is a non-profit organization with a global presence that brings together academics, civil society, government officials, and industry to evaluate the societal, policy, and legal implications of data uses, identify the risks and develop appropriate protections. We are optimists who believe technology and data can benefit society and improve lives if the right laws, policies, and rules are in place.

FPF Europe maintains strong partnerships across the EU through its convenings and knowledge-sharing with policymakers and regulators. This transatlantic engagement helps regulators, policymakers, and staff at European Union data protection authorities better understand the technologies at the forefront of data protection law. By building this bridge between European and U.S. data protection cultures, FPF hopes to build a common data protection language.

Brussels Privacy Hub

At the **Brussels Privacy Hub (BPH),** we believe strongly in the relevance and importance of data protection and privacy law, particularly in light of the challenges posed by the rapid development of technology and globalization. We also believe that fresh and innovative thinking based on multidisciplinary research is necessary to meet these challenges. The BPH thus brings together scholars from a wide array of disciplines who collaborate with the private sector, policymakers, and NGOs to produce cutting-edge research. We believe in network-building and have built a strong network of contacts with leading privacy researchers both in and outside the EU. The BPH's main goals are to produce privacy research of the highest quality; bring together leading thinkers from around the world; and foster an interchange of ideas among privacy stakeholders in a climate of intellectual openness.

Table of Contents

1. Introduction

On November 16, 2021, the Future of Privacy Forum (FPF) and the Brussels Privacy Hub (BPH) of Vrije Universiteit Brussel (VUB) hosted the *Brussels Privacy Symposium 2021 – The Age of AI Regulation: Global Strategic Directions,* convened by Jules Polonetsky, CEO of FPF, Christopher Kuner and Gianclaudio Malgieri, Co-Chairs of the Brussels Privacy Hub. The Symposium brought together policymakers, academic researchers, civil society organisations and industry leaders from the European Union (EU), the Organization for Economic Cooperation and Development (OECD), the United States, Brazil, and Singapore to discuss the most recent trends in the governance of Artificial Intelligence (AI), with a focus on addressing the risks posed by AI systems to fundamental rights, while fostering their responsible development and uptake.

Most notably, the Symposium's panelists debated the proposal for a legal framework that the European Commission (EC) published in April 2021 (AI Act), a first-of-its-kind comprehensive law for AI systems, which comprises a risk-based approach by scaling legal obligations to the severity of risks that specific AI systems pose. Furthermore, speakers drew comparisons between the proposed EU model and different approaches to AI regulation that are surfacing elsewhere – such as the US, Brazil, Singapore and China. The keynote panel, which covered the EU's road ahead to the proposed AI Act and was moderated by Gianclaudio Malgieri, BPH Co-Director and Associate Professor of Law at EDHEC Augmented Law Institute (Lille), counted on:

- Brando Benifei, Member of the European Parliament, President of the Spinelli Group
- Lucilla Sioli, Director for Artificial Intelligence and Digital Industry (CNECT.A), Directorate-General CONNECT at the European Commission

The following panel saw a *Global Comparative Discussion on Approaches to AI Regulation, Governance and Oversight,* moderated by Dr. Gabriela Zanfir-Fortuna, Vice President for Global Privacy at FPF and Affiliated Researcher at the VUB's Research Group on Law, Science, Technology & Society (LSTS). Speakers included:

- Simon Chesterman, Dean and Provost's Chair Professor of the National University of Singapore Faculty of Law and Senior Director of AI Governance at AI Singapore
- Luca Belli, Professor of Internet Governance and Regulation at Fundação Getúlio Vargas (FGV) Law School
- Audrey Plonk, Head of Digital Economy Policy Division – Directorate for Science, Technology and Innovation, OECD
- Elham Tabassi, Chief of Staff, Information Technology Laboratory of the U.S. National Institute of Standards and Technology (NIST)

The last panel was titled *Should Certain Uses of AI Be Banned?,* and it was moderated by Ivana Bartoletti, Global Chief Privacy Officer at Wipro and Co-Founder of the Women Leading in AI Network. Speakers included:

- Theodore Christakis, Professor of International and European Law at University Grenoble Alpes
- Frank Pasquale, Professor of Law at Brooklyn Law School

- Cornelia Kutterer, Senior Director, EU Government Affairs, AI, Privacy and Digital Policies, Microsoft
- Ursula Pachl, Deputy Director General at the European Consumer Organisation (BEUC)

Jules Polonetsky, Christopher Kuner, and Gianclaudio Malgieri made brief opening remarks to welcome the attendees to the virtual symposium. Polonetsky stressed the importance of keeping up with developments in the regulation of privacy and data protection across the globe, which the 2021 Symposium was a reflection of. Kuner echoed Polonetsky's remarks, adding that regulation in this space is becoming increasingly complex. Malgieri noted that the EC's proposed AI Act is shaping the global discussion around the regulation of the AI, hence the Symposium's agenda swirls around the proposed text.

2. Keynote Panel - The EU's Road to an AI Act: Views From the Co-Legislators

The Symposium began with a keynote panel with two leading figures in the AI Act legislative procedure: Lucilla Sioli, Director for "Artificial Intelligence and Digital Industry" within Directorate-General CONNECT at the European Commission who was involved in drafting the AI Act Proposal; and Brando Benifei, a Member of the European Parliament (MEP) who was appointed as a lead rapporteur for the AI Act at the level of one of the EP's leading committees – the Committee on Internal Market and Consumer Protection (IMCO).

According to the April 2021 EC Communication released upon the publication of the AI Act, the proposed legislative instrument has two different aims: "addressing the risks associated with specific AI applications in a proportionate manner and of promoting the uptake of AI" across the EU. The Communication adds that the AI Act Proposal "puts forward rules to enhance transparency and minimise risks to safety and fundamental rights before AI systems can be used in the EU. Its architecture is based on a number of core components which, as a whole, build a proportionate and risk-based European regulatory approach."

Malgieri invited Sioli to present the structure of the AI Act and its strengths to pursue its intended goals. The EC official highlighted that the Proposal constitutes an advancement from the principles – and ethics-based soft law approaches that have dominated the AI regulation landscape thus far. According to the speaker, the Proposal is an attempt from the EC to consolidate and codify broadly discussed principles into binding rules, which will hopefully inspire legislators in other jurisdictions. She described the Proposal as a horizontal, cross-sector instrument, intended to give legal certainty to AI systems' developers, and a reflection that many issues stemming from untrustworthy and opaque AI systems are not context or sector specific.

2.1 AI CONFORMITY ASSESSMENTS' PROS AND CONS

Sioli also noted that the AI Act includes a certification mechanism for AI systems, which will allow providers to affix a CE marking to their AI systems once they have passed a conformity assessment, with two notable advantages: (i) reliance on organisations – such as market surveillance authorities and conformity bodies – that already exist and are active in some sectors in the EU; and (ii) creation of a level-playing field, also for non-EU AI system providers that wish to market their solutions in the EU. The speaker explained that conformity assessments will be carried out either by third-parties (notified bodies) where high-risk AI systems are integrated into products that are regulated by sectoral product safety laws (Annex II), or by providers of standalone high-risk AI systems themselves (Annex III). However, she took note of the EP's reservations with regards to the latter scenario, which would consist of an innovative self-certification system. Sioli added that AI developers should be able to rely on harmonised EU technical standards for carrying out their self-assessment, and that the EC will contribute to the swift development of such standards.

Then, the moderator asked Benifei to share initial reactions and critical thoughts regarding the EC's Proposal, as well as about specific improvements he would like to see to the text. The MEP started by clarifying that the EP was, at the time of the Symposium, yet to decide on the model of cooperation between the Parliamentary Committees involved in the AI Act, and that there had been only informal discussions on the Proposal's substance at EP level. He congratulated the EC for its degree of ambition with the proposed AI Act, and for drawing inspiration from the EP's 2020 reports on AI applications in different contexts. This included not making a black-and-white distinction between AI systems that needed no regulation from others that required significant legal restraints, with the speaker highlighting that AI systems that pose intermediate levels of risk to fundamental rights would face increased transparency duties under the EC's wording.

Diving into his concerns with the Proposal, Benifei stressed that the certification architecture and the list of high-risk AI systems proposed by the Commission may need to be revised during the EU's legislative procedure. With regards to the former issue, the speaker gave the example of usages of high-risk AI systems that would only be uncovered pursuant to a market surveillance authority investigation, in cases where providers conducted legally permitted self-certification. To illustrate his concern, Benifei pointed to the Dutch government's use of its System Risk Indication (SyRi) algorithm, that built risk profiles of individuals to arguably detect various forms of fraud and assist in taking impactful decisions – such as removing children's custody from their parents – which was deemed unlawful by the District Court of The Hague in February 2020. He also mentioned cases where persons in the U.S. who are entitled to welfare benefits are being algorithmically excluded. According to Benifei, these examples show the need of having stronger ex ante external verifications of AI systems that impact fundamental rights, and that self-certification will often not be enough.

To a question on whether certain notified bodies under the AI Act Proposal – which will be in charge of third-party conformity assessments and issuing certifications in those cases – may often lack expertise in assessing uses of AI systems in specific fields (e.g. healthcare), Sioli observed that such bodies will be functioning at the national level and that it will be EU Member States' responsibility to equip them with proper resources and knowledge. She admitted that notified bodies in some EU countries will be more specialised in certain areas and that an investment in the education of existing notified bodies – but also on the creation of such bodies in Member States that do not have them – will be needed. Sioli disclosed that the EC will devote efforts to assisting Member States in this respect once the final text of the AI Act is approved, in the 2 years before it

eventually becomes effective across the EU. In any case, Sioli stated that there is also the possibility of opening legal leeway for notified bodies to outsource parts of their assessment to specialised laboratories, universities, and Test and Experimentation Facilities (TEFs), which the Commission wishes to launch in a variety of areas (e.g. manufacturing, healthcare, agriculture, and Smart Cities).

Additionally, some AI system deployers (i.e., users) may become concerned about the authenticity of CE markings that will be affixed to AI products, as raised by a member of the audience. On this note, Sioli expressed that she was generally not concerned with the authenticity of watermarking that takes place within the EU, but rather about watermarking that is carried out in third countries, following cases of fraudulent CE marking of COVID-19 protective facemasks outside of EU borders. The speaker stressed that the AI Act Proposal did not advance an adequacy framework for certifications issued outside of the EU, which means that initially all certifications would need to be issued within EU borders, with a view to ensure their authenticity.

2.2 PROHIBITED AI PRACTICES: DOES THE PROPOSED AI ACT GO FAR ENOUGH, OR TOO FAR?

The MEP also referred to the list of prohibited AI practices under Article 5 of the Commission's proposal, arguing that the exceptions to the ban on the use of real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes could be interpreted too broadly by the judiciary and law enforcement agencies. For the speaker, making the scope of such exceptions more precise, or introducing a clearer ban, would be preferable options. Benifei made similar arguments on Article 5(1)(c)'s social scoring ban, stating that players could try to circumvent the prohibition, by using AI systems to otherwise rank citizens hierarchically or to classify them as "good" or "bad."

Benifei was also concerned about the recent trend of introducing AI systems in workplace management. For the panelist, the AI Act text should include participatory rights for social partners where AI systems are used in such contexts, and to include more robust limitations on their usage to prevent abuses against workers.

In reaction to the MEP's comments, moderator Malgieri asked Sioli whether she believed that the list of prohibited AI practices was fit for purpose, in particular given the arguably narrow scope of forbidden social scoring, or AI systems that exploit individuals' specific vulnerabilities or that deploy subliminal techniques to manipulate persons' behaviours in a way that may lead to physical or psychological harm. Are there reasons for excluding other vulnerabilities or economic harms from the remit of Article 5's bans, and is there room for improving the list of prohibitions?

The EC's representative argued that the prohibited AI practices list was comprehensive and precise enough, also having regard to its practical future implementation. According to Sioli, the reason for specifying age- and disability-related vulnerabilities was the difficulty that AI providers and users would have in identifying other types of vulnerabilities when developing or deploying AI solutions. As per the rationale of not covering economic harms in the prohibition, Sioli pointed to the fact that the AI Act builds on existing and upcoming EU legal standards that seek to protect consumers in that regard, such as the Unfair Commercial Practices Directive (UCPD), the General Data Protection Regulation (GDPR), and the Digital Services Act (DSA).

Benifei responded that the list of prohibitions represents one of the thorniest issues that the EP will need to discuss and align on. However, he pointed to the set of principles that the EP approved in the recent non-binding resolution against the use of AI by law enforcement in public spaces as a desired source of inspiration for the future debates on the AI Act Proposal. According to the speaker, said resolution reveals that the EP has a more restrictive view on the use of biometric identification systems than the one forwarded by the Commission, which should feed on the Parliament's future AI Act amending work. Benifei also highlighted that negotiations with the Council on this matter during trialogues may be difficult, given EU Member States' expected, looser position. He also warned against indirectly legitimising social scoring practices by public bodies and private companies if they are not included in the Article 5 list. On the argument that existing legislation – such as the GDPR – already protects individuals against such practices, Benifei stressed the importance of enforcing such laws to render them effective.

Malgieri brought to the table the European Data Protection Supervisor's (EDPS) and the European Data Protection Board's joint opinion on the AI Act, by pointing to their critical stances on the Proposal's arguably soft approach to many automated facial recognition (AFR) and "pseudo-science" emotion detection practices. He stated that the EP had taken a clear position on AFR through its non-binding resolution on the usage of AI systems for law enforcement, but the same was yet to happen on AI-powered emotion detection, which is currently considered in the Proposal as having limited risks to fundamental rights and thus only requiring transparency from such AI systems' users towards exposed natural persons (Article 52). Malgieri added that individuals are not given a right to opt-out from such exposure via the AI Act, but that emotion detection practices may be considered to have "high-risk" if they fall under the Annex III use cases (e.g. if they are used in the school context).

In reply, Sioli stated that the EC did not propose banning emotion recognition AI, as it can be used for benevolent purposes, with good examples in the healthcare context and in connected cars. However, she conceded that some of its usages may affect individuals' autonomy and may be perceived by them as negatively intrusive, hence the importance of letting the latter know that they are being exposed to such techniques. For Sioli, this will empower individuals to decide whether or not to engage with businesses or services that deploy emotion detection technologies. Benifei shared most of Sioli's views but took a slightly less optimistic view of emotion detection practices. He stated that the EP will look carefully into whether transparency obligations are sufficient to tackle the risks arising from such AI systems, in particular when they are used by public bodies.

Sioli also addressed the question of whether introducing human intervention in AI-powered decision-making systems that would fall under the Article 5 list of prohibitions would enable players to circumvent such prohibitions. As the Proposal regulates the use of AI systems, the speaker noted that practices that do not rely on such usage at all would not be covered by the former's prohibitions. In any case, Sioli highlighted that the draft text promotes human intervention in otherwise fully automated decision-making systems, to provide for a meaningful oversight of AI systems.

2.3 ADDITIONAL SAFEGUARDS FOR "HIGH-RISK" AI SYSTEMS

Sioli underlined that the EC has tried to make the AI Act's text future-proof, by allowing flexible updates to the definition of "AI system" in Annex I and the list of high-risk AI use cases in Annex III of the Proposal, through EC delegated acts. The speaker also positively mentioned that the Proposal's list of AI systems use case applications removes the burden from AI providers to assess the abstract level of risk that the solution could pose to fundamental rights when determining the scope of their obligations under the AI Act.

Malgieri noted that the AI Act Proposal advanced a number of safeguards that AI providers and users would need to put in place during the development and deployment of high-risk AI systems. Among those, the moderator highlighted mandatory and by-design human oversight, risk management and data governance processes, which would help AI providers and users identify and contextualise the AI system's gaps. Nonetheless, Malgieri pointed out that the Proposal did not create rights and redress mechanisms for individuals who are targeted by high-risk AI systems, which some argue is due to the nature of the AI Act as a product safety law and not a fundamental rights-centered diploma.

The EC official replied that it was not the Commission's aim to introduce novel rights for individuals with the AI Act, but rather create a mechanism to minimise possibilities of fundamental rights distortions through AI systems. According to Sioli, ex ante conformity assessments as regulated under the Proposal, together with post-market surveillance by competent bodies to address eventual harms, are adequate means to reduce such risks. She added that regulators and courts alike are given rights to access relevant information about the AI systems they will scrutinize in specific cases.

In turn, MEP Benifei took the view that it would be legally challenging to introduce novel individual rights in the AI Act, given the nature of the Regulation at stake. Nonetheless, he argued that individuals should be given avenues for redress and obtaining compensation for damages they may suffer from the usage of AI systems. Benifei also called for enlarging the scope of *ex ante* third-party conformity assessments and the number of cases where they are required, which would not amount to disproportionate "red tape" for high-risk AI systems providers.

On the question of whether including biometric identification and categorisation AI systems in the Annex III high-risk list could have the effect of generally legitimising their use, Sioli underlined that the use of such systems in the private sector is already regulated – and, to an extent, forbidden – under the GDPR. According to the speaker, the Proposal's heavier focus on the use of such technologies for law enforcement purposes was due to the lack of clarity resulting from the Law Enforcement Directive on the matter. By placing them in Annex III as "high-risk" AI systems - to the extent that they are not forbidden under Article 5 as 'real-time' practices carried out in public places – the Commission recognises that remote biometric identification technologies can be useful for police forces, but also problematic if they are biased and facilitate group discrimination. Therefore, Sioli concluded, it is fundamental to ensure that such systems work well, which is the aim of the Proposal. Nonetheless, the panelist revealed that some EU Member States still claim that the Commission's approach would place too heavy restrictions on law enforcement authorities that wish to leverage such tools to fight crime.

Regarding doubts about the effectiveness of a closed "high-risk" AI systems list to address future risks stemming from new usages of AI technologies, Sioli stressed that the European Commission will have the possibility of updating the initially approved list. However, the EC official stated that

there is a need to provide legal certainty to the AI ecosystem, which means that the list cannot be amended too often. Sioli underlined that the Commission will continuously consult the new European AI Board, market surveillance authorities and experts invited by the Board (including academic researchers and fundamental rights organisations) about the need of updating the Annex III list.

As for the timeline of adoption of the AI Act, Benifei foresaw that the EP would be able to agree its position on the AI Act in the first semester of 2022, and that a deal between the EP and the Council of the EU (Council) on the final text could be struck by Spring 2023.

3. Global Comparative Discussion on Approaches to AI Regulation, Governance, and Oversight

The second panel of the Symposium concerned legal and self-regulatory frameworks emerging around the world for the management, development and use of AI systems. The panelists were drawn from three different continents and from the OECD to discuss the differences and similarities between their home jurisdictions' strategies for the governance of AI.

3.1 THE BRAZILIAN BILL & AI IN LATIN AMERICA

The panel began with moderator Dr. Gabriela Zanfir-Fortuna asking Luca Belli to describe the key features of a recent Brazilian proposal: the Brazilian Artificial Intelligence Bill (Bill N. 21/2020, original in Portuguese, unofficial translation by Professor Belli, hereafter the "AI Bill"), which passed the Brazilian Chamber of Deputies on 29 September, 2021, and is currently under consideration before the Brazilian Senate.

Belli began by noting that the new AI bill has been seen as a surprise by many in Brazil, because the bill was adopted very quickly by the Chamber of Deputies, less than six months after the government's April adoption of the National AI Strategy (original in Portuguese only). Belli further noted that, with the original bill introduced before the AI Strategy was finalized, and much regular legislative activity disrupted by the pandemic, some policymakers are concerned that there has not been sufficient opportunity for public debate on the AI bill. The speaker characterized the AI Bill as a 'very light touch' and principles-based law – with only 16 Articles – and as a 'remarkable difference' from the 'very detailed approach' adopted by the EU institutions. He noted that the AI Bill's rapporteur indicated the law draws inspiration from the OECD Principles on AI.

Belli argued that the light touch approach may result in some significant negative externalities to the functioning of the regulation should it become law. Particular concerns include the risk that because so much of the "practice" of the bill will be left to other governmental bodies such as the judiciary, it may be difficult to find the institutional expertise to implement with consistency. The panlist explained that, as existing judges and government agencies do not have preexisting personnel dedicated to understanding AI, the bill may become a bottleneck, as it does not create a specialized agency or authority for regulating AI.

Finally, Belli noted that a major criticism of the bill is that it creates a 'subjective' responsibility for developers and users of AI to determine where AI systems are risky, rather than imposing 'objective' requirements on certain types of AI systems or types of processing activity for enhanced monitoring from the government or regulators. Belli argued this structure places the onus on individuals who are targeted by AI systems to identify and prove where the latter are not functioning well, have victimized them, or have created a discriminatory outcome, a task that average consumers do not have the appropriate knowledge and resources to carry out.

Later in the panel, Dr. Zanfir-Fortuna asked Belli about the interaction between new AI regulatory proposals like Brazil's AI Bill and existing general data protection laws such as the LGPD, and whether there was any consistent approach to the issue across the broader Latin American data protection landscape. Belli identified two broad tendencies in the Latin American space: (i) a positive one, where Latin American countries have shared key principles in data protection and AI rules (e.g., fairness, transparency, and security); and (ii) a negative one, where AI is very much utilized for public safety purposes – including facial recognition –, which are generally excluded from the scope of data protection laws, along with national security uses. On a further negative note, Belli expressed concern that Latin American regulators and authorities are chronically under-resourced. For example, he noted that the Brazilian data protection authority (ANPD) only employs 36 people, which highly contrasts with the Chinese Cyberspace Administration's (CAC) 60,000 employees.

Addressing the interaction of the AI Bill and existing general data protection obligations, Belli noted that the LGPD must be applied in conjunction with the AI Bill, which is something that the latter mentions. He pointed to Article 5 of the AI Bill, which reiterates the LGPD obligation to have a clear description of data processing – specifically on what is related to AI –, explaining that both laws have shared transparency and explainability obligations.

3.2 THE NIST APPROACH & TRUSTWORTHY AI FRAMEWORK

Dr. Zanfir-Fortuna then turned to Tabassi, asking her to briefly describe the United States' self-regulatory approach to AI as exemplified by the NIST Trustworthy AI Program. Tabassi characterized the NIST framework as "more technical" than other, principles-based methods for regulating AI. In regards to NIST's efforts going forward, Tabassi indicated that a key goal of the agency is to work with both industry and governmental policymakers to develop a shared taxonomy across all of the various actors in the AI space, that can accompany developing NIST standards on AI trustworthiness and risk management. She argued that trustworthiness and risk management are inherently related in the AI space, as when a system's risk is effectively minimized, its trustworthiness increases.

With regard to the development of the trustworthy AI Framework, which NIST seeks to complete by January 1, 2023, Tabassi described NIST's main goal as the challenge of translating AI principles to practice. In this regard, she gave the example of translating topics such as "harm to individuals" or "harm to privacy" into distinct criteria and characteristics that could be designed and implemented in the development of AI systems and used to test those systems. Tabassi identified several core goals for AI systems the trustworthy AI and the taxonomy of risk standards seek to effectuate. These include: (i) accuracy, (ii) resiliency vis-à-vis different vulnerabilities, (iii) privacy preservation, (iv) robustness (in different contexts and environments), (v) reliability, (vi) safety, (vii) interoperability, and (viii) explainability AI systems.

Finally, Tabassi noted that as a non-regulatory agency, any NIST standard will be voluntary rather than mandatory, so the main goal of the agency is to create an 'evidence-based' standard that will contribute to other policymakers' efforts. This, in part, drives the NIST process of developing standards and principles in consultation with a variety of stakeholders, including researchers, technology developers, other industry participants, academics, and regulators.

Later in the panel, Tabassi was asked to contrast the proposed NIST taxonomy with similar classification efforts around the world (such as the conformity assessment provisions in the proposed EU AI Regulation). The panelist re-emphasized that NIST's goal was to engage with stakeholders across the AI space – meaning other regulators and policymakers as well as industry actors – and that the agency would view a taxonomy directly contrary to other emerging frameworks as a failure. She described NIST's goal as 'to align as much as possible' with other frameworks, while 'building on' what is achieved in other parts of the AI regulatory space.

3.3 CHARACTERIZING THE ASIAN APPROACH

Third, Dr. Zanfir-Fortuna asked panelist Simon Chesterman to characterize the landscape in Singapore and the broader Southeast Asia (SEA) region, specifically mentioning the self-regulatory AI Framework in Singapore and the emerging regulation of algorithmic decision-making in China.

Chesterman began by highlighting a commonality in the approach that appears to be emerging in Asia, though this approach is distinct from the EU and U.S. strategies. The speaker argued that while the EU is pursuing a primarily 'rights-based' strategy about protecting individuals in the EU, and the U.S. focused on a more self-regulatory space given the prominence of industry players' standard-setting activities, in Asia, it appears that governments are the dominant players independent of industry pressure or consumer advocacy. The panelist argued that many Asian jurisdictions are primarily concerned with taking action to reap the economic benefits of AI and are more cautious about possibly constraining innovation or driving development elsewhere. As an example, he cited the Singapore strategy for AI, the Model Artificial Intelligence Governance Framework, which is a principles-level document.

Chesterman cited the ongoing policy problem as a tension between the ease of creating controls around emerging technology before it is established (but running the risk of doing so without sufficient knowledge of what to make policy about) and the difficulty of creating new regulations once a technology has matured enough that the risks associated with it are more readily apparent. He characterized the EU approach as proposing regulation much harder and earlier, in distinct contrast to what seems to be emerging in Asia.

Later in the panel, Dr. Zanfir-Fortuna asked Chesterman to describe the 'intrinsic limits' of the law for regulating AI (as described in his recently published book). The speaker summarized his theory in two points: one was the idea that for the most part, new laws are not needed for AI. Chesterman argued that there is a "disturbing" tendency to assume that new laws are needed to deal with AI technologies, when really it would be better to consider whether and how a given use of AI is already covered under existing laws; another were three overarching reasons to regulate AI outside of normal law, when existing law is 'stretched' by the implementation of new AI technologies, namely: (i) minimizing AI systems' harms and optimizing their results (e.g., ensuring the safety of autonomous vehicles); (ii) drawing moral redlines where policymakers decide that humans should

make certain decisions, but not necessarily because humans will make better decisions than AI (e.g., laws prohibiting fully autonomous weapons); and (iii) governing some situations where a particular human needs to make a decision (e.g., public sector positions where a judge or a legislator needs to make a decision because they are empowered by society to do so and are accountable for the results).

3.4 THE MULTIJURISDICTIONAL APPROACH AT OECD

Lastly, Dr. Zanfir-Fortuna invited Plonk to summarize how the OECD has approached the need to work with and synthesize the different jurisdictional approaches to AI as a multinational organization. The last speaker of the panel began by noting that the OECD's work has emerged from an apparent moment of consensus among OECD members that AI technology needed a human-centric, rights-based approach. With that in mind, the OECD principles are an early intergovernmental standard, though remaining at a high level.

Outside of such AI principles, Plonk described three ongoing OECD efforts: (i) developing a framework for classifying AI systems. This Framework will ideally be a baseline standard for governments and policymakers to think about and classify what an AI system does, and to enable actors to better assess the risk proposed by the system. It recognizes a major commonality between AI regulatory efforts: that most feel that such efforts should be risk-based, even if they differ on the details of how to accomplish that goal in practice; (ii) creating tools for trustworthy AI, designed to operate with the OECD AI Principles. In this regard, Plonk underlined that the OECD has released an analysis that classifies trustworthiness tools as either technical, procedural, or educational. Tools are cataloged on the OECD AI Policy Observatory, with a view to facilitating access for policymakers and practitioners to information; (iii) developing a global AI incidents tracker. The OECD intends to, using committees and working groups to investigate 'AI incidents,' create an interoperable and harmonized approach for tracking incidents with AI systems.

Later in the panel, Dr. Zanfir-Fortuna asked Plonk whether there was a definitional distinction between "AI" and "Machine Learning" emerging in the new regulatory space. The panelist responded that at the moment, the terms were often used interchangeably, capturing both more general machine learning technologies as well as theoretical 'true' artificial intelligence. Plonk noted that settling these questions of terminology and increasing precision in language around AI is part of the OECD's ongoing taxonomy efforts, though the AI classification framework is less about the definition of terms and more about "trying to figure out what a system is doing." Plonk provided four major elements that make up the latter issue: (i) the **context** of the system: what is the sector where the system is deployed? What is the scale of its deployment? Who interacts with it? How much choice do they have? (ii) the **data and inputs:** what data flows from the outside context into the system? Where does the data come from? What kind of data is used in the system? Are there rights attached to the data? (iii) features of **the AI model:** the computational representations of everything that takes place in the model's context. What are the technical aspects? What are the types of models and/or types of training methods in the system? (iv) what **tasks** does the AI system perform? What is the system's output (e.g., recommendations, predictions, automated decisions)?

4. Should Certain Uses of AI Be Banned?

The third and last panel of the Symposium discussed the issue of banning specific uses of AI systems. The panel's moderator, Ivana Bartoletti, highlighted that the debate on 'to ban or not to ban' constitutes one of the most controversial issues of the proposed AI Act. Bartoletti referred to examples such as facial recognition technology, algorithmic management in the context of work and employee's performance and noted that the big question lies on whether the classification of certain uses of AI as 'high risk' will de facto legitimise these uses and open the gates for their acceptance. Bartoletti invited Theodore Christakis as a first speaker to express his views on facial recognition technology, by drawing the lines with biometric technologies and also by considering how these technologies shape and change the way that people exist in public spaces.

4.1 AUTOMATED FACIAL RECOGNITION: A BAN OR A "USAGE-BY-USAGE" APPROACH?

Christakis began by presenting two opposing views as to whether Automated Facial Recognition (AFR) should be prohibited. On the one hand, there are strong voices in the civil society against AFR (for example, the Reclaim Your Face campaign), while on the other hand, there are voices (mainly coming from the security field) that support the use of AFR, as long as necessary safeguards are in place. Christakis stressed the importance of having a debate on AFR given that current discussions revolve around societal choices with regard to these technologies. After quoting the French DPA (CNIL), the speaker stressed that in order to engage in a fruitful debate on AFR there is a need to clarify its meaning. In this regard, he argued that it is vital to avoid confusion between the multiple and different uses behind the catch-all term 'AFR', given the different issues raised by each use case. He thus recommended taking a "usage-by-usage" approach to regulating AFR, considering the risks that arise in specific contexts in which AI systems are used.

Following this line of argumentation, Christakis presented the results of a project run by the Chair on the Legal and Regulatory Implications of Artificial Intelligence from the MIAI Grenoble Alpes. The project's aim is to map the current multiple uses of AFR in public spaces in Europe and to propose a classification and, to the extent available, accurate information about past, existing or planned projects on AFR in public spaces. The panelist explained that two tools have been developed in the context of this project:

(i) the first is a classification of AFR on the basis of technical characteristics and functionalities but also on the basis of specific applications, which has identified 5 technical functionalities and 12 applications of AFR. Christakis argued that such results illustrate that banning some particular uses of AFR does not mean that all AFR uses in public spaces should be prohibited. He gave the example of the French PARAFE system, a one-to-one face verification system placed at French airports which has been approved by the CNIL, after verifying that adequate data protection safeguards were in place;

(ii) the second is a questionnaire with three sets of questions: one focusing on the technical details of identified use cases in Europe; another dedicated to the protection of human rights and data processing principles (e.g., legal bases used by controllers, positions taken by national courts and DPAs, necessity and proportionality, etc.); and a last one covering whether the data controller considered accountability and transparency requirements, and carried out a DPIA before implementing the system.

Christakis added that, on the basis of such tools, the project presents more than 100 European use cases of AFR in public spaces. From those, deeper focus was devoted to 25 use cases and associated positions taken by DPAs and Courts, as well as to DPIAs and their proposed methodologies. Christakis shared a few highlights from the project's findings, which will soon become publicly available.

1. **Existing law seems to suffice to deal with at least some groups of use cases,** such as 'face verification' cases. Christakis mentioned that, while face verification within the PARAFE system passed the CNIL's scrutiny, the French DPA found that the use of face verification in two Southern France high schools did not comply with the GDPR, given the lack of valid data subject consent. According to the speaker, this does not mean that face verification should be banned in general, but rather examined on a case-by-case basis.
2. **The existing regulatory framework needs further precision, development and interpretative guidance.** Here, the example brought by Christakis was Article 10 of the Law Enforcement Directive (LED), which prohibits the processing of biometric data unless certain conditions are met, including strict necessity and authorisation by EU Member State law. In this context, Christakis observed the problematic divergences between national laws, which could be bridged with additional DPA guidance.
3. **There are significant discrepancies between DPAs' and national Courts' decisions in Europe, especially concerning Article 9 GDPR.** The speaker observed that many private actors in Europe have relied on the Article 9(2)(g) GDPR exception to escape the Article 9(1) prohibition on the processing of biometric data. The former allows controllers to process biometric data if the 'processing is necessary for reasons of substantial public interest', and controllers have used this to enforce entry bans on this premises using AFR. Christakis highlighted the Mercadona case in Spain - where both the Spanish DPA (AEPD) and the Barcelona Court of Appeals established that the retailer could not rely on the Article 9(2)(g) exception - and another restrictive position from the Dutch DPA when assessing the use of AFR in a supermarket to prevent shoplifting. He then pointed to cases in the UK, where data controllers have systematically used AFR for the same purpose by invoking the Article 6(1)(f) GDPR 'legitimate interest' legal ground to process biometric data, although this is not an exception under Article 9(2) GDPR. Lastly, Christakis added references to conflicting decisions from the French and Danish DPAs on whether using AFR to enforce football stadium bans was allowed under the exception, with only the latter replying positively.

4.2 LICENSING OF AI SYSTEMS: THE *TERTIUM QUID* BETWEEN BANNING AND NOT BANNING

Bartoletti invited the second speaker of the panel, Frank Pasquale, to share his thoughts on a possible ex ante licensing mechanism for AI systems that could shift the focus away from the debate of whether to ban certain AI practices. Pasquale suggested looking at licensing schemes as a *tertium quid* that could come between the options of banning and not banning specific AI systems. The speaker expressed his disappointment when seeing how legislative initiatives' big aspirations are finally not met in light of technology's fast pace, in what Hartmut Rosa calls the 'social acceleration of time.'

Regarding the proposed AI Act, Pasquale listed its four risk categories for regulating AI systems: (i) **unacceptable risk,** including AI systems that aim to manipulate individuals by exploring their age – or disability – related vulnerabilities in a manner that causes or is likely to cause psychological or physical harm. An example would be an AI system that targeted cosmetics advertisements to people at specific moments in which they would feel less attractive. This first category also includes AI-powered social scoring systems that are unjustifiably detrimental to the targeted individuals, inasmuch that such systems would try to commensurate one's trustworthiness or value over different areas of life. Pasquale also referred to indiscriminate facial recognition by police forces as a banned AI practice under Article 5 of the AI Act; (ii) **high risk,** with examples such as credit scoring, AI used in critical infrastructure by judges, public administration or police, products under safety regulations, among others. The speakers stressed that AI systems that fall under this category must comply with strict obligations from their design stage onwards, but that concerns have been expressed that standards bodies which the AI Act burdens with pre- and post-marketing check may have the necessary skills or resources to fulfill such duties; (iii) **limited risk,** with use cases such as deep fakes, emotion recognition technology, and chatbots; and (iv) **minimal risks,** which are all the remaining AI systems. Pasquela observed that the AI Act encourages the industry to approve voluntary codes of conduct to self-regulate AI systems belonging to this category.

Another point raised by Pasquale was that, instead of chasing AI systems after they are put into use or placed on the market, regulators should be encouraged to assess AI systems' proposed data collection, analysis and uses at a preliminary stage. According to the panelist, such an approach would have two advantages: (i) added transparency, which would help the public to understand how AI systems shape or judge them. This would be enabled by a public record of licensed AI systems, like currently exists for patents; and (ii) preventing the most troubling AI applications from reaching the market. While admitting that the idea of licensing all AI systems seems utopian, Pasquale proposed to initially focus licensing efforts on specific use cases, such as high-risk AI systems that use the most data, very large online platforms and very high-risk AI-powered activities (e.g., modelling health conditions or suicidality). According to the speakers, discovering the areas in which an *ex ante* (instead of an ex post) model of oversight is preferable should be part of AI policymakers' agenda.

Another phenomenon that Pasquale argued was missing from these discussions was the interaction between different AI systems. While separately AI systems may not pose substantial risks, they may be problematic when jointly used. As an example, he mentioned credit scoring that is initially used to determine the conditions of someone's loan, but then ends up being leveraged by car and home insurance companies for their own purposes. Pasquale also highlighted that data coming from private sources to feed AI systems should meet the same standards as data coming from public sources, as the divide between both is often blurry.

4.3 CAN CORPORATE POLICIES AND STANDARDS FOR RESPONSIBLE AI INSPIRE THE AI ACT?

Kutterer started her intervention by presenting Microsoft's high-level approach to ensuring the development of responsible AI. In this context, the speaker highlighted specific actions undertaken by Mcrosoft over the years with this aim, starting with its ethical principles-infused internal AI development standard, which draws inspiration from national and international ethical frameworks. She added that, since 2019, the company has managed to operationalise such principles by

overcoming the socio-technical challenges around AI through the definition of specific objectives for engineering teams to achieve accountability, transparency, and inclusion when designing AI solutions. Kutterer also explained that Microsoft's internal standards allows the company to detect and mitigate potential risks during the development phase of AI systems. According to the speaker, this exercise impacts the way Microsoft contractually deals with its customers on the AI systems' deployment side, and to adopt a holistic approach to AI risk mitigation.

Kutterer also stressed Microsoft's efforts to responsibly develop and to call for regulating AFR. In this space, the panelist mentioned the six guiding principles that Microsoft has adopted when building such solutions, namely fairness, transparency, accountability, non-discrimination, notice-and-consent and lawful surveillance. Kutterer also stressed the importance of assessing the regulation of AFR with a broader lens, not narrowly focused on the upcoming AI Act. In this respect, the speaker mentioned the EU's robust human rights legal landscape, which is currently capable of addressing the challenges raised by new technologies. An example brought forward by Kutterer of concrete judicial application of such rules was the European Court of Human Rights' Malone Case, which focused on necessity, proportionality, and appropriate safeguards as guiding principles for the development and deployment of technology. She also noted that there was room for improvement when it came to transparency of AFR usage by law enforcement agencies (e.g., lack of publication of annual reports) and to technical details applicable to AFR, which could evolve and provide more robust safeguards through standardisation.

On the EU's regulatory efforts, Kutterer argued that the regulatory focus of the AI Act should be on specific AI use cases, and the way AI systems are actually executed. This approach would be most efficient in mitigating the risks resulting from AI. In this line, the speaker pointed to Microsoft's transparency notices that inform customers about the responsible deployment of an AI system, its guidance on image quality issues and on how environmental factors may impact the result of AFR. Kutterer highlighted that Microsoft more thoroughly scrutinises sensitive AI use cases, which are largely aligned with those cases enumerated in Annex III of the AI Act. With regard to such a list, the panelist welcomed the European Commission's proposal, which she called a good starting point. Nonetheless, Kutterer noted that the risks which the Commission aimed to target are still not clear, notably, whether the focus is preventing discrimination, promoting equity, or both. Kutterer praised the AI Act's approach to AI risk mitigation, even though she would welcome further development of obligations for AI users, to apply at the AI systems' deployment.

4.4 A CONSUMER LAW PERSPECTIVE ON THE PROPOSED AI ACT

As the last panelist taking the floor and representing the civil society organisation BEUC, Ursula Pachl provided a consumer perspective to the discussion on the proposed AI Act. She noted that the draft AI Act does not provide a high level of consumer protection against the harms that may be caused by AI systems. Pachl's position was based on five key arguments: (i) the AI Act uses a concept of 'user' which differs from the concept of consumer or end-user, rather focusing on institutional or business users of AI systems; (ii) consumer protection is not an explicit part of the proposal's legislative objectives, which could reveal that the AI Act is not fit for purpose under the EU's Better Regulation framework; (iii) economic harm does not figure in the Article 5 list of prohibited AI practices, which focuses only on physical or psychological harm; (iv) although the AI Act is a horizontal framework, it does not provide horizontal rights for consumers, such as a right to complain and to obtain judicial redress, including through representative actions; and (v)

the proposed AI Act cannot serve as a standalone global standard, given that its application and effectiveness presupposes the understanding and implementation of other pre-existing legal frameworks, (e.g., the GDPR, consumer protection and product liability law). According to Pachl, this does not guarantee legal certainty for businesses nor protection for consumers.

Pachl revealed that BEUC favors having a list of prohibited practices under the AI Act. When referring to remote biometric identification, the speaker stressed that Article 9 GDPR was not enough to ensure citizens' protection, given the wide margin of discretion it leaves to EU Member States. Therefore, BEUC has called for a total ban of remote biometric identification systems, including when they are used by private entities. Additionally, Pachl highlighted social scoring as a very problematic practice, whose use by private entities is currently not covered by the draft AI Act. In this regard, the panelist stated that consumer law fell short of effectively protecting individuals against such practices, therefore calling for stronger protection within the upcoming instrument. For Pachl, the same goes for emotion recognition practices, which are not adequately tackled by mere added transparency duties incumbent upon AI providers, as they are currently provided under the proposed AI Act.

Pachl also stressed that the proposed AI Act has a very broad material scope, as the definition of AI systems was encompassing. However, she opined that its concrete requirements were very narrow, save for a few high-risk AI systems. Such a use case approach may, according to the speaker, turn out to be problematic, as certain practices (such as credit scoring) are burdened with significant requirements, while very similar ones (such as insurance scoring) which ought to be strongly regulated are not. Pachl added that it could be interesting to discuss the possibility of combining the AI Act's use case approach with a more principles-based one, inspired by the GDPR. In her view, having overarching principles applicable to the design and deployment of AI systems as a default, without the need to articulate with sectoral legislation, would be a better approach to regulating AI systems.

Lastly, Pachl criticised the Commission's choice to follow the New Legislative Framework (NLF) approach when drafting the AI Act. As she qualified the NLF as an old and outdated approach that introduces an accreditation system, being largely based on technical standardisation, which may not be fit to regulate new technologies. The speaker concluded that, while the NLF guarantees the free flow of products in the European market, it does not address the impact that AI systems have on fundamental rights and freedoms of individuals.

5. Conclusions

From the fruitful discussion during the 2021 Brussels Privacy Symposium, we have learned important lessons and obtained meaningful insights on the status of AI regulation in Europe and around the world. Different audiences and stakeholders have varying expectations about what new laws covering AI should achieve, and this was reflected in each speaker's contributions.

During the Keynote Panel on the EU's AI Act, we heard that:

1. Co-rapporteur Brando Benifei predicted the European Parliament would be able to agree its position on the AI Act in the first semester of 2022, and that **a deal between the Parliament and the Council of the EU (Council) on the final text could be struck by Spring 2023.**

2. **Regarding AI certification schemes**, the European Commission will contribute to the swift development of harmonised EU technical standards that AI developers may leverage to carry out their conformity self-assessments under the AI Act. It will also seek to ensure that national authorities who are responsible for third party conformity assessments and post-market surveillance have the adequate resources and expertise to fulfil their responsibilities. In turn, Benifei worried that certain high-risk AI systems would only reach the public eye following a market surveillance authority investigation, in cases where providers conducted legally permitted self-certification, arguing for stronger ex ante external verifications of AI systems that impact fundamental rights.
3. **On this list of prohibited AI practices under Article 5 of the AI Act Proposal**, European Commission's Sioli stressed the difficulty that AI providers and users would have in identifying individuals' vulnerabilities which were not age – or disability – related when developing or deploying AI solutions, hence the narrow scope of the provision's prohibition. From the EP's side, Benifei would favor a more restrictive approach on remote biometric identification and social scoring systems, thereby avoiding indirectly legitimising practices that are currently not covered by Article 5.
4. **Focusing on the list of "high-risk" AI systems and their associated legal requirements**, the Commission representative argued that having a flexible list of systems that will be subject to strict conformity assessments and post-market surveillance is fit to address the potential harms and mitigate potential risks to fundamental rights deriving from AI which is not otherwise prohibited. Benifei called for introducing redress and compensation for individuals to seek damages they might suffer from the usage of AI systems, as well as for enlarging the scope of *ex ante* third-party conformity assessments.

From the panel in which panelists had a comparative discussion on different approaches to AI regulation and oversight across several jurisdictions, we concluded that:

1. While there are significant differences in AI approaches, a number of high-level **similarities** at the principles level are beginning to emerge. These mostly concern explainability and transparency of, and responsiveness to individual concerns over AI systems.
2. There appears to be an **increasing consensus that the best regulatory approach for managing AI is to do so on a risk-based model**, with particular requirements to be imposed on 'high risk' AI systems. However, there are diverging methods to measure and respond to the risks associated with AI systems, and even whether risk should be primarily determined by the general features of an AI system or the particular purposes or uses to which it will be placed.
3. There are significant underlying philosophical distinctions between the approaches to AI regulation emerging in the EU, US, Asia, and Latin America regions.
4. As these regulatory models mature, **one key challenge will be ensuring that all the players mean the same thing when using the same terms**. Differences in taxonomy (e.g., whether and to what extent "AI" includes particular machine learning technologies) will become critical.

Lastly, in the panel which discussed the pros and cons of banning certain AI practices or systems in the AI Act, stakeholders expressed that:

1. A fruitful debate on **Automated Facial Recognition (AFR)** requires a clarification of the term, an understanding of what the technology is, and the multiple different uses it can be deployed for. A usage-by-usage approach is thus highly recommended. Some existing regulatory frameworks that deal with the matter (e.g., the EU's Law Enforcement Directive) need further development and interpretation. Lastly, there are significant divergences between DPAs and national Courts in Europe – especially concerning Article 9 GDPR – when ruling over cases involving AFR, which complicates the robust protection of individuals with regard to the technology.
2. An approach of requiring the ***ex ante*** **licensing of AI systems** by a competent body could avoid the regulatory dichotomy between strictly prohibited and other risky AI systems. This would guarantee transparency towards the public and block the development or marketing of some problematic AI systems at an early stage.
3. Existing corporate policies on AI development can inspire the EU legislator, notably where they define **specific objectives for engineering teams to achieve accountability, transparency, and inclusion when designing AI solutions**.
4. The AI Act may fall short of providing a high-level of consumer protection against the harms that may be caused by AI systems, if it does not cover protection against unacceptable AI practices that generate economic harm and if it does not provide for appropriate redress rights for individuals who have been targeted by AI. **Civil society would also like to see stricter prohibitions** on remote biometric identification, use of social scoring systems, and emotion recognition practices.

To learn more about FPF in Europe, please visit fpf.org/about/eu.

1400 Eye Street, NW, Suite 450
Washington, DC 20005

fpf.org