

accessnow.org

EUROPE'S APPROACH TO ARTIFICIAL INTELLIGENCE: HOW AI STRATEGY IS EVOLVING

 accessnow

Access Now defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, legal interventions, and convenings such as RightsCon, we fight for human rights in the digital age.

EUROPE’S APPROACH TO ARTIFICIAL INTELLIGENCE: HOW AI STRATEGY IS EVOLVING

DECEMBER 2020

TABLE OF CONTENTS

I. INTRODUCTION TO THE AIMS OF THE REPORT	2
II. THE EUROPEAN UNION APPROACH IN A GLOBAL CONTEXT	3
How has the EU’s approach to AI influenced the global debate?	3
Where has “Trustworthy AI” been adopted?	5
Europe	6
The Americas	8
Asia Pacific	10
Private Sector & Standards Bodies	12
International Institutions	13
Critiques of the European Union approach	15
Council of Europe and other critiques from a human rights standpoint	19
Responses to the AI Whitepaper & next steps	22
III. HUMAN RIGHTS AND ETHICS AS GOVERNANCE FRAMEWORKS FOR AI	26
IV. NATIONAL STRATEGIES IN THE EU	30
High-level update on the status and trends in national strategies in the EU	30
Analysing the interaction between national strategies and EU level positioning	32
On a European regulation	33
Comments on the White Paper	34
Other issues	34
National AI strategies	36
Takeaways from the final stakeholder roundtable	37
Facial recognition as a case study in EU-national level interaction	40
V. CONCLUSION: AI POLICY STRATEGIES — WHAT HAS WORKED AND WHAT HAS NOT	43
VI. ANNEX - AGENDAS & PARTICIPANT LISTS	46

I. INTRODUCTION TO THE AIMS OF THE REPORT

In November 2018, Access Now, in collaboration with the Vodafone Institute, launched a report, *Mapping Regulatory Proposals For Artificial Intelligence In Europe*, to map and analyse strategies and proposals for regulation of artificial intelligence (AI) in Europe¹. The report covered regional strategies from the European Union and the Council of Europe as well as national plans from several member states including France, Finland, Germany, and Italy. Additionally, it laid down criteria to assess AI strategies to make sure that the development and deployment of AI is individual-centric and human rights-respecting.

One year later, we wanted to find out how the debate on AI governance had progressed, and what Member State stakeholders thought of the proposals and initiatives coming from the EU level. Access Now and the Vodafone Institute therefore organised four roundtable discussions and conducted a number of individual stakeholder interviews on the topic of artificial intelligence and human rights, to bring together key stakeholders from a number of EU countries and regions to discuss how we can work together to ensure that the design, development, and deployment of AI-assisted technologies in Europe are human centric and respect human rights.

These multi-stakeholder roundtables included government representatives, representatives from the private sector, civil society organisations, and academics (a full list of participants and the agendas can be found in the Annex to this report).

The main objective was to discuss the role of Member States and other national stakeholders on the one hand, and the role of EU institutions on the other. The roundtables were held in Berlin and Helsinki in November 2019, while the second part of the series was held online due to the COVID-19 pandemic during the summer and fall of 2020, gathering stakeholders from Central European countries (Czech Republic, Poland, and Hungary), Spain, and France.

These roundtables were an opportunity to gather feedback on the scope and the regulatory approaches of the European Union on AI, including, for the second part of the series, feedback on the European Commission's *White Paper on Artificial Intelligence: a European*

¹ See *Mapping artificial intelligence strategies in Europe: a new report by Access Now*, available at <https://www.accessnow.org/mapping-artificial-intelligence-strategies-in-europe/>.

*approach to excellence and trust*² (“the White Paper”). The roundtables were held under Chatham House rule.

Building on the insights gained from these roundtables, and from an analysis of recent policy initiatives related to AI governance, this report aims to provide an overview of where the debate on AI and human rights stands two years after our original report. To do so, we start by looking at the impact of the European Union approach to AI governance, in particular the idea of “Trustworthy AI”, and the ideas put forward in February 2020 in the Commission’s White Paper.

We then provide an overview of the debate about ethics and human rights as frameworks for AI governance, and look at the calls to ban certain applications of AI, which have increased in the past six months. Finally, we provide an overview of the key takeaways from our roundtable series, including an overview of the closing workshop organised on 27 October, 2020.

II. THE EUROPEAN UNION APPROACH IN A GLOBAL CONTEXT

How has the EU’s approach to AI influenced the global debate?

Both the United States and China have tried to assure their dominance in AI development and deployment: the former by allowing its market-driven, venture capitalist culture to flourish relatively uninhibited, and the latter in a more top-down, statist fashion as part of its overall industrial strategy. While also wishing to promote AI development and deployment, the European Union has attempted to take the global lead in the governance of artificial intelligence, aiming to “define its own way, based on European values, to promote the development and deployment of AI”³.

The European Union’s attempt to ground its approach on European values has resulted in a number of significant documents: the High Level Expert Group (HLEG) on AI’s *Ethics*

² See *White Paper on Artificial Intelligence: a European Approach to Excellence and Trust*, available at https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

³ *Ibid.* p. 1

*Guidelines for Trustworthy AI*⁴ and their *Policy and Investment Recommendations*⁵, and the European Commission's *White Paper on artificial intelligence - A European approach to excellence and trust*⁶. While the term "European values" is often used in political statements, it is important to note that those values are underpinned by the principle of rule of law and fundamental rights enshrined by EU Treaties and the EU Charter of Fundamental Rights. These values are built upon binding and enforceable rights.

To position itself at the forefront of the global debate around AI regulation, the EU wants to rely on its market and regulatory power in the same vein as the recent General Data Protection Regulation (GDPR) with its two-fold objective (protection of personal data and free flow of personal data), by setting industry standards, building trust, and ensuring legal clarity and public legitimacy in AI-based applications. Just as the GDPR has set a high standard for data protection regulation, the hope is that the EU approach to AI governance will set a strong precedent for others to follow.

Indeed, the World Economic Forum (WEF) has warned about the disruption that this upcoming regulation could cause, and advised that "companies should preemptively introduce a sound vetting process for AI products and services to experience the least disruption"⁷. This acknowledgment shows that the EU's focus on good governance could give it a serious role in determining how AI is developed; as Nathalie Smuha has pointed out, the "*first-mover advantage* that can be gained from setting the standards means the *race to AI* has also become a *race to AI regulation*"⁸.

Strategies at EU and national level also focus on international and European cooperation, as a matter of policy or research, with calls for the creation of research centres across Europe and the creation of the AI alliance⁹, a multistakeholder forum. In terms of policy, the approach put forward by the European Commission aims at tackling "technological, ethical,

⁴ See *Ethics Guidelines for Trustworthy AI*, available at

<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

⁵ See *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*, available at

<https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>.

⁶ See *White Paper on Artificial Intelligence: a European Approach to Excellence and Trust*, available at

https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

⁷ See *Regulation could transform the AI industry. Here's how companies can prepare*, available at

<https://www.weforum.org/agenda/2020/10/ai-ec-regulation-could-transform-how-companies-can-prepare/>.

⁸ See *Europe's approach to AI governance: time for a vision*, available at

<https://www.friendsofeurope.org/insights/europes-approach-to-ai-governance-time-for-a-vision/>.

⁹ See *The European AI Alliance*, available at <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>.

legal, and socio-economic aspects to boost the EU's research and industrial capacity and to put AI at the service of European citizens and economy”¹⁰.

In this first section, we will examine where this EU approach has been successful and where it has fallen short, both domestically and globally, and look at some of the criticisms that have been raised against it. We will begin by looking at the impact of the HLEG's work, and then look at how the Commission's White Paper has been received in the global debate about AI governance.

Where has “Trustworthy AI” been adopted?

The most visible trademark of the EU approach to AI governance is arguably the concept of “Trustworthy AI”. According to the HLEG's *Ethics Guidelines*, trustworthiness in AI means that an AI system should be:

1. **lawful** - respecting all applicable laws and regulations
2. **ethical** - respecting ethical principles and values
3. **robust** - both from a technical perspective while taking into account its social environment

This is, of course, very general, and the guidelines received criticism for not specifying what it means for AI systems to be lawful given that there is a lack of clarity around which laws and regulations currently apply¹¹. The guidelines do provide more detail on what it means for an AI system to respect ethical principles and values, outlining seven “key requirements that AI systems should meet in order to be deemed trustworthy”:

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and data governance
4. Transparency
5. Diversity, non-discrimination, and fairness
6. Societal and environmental well-being
7. Accountability

¹⁰ See European Commission's webpage on Artificial Intelligence, available at <https://ec.europa.eu/digital-single-market/en/artificial-intelligence>.


¹¹ See *Laying down the law on AI: ethics done, now the EU must focus on human rights*, available at <https://www.accessnow.org/laying-down-the-law-on-ai-ethics-done-now-the-eu-must-focus-on-human-rights/>.

Most recently, the HLEG launched the final version of the *Assessment List for Trustworthy AI* (ALTAI), which is available as a PDF document or as an interactive web-based tool¹². To look at the impact that these documents, and the ideas they promote, have had, we can begin by looking at where the term and the concept of trustworthy AI have been adopted.

One obvious place in which the EU approach, and particularly the concept of Trustworthy AI, has had a major impact is in the national AI strategies of EU member states. In their report, *National Artificial Intelligence Strategies and Human Rights: A Review*¹³, Global Partners Digital note that most of the EU Member State national strategies make explicit reference to human rights, and for those that do not, “human rights are often assumed to form the foundation of policy whether or not it is explicitly stated”. At the same time, the report notes that merely acknowledging human rights does not amount to providing proper provisions for protecting them. But beyond this acknowledgment of human rights as a baseline, how many of the strategies make explicit reference to the concept of Trustworthy AI?

Europe

Out of the 17 national strategies published as of today¹⁴, five explicitly mention “Trustworthy AI”, while only one Member State, Malta, fully integrates the seven requirements of the guidelines.



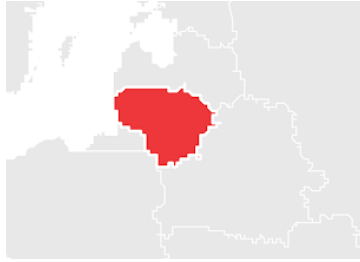
In its own ethical framework, *The Malta Ethical AI Framework, Towards Trustworthy AI*¹⁵, Malta also included “illustrative leading control practices for AI, first at the governance-level and then for each of the Trustworthy AI Requirements”. With this framework, the Maltese government is at the forefront of the implementation of Trustworthy AI in the EU.

¹² See ALTAI - The Assessment List on Trustworthy Artificial Intelligence, available at <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence>.

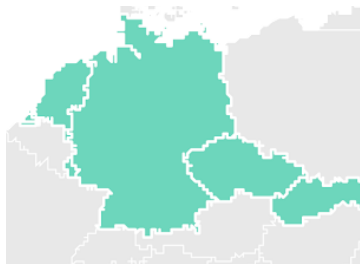
¹³ See *National Artificial Intelligence Strategies and Human Rights: A Review*, available at <https://www.gp-digital.org/publication/national-artificial-intelligence-strategies-and-human-rights-a-review/>.

¹⁴ See *AI Watch, National strategies on Artificial Intelligence, A European perspective in 2019*, available at https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119974/national_strategies_on_artificial_intelligence_final_1.pdf (last update on 25 February 2020 but still accurate as of 23 July 2020 based on desk research).

¹⁵ See *Malta, Towards Trustworthy AI, Malta's Ethical AI Framework*, available at https://malta.ai/wp-content/uploads/2019/10/Malta_Towards_Ethical_and_Trustworthy_AI_vFINAL.pdf.



Lithuania also features the concept, and details two components of Trustworthy AI: “(1) ethical purpose — it should respect fundamental rights, applicable regulation, and core principles and values and (2) it should be technically robust and reliable since, even with good intentions, a lack of technological mastery can cause unintentional harm”¹⁶.



The Netherlands¹⁷, Slovakia¹⁸, and Cyprus¹⁹ only mention the concept of Trustworthy AI without going into details. Other Member States mention the need for AI to be trustworthy, amongst other criteria (Luxembourg)²⁰ or use related language such as “responsible and trusted AI” (Czech Republic) or “trust in AI” (Germany).



Outside of the EU, Norway has also included an entire chapter on Trustworthy AI, noting, for example, that supervisory authorities must be empowered to “ensure compliance with the principles for responsible and trustworthy artificial intelligence”²¹.

¹⁶ See *Lithuanian Artificial Intelligence Strategy, a A vision of the future*, available at <http://kurklt.lt/wp-content/uploads/2018/09/StrategyIndesignpdf.pdf>.

¹⁷ See *Strategic Action Plan for Artificial Intelligence*, available at <https://www.government.nl/documents/reports/2019/10/09/strategic-action-plan-for-artificial-intelligence>.

¹⁸ See *Action plan for the digital transformation of Slovakia for 2019-2022*, available at <https://www.mirri.gov.sk/wp-content/uploads/2019/10/AP-DT-English-Version-FINAL.pdf>.

¹⁹ See *Cyprus AI Strategy*, available at https://ec.europa.eu/knowledge4policy/sites/know4pol/files/cyprus_ai_strategy.pdf.

²⁰ See *Artificial Intelligence: a strategic vision for Luxembourg*, available at https://digital-luxembourg.public.lu/sites/default/files/2020-09/AI_EN_0.pdf.

²¹ See *National Strategy for Artificial Intelligence*, available at https://www.regjeringen.no/contentassets/1feb3bb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf

The Americas



In 2017, Canada became the first country to produce an AI strategy²² (at least in this most recent wave of AI research, as there were much earlier initiatives, such as Japan's *Fifth Generation Computer Project*²³). Given that it precedes the EU *Ethics Guidelines for Trustworthy AI*, it does not make mention of the concept of Trustworthy AI, although it does promote the responsible use of AI through a number of initiatives, such as the *Directive on Automated Decision Making*²⁴ and the pioneering *Algorithmic Impact Assessment (AIA)* tool²⁵.

Following a call launched by the French President and the Canadian Prime Minister in the June 2018 French-Canadian Declaration on Artificial Intelligence, Canada and France spearheaded the launch of the *Global Partnership on AI (GPAI)*, an initiative aiming to “support the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values, as elaborated in the OECD Recommendation on AI”²⁶. With the involvement of the EU, and the collaboration of OECD, it seems highly likely that Trustworthy AI and other elements of the EU approach will have a strong influence here.

²² See *CIFAR Pan-Canadian Artificial Intelligence Strategy*, available at <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>.

²³ See Fifth generation computer, available at https://en.wikipedia.org/wiki/Fifth_generation_computer.

²⁴ See *Directive on Automated Decision-Making*, available at <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592>.

²⁵ See *Algorithmic Impact Assessment*, available at <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>.

²⁶ The full list of countries involved is Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, the Republic of Korea, Singapore, Slovenia, the United Kingdom, the United States of America, and the European Union. See *Joint Statement from founding members of the Global Partnership on Artificial Intelligence*, available at <https://www.canada.ca/en/innovation-science-economic-development/news/2020/06/joint-statement-from-founding-members-of-the-global-partnership-on-artificial-intelligence.html>.



The United States has published a number of documents outlining its strategy on AI, including the Presidential Document, *Maintaining American Leadership in Artificial Intelligence*²⁷, and other documents that are gathered in the list of resources, *Artificial Intelligence for the American People*²⁸. A number of these documents mention trustworthiness as a desirable characteristic of AI system, and according to *Reuters*, the “Trump administration said agencies should ‘promote trustworthy AI’ and ‘must consider fairness, non-discrimination, openness, transparency, safety, and security’”²⁹.



While many other countries in the region have published AI strategies³⁰, such as Mexico³¹, they have not explicitly taken up the language and concepts of the EU approach.

²⁷ See *Maintaining American Leadership in Artificial Intelligence*, available at <https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence>.

²⁸ See *Artificial Intelligence for the American People*, available at <https://www.whitehouse.gov/ai/>.

²⁹ See *White House proposes regulatory principles to govern AI use*, available at <https://www.reuters.com/article/us-tech-ces-ai-white-house/white-house-proposes-regulatory-principles-to-govern-ai-use-idUSKBN1Z60GL>.

³⁰ See *National AI policies popping up across South America*, available at <https://www.bnamericas.com/en/news/national-ai-policies-popping-up-across-south-america--marketing-or-strategy>.

³¹ See *Towards an AI strategy in Mexico: Harnessing the AI Revolution*, available at https://7da2ca8d-b80d-4593-a0ab-5272e2b9c6c5.filesusr.com/ugd/7be025_e726c582191c49d2b8b6517a590151f6.pdf.

Asia Pacific



New Zealand (Aotearoa) has taken up the EU language in its *Trustworthy AI in Aotearoa AI Principles*³². In addition to using the term “Trustworthy AI”, the New Zealand principles also echo the EU approach in foregrounding human rights, noting that the human rights framework “provides a ready-made, internationally tested, and legitimate framework of civil, political, economic, cultural and social values, addressing both individual and collective concerns”.



Australia's initial discussion paper, *Artificial Intelligence: Australia's Ethics Framework*³³, did not adopt the term Trustworthy AI and proposed a set of principles that were far more utilitarian than the human-rights-based approach promoted by the EU. However, a number of responses to the consultation urged greater consideration of human rights and a move away from the original utilitarian approach.

Access Now, for example, made a submission to the consultation process on this draft ethics framework³⁴, recommending that it revise several of its principles and move to a human-rights-based approach. The Law Council of Australia made several recommendations for the Australian principles to adopt the EU approach, including a recommendation for a “requirement that ethics and rule of law principles be included ‘by design’” and that the “use of AI systems for ‘scoring’ of citizens should be restricted, to avoid

³² See *Trustworthy AI in Aotearoa, AI Principles*, available at <https://aiforum.org.nz/wp-content/uploads/2020/03/Trustworthy-AI-in-Aotearoa-March-2020.pdf>.

³³ See *Australia's AI Ethics Principles*, available at <https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>.

³⁴ See *Access Now submission to the Department of Industry, Innovation and Science's Paper Artificial Intelligence: Australia's Ethics Framework*, available at <https://www.accessnow.org/cms/assets/uploads/2020/02/Access-Now-submission-to-the-Department-of-Industry-Innovation-and-Science%E2%80%99s-Paper-Artificial-Intelligence-Australia%E2%80%99s-Ethics-Framework.pdf>.

the undermining of human rights”, in all cases citing the EU guidelines as a positive example. Following the consultation, the revised principles³⁵ now make more explicit mention of the need for AI systems to respect human rights³⁶.



Singapore, in its paper, *Smart Nation Singapore, National Artificial Intelligence Strategy: Advancing our Smart Nation Journey*, identifies five key enablers for its proposed AI ecosystem, one of which is a Progressive and Trusted Environment, which aims to “strengthen trust in AI technologies to enable an environment for test-bedding, developing, and deploying AI solutions”³⁷. The three other enablers are not explicitly linked to the EU’s trustworthy framework (they are: Triple Helix Partnership between the Research Community, Industry and Government, AI Talent and Education, and Data Architecture).



Even the Chinese government, whose use of AI in mass surveillance and in the oppression of the Uighur minority has become shorthand for “unethical” AI, has made some moves to promote AI principles³⁸. The *Beijing AI Principles*, published by the Beijing Academy of Artificial Intelligence (BAAI), an organisation backed by the Chinese Ministry of Science and Technology and the Beijing municipal government, outline seven principles for AI research and development, one of which is to “Be Ethical” and which makes mention of trustworthiness:

³⁵ See *Australia’s AI Ethics Principles*, available at

<https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles>.

³⁶ The Human Rights Commissioner of Australia has been carrying out a consultation and publication process about “protecting and promoting human rights amid the rise of new technologies”. More details are available here https://tech.humanrights.gov.au/?_ga=2.185828841.1381358398.1604322531-896747033.1604322531.

³⁷ See *National AI Strategy: The next key frontier of Singapore’s Smart Nation Journey*, available at <https://www.smartnation.gov.sg/why-Smart-Nation/NationalAIStrategy>.

³⁸ See *One Month, 500,000 Face Scans: How China is Using A.I. to Profile a Minority*, available at <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>

“AI R&D should take ethical design approaches to make the system trustworthy. This may include, but not limited to: making the system as fair as possible, reducing possible discrimination and biases, improving its transparency, explainability, and predictability, and making the system more traceable, auditable, and accountable”³⁹.

Private Sector & Standards Bodies

The EU approach to AI governance has also had an impact on a number of companies. Mozilla, for example, adopted the term midway through 2019⁴⁰, and now use it as an umbrella term for their work on artificial intelligence. IBM has adopted “Trusting AI” as one of its three pillars of AI research alongside “Advancing AI” and “Scaling AI”. This includes the development of technical tools to check AI systems for bias in its AI Explainability 360 Open Source Toolkit⁴¹. Fujitsu, as part of its focus on human-centric ICT, has outlined principles for human-centric AI, which mention that the company “will seek trustworthy AI through considering fairness and safety to prevent discrimination and harm”.

Deloitte has put forward its own Trustworthy AI Framework which “aims to help businesses increase brand equity and trust, which can lead to new customers, employee retention, and more customers opting in to share data”⁴². In July 2019, Vodafone also launched its Artificial Intelligence Framework⁴³, which is based on the idea of Trustworthy AI, and aims to promote: transparency and accountability; ethics and fairness; privacy and security; human rights, diversity, and inclusion; and to ensure an equitable transition to AI and contribute to building an inclusive digital society. In addition to Microsoft’s work on Responsible AI, Microsoft

³⁹ See *Beijing AI Principles*, available at <https://www.baai.ac.cn/news/beijing-ai-principles-en.html>. For a comparison of the Chinese and EU approaches, see *Comparing China’s and EU’s Artificial Intelligence Strategies*, available at <https://chinaobservers.eu/comparing-chinas-and-eus-artificial-intelligence-strategies/>.

⁴⁰ See *Mozilla’s Approach to Trustworthy Artificial Intelligence (AI)*, available at <https://foundation.mozilla.org/en/blog/mozillas-approach-to-trustworthy-artificial-intelligence-ai/>.

⁴¹ See *IBM Trusting AI*, available at <https://www.research.ibm.com/artificial-intelligence/trusted-ai/>.

⁴² See *Deloitte Introduces Trustworthy AI Framework to Guide Organizations in Ethical Application of Technology in the Age of With*, available at <https://www2.deloitte.com/us/en/pages/about-deloitte/articles/press-releases/deloitte-introduces-trustworthy-ai-framework.html>.

⁴³ See *Vodafone launches Artificial Intelligence framework*, available at <https://www.vodafone.com/perspectives/blog/vodafone-launches-artificial-intelligence-framework>.

Research has a project with MIT's Computer Science & Artificial Intelligence Lab on Trustworthy & Robust AI Collaboration (TRAC)⁴⁴.

In November 2016, as one of the first industry initiatives, Google, Facebook, Amazon, IBM, and Microsoft announced the launch of the Partnership on AI “to advance public understanding of the sector, as well as coming up with standards for future researchers to abide by”⁴⁵. The Partnership has evolved towards a multistakeholder forum to bring together industry members, civil society, and others to “conduct research, organize discussions, share insights, provide thought leadership, consult with relevant third parties, respond to questions from the public and media, and create educational material that advances the understanding of AI technologies including machine perception, learning, and automated reasoning”⁴⁶.

Standards bodies have also integrated the concept of Trustworthy AI in their work. The Institute of Electrical and Electronics Engineers (IEEE), for example, published their *Trustworthy AI Development Guidelines for Human System Interaction*⁴⁷, and the International Organization for Standardization, an independent, non-governmental international organisation with a membership of 165 national standards bodies, has published a technical report on *Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence*⁴⁸.

International Institutions

Perhaps the biggest impact of the EU approach can be seen in the use of the term “Trustworthy AI” in the AI ethics principles developed by the Organisation for Economic Co-operation and Development (OECD). Published only a month after the AI HLEG guidelines, the OECD principles use the concept of Trustworthy AI and arguably converge to a large extent with the AI HLEG guidelines. As Nathalie Smuha has noted, this is hardly surprising

⁴⁴ See *Trustworthy & Robust AI Collaboration (TRAC): A Microsoft Research & MIT CSAIL Collaboration*, available at <http://trac.csail.mit.edu/>.

⁴⁵ See ‘Partnership on AI’ formed by Google, Facebook, Amazon, IBM and Microsoft, available at <https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms>.

⁴⁶ Access Now resigned from the Partnership on AI in October 2020. See *Access Now resigns from Partnership on AI*, available at <https://www.accessnow.org/access-now-resignation-partnership-on-ai/>.

⁴⁷ See *Trustworthy AI Development Guidelines for Human System Interaction*, available at <https://ieeexplore.ieee.org/document/9142644/authors#authors>.

⁴⁸ See *Towards a Trustworthy AI*, available at <https://www.iso.org/news/ref2530.html>.

given that the group that developed the OECD principles shared a number of experts with the AI HLEG and also included the European Commission itself⁴⁹.

As Smuha further notes, the influence on the OECD principles did much to ensure the global impact of the EU approach, as the principles were adopted by 42 countries and formed the basis for a G20 declaration that also used these same principles and included some countries that had not adopted the OECD principles. Some hailed the principles as a success for getting the US onboard, with *Politico* noting that this “marks the first time that the United States — home to some of the world's largest and most powerful tech companies — has endorsed international guidelines for the emerging technologies”⁵⁰. (Regardless of its endorsement of these guidelines, however, the US is obliged to respect commitments to international human rights law as it relates to emerging technologies and the numerous United Nations resolutions that include language on emerging technologies.)

At the same time, critics have noted that the OECD principles, while “derived” from the AI HLEG guidelines, are significantly vaguer and arguably weaker⁵¹. Given that human rights organisations, such as Access Now, had already criticized the AI HLEG guidelines themselves for being too weak, we should temper any optimism about the global impact of the EU approach.

From an optimistic perspective, it would seem that the concept of Trustworthy AI has become mainstream, and we now need to see how it can be put into practice. From a critical perspective, the spread of the term “Trustworthy AI” across the world may indicate nothing more than the success of a branding exercise instead of providing accountability for the content of those principles. The major risk is that ethics guidelines are simply insufficient to effectively draw red lines regarding the use of AI. As Access Now noted when the AI HLEG guidelines were released, we need the EU to “lay down what Europe’s red lines are to prevent

⁴⁹ See Nathalie Smuha, (2019), *The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence*, Computer Law Review International Vol. 20; iss. 4; pp. 97 –106. P. 17, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3443537.

⁵⁰ See *US to endorse new OECD principles on artificial intelligence*, available at <https://www.politico.eu/article/u-s-to-endorse-new-oecd-principles-on-artificial-intelligence/>.

⁵¹ See Nathalie Smuha, (2019), *The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence*, Computer Law Review International Vol. 20; iss. 4; p. 17, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3443537.

the development or deployment of AI in certain areas, and to see how we can ensure that Trustworthy AI is not just an empty brand name”⁵².

Critiques of the European Union approach

The European Commission's efforts to drive the AI debate — including the establishment of the HLEG and the produced documents — have stirred a lot of attention and provoked strong reactions from different stakeholders.

On the positive side, Digital Europe (a trade association representing industry in Europe⁵³) that was represented in the HLEG by Digital Europe's Director General Cecilia Bonefeld-Dahl welcomed the HLEG's approach by praising the composition of the group (“[t]he result of the HLEG is a breakthrough in the sense that it is the outcome of a very diverse multi-stakeholder group with members from all types of backgrounds”) and pointed out that “[w]e need to get it right in order to drive European innovation and welfare and to avoid the risks of misuse of AI. We outline the common European values and principles that AI should respect”.

The law firm, DLA Piper, pointed out in its briefing on the EU's *Ethics Guidelines for Trustworthy AI* (“Ethics Guidelines”) in April 2019, that “[t]he Guidelines are not legally binding and do not replace any current or future regulations applicable to AI systems. However, they are important for companies and other entities and persons developing, deploying, or using AI, as the requirements and the framework provided by the Guidelines are likely to be an important point of reference for the policy-makers and legislators at the EU and national level working on the future legislative and regulatory frameworks for AI”⁵⁴.

BSA | The Software Alliance expressed support for the Commission's approach in its submission to what were then draft Ethics Guidelines, “[t]he formation of the HLEG is a unique opportunity for Europe's leading experts from industry, academia, and civil society to help the European Commission develop a ‘coordinated approach to make the most of the opportunities offered by AI and to address the new challenges that it brings.’ We agree with the Commission that the success of such a framework will turn in large part on whether it fosters an ‘environment of trust and accountability around the development and use of AI’”.

⁵² See *Laying down the law on AI: ethics done, now the EU must focus on human rights*, available at <https://www.accessnow.org/laying-down-the-law-on-ai-ethics-done-now-the-eu-must-focus-on-human-rights/>.

⁵³ See Digital Europe About us, available at <https://www.digitaleurope.org/about-us/>.

⁵⁴ See *EU Policy & Regulatory Alert - EU Publishes Artificial Intelligence Ethics Guidelines*, available at <https://www.dlapiper.com/en/belgium/insights/publications/2019/04/eu-publishes-artificial-intelligence-ethics-guidelines/>.

The industry association maintained its positive opinion about the Commission's initiative in publishing its White Paper⁵⁵, and “BSA supports risk-based approaches to AI governance that are informed by existing law, and account for context-specific considerations in determining whether specific applications of AI should be regulated. BSA therefore welcomes the Commission's decision to adopt such an approach as the foundation for the AI White Paper”⁵⁶.

Facebook also expressed broad support for the overall framing of the EU's White Paper. The company points out that “just like other emerging technologies, AI also raises unique policy and legal challenges — hard questions about how to ensure that the growing number of AI systems that help us make important decisions are fair, transparent, accountable, and privacy-protecting. Which is why we are glad to see that the European Union, which has already proved itself to be a leader in technology regulation with its influential General Data Protection Regulation, is prioritizing those questions”⁵⁷.

In addition to making a positive impact on the global conversation on AI governance and garnering some positive feedback from EU stakeholders, the EU approach has also faced criticism — both internal and external — right from the beginning. Critics have highlighted the following issues: lack of sufficient measures to protect fundamental rights; overrepresentation of industry voices on the expert group, leading to a watering down of restrictions; lack of representation of affected people and communities; lack of consideration and resources to compensate for varying resources and capacities between stakeholders in the group; expertise and background of members are under-utilised or mismatching tasks; and the claim that the focus on ethics was being used a way to dodge regulation.

The following quotes and materials serve as concrete examples for how these deficiencies were voiced by different members of the AI community.

One example of this criticism came from a group of HLEG members, including Access Now's Fanny Hidvégi. Hidvégi and two other HLEG members — Ursula Pahl, Deputy Director General of BEUC and Chiara Giovannini, Deputy Secretary General of ANEC — pointed out

⁵⁵ See *White Paper on Artificial Intelligence: a European Approach to Excellence and Trust*, available at https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

⁵⁶ See *BSA submission to the European Commission Consultation on the White Paper on Artificial Intelligence*, available at <https://www.bsa.org/files/policy-filings/061220euwhiteppaerai.pdf>.

⁵⁷ See *Collaborating on the future of AI governance in the EU and around the world*, available at <https://ai.facebook.com/blog/collaborating-on-the-future-of-ai-governance-in-the-eu-and-around-the-world/>.

that ethics guidelines are not sufficient to ensure the protection of fundamental rights (even when ethical principles are “based on fundamental rights”), and asked the European Commission to undertake a number of steps, including carrying out a “comprehensive mapping of existing legislation that applies to AI development and deployment, and an identification of legal uncertainties and gaps”⁵⁸. The group also criticised industry dominance in HLEG, which out of 56 experts in total⁵⁹, contained 37 industry representatives (the remaining members included 18 academics, only four civil society representatives, and a number of other stakeholders).

In an article on the HLEG’s *Policy and Investment Recommendations* (PIR)⁶⁰, Michael Veale⁶¹, a scholar at University College London and the Alan Turing Institute whose research focuses on the “intersections of emerging digital technologies, Internet and data law, technology policy and human–computer interaction”, further noted the lack of “low-level” experts: “Seeing a greater role for ethicists of technology or conscientious engineers as the correct response to injustice exacerbated by technology but not, at its root, caused by it risks further marginalising those with the clearest view of on-the-ground issues and the closest connection and legitimacy to affected communities. The HLEG is notable by its exclusion of such voices, seeing expertise in artificial intelligence as the domain of technical researchers, generalist ethics and governance scholars, industry lobby groups”⁶².

While Thomas Metzinger, an academic and a HLEG member⁶³, acknowledged that the Ethics Guidelines produced by the HLEG were “currently the best globally available platform for the next phase of discussion”, he was sharply critical of a number of points. In addition to echoing the criticism of industry dominance, Metzinger highlighted industry resistance to the idea of drawing red lines to prohibit certain uses of AI, claiming that he was asked by the

⁵⁸ See *AI ethics guidance a first step but needs to be transformed into tangible rights for people*, available at <https://www.accessnow.org/ai-ethic-guidance-a-first-step-but-needs-to-be-transformed-into-tangible-rights-for-people/>.

⁵⁹ Although the group contained only 52 experts at any one time, some members left and were replaced at various times. For an overview of the members, see *High-Level Expert Group on Artificial Intelligence*, available at <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.

⁶⁰ Access Now’s Fanny Hidvégi opted out from endorsing the final Sectoral Considerations on Policy and Investment Recommendations for Trustworthy AI made by the HLEG, available at <https://futurium.ec.europa.eu/en/european-ai-alliance/document/ai-hleg-sectoral-considerations-policy-and-investment-recommendations-trustworthy-ai>.

⁶¹ See Michael Veale’s bio, available at <https://www.ucl.ac.uk/laws/people/dr-michael-veale>.

⁶² See Michael Veale, *A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence*, (2020) *European Journal of Risk Regulation*, doi:10/ggjdjs, available at <https://osf.io/preprints/lawarxiv/dvx4f/download>.

⁶³ See *Ethics washing made in Europe*, available at <https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html>.

HLEG president “whether we could remove the phrase ‘non-negotiable’ from the document” and that “many industry representatives and group members interested in a “positive vision” vehemently insisted that the phrase ‘Red Lines’ be removed entirely from the text”. This led to Metzinger claiming that the document was an example of “ethics washing”, where “[i]ndustry organises and cultivates ethical debates to buy time – to distract the public and to prevent or at least delay effective regulation and policy-making”.

In line with other criticisms of the focus on AI adoption/uptake in the EU approach, Veale also noted that the PIR includes a long list of recommendations without a clear prioritisation, but among the suggestions for different types of AI applications, it “foregrounds AI as a technological solution to completely inappropriate issues without considering the capacities needed to understand the problems which AI could potentially be applied to”. Furthermore, he points to the flawed assumption that the harms caused by AI systems are due to “ethical oversights”, which makes these harms “appear like bone fide oversights that ethicists might be able to highlight rather than intrinsic parts of business models which disregard their effects on societies and environments”⁶⁴.

Finally, Veale argues that taken as a whole, the process to develop AI ethics recommendations may serve to “further sideline and marginalise community and domain voices, and seek to reify an elite club of AI and society experts to the detriment of those with connection to harms and issues that technologies exacerbate”⁶⁵.

In addition to criticisms regarding the weakness or insufficiency of the ethics guidelines and other aspects of the EU approach, there were some voices who lamented that the guidelines were already too strict. The Centre for Data Innovation, a Brussels-oriented offshoot of the US Information Technology and Innovation Foundation, claimed that even the voluntary measures proposed by the ethics guidelines would hamper innovation⁶⁶.

The relative “strictness” of the EU approach also seemed to have worried some policymakers in other countries. In 2020 in the United States, the White House released an updated AI

⁶⁴ See Michael Veale, *A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence*, (2020) *European Journal of Risk Regulation*, doi:10/ggjdjs, available at <https://osf.io/preprints/lawarxiv/dvx4f/download>.

⁶⁵ *Ibid.*

⁶⁶ See *Recommendations to the EU High Level Expert Group on Artificial Intelligence on its Draft AI Ethics Guidelines for Trustworthy AI*, available at <http://www2.datainnovation.org/2019-hleg-ai.pdf>.

policy guidelines that strongly advocated against heavy-handed regulation⁶⁷. The White House Office of Science and Technology released this statement at the time:

Europe and our allies should avoid heavy handed innovation-killing models, and instead consider a similar regulatory approach. The best way to counter authoritarian uses of AI is to make sure America and our international partners remain the global hubs of innovation, shaping the evolution of technology in a manner consistent with our common values⁶⁸.

The incoming Biden-Harris administration has not detailed its exact plans for AI, but “the Democrat’s campaign indicated that it considers general scientific research and development to be crucial to the nation”⁶⁹. It’s yet to be seen what policies the new administration will implement on issues related to AI⁷⁰.

While no one would favour regulation that puts an unnecessary burden on any actor without a specific objective or that creates a competitive disadvantage, the fact that policymakers in the US, a country purportedly “winning the AI race”, have voiced concerns about EU regulation on AI indicates that it is expected to have real impact on a global scale. If American or Chinese companies want to sell their AI-based products and services in the EU, regulation could force them to comply with EU standards, and could lead to enhanced rights and controls for users and consumers worldwide. Such an impact has already been seen with the GDPR, and some companies, such as Microsoft, have made a commitment⁷¹ to roll out GDPR-inspired rights not only in Europe but worldwide.

Council of Europe and other critiques from a human rights standpoint

Some of the most stringent criticisms of the EU approach have come from a human rights standpoint. In addition to Access Now’s work on AI and human rights, including the Access Now and Amnesty International-led *Toronto Declaration on protecting the right to equality and*

⁶⁷ See *Memorandum for the Heads of Executive Departments and Agencies*, available at <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>.

⁶⁸ See *White House urges federal agencies and European allies to avoid overregulation of AI*, available at <https://venturebeat.com/2020/01/06/white-house-urges-federal-agencies-and-european-allies-to-avoid-overregulation-of-ai/>.

⁶⁹ See *What a Biden-Harris administration means for artificial intelligence*, available at <https://fortune.com/2020/11/10/biden-harris-administration-artificial-intelligence/>.

⁷⁰ See *Official Campaign Website Battle for the Soul of the Nation*, available at <https://joebiden.com/joes-vision/>.

⁷¹ See *Microsoft’s commitment to GDPR, privacy and putting customers in control of their own data*, available at <https://blogs.microsoft.com/on-the-issues/2018/05/21/microsofts-commitment-to-gdpr-privacy-and-putting-customers-in-control-of-their-own-data/>.

non-discrimination in machine learning systems (the “Toronto Declaration”)⁷², organisations such as ARTICLE 19⁷³, Human Rights Watch⁷⁴, and others have put out a number of reports analysing aspects of AI governance from a human rights standpoint⁷⁵.

A common theme throughout this work on human rights and AI governance has been that ethics guidelines are insufficient to mitigate the harms caused by the use of AI systems. Rather than use the HLEG approach of developing ethical principles “based on” fundamental rights, critics argue for applying human rights standards and processes themselves, such as mandatory human rights impact assessments (HRIAs), to the governance of AI systems. The HLEG tool for the self-assessment of the Trustworthy Ethics Guidelines does not correspond to or meet the criteria for such requirement.

The Council of Europe has been very active in the debate on AI and human rights, producing a number of important reports and recommendations. On 8 April, 2020, the Committee of Ministers adopted recommendations to Member States on the human rights impacts of algorithmic systems⁷⁶, based on the work of the Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), for which Access Now provided comments⁷⁷.

The Committee of Ministers acknowledges that there are “significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life, such as regarding the right to a fair trial; the right to privacy and data protection; the right to freedom of

⁷² See *The Toronto Declaration Protecting the right to equality and non-discrimination in machine learning systems*, available at <https://www.torontodeclaration.org/declaration-text/english/>.

⁷³ See *Privacy and Freedom of Expression In the Age of Artificial Intelligence*, available at <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf>.

⁷⁴ See *UK: Automated Benefits System Failing People in Need*, available at <https://www.hrw.org/news/2020/09/29/uk-automated-benefits-system-failing-people-need>.

⁷⁵ See *The Toronto Declaration, Protecting the right to equality and non-discrimination in machine learning systems*, available at <https://www.torontodeclaration.org/declaration-text/english/>. See also Mark Latonero’s report, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, available at https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

⁷⁶ See recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, available at <https://rm.coe.int/09000016809e1154>.

⁷⁷ See *Comments on the draft recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems by Access Now and the Wikimedia Foundation*, available at <https://www.accessnow.org/cms/assets/uploads/2019/10/Submission-on-CoE-recommendation-on-the-human-rights-impacts-of-algorithmic-systems-21.pdf>.

thought, conscience and religion; the right to freedom of expression; the right to freedom of assembly; the right to equal treatment; and economic and social rights”.

It urges the 47 member states of the Council of Europe to “develop legislative and regulatory frameworks that foster an environment where all actors respect and promote human rights and seek to prevent possible infringements”. The Council of Europe’s guidelines address both Member State obligations and the responsibilities of private sector actors, recalling general human rights obligations but also providing recommendations on data management, analysis and modeling, transparency, accountability, and precautionary measures.

In September 2019, the Council of Europe established the Ad hoc Committee on Artificial Intelligence of the Council of Europe (CAHAI), of which Access Now is an observer. Its mission is to engage in broad multi-stakeholder consultations to examine the feasibility of a legal framework for the development, design, and application of artificial intelligence, based on Council of Europe’s standards on human rights, democracy, and the rule of law. During the stakeholder consultation, the CAHAI is taking stock of the positions of Member States but also of civil society, corporations, and other international organisations.

The CAHAI Policy Development Group is currently preparing a Draft Feasibility Study in view of a future Council of Europe legal framework on AI. At the time of writing, the study is in a draft stage, and will be finalised and formally sent to the CAHAI plenary in view of its submission and examination by the CAHAI during its Plenary meeting on 15-17 December, 2020.

Also from Council of Europe, on 22 October, 2020, the Parliamentary Assembly of the Council of Europe (PACE) adopted seven reports concerning the impact of AI: the need for democratic governance of AI⁷⁸; the role of AI in policing and criminal justice systems⁷⁹; preventing discrimination caused by AI⁸⁰; ethical and legal frameworks for the research and development

⁷⁸ See *Need for democratic governance of artificial intelligence*, available at <https://pace.coe.int/en/files/28742/html>.

⁷⁹ See *Justice by algorithm - the role of artificial intelligence in policing and criminal justice systems*, available at <https://pace.coe.int/en/files/28723/html>.

⁸⁰ See *Preventing discrimination caused by the use of artificial intelligence*, available at <https://pace.coe.int/en/files/28715/html>.

of neurotechnology⁸¹; AI and health care⁸²; consequences of AI on labour markets⁸³; and legal aspects of “autonomous vehicles”⁸⁴. The PACE committee also proposed that the Committee of Ministers support the elaboration of a “legally binding instrument” governing AI, possibly in the form of a Convention⁸⁵.

Responses to the AI Whitepaper & next steps

Published in February 2020, the *White Paper on Artificial Intelligence: a European approach to excellence and trust*⁸⁶ (“the White paper”) presents a European framework for AI, built upon the EU’s previous work on AI such as the HLEG’s *Ethics Guidelines for Trustworthy Artificial Intelligence*.⁸⁷

The White paper presents policy options to promote the uptake of AI and to address “the risks associated with certain uses of this new technology”. As well as outlining actions to foster the development and the adoption of AI, the White Paper presents a new regulatory framework to address specific concerns about AI, embracing a risk-based approach focusing on high-risk applications.

In concrete terms, this would mean adding legal requirements only to applications of AI identified as high-risk according to two cumulative criteria based on the sector and the use and impact of the AI system. The first criterion is sectoral: it considers whether “the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur”. The second relates to the likelihood of risk: considering whether “the AI application in the sector in question is, in addition, used

⁸¹ See *The brain-computer interface: new rights or new threats to fundamental freedoms?*, available at <https://pace.coe.int/en/files/28722/html>.

⁸² See *Artificial intelligence in health care: medical, legal and ethical challenges ahead*, available at <https://pace.coe.int/en/files/28737/html>.

⁸³ See *Artificial intelligence and labour markets: friend or foe?*, available at <https://pace.coe.int/en/files/28738/html>.

⁸⁴ See *Legal aspects of “autonomous vehicles”*, available at <https://pace.coe.int/en/files/28721/html>.

⁸⁵ See *Establishing a ‘legally binding instrument’ for democratic governance of AI*, available at <https://www.coe.int/en/web/artificial-intelligence/-/mettre-en-place-un-instrument-juridiquement-contraignant-pour-une-gouvernance-democratique-de-l-ia>.

⁸⁶ See *EU White Paper on Artificial Intelligence: a European approach to excellence and trust*, available at https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

⁸⁷ See *Ethics Guidelines for Trustworthy AI*, available at <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

in such a manner that significant risks are likely to arise” in acknowledgment of that fact “that not every use of AI in the selected sectors necessarily involves significant risks”⁸⁸.

In addition to optional general approaches such as clarifying existing legislation, for the applications not classified as high-risk, the paper proposes a voluntary labelling scheme as a potential measure (Section G). The White Paper also notes, however, that “there may also be exceptional instances where, due to the risks at stake, the use of AI applications for certain purposes is to be considered as high-risk as such – that is, irrespective of the sector concerned and where the below requirements would still apply”⁸⁹.

A number of stakeholders, from civil society organisations⁹⁰ and individual researchers⁹¹ to the European Data Protection Supervisor (EDPS)⁹², have raised criticisms against this risk-based approach in their responses to the consultation on the White Paper. The EDPS, for example, has argued that the proposed risk-based approach is “too narrow, as it would seem to exclude individuals from being adequately protected from AI applications that could infringe on their fundamental rights”⁹³.

In Access Now’s response, we argued that the Commission has reversed its priorities by adopting a risk-based approach: the primary objective of a regulation on AI should be protect and promote fundamental rights enshrined in the Charter, to avoid individual and societal harms, not to promote AI uptake and then to try and mitigate any harms caused.⁹⁴

The European Digital Rights (EDRI) network, a network of 44 NGOs (including Access Now), as well as experts, advocates, and academics working to defend and advance digital rights across Europe and beyond, responded to the White Paper consultation with a call “for

⁸⁸ See *EU White Paper* p. 17, available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

⁸⁹ *Ibid.* p.18

⁹⁰ See *Can the EU make AI “trustworthy”? No - but they can make it just*, available at <https://edri.org/our-work/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/>.

⁹¹ See *The EC’s risk based approach to AI regulation is inadequate, here’s why*, available at https://medium.com/@hello_95259/the-ecs-risk-based-approach-to-ai-regulation-is-inadequate-here-s-why-6fd6da4d5aba.

⁹² See *Opinion 4/2020: EDPS Opinion on the European Commission’s White Paper on Artificial Intelligence - A European approach to excellence and trust*, available at https://edps.europa.eu/sites/edp/files/publication/20-06-19_opinion_ai_white_paper_en.pdf.

⁹³ *Ibid.* p.11-12.

⁹⁴ See *Access now’s submission to the Consultation on the “White Paper on Artificial Intelligence - a European approach to excellence and trust”*, available at https://www.accessnow.org/cms/assets/uploads/2020/06/EU-white-paper-consultation_Access_Now_June2020.pdf.

fundamental rights to be prioritised in the regulatory proposal for all AI systems, not only those categorised as ‘high-risk’”, and also argued that “AI regulation should avoid creating loop-holes or exemptions based on sector, size of enterprise, or whether or not the system is deployed in the public sector”⁹⁵. The Polish NGO Panoptykon, also an EDRI member, called for an approach “based on mandatory HRIA [human rights impact assessments] for all AI applications that may affect humans, combined with public disclosure obligations”⁹⁶. AlgorithmWatch, a German NGO, called for “robust, legally-binding data access frameworks, focused explicitly on supporting and enabling public interest research and in full respect of data protection and privacy law”⁹⁷.

On the other side of the debate, 14 Member States (including France and Poland) led by Denmark, sent a position paper to the Commission which calls on the EU to “avoid setting burdensome barriers and requirements which can be a hindrance for innovation”⁹⁸. They caution against over-regulation and plead for an approach that puts innovation front and centre. One of the key worries they cite is that the European Commission’s proposed risk-based approach to regulating AI will end up classifying too many AI systems as high-risk. They argue for an “objective methodology” in assessing the risk of such systems, and suggest the risk-classification “should make the category of high-risk AI the exception rather than the rule”.

In response to this paper, Access Now and EDRI published a response which agreed on the need for an objective methodology, but noted that “the ultimate goal of such a methodology should not be to limit the number of AI systems classified as high risk” because the reason why “we need to identify risks at all is to better protect our rights — not to make things easier for companies at any cost”⁹⁹.

The White Paper also came under fire for dropping consideration of a ban or moratorium on certain applications of AI such as facial recognition. Despite including the option of a

⁹⁵ See *Can the EU make AI “trustworthy”? No - but they can make it just*, available at <https://edri.org/our-work/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/>.

⁹⁶ See *Panoptykon Foundation’s submission to the consultation on the ‘White Paper on Artificial Intelligence - a European approach to excellence and trust’*, available at https://panoptykon.org/sites/default/files/stanowiska/panoptykon_ai_whitepaper_submission_10.06.2010_fin_al.pdf.

⁹⁷ See *AlgorithmWatch Our response to the European Commission consultation on AI*, available at <https://algorithmwatch.org/en/response-european-commission-ai-consultation/#H>.

⁹⁸ See *Innovative and Trustworthy AI: Two Sides of the Same Coin*, available at <https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf>.

⁹⁹ See *Attention EU regulators: we need more than AI “ethics” to keep us safe*, available at <https://www.accessnow.org/eu-regulations-ai-ethics/>.

moratorium in an early leaked draft of the White Paper¹⁰⁰, potentially “a time-limited ban on the use of facial recognition technology in public spaces”, no mention of a potential ban was made in the final version.

Despite this omission, there are indications that a moratorium, or even a ban, may be back on the agenda for certain key figures and bodies in the EU. Margrethe Vestager, the European Commission's Vice-President for Digital policy, warned that applications of AI such as predictive policing are “not acceptable” in the EU¹⁰¹. In a similar spirit, Wojciech Wiewiórowski, the European Data Protection Supervisor, has announced that he is aiming to convince the European Commission to institute a moratorium on the use of facial recognition and other biometric surveillance technology in public spaces¹⁰².

As critics have pointed out, the EU can perfectly well decide to divest from or ban particular applications of AI technology while remaining competitive in others. AI development is not a zero-sum game where we must embrace all AI applications or none; it is possible for the EU to remain competitive in certain branches and applications of AI while having the maturity and foresight to refuse to develop and deploy other branches and applications that threaten or violate our rights that are laid down in EU Charter of Fundamental Rights.

As Access Now and other members of the EDRi network argue, applications of AI such as facial recognition pose such a threat to our rights that precaution must be put before innovation and competitiveness; in these cases, we need red lines rather than risk mitigation. If the EU wants to promote the idea of Trustworthy AI, it must demonstrate the resolve not to pursue the development and deployment of AI applications that undermine fundamental rights. The upcoming AI regulation offers a unique opportunity for the European Commission to show true leadership in AI governance.

¹⁰⁰ See *LEAK: Commission considers facial recognition ban in AI 'white paper'*, available at <https://www.euractiv.com/section/digital/news/leak-commission-considers-facial-recognition-ban-in-ai-white-paper/>.

¹⁰¹ See *Vestager warns against predictive policing in Artificial Intelligence*, available at <https://www.euractiv.com/section/digital/news/vestager-warns-against-predictive-policing-in-artificial-intelligence/>.

¹⁰² See *EU data watchdog to 'convince' Commission to ban automated recognition tech*, available at <https://www.euractiv.com/section/digital/news/eu-data-watchdog-argues-for-moratorium-on-recognition-technology/>.

III. HUMAN RIGHTS AND ETHICS AS GOVERNANCE FRAMEWORKS FOR AI

Beginning in 2016, there was a pronounced boom in AI ethics guidelines. A 2019 study by Anna Jobim et al., *The global landscape of AI ethics guidelines*¹⁰³, identified “84 documents containing ethical principles or guidelines for AI[...] with 88% having been released after 2016”. The inventory of AI ethics guidelines compiled by AlgorithmWatch lists over 160 such guidelines, and it is constantly being updated¹⁰⁴.

But what trends have emerged in this proliferation of guidelines, and has this enthusiasm for ethics guidelines been matched by any real impact in governing AI? In a study from Harvard University's Berkman Klein Center for Internet & Society, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*¹⁰⁵, the authors analysed a sample of 36 sets of AI principles, and note a “convergence” around eight key themes the various sets of principles share:

- Privacy
- Accountability
- Safety and Security
- Transparency and Explainability
- Fairness and Non-discrimination
- Human Control of Technology
- Professional Responsibility
- Promotion of Human Values

Although the authors note that these principles could be seen to represent a “normative” core, they caution against drawing any overly optimistic conclusions from this apparent convergence, noting that the impact of any set of AI ethics principles is not likely to be very great, given that it will largely “depend on how it is embedded in a larger governance ecosystem, including for instance relevant policies (e.g. AI national plans), laws, regulations, but also professional practices and everyday routines”.

¹⁰³ See *The global landscape of AI ethics guidelines*, available at <https://www.nature.com/articles/s42256-019-0088-2>.

¹⁰⁴ See *AI Ethics Guidelines Global Inventory*, available at <https://inventory.algorithmwatch.org/>.

¹⁰⁵ See *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, available at <https://dash.harvard.edu/handle/1/42160420>.

Indeed, it is this “lack of teeth”¹⁰⁶ that forms the basis for many of the criticisms levelled against AI ethics guidelines like the *Ethics Guidelines for Trustworthy AI* produced by the European Commission’s High-Level Expert Group on AI. As critics have noted, ethics guidelines have no real normative force; if a company or government violates a principle from a set of ethics guidelines, there are typically no enforcement mechanisms that could ensure compliance, and no established redress mechanisms.

As we note previously in this report, such criticisms of ethics guidelines have culminated in accusations of “ethics washing”, a term which refers to the use of ethics “as an acceptable façade that justifies deregulation, self-regulation or market driven governance, and is increasingly identified with technology companies’ self-interested adoption of appearances of ethical behavior”¹⁰⁷.

Responding to this criticism of ethics guidelines for AI governance, many have advocated for applying a human rights framework. One of the earliest initiatives advocating the application of the human rights framework to AI was the 2018 *Toronto Declaration - Protecting the right to equality and non-discrimination in machine learning systems*¹⁰⁸. Since then, there has been an ever-increasing amount of work in this area, with academics, civil society organisations, and international bodies all publishing work on human rights and AI.

Whereas voluntary ethics guidelines leave large scope for those developing and deploying AI systems to interpret what different principles mean, the international human rights framework has established mechanisms for resolving such ambiguities, and enforcing compliance, even if these processes have not always been without issues. As the authors of the *Principled Artificial Intelligence* study noted¹⁰⁹, the established mechanisms of the human rights framework could help in cases where ethical principles such as “fairness” are subject to conflicting interpretations, and provide “solutions for complex situations in which separate principles within a single document are in tension with one another”.

¹⁰⁶ See *Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence*, available at <https://www.article19.org/resources/governance-with-teeth-how-human-rights-can-strengthen-fat-and-ethics-initiatives-on-artificial-intelligence/>.

¹⁰⁷ See *From Ethics Washing to Ethics Bashing: A View on Tech Ethics from Within Moral Philosophy*, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3513182.

¹⁰⁸ See *The Toronto Declaration Protecting the right to equality and non-discrimination in machine learning systems*, available at <https://www.torontodeclaration.org/declaration-text/english/>.

¹⁰⁹ See *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, available at <https://dash.harvard.edu/handle/1/42160420>.

Two notable applications of the international human rights framework to issues of AI governance can be found in the reports published by former UN Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, and the current UN Special Rapporteur on Contemporary Forms of Racism, E. Tendayi Achiume.

In Alston's report on digital technology, social protection, and human rights¹¹⁰, he notes how the digitisation of the welfare state risks us “stumbling, zombie-like, into a digital welfare dystopia” where AI and other technologies are used to “automate, predict, identify, surveil, detect, target and punish” vulnerable people rather than help them. The report ends with “a call for the regulation of digital technologies, including artificial intelligence, to ensure compliance with human rights and for a rethinking of the positive ways in which the digital welfare state could be a force for the achievement of vastly improved systems of social protection”.

E. Tendayi Achiume, in her report *Racial discrimination and emerging digital technologies: a human rights analysis*¹¹¹, demonstrates how a human rights analysis of emerging digital technologies can and must centre racial discrimination. Further, she notes that ethical assessments and tweaking of algorithms to comply with fairness criteria will not be enough in certain cases, and that “in some cases the discriminatory effect of digital technologies will require their outright prohibition”.

In response to Achiume's report, a number of digital rights NGOs, including Access Now, released a statement on *Interventions to Mitigate the Racially Discriminatory Impacts of Emerging Tech*¹¹². Among other points, the NGOs noted that technologies that are demonstrably likely to cause racially discriminatory harm (such as facial recognition and predictive policing) must be banned outright.

Indeed, Achiume's report and the NGO statement highlight two important recent developments in the discussion around AI governance: first, the need to centre concerns about racial discrimination; second, the increasing realisation that certain applications of AI must be banned to protect human rights.

¹¹⁰ See *Report of the Special rapporteur on extreme poverty and human rights to the 74th session to the UN General Assembly*, available at <https://undocs.org/A/74/493>.

¹¹¹ See *Emerging digital technologies entrench racial inequality, UN expert warns*, available at <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26101>.

¹¹² See *Joint Civil Society Statement: Interventions to mitigate the racially discriminatory impacts of emerging tech including AI*, available at <https://www.amnesty.org/en/latest/research/2020/07/joint-statement-on-interventions-to-mitigate-the-racially-discriminatory-impacts-of-emerging-tech-including-ai/>.

Regarding the first point, Achiume and others have noted how discussions of systemic and structural racism have largely been marginal in both AI ethics and in the human rights discussion of AI, which have instead tended to focus on more superficial issues such as bias in datasets. She further noted that:

*The deaths of George Floyd and countless others have prompted a transnational uprising against systemic racism in law enforcement [...] Part of the human rights response must include greater scrutiny of how the design and use of digital technologies is further entrenching this systemic racism.*¹¹³

Achiume echoes much recent work on AI governance from a human rights perspective when she proclaims that ensuring racial justice and the protection of human rights will require prohibiting certain applications of AI. The European Digital Rights (EDRI) network, of which Access Now is a member, have called for a ban on facial recognition technologies that enable mass surveillance¹¹⁴. As they point out, such systems violate human rights in such an egregious manner that they must be banned outright.

Looking at the EU's *White Paper on Artificial Intelligence: a European approach to excellence and trust* ("the White Paper") from this perspective, we see that the EU approach has the potential to lack teeth. As we have highlighted, although the idea of a moratorium appeared in a leaked early draft of the White Paper¹¹⁵, it disappeared from the final version, which proposes far weaker regulatory options. If applications of AI like facial recognition and predictive policing pose an extreme threat to people's rights, the EU approach to AI governance ought to consider not only the option of a moratorium (a pause until certain conditions are met), but a ban of certain use cases.

Beyond the White Paper and future AI regulation, there are other developments in the EU which could have an impact on how AI is regulated. In advance of a meeting in Brussels on 22 September, 2020, where the European Parliament and EU Member States expected to make progress on an agreement on whether to strengthen lax surveillance export rules, Amnesty International published a report outlining how "European tech companies risk fuelling widespread human rights abuses by selling digital surveillance technology to China's public

¹¹³ See *Emerging digital technologies entrench racial inequality, UN expert warns*, available at <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26101>.

¹¹⁴ See *Ban biometric mass surveillance!*, available at <https://edri.org/blog-ban-biometric-mass-surveillance/>.

¹¹⁵ See *LEAK: Commission considers facial recognition ban in AI 'white paper'*, available at <https://www.euractiv.com/section/digital/news/leak-commission-considers-facial-recognition-ban-in-ai-white-paper/>.

security agencies”¹¹⁶. The final text is still under discussion, but on 15 October, 2020, *Politico* reported that the “European Union is finalizing a plan to toughen export controls on technologies used for espionage and surveillance outside the bloc”¹¹⁷. While the institutional negotiations are still ongoing, the text will almost certainly fall short of providing adequate human rights protections^{118, 119}.

As Access Now and others have pointed out, the EU can choose to pursue some applications of AI technology and not others. Indeed, it is misleading to assume that all applications of AI can be made compatible with European values, when some applications inherently threaten human rights. The European Union must have the maturity to dismiss simplistic narratives about “keeping up in an AI race” with China and the US, and carve out a EU approach that puts European values and human rights first, including by banning applications that conflict with those values and rights.

IV. NATIONAL STRATEGIES IN THE EU

High-level update on the status and trends in national strategies in the EU

One of the key priorities of the *Coordinated Plan on Artificial Intelligence*¹²⁰, published in December 2018, was to encourage Member States to develop their national AI strategies by the end of 2019, outlining investment levels and implementation measures.

In parallel to this ongoing effort, the European Commission published its *White Paper on Artificial Intelligence: a European approach to excellence and trust* (“the White Paper”) in

¹¹⁶ See *EU companies selling surveillance tools to China's human rights abusers*, available at <https://www.amnesty.org/en/latest/news/2020/09/eu-surveillance-sales-china-human-rights-abusers/>.

¹¹⁷ See *Europe to crack down on surveillance software exports*, available at <https://www.politico.eu/article/europe-to-curtailed-spyware-exports-to-authoritarian-countries/>.

¹¹⁸ See *Human rights organisations call to strengthen the European Commission position on dual use recast*, available at <https://www.accessnow.org/cms/assets/uploads/2020/06/Joint-Letter-to-Commission-on-dual-use-recast-June-2020.pdf>.

¹¹⁹ The trilogue negotiation was closed in November 2020. “The German presidency of the Council and European Parliament representatives today reached a provisional political agreement on a revised regulation setting out the EU regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items.”, See *New rules on trade of dual-use items agreed*, available at <https://www.consilium.europa.eu/en/press/press-releases/2020/11/09/new-rules-on-trade-of-dual-use-items-a-greed/>.

¹²⁰ See *Coordinated Plan on Artificial Intelligence*, available at https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en.

February 2020. Since Access Now and the Vodafone Institute published the report on *Mapping Regulatory Proposals For Artificial Intelligence In Europe* in November 2019¹²¹, a number of other mapping reports¹²² have been drafted and show that most Member State AI strategies centre on how to support AI research and development, skills development, and infrastructure.

Most strategies do make explicit reference to some sort of ethical framework, whether that is merely a commitment to “human-centred” approaches or an explicit mention of human rights. Using the language of human rights may be a first step, but what truly matters is whether states make a real commitment to ensuring that human rights standards are followed.

In their report, *National Artificial Intelligence Strategies and Human Rights: A Review*¹²³, Global Partners Digital and the Stanford Cyber Policy Center took a close look at a number of national strategies and revealed several issues:

- A lack of clear goals and indicators of success or lack of specific policy commitments.
- An exclusive focus on AI in the future, and a lack of any landscaping or assessment of the current status of AI and its existing impacts.
- A focus on government exclusively, and a failure to set out how other stakeholders would be involved in the implementation of the strategy.
- A failure to take into account the need for international coordination and engagement.

The lack of a multistakeholder approach and predominant focus on AI uptake and investment could limit the relevance and acceptance of national AI strategies, putting their success in jeopardy. One possible reason these strategies lack clear and proper human rights commitments is that Member States may be anticipating and waiting for the Commission to tackle the subject. However it is not clear the Commission will take this kind of action, as

¹²¹ See *Mapping artificial intelligence strategies in Europe: a new report* by Access Now, available at <https://www.accessnow.org/mapping-artificial-intelligence-strategies-in-europe/>.

¹²² See, for example, the report by AI Watch, *National strategies on Artificial Intelligence A European perspective in 2019*, available at https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119974/national_strategies_on_artificial_intelligence_final_1.pdf.

¹²³ See *National Artificial Intelligence Strategies and Human Rights: A Review*, available at https://www.gp-digital.org/wp-content/uploads/2020/04/National-Artificial-Intelligence-Strategies-and-Human-Rights%E2%80%94A-Review_April2020.pdf

critics point out that its own White Paper fails to map out the pathway to ensure adequate protection of rights against the risks of AI systems.

Analysing the interaction between national strategies and EU level positioning

As we explain in the introduction to this report, Access Now and the Vodafone Institute organised four roundtable discussions and conducted a number of individual stakeholder interviews on the topic of artificial intelligence and human rights, gathering key stakeholders from a number of EU countries and regions to discuss how to ensure that the design, development, and deployment of AI-assisted technologies in Europe are human centric and respect human rights.

These multi-stakeholder roundtables included government representatives, representatives from the private sector, civil society organisations, and academics (see a full list of participants in the Annex).

Our main objective was to discuss the role of Member States and other national stakeholders on the one hand, and the role of EU institutions on the other. We held the first roundtables in Berlin and Helsinki in November 2019, and the second set online due to the COVID-19 pandemic during the summer and fall of 2020. These roundtables gathered stakeholders from Central European countries (Czech Republic, Poland, and Hungary), Spain, and France.

The roundtables were an opportunity to gather feedback on the scope and regulatory approaches to AI in the EU , including, in the second part of the series, feedback on the EU Commission's White Paper on AI. The roundtables were held under Chatham House rule.

Although participants discussed fear of over-regulation, the majority of the regional stakeholders in the roundtable embraced and advocated for some EU intervention on AI. What differed in their opinions was the appropriate scope of such an intervention and its potential feasibility and success.

Whether it was through a comment on the White Paper, or to address a general issue that would need to be resolved, participants raised several points they deemed important to ensure a useful regulation and way forward.

On a European regulation

While most participants welcomed some kind of regulation at the EU level, one of the recurring issues was that the process proposed by the Commission did not seem to be evidence-based. Similarly, critics noted the lack of any preparatory legislative mapping and assessment of existing EU legislation.

Some stakeholders argued that existing EU and national legislation already apply to automated decision-making systems or AI, and some saw existing horizontal regulations, such as the GDPR, as sufficient. Even for those who disagreed, they believed that such assessments are still useful to identify and fill existing legislative gaps.

On this note, participants discussed the gaps in the GDPR, especially on automated decision-making, as well as the danger of re-opening the GDPR to address these gaps instead of simply writing complementary regulation.

On the same issue, regional stakeholders pointed out a lack of concrete local EU AI application cases, as evidence on which to base a potential legislation but also as a monitoring tool. Due to the complexity of the impact of AI, research and evidence would be needed as a first step to any regulation. There was a wish to go deeper than the overly general level of the current ethics/human rights debate. Participants also noted that the EU was in an adequate position to develop its own positive AI cases.

On the specifics of an EU regulation, government representatives agreed that European values and fundamental rights should be at the core of efforts, as it also serves the instrumental goal of creating trust in the EU AI market. In this vein, some participants insisted on the importance of the balance between human rights considerations and innovation, and voiced their concerns about overregulation, especially regarding startups and small and mid-size enterprises (SMEs).

While participants generally preferred a horizontal approach to tackle human rights issues, there were strong voices advocating for a sectoral, vertical approach. Some stakeholders were also worried that a horizontal regulation could be hard to contextualise. On the question of liability, there was strong skepticism that the EU could achieve a joint position, regarding the fragmentation of national laws.

Comments on the White Paper

Reception of the **risk-based approach** proposed in the White Paper was mixed. Participants had quite differing views and joint conclusions could not be reached.

Some participants voiced their opposition, citing the gaps a binary high risk vs. low risk approach could leave, especially from a human rights perspective. In particular, participants raised the issue of a potentially wide interpretation of what is considered low risk, what the human rights enforcement would be for low risk applications, and how to address risk for applications in the grey area between the binary categorisation.

A number of participants noted that automated decision-making should always be considered high risk. On the other hand, there was also strong support for maintaining the categories of low and high risk applications but with no *ex-ante* regulatory obligations, enough support for self-assessment, business incentive, and investment and innovation orientation.

On **voluntary labelling**, some participants showed strong support and others believed voluntary labelling and self assessment have the potential to work depending on how civil liability is handled. Several participants noted the missing incentive to participate in such schemes, while others were reluctant to support it, citing “ethics washing” and need for external audits.

If the schemes were to be used, participants highlighted the need for transparency and for the information to be useful, in order for consumers to make informed, fact-based choices. In this regard, the complexity of the impact of AI and lack of evidence was mentioned.

Some participants mentioned that the White Paper was missing accessible assessments on the trustworthiness of AI, especially for SMEs, while others pointed out the lack of balance between transparency, openness, and confidentiality of the systems.

Other issues

Concerns around the **definition of AI** were raised in every roundtable discussion, both as a general issue and specifically regarding the definition adopted in the EU Commission's White Paper.

Participants discussed the various angles for looking into AI (“ADM”, “algorithms”, “machine-learning”, etc), and the way they each impact or shape the debate, as well as the surrounding hype and use of certain buzzwords and narratives (such as “the race for AI”) as opposed to more down-to-earth terminologies and initiatives.

While a definition of AI was considered essential regarding the scope of a potential regulation, most participants criticised the definition of AI adopted in the White Paper, in particular, as being too narrow. Some participants advocated instead for a dynamic approach, such as UNESCO's definition¹²⁴, based on the functions of AI rather than what it is.

Another common concern among participants of the roundtables was **transparency** as a way to attain trustworthy AI or to allow oversight and monitoring. Transparency would be needed to evaluate the legitimacy and the discrimination potential of ADM systems.

A wide range of stakeholders explained that increased transparency in the training, deployment, and procurement process of AI would be beneficial for them. This concerns not only the people affected by AI systems, but also the creators, sellers, and distributors of AI products, as well as oversight mechanisms.

A first step noted by participants is the opening of public sector algorithms. Most stakeholders supported the idea of establishing public **AI registers**, as Access Now advocated in our White Paper submission¹²⁵. However, some participants questioned the usefulness of transparency and whether it would actually lead to trust. Moreover, some criticised the unrealistic narrative about capacity and transparency, asserting that not everyone needs to understand AI systems but at least some people should have the means to understand the algorithms.

The different layers of transparency were acknowledged (oversight body, users, governments) while it was suggested that the solution could be at the EU level with mandatory transparency obligations, on top of national measures.

¹²⁴ See *Recommendation on the Ethics of Artificial Intelligence*, available at <https://unesdoc.unesco.org/ark:/48223/pf0000373434>.

¹²⁵ See Access Now's submission to the Consultation on the “White Paper on Artificial Intelligence - a European approach to excellence and trust”, available at https://www.accessnow.org/cms/assets/uploads/2020/06/EU-white-paper-consultation_Access_Now_June2020.pdf.

Some stakeholders noticed that the COVID-19 crisis has had a positive impact in transparency efforts, as citizens and civil society organisations became more aware and concerned about their data, putting pressure on the government to be more transparent, such as pushing them to release the code of COVID-19 tracking apps. On the other hand, other stakeholders pointed out that the crisis has also exacerbated the citizens' lack of trust in new technologies and in public administration.

In response to this concern and as a general issue, participants discussed the lack of **education**, competencies, and **digital literacy** as an obstacle to the development of AI. They regret the lack of experience and knowledge about working with data, as there is no information in the media, nor a systematic education of society on data, AI, and technologies in general. Most stakeholders agreed that digital education is key to enabling citizens to make informed choices and ultimately to build or restore trust in AI systems and in public administration. On a more specialised level, some stakeholders advocated for ethics and legal aspects modules to be taught in data science courses.

Stakeholders also mentioned the current **digitalisation** process throughout Europe, and how there needs to be more work on data, such as data usage and data governance, as well as on enforcement of the GDPR, before we can tackle AI itself.

National AI strategies

A common worry among stakeholders is the impact of the COVID-19 pandemic and economic crisis and the resulting shifting of allocation of time, research, and resources for the development and implementation of the national AI strategies.

Meaningful consultation of different stakeholders was a priority across stakeholders but the level of satisfaction with existing processes varied among participants.

The preference, among stakeholders from Central European countries, for a **potential oversight body on the national level**, highlighted the need for both political coordination and a broad understanding of legal, societal, and technical impacts of AI. Stakeholders from Spain, on the other hand, tended to support an oversight body on the EU level.

A common component of the national AI strategies is the focus on **research and development**, and the cooperation between Member States in the EU. At the same time,

most strategies also include the ambition for a Member State to become a center of innovation, with the creation of AI research centres and the wish to attract and retain talent.

When discussing national strategies concretely, some participants welcomed the ambition of the Czech strategy¹²⁶, but others were disappointed in the lack of concrete implementation since its publication. Similarly, there were expectations for national strategies that have yet to be published, such the Spanish AI strategy.

In contrast, at the time of the roundtable events, the implementation process of the German strategy was focusing on digitisation of the administration as a groundwork and introducing the widespread use of AI in public administration was not yet on the table, while the policy debate was centered on the German data strategy and the then-recent paper of the Data Ethics Commission. At the time of the roundtables, Finland was reviewing its law on the public sector's use of automation and looking into developing a mandatory due diligence legislation.

In Spain, the succession of governments and the COVID-19 crisis threw a spanner in the works of the creation of a Spanish AI policy and so far, only a Research, Development, and Innovation (RDI) strategy has been published. In this context, and mentioning the existing data infrastructure and the lack of big AI companies, some stakeholders deplored that Spain was not at the forefront of AI.

At the regional level, the autonomous regions of Valencia¹²⁷ and Catalonia¹²⁸ published their own AI strategies in 2019.

Takeaways from the final stakeholder roundtable

On 27 October, 2020, Access Now and the Vodafone Institute held the closing workshop of this roundtable series. We invited stakeholders from the earlier Member State events to present their perspectives on AI governance in the EU, and invited representatives from the European Parliament and European Commission, as well as representatives of

¹²⁶ See *National Artificial Intelligence Strategy of the Czech Republic*, available at https://www.mpo.cz/assets/en/guidepost/for-the-media/press-releases/2019/5/NAIS_eng_web.pdf.

¹²⁷ See *Estrategia de Inteligencia Artificial de la Comunitat Valenciana*, available at http://www.presidencia.gva.es/documents/80279719/169117420/Dossier_en.pdf/c943f4aa-2822-4c5e-a3db-63a45cca5bf5.

¹²⁸ See *Catalonia's Artificial Intelligence Strategy*, available at https://participa.gencat.cat/uploads/decidim/attachment/file/932/Document-Bases-Estrategia-IA-Catalunya-E_Nversion.pdf.

Brussels-based civil society organisations, both to listen and to give an update on the debate at EU level. Many of the topics raised during the original roundtable discussions surfaced again, but the discussion also led to some novel ideas, which we will briefly summarise here.

At numerous points, the topic of conversation turned to the **difficulty of coordination between the multiple levels of governance**, both within Member States (such as between regions, or on a city level) and between Member States and EU-level institutions. Some stakeholders pointed to the reality and importance of leadership in regions, as for example in Spain, where despite the continued lack of a national AI strategy, the Valencian and Catalan regions have taken the lead and produced their own strategies. Attention was also drawn to grassroots initiatives such as the European Laboratory for Learning and Intelligent Systems (ELLIS) which are being set up to push for research excellence in AI in Europe.

¹²⁹

The important **role of city authorities** was also raised, with the examples of Helsinki and Amsterdam, both of whose administrations have recently launched “AI registers” to document the use of AI and automated decision making systems in their cities.¹³⁰ The need for transparency was underlined multiple times, and it was noted that in the past two years the use of AI and ADM systems has increased dramatically in Europe, with little to no oversight.¹³¹ In addition to measures such as registers, it was noted that meaningful transparency may require legally binding data access frameworks for public watchdogs, researchers, and those affected by systems to be able to have effective oversight and auditing capabilities.

Beyond just transparency, the need was discussed for the public sector to do more and better outreach to citizens. While public authorities are currently thinking about using AI and ADM to offer better public services, more thought needs to be given to how they will communicate about the use of ADMs or how to involve citizens in decision-making processes.

A number of stakeholders noted that **opportunities for introducing effective (and perhaps not so effective) governance mechanisms for AI** can be found in places other than the Commission's proposed AI regulation. In addition to mentioning the Data Governance Act,

¹²⁹ See *Ellis Society - European Laboratory for Learning and Intelligence Systems*, available at <https://ellis.eu/>.

¹³⁰ See *Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI*, available at <https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/>.

¹³¹ See *AlgorithmWatch Automating Society Report 2020*, available at <https://automatingsociety.algorithmwatch.org/>

which promises to have a significant impact on data governance more broadly but also AI development and use within the EU, there was a discussion of how various national regulatory initiatives related to cybersecurity and medical devices could impact AI governance. Moreover, the certification procedures, technical norms, and best practices and licencing schemes for “operators” of AI systems were discussed.

The **definition of AI** was again raised as a problem. Some stakeholders noted the advantages of using automated decision making as a term, as it focused on how the technology impacts governance structures rather than focusing on any specific technique such as machine learning.

From the update on the discussion in Brussels, during which we heard the latest developments from both the European Parliament and Commission, a number of key topics emerged. The topic of the **risk-based approach** put forward by the Commission's White Paper was again raised. While it was reaffirmed that the risk-based approach remains central to the Commission's thinking on the topic, there was much discussion of how to define high risk AI, and alternative models of assessing risk were put forward by different stakeholders.

The problem of **protecting fundamental rights while not stifling innovation** was discussed, with some stakeholders critiquing the dominant narrative around innovation. In particular, it was noted that we should pay attention to who stands to benefit from innovation, and who is likely to be put at risk by the increased use of AI systems. There was a discussion of how AI and ADM can reinforce structural racism and inequalities, and how under the current model, there is a disproportionate burden on individuals and civil society organisations to demonstrate violations of their rights.

Criticism was given of approaches that prioritise technical fixes to AI-related problems, such as “debiasing datasets”, an approach that is unlikely to properly address the issue of structural racism. Instead, the need was underlined for robust governance approaches to address problems in their full complexity and allow marginalised groups to have meaningful input into systems that impact them.

Finally, there was a discussion about possible red lines, and outright bans, on certain applications of AI. It was indicated that the question of red lines is under serious discussion on multiple sides, and that the possibility of a ban remains on the table, especially regarding so-called remote biometric identification systems, and the use of AI in sensitive domains such as criminal justice.

Facial recognition as a case study in EU-national level interaction

Among the wide range of AI applications, facial recognition is the most hotly debated in the EU. Pilot projects and the testing of systems at national and local level are widespread and have taken place outside the public debate, without guarantees of legality, transparency, safeguards, or accountability.

Under the GDPR and the Law Enforcement Directive (LED), the processing of biometric data must meet strict criteria. Other than for law enforcement purposes, the use of biometric data for identification purposes is generally prohibited unless explicit consent is given.

Access Now reported on two cases in Europe where facial recognition was trialed in schools for monitoring purposes, in France and in Sweden in 2019¹³². In both cases, the national data protection authorities intervened to put a stop to the experiments, deeming them unlawful under the GDPR, even though, in France, a data protection impact assessment (DPIA) was conducted beforehand.

It seems therefore that there are use cases where the scale of the rights violations to privacy and data protection are so significant that even safeguards and DPIAs cannot make the use compliant with the EU's fundamental rights laws.

In addition, enforcement of data protection laws relies on under-funded public authorities with limited powers¹³³. The Swedish DPA only heard about the facial recognition testing through the media, and the French DPA's opinion is not binding.

The current regulatory and enforcement framework therefore seems to fall short in preventing Member States from deploying unlawful biometric mass surveillance systems, especially in grey areas or when involving private actors.

This last point can be illustrated by the case of PimEyes, a Polish search engine with a database of over 900 million faces, including content from social media and porn sites.

¹³² See *In the EU, facial recognition in schools gets an F in data protection*, available at <https://www.accessnow.org/in-the-eu-facial-recognition-in-schools-gets-an-f-in-data-protection/>.

¹³³ See *Two years under the EU GDPR*, available at <https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf>.

Similar to the scandalous US start-up Clearview AI,¹³⁴ whose customers included companies, governments, and police authorities, PimEyes differs in that it offers its services to everyone. Netzpolitik.org, a German journalist platform, investigated the firm's potential for abuse and the threat the business poses to our fundamental rights¹³⁵.

Without that investigative work, we might never have heard of PimEyes, just as the Swedish DPA might never have heard of the school experiment. It is evident that despite strong data protection laws, private entities in the EU are still creating and commercialising dangerous products.

Accordingly, Access Now joined the call by European Digital Rights (EDRi) network for a ban on biometric mass surveillance across the European Union¹³⁶. EDRi calls on EU Member States, as well as the European Commission, as the guardian of the EU's fundamental rights treaties and its competency with regard to European borders, to permanently stop all biometric processing in public and publicly accessible spaces, wherever it has the effect or potential effect to establish mass surveillance.

This particular application of AI demonstrates the need to review whether existing laws and enforcement are sufficient in light of the significant threat some AI applications pose to our fundamental rights.

Due to the wide range of applications of AI that we may not be aware of, and the freedom Member States and local authorities have in funding, developing, and deploying such systems, it is imperative to implement an effective rights-based approach.

In contrast to what is happening in Europe and despite the lack of strong federal data protection legislation in the US, local legislators have been proactive in imposing bans on use of technology they deem dangerous, such as facial recognition. In total, 13 cities¹³⁷, including three cities in California (San Francisco, Oakland, and Berkeley), six in Massachusetts (including Boston), and the city of Portland, Maine, have placed tight restrictions on the use of the controversial technology.

¹³⁴ See *The Secretive Company That Might End Privacy as We Know It*, available at <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>.

¹³⁵ See *PimEyes - A Polish company is abolishing our anonymity*, available at <https://netzpolitik.org/2020/pimeyes-face-search-company-is-abolishing-our-anonymity/>.

¹³⁶ See *Ban Biometric Mass Surveillance, A set of fundamental rights demands for the European Commission and EU Member States*, available at <https://edri.org/wp-content/uploads/2020/05/Paper-Ban-Biometric-Mass-Surveillance.pdf>.

¹³⁷ See https://twitter.com/Matt_Cagle/status/1290485013331337217?s=20.

The US states of Oregon and New Hampshire have also enacted bans on use of facial recognition technologies in police body cameras, while California has had a three-year moratorium in place since January 2020¹³⁸. Apart from local policies, major music festivals in the US also pledged not to use the technology, responding to human rights advocacy and public pressure¹³⁹. In June 2020, IBM, Amazon, and Microsoft voluntarily introduced moratoriums on the use of their facial recognition products by law enforcement¹⁴⁰. But these announcements should be taken with a pinch of salt, as these companies might still be lobbying behind closed doors to stop legislative bans. Amazon¹⁴¹ spent a large amount of money to try to quash the recent bill passed in Portland, Oregon that bans the use of facial recognition technology for both the public and private sector, perhaps the strongest ban so far¹⁴².

While facial recognition applications are fought and debated on the policy level or in local politics and campaigns, in some cases and in other countries, legal challenges are the driving force. In Brazil, in the case *IDEC vs. Via Quatro*, the installation and use of an AI crowd analytics system in the metro system of Sao Paulo, purporting to predict the emotion, age, and gender of metro passengers without processing personal data, is being challenged in court, with Access Now submitting an expert opinion in the case.¹⁴³ Cases like this one could have a global impact and set a precedent in the fight against pseudoscience and use of invasive technologies that violate human rights.

Returning to the EU, there are currently two major campaigns to ban facial recognition. The first, *Ban Facial Recognition Europe*, calls for “**the permanent ban of Facial Recognition**

¹³⁸ See *California Governor Signs Landmark Bill Halting Facial Recognition on Police Body Cams*, available at <https://www.aclunc.org/news/california-governor-signs-landmark-bill-halting-facial-recognition-police-body-cams>.

¹³⁹ See *Opinion: How Artists And Fans Stopped Facial Recognition From Invading Music Festivals*, available at <https://www.buzzfeednews.com/article/evangreer/stop-facial-recognition-music-festivals-concerts>.

¹⁴⁰ See *IBM, Microsoft And Amazon Not Letting Police Use Their Facial Recognition Technology*, available at <https://www.forbes.com/sites/larrymagid/2020/06/12/ibm-microsoft-and-amazon-not-letting-police-use-their-facial-recognition-technology/>.

¹⁴¹ See *Amazon Spent \$24,000 To Kill Portland's Facial Recognition Ban*, available at https://www.vice.com/en_us/article/g5p9z3/amazon-spent-dollar24000-to-kill-portlands-facial-recognition-ban.

¹⁴² See *Portland, Oregon, passes toughest ban on facial recognition in US*, available at <https://www.cnet.com/news/portland-passes-the-toughest-ban-on-facial-recognition-in-the-us/>.

¹⁴³ See *Facial recognition on trial: emotion and gender “detection” under scrutiny in a court case in Brazil*, available at <https://www.accessnow.org/facial-recognition-on-trial-emotion-and-gender-detection-under-scrutiny-in-a-court-case-in-brazil/>.

used for identification and profiling in all of Europe”. The second, *Reclaim your Face* , launched by a number of civil society organisations who are members of the European Digital Rights (EDRI) network, seeks to ban biometric mass surveillance. Launched on 12 November, 2020 in Czechia, Serbia, Greece, and Italy, the campaign will soon expand to more Member States.¹⁴⁴

V. CONCLUSION: AI POLICY STRATEGIES — WHAT HAS WORKED AND WHAT HAS NOT

Since the European Commission published its communication, *Artificial Intelligence for Europe*, on 25 April, 2018¹⁴⁵, the debate on AI governance has progressed in some senses and stagnated in others.

On a positive note, more and more governments, both within the EU and around the world, have formulated strategies for AI. As noted above, the EU approach has had a significant impact on many of those strategies, as well as influencing the various guidelines and sets of principles put out by companies, standards bodies, and international institutions. The idea that AI needs to be trustworthy has become commonplace in discussions of AI governance, although debate remains about what that means in practice.

There have also been a number of positive initiatives to move us from abstract, high-level discussions about principles to more concrete, actionable measures. On the technical side, we have seen important developments for documentation of AI systems, such as *Model Cards for Model Reporting*¹⁴⁶ and *Datasheets for Datasets*¹⁴⁷, which have set a high standard for developers of AI systems. Regarding transparency, we have seen two cities in the EU experiment with public registers to be transparent about their use of AI / automated decision making¹⁴⁸.

¹⁴⁴ See Reclaim Your Face, available at <https://reclaimyourface.eu/>.

¹⁴⁵ See European Commission's Communication Artificial Intelligence for Europe, available at <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>.

¹⁴⁶ See *Model Cards for Model Reporting*, available at <https://arxiv.org/abs/1810.03993>.

¹⁴⁷ See *Datasheets for Datasets*, available at <https://arxiv.org/abs/1803.09010>.

¹⁴⁸ See *Amsterdam and Helsinki become first cities to launch AI registers explaining how they use algorithms*, available at <https://thenextweb.com/neural/2020/09/28/amsterdam-and-helsinki-become-first-cities-to-launch-ai-registers-explaining-how-they-use-algorithms/>.

On the topic of Trustworthy AI in particular, we have seen the development of technical mechanisms for its implementation, as outlined in the paper *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims*¹⁴⁹, and the HLEG's publication of the ALTAI tool (The Assessment List on Trustworthy Artificial Intelligence)¹⁵⁰.

There has also been growing recognition that voluntary ethical principles won't be enough to protect people from the impact of AI systems. As outlined in Section III of this report, this has led to increased advocacy for adoption of the international human rights framework in AI governance, and to increased calls to prohibit or ban certain applications of AI, such as remote biometric identification, which are deemed to be incompatible with the exercise and protection of fundamental rights¹⁵¹.

Among EU Member State stakeholders, there is great anticipation of the long-awaited AI regulation¹⁵², although there is significant divergence regarding what stakeholders want to see in it. While a number of Member States have explicitly called for a "light-touch" regulation to avoid hampering innovation¹⁵³, other stakeholders, especially from civil society, are looking to see the Commission take the lead in imposing legal obligations across the board, and to ban certain applications of AI.

While acknowledgment of the risks posed by AI systems has become mainstream, the political will to take the necessary measures to prevent them is still lacking. Across the world people, communities, and civil society organisations are signing petitions, protesting, and taking legal action to protect themselves from the proliferation of AI-driven surveillance tools and other harmful applications of AI and automated decision making¹⁵⁴.

¹⁴⁹ See *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims*, available at <https://arxiv.org/abs/2004.07213>.

¹⁵⁰ See *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, available at <https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>.

¹⁵¹ See *Ban biometric mass surveillance!*, available at <https://test.edri.org/our-work/blog-ban-biometric-mass-surveillance/>.

¹⁵² See the timeline for the legislative agenda on Artificial Intelligence - ethical and legal requirements, available at <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements>.

¹⁵³ See *Innovative and Trustworthy AI: Two sides of the Same Coin*, available at <https://em.dk/media/13914/non-paper-innovative-and-trustworthy-ai-two-side-of-the-same-coin.pdf>.

¹⁵⁴ See *Black Lives Matter could change facial recognition forever - if Big Tech doesn't stand in the way*, available at <https://www.washingtonpost.com/technology/2020/06/12/facial-recognition-ban/>.

We have seen how even simple algorithms, such as that used in the UK's A-Level grading fiasco, can amplify unfair and discriminatory outcomes and mobilize people to demand justice amid chants of “fuck the algorithm”¹⁵⁵. If people do not see that measures are being taken to protect them from AI and ADM systems, then the idea of “Trustworthy AI” will be doomed from the start.

If we want AI to deliver on its potential benefits to society, EU policy and strategy choices must show that the government is putting people and their rights ahead of AI innovation at any cost.

For more information:

Project Lead:

Fanny Hidvégi (fanny@accessnow.org)

Authors:

Daniel Leufer (daniel.leufer@accessnow.org)

Laureline Lemoine



Access Now (<https://www.accessnow.org>) defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, legal interventions, and convenings such as RightsCon, we fight for human rights in the digital age.

With the support of the Vodafone Institute



**Vodafone Institute
for Society and
Communications**

¹⁵⁵ See *Why 'Ditch the algorithm' is the future of political protest*, available at <https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics>.

VI. ANNEX - AGENDAS & PARTICIPANT LISTS

The original intent of this project was to conduct physical roundtable discussions in Germany, Finland, Czech Republic, Spain, and Brussels. Due to the COVID-19 pandemic, only those in Germany and Finland were able to take place physically. The events in the Czech Republic and Brussels were replaced by online events, and the Spain event was replaced by individual stakeholder interviews. Below is a list of all individuals consulted during these events and calls:

Berlin roundtable:

Daniela Kolbe, Commission on Artificial Intelligence

Peter Parycek, Fraunhofer FOKUS Institute, ÖFIT

Miika Blinn, Verbraucherzentrale Bundesverband (VZBV)

Carla Hustedt, Bertelsmann Foundation

Sybille Gabler, German Institute for Standardisation (DIN), Head of Government Relations

Clemens Otte, Germany Industry Federation, BDI

Nabil Alsabah, BITKOM

Daniel Krupka, Gesellschaft für Informatik

Marianna Rusche, Enquete Commission on AI

Oskar Schumacher, assistant of prof. Thomas Wischmeyer, of the Data Ethics Commission

Caitlin Corrigan, Institute for Ethics in Artificial Intelligence, IEAI - TU Munich

Matthias Spielkamp, AlgorithmWatch

Lisa Gutermuth, Ranking Digital Rights

Orsolya Reich, Civil Liberties Union for Europe

Helsinki roundtable:

Prof. Minna Ruckenstein, University of Helsinki

Janne Järvinen, Vice President, Data-driven Solutions, VTT

Linda Piirto, Senior Advisor in CSR and business and Human Rights at the Finnish Ministry of Employment and Economy

Timo Hankala, Human Rights Adviser, Finnvera

Sirpa Rautio, Director of the Finnish Human Rights Centre

Päivi Luostarinen, Finnish Ambassador at the Finnish Embassy in London

Jarmo Sareva, Ambassador for Innovation at the Ministry for Foreign Affairs (MFA) of Finland

Rauno Merisaari, Ambassador on Human Rights and Democracy for the Ministry for Foreign Affairs of Finland

Elias Aarnio, Deputy Chairman of Electronic Frontier Finland

Antti 'Jogi' Poikola, MyData International Lead

Teemu Ropponen, General Manager, MyData Global.

“Prague” roundtable (virtual)

Robert Kroplewski (PL), Minister of Digital Affairs' expert and representative for information society
Jan Mica (CZ), Head of Unit, European Digital Agenda Unit, Section for European Affairs
Renata Paleń (PL), Minister's Counsel at the Ministry of Digital Affairs
Krzysztof Izdebski (PL), Policy Director at ePaństwo Foundation
Josef Šmída (CZ), Open Society Fund and Code for Czechia
Sandor Lederer (HUN), co-founder and director of K-Monitor

Eva Fialová (CZ), Attorney and researcher
Alžběta Krausová (CZ), Expert in AI Law, Member of European Commission and OECD AI Expert Groups, Legal Scholar at Czech Academy of Science.
Marek Havrda (CZ), Director of AI policy and Social Impact at GoodAI
Sara Boutall (CZ), Co-Founder of Innovation Disrupt House and Communication
Jan Klesla (CZ), National Coordinator for European AI Centres, Ministry of Industry and Trade of the Czech Republic

Spain 1-1 stakeholder consultations:

Karma Peiro, Data Journalist & Co-director of the Visualization and Transparency Foundation, Spain
David Cabo, founder of CIVIO
Ana Berenguer, Director General for the President of the Valencian Region
Carlos Castillo, Professor of Computer Science, Universitat Pompeu Fabra, Barcelona
Lorena Jaume-Palasi, founder and CEO of The Ethical Tech Society
Amparo Alonso Betanzos, computer scientist and president of the Spanish Association for

Artificial Intelligence, professor at University of A Coruña
Simona Levi, founder, X-NET
Idoia Salazar & V. Richard Benjamins, Observatorio del impacto social y ético de la inteligencia artificial (ODISEIA)
Nuria Oliver, Commissioner for the President of the Valencian Region on AI Strategy and Data Science to fight COVID-19, Spain

Brussels roundtable:

Speakers

Nuria Oliver, Commissioner for the President of the Valencian Region on AI Strategy and Data Science to fight COVID-19, Spain
Krzysztof Izdebski, Policy Director of ePaństwo Foundation, Poland
Meeri Haataja, CEO & co-founder of Saidot, Finland

Friederike Reinhold, senior policy advisor for AlgorithmWatch, Germany
Veronika Žolnerčíková, CyberSecurity & CyberCrime Center of Excellence at Masaryk University (C4E), Co-creator of Czech National strategy on AI, Czech Republic

Karma Peiro, Data Journalist & Co-director of the Visualization and Transparency Foundation, Spain
Sarah Chander, Senior Policy Advisor at European Digital Rights (EDRI), Belgium
Hanna Zinner, Artificial Intelligence and Digital Industry, DG CNECT, European

Commission Marcel Kolaja, European Parliament Vice President and member of the Czech Pirate Party
Andreas Hartl, Head of Division, Strategy Artificial Intelligence, Data Economy, Blockchain, Federal Ministry for Economic Affairs and Energy, Germany

Audience

Jim Dratwa, Team Leader, European Group on Ethics, European Commission
Killian McDonagh Dit, Directorate-General for Justice and Consumers, European Commission
Anna Moscibroda, Directorate General for Justice and Consumers, European Commission
Zoi Kardasiadou, Directorate General for Justice and Consumers, European Commission
Aimilia Givropoulou, assistant to MEP Patrick Breyer
Anne van Heijst, assistant to MEP Liesje van Schreinemacher

Despoina Riga, assistant to MEP Anna-Michelle Assimakopoulou
Natalia Joanna Boniecka, assistant to MEP Andrzej Halicki
Georgios Theodotou, assistant to MEP Elena Kountoura
Matt Mahmoudi, Researcher/Advisor on Artificial Intelligence & Human Rights at Amnesty International
Ella Jakubowska, Policy & Campaigns on biometrics at European Digital Rights

Sample Agenda



**Roundtable on Artificial Intelligence and Human Rights
hosted by Access Now and the Vodafone Institute**

Agenda

30 mins Welcome by Access Now and the Vodafone Institute

- Paus Inger, Executive Director, Vodafone Institute
- Matt Allison, Senior Public Policy Manager, Vodafone
 - Presenting the Vodafone AI Framework
- Fanny Hidvégi, European Policy Manager, Access Now
 - Introducing the agenda and objective

60 mins Tour de table:

- Name & affiliation
- Guiding questions
 - What does your organisation do regarding automated decision-making systems, machine learning or artificial intelligence more broadly?
 - What are your organisations' main priorities about AI?
 - How is accountability considered in your organisation's work and mission in relation to automated decision-making systems?

20 mins Coffee break

45 mins What should the European Union do next for "Trustworthy AI"?

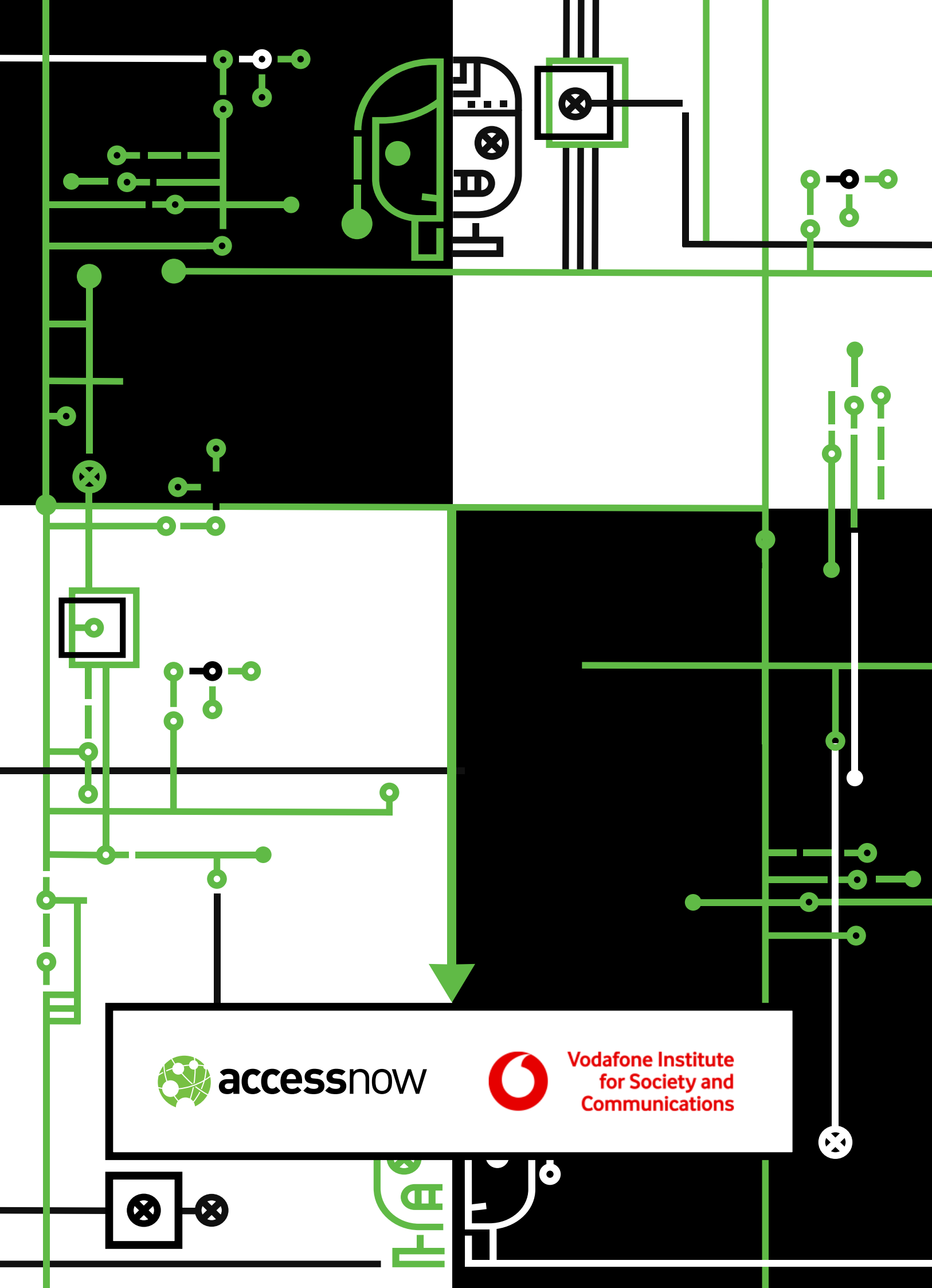
- Access Now presents the state of play for the EU AI Ethics Guidelines, the Policy and Investment Recommendations, the current status of the "100 day AI regulation", the EU White Paper on AI and Access Now's recommendations
- Participants share their feedback on the recommendations
- Participants share their expectations for EU institutions in the development of AI policies

45 mins **Where do participants see a need or space for member state actions?**

- Presentation by selected government representatives on the state of play in the region
- Participants discuss:
 - What are the current applications of AI in your organisation or under your oversight (eg. local pilot projects)?
 - Do you see a gap between existing legislative frameworks and what's necessary for "trustworthy AI"?
 - What is the role of regional and national AI strategies and policy initiatives?

10 mins **Closing round**

- Summarising the takeaways
 - Explaining the next steps about the Brussels roundtable and the project outcomes
-



accessnow



**Vodafone Institute
for Society and
Communications**