THE ATHENS ROUNDTABLE

ARTIFICIAL INTELLIGENCE AND THE RULE OF LAW

2022 4TH EDITION REPORT

prepared by:
The Future Society

The premier international, multi-stakeholder forum on Artificial Intelligence and the Rule of Law, particularly within the prism of legal, judicial, and compliance systems.

Organizer

THE FUTURE SOCIETY

In partnership with

Patrick J McGovern
F O U N D A T I O N

Co-hosts

European Parliament
STOA Panel for the Future
of Science and Technology

unesco
United Nations
Educational, Scientific
and Cultural Organization

OECD

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IEEE

CRAVATH

aws

Supporting Organizations

Jain
Family
Institute

Arnold&Porter

Debevoise
&Plimpton

www.aiathens.org

FOREWORD

Under the patronage of the President of the Hellenic Republic Aikaterini Sakellaropoulou, we were honored to reconvene key stakeholders from the three branches of government, industry representatives, civil society leaders, and academic experts at the fourth edition of The Athens Roundtable on AI and the Rule of Law.

The year of 2022 brought significant changes in the AI landscape. Notably, we witnessed an explosion of the deployment of large generative AI models for a multitude of applications. The first reactions to the swift impact of diffusion models on the creative industry – with legal disputes over copyrights and fears over the future of visual arts – and of language models in education – with some institutions banning the use of advanced AI chatbots – only begin to surface the myriad of governance challenges ahead. The use of AI on the ground for surveillance and in conflicts also marked 2022, with the illegal and unprovoked invasion of Ukraine. In parallel, the AI regulatory landscape has also rapidly evolved. The intensely debated regulatory regime in the EU has already started to influence global value chains. Meanwhile, organisations such as UNESCO, OECD and the US-EU Trade and Technology Council have advanced the enforcement of governance mechanisms and processes that work across borders and sectors, as will be explored in this report.

Amidst this rapidly evolving landscape, the fourth edition of The Athens Roundtable brought together key decision-makers and participants from 112 countries for an hybrid action-oriented dialogue, held over 2 days, shaping the future of AI governance with a focus on implementation and enforcement.

In the interim period since the previous edition, the working groups of The Athens Roundtable have made concrete progress towards their objectives. The Working Group on Judicial and Legal Education on AI ideated the [MOOC on AI and the Rule of Law](#), launched by The Future Society and UNESCO, which enhanced the AI capacities of over 4,500 judicial operators in its first cohort; the Working Group on Interoperable Benchmarking of AI Systems commenced work on a Proposal for Advancing the Responsible Use of AI through Benchmarks; and the Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of Artificial Intelligence published a [Manifesto on Enforcing Law in the Age of AI](#), which put forward 10 concrete recommendations to address the key enforcement challenges shared across transatlantic stakeholders.

In December, we were able to gather in the European Parliament and online, serving the international community as a venue for collective intelligence, coalition building, and participative dialogue.

We are grateful for our partners – the Patrick J. McGovern Foundation, the European Parliament, UNESCO, OECD, and the Council of Europe, IEEE, Cravath, Amazon Web Services, the Jain Family Institute, Arnold & Porter, and Debevoise & Plimpton – and all participants for their continued support to this year's edition of The Athens Roundtable. We hope the upcoming edition continues to address the most urgent issues and catalyze initiatives that consolidate progress in AI governance.

Best regards,

Nicolas Miailhe
Founder & President, The Future Society

CONTENTS

EXECUTIVE SUMMARY

On December 1st and 2nd, 2022, the fourth edition of The Athens Roundtable on Artificial Intelligence and the Rule of Law convened a diverse community of policymakers, legal experts, judicial operators, civil servants, industry leaders, and civil society representatives from 112 countries. Hosted physically at the European Parliament in Brussels, participation included a selective group of over 120 onsite participants in addition to a global audience of over 1,000 practitioners who were able to access the dialogues in English, French, and Spanish.

Participant Statistics:

120
DAY 1
ONSITE

571
DAY 1
ONLINE

59
SPEAKERS

579
DAY 2
ONLINE

Co-founded by The Future Society in 2019, The Athens Roundtable is the premier international, multistakeholder AI governance forum within the prism of legal, judicial, and compliance systems. The fourth edition focused on the implementation and enforcement of laws, regulations, standards, and policies across the industrial value chain of AI systems. In keeping with technological and regulatory trends, topics spanned the forthcoming EU AI Act, generative AI, and coordinating international efforts to measure and evaluate AI systems.

This year's Roundtable was held under the auspices of H.E. the President of the Hellenic Republic Ms. Aikaterini Sakellaropoulou and was organized in partnership with the Patrick J. McGovern Foundation. Proudly, the event was co-hosted by prominent intergovernmental organizations, including the European Parliament, UNESCO, OECD, and the Council of Europe, as well as leading AI institutions and firms, including IEEE, Cravath, and Amazon Web Services. It was generously supported by the Jain Family Institute, Arnold & Porter, and Debevoise & Plimpton.
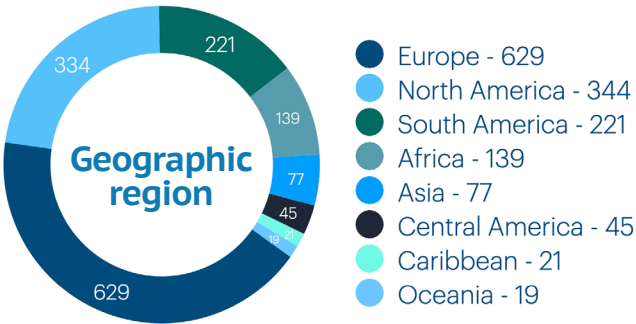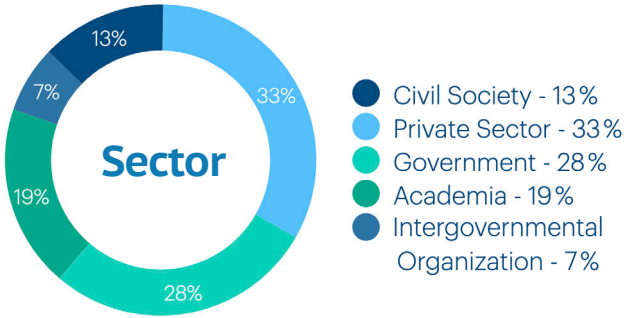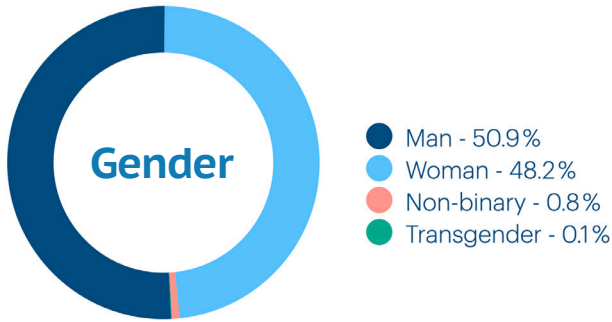
The dialogue convened over 50 speakers across 30 engaging sessions, including keynote speeches, fireside chats, plenary panel, and expert remarks. The public two-day event generated valuable insights into opportunities for further institutional innovation and international coordination to ensure effective AI governance. In addition, the program included a pioneering closed-door discussion, at the European Parliament, with industry leaders and policymakers on the governance of generative AI. The program drilled into how advanced AI systems have opened new frontiers in research

and productivity but also pose novel and large-scale risks. This year's discussions put forward concrete steps to increase safety and uphold the rule of law amid the uncertainties of General Purpose AI Systems' (GPAIS) impact.

Across sessions, the focus was on the implementation and enforcement of AI laws, regulations, standards, and policies throughout the industrial value chains of an AI system. Conversations covered the impact of AI technologies on human rights and democratic values, the substantial headway on the development of legal and normative frameworks

Day 1 Video Recordings

Day 2 Video Recordings

Gender
- Man - 50.9%
- Woman - 48.2%
- Non-binary - 0.8%
- Transgender - 0.1%

Sector
- Civil Society - 13%
- Private Sector - 33%
- Government - 28%
- Academia - 19%
- Intergovernmental Organization - 7%

Geographic region
- Europe - 629
- North America - 344
- South America - 221
- Africa - 139
- Asia - 77
- Central America - 45
- Caribbean - 21
- Oceania - 19

Data based on voluntary information provided by registrants.

to mitigate AI risks, enforcement, and how these frameworks can complement one another. For instance, experts provided insights into the legal, political, and technical red lines within the EU AI Act, the implementation of normative frameworks including the UNESCO Recommendation on the Ethics of AI, and the operationalizations of the US Bill of Rights.

The summary below provides a glimpse into the discussions' takeaways presented throughout the report, divided by this edition's three main themes: **enforcement, institutional innovation, and AI industrial value chains.**

ENFORCEMENT

The fourth edition explored the challenges transpiring in the *implementation* versus the *enforcement* of AI laws, regulations, and policies. Across sessions, it was highlighted that enforcement mechanisms must provide clarity to the market about compliance and shape developers', deployers', investors' and regulators' behaviors towards a swift adoption of trustworthy AI. Some actionable guidance includes investing in building AI capacities of actors in the public sector and leveraging regulatory sandboxes to help anticipate actors' behaviors and regulatory needs.

INSTITUTIONAL INNOVATION

Methodologically robust, evidence-based frameworks and a socio-technical approach to measurement and evaluation are crucial to institutional innovation in the field of AI. The discussions confirmed the urgency in redesigning governance structures, frameworks, processes, and standards for governance due to the increasing capabilities and disruption of large-language models (LLMs) such as ChatGPT.

Normative and policy instruments – terms of service, contracts, codes of conduct, guidelines – were invoked as useful tools to preempt uncertainties of generative AI products, such as with copyright, accountability, and civil liability.

Capacity building, technical competence, legal basis, and market fragmentation were identified as major challenges to developing and upscaling institutional innovation for trustworthy AI uptake.

THE AI INDUSTRIAL VALUE CHAINS

AI industrial value chains are rapidly increasing their economic power and establishing new dynamics that will define the future in AI-enabled societies. Aligning value and risk-sharing configurations across the different actors of AI value chains (researchers, data providers, platform providers, developers, solution providers, and end-users) is time-sensitive. Speakers highlighted the need to correct a misalignment in compliance incentives between the executive and technical levels of organizations.

Internal soft-law governance mechanisms – principles, guidelines, codes of conduct, standards – should be leveraged for preemptive compliance across jurisdictions. Discussions underscored the importance of a multi-stakeholder approach in developing such mechanisms.

KEYNOTE

Vilas Dhar
President, Patrick J. McGovern Foundation

▣ Watch video

Vilas Dhar opened the fourth edition with a timely discussion on AI and the rule of law, highlighting the major governance gaps unveiled by the rise of increasingly capable AI models in the past year. Despite growing awareness of AI impacts on the ground, some institutions have fallen behind on their duty to uphold fundamental rights.

Dhar proposed three interventions to **redirect the current state of the digital age towards a rights-based approach centered on human dignity**.

First, it is crucial to **reconceptualize institutions** for an AI-enabled future so as to curb rising socioeconomic inequalities. Second, sharing an example of indigenous communities building AI tools to further advance their forest-dwelling traditions, he advocated for **expanding the conversational space in AI governance** to a wider and broader global community. Inclusion efforts should be prioritized to better assess the opportunities and the impact of emerging technologies on the ground while AI governance frameworks are incipient.

Finally, he suggested a cultural shift from the idea that new market opportunities are always coupled with new vulnerabilities. Instead, Dhar proposed we should **design a world that cultivates new economic opportunities while upholding a bedrock of human equity**.

PANEL   The Path Towards an Enforceable EU AI Act

▣ Watch video

Eva Maydell, *Member of the European Parliament*

Katherine Forrest, *Partner, Cravath, Swaine and Moore LLP; Former U.S. District Judge for the Southern District of New York*

Werner Stengg, *Digital Expert, Cabinet of Executive Vice-President Margrethe Vestager, European Commission*

Juraj Čorba, *Digital Regulation & Governance Expert, Ministry of Investments, Regional Development and Informatization of the Slovak Republic*

Eva Kaili, *Vice-President, European Parliament*

Nicolas Moës, *Director, European AI Governance, The Future Society*

Ján Hargaš, *Deputy Minister of Investments, Regional Development and Informatization of the Slovak Republic (livestreamed opening statement)*

The first panel of this year's edition – moderated by **Nicolas Moës** – brought together institutional and geographical perspectives for a discussion on the practical challenges for the EU AI Act enforcement and international implications. European perspectives from the Commission – represented by **Werner Stengg** – and from the Parliament – represented by **Eva Maydell** and **Eva Kaili** – converged towards a call to action for the adoption of international standards and clear definitions within the EU AI Act.

The Slovakian Minister **Ján Hargaš** and digital expert **Juraj Čorba** presented the challenges to national enforcement and the need for international coordination. They stressed that obligations at the EU level should be compatible with forthcoming treaties such as the convention being prepared by the Council of Europe[1].

In addition, the former NYC Judge and lawyer **Katherine Forrest** provided a U.S. perspective for AI governance and touched upon the challenges and limitations to the international impact of the EU AI Act. Notably, she pointed out the complexities of fragmented legal systems, which must be urgently overcome to ensure effective enforcement of AI governance mechanisms and protect rights. In addition to jurisdictional difficulties, regulators and judicial actors must increase constant foresight efforts to reduce uncertainties around an ever-changing AI landscape.

With inputs from interactive polls, moderator Nicolas Moës and panelists explored how current premises of the draft legislation might pose a challenge to enforcement. For instance, the lack of human resources with the specialized AI skill sets in the public sector is expected to impact enforcement at the national level. Second, a future-proof approach will require overcoming divergences regarding the framing of important concepts such as the definition of high-risk AI systems. Therefore, such challenges also present timely opportunities for multi-level institutional coordination.

---

1. For an overview of the Council of Europe's work, see "Panel: Operationalizing International Agreements on AI".

TAKEAWAYS

- The most crucial elements for an effective enforcement of the AI Act:
    - strong EU-level coordination.
    - public and private buy-in from an early stage.
    - regulations that are designed to be practically enforceable and are in line with existing (and anticipated future) enforcement capabilities.
    - the incorporation of lessons learned from other regulatory endeavors in the EU.

- Looking into past regulatory endeavors in the EU can help anticipate and address operationalization challenges with AI regulations. For instance, GDPR enforcement has faced important limitations:
    - The GDPR's "one-stop shop" mechanism has been considered a bottleneck for enforcement, due to the limited competencies of national authorities outside of company headquarters' territories. Transnational AI regulations should preempt the legal obstacles for different jurisdictional levels.
    - The vast majority of data protection cases in Europe are litigated at a local jurisdictional level. Differences in procedural law across local jurisdictions have proven to be an obstacle to cooperation among countries.

A Manifesto for Enforcing Law in the Age of AI

Paul Nemitz, *Principal Adviser, Directorate-General for Justice and Consumers, European Commission*

Francesco Lapenta, *Founding Director of the John Cabot University Institute of Future and Innovation Studies*

Rebecca Slaughter, *Commissioner, Federal Trade Commission*

Watch video

In this fireside chat, speakers examined the recommendations presented in A Manifesto for Enforcing Law in the Age of AI. Published earlier this year by the Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of AI, the manifesto calls for the "effective and legitimate enforcement" of laws concerning AI systems by governments on both sides of the Atlantic. The moderator, **Francesco Lapenta**, presented the main recommendations:

- fostering international coordination efforts to strengthen enforcement and compliance.
- providing metrics, standards, and trustworthy AI tools.
- building capacities across institutions and empowering a wide range of stakeholders.
- establishing clear prohibitions and sanctions to illegal behavior.
- upholding public interest over countervailing proprietary approaches.

FTC Commissioner **Rebecca Kelly Slaughter** provided valuable insights on how different subject matters are intertwined when AI is deployed, making institutions with a broader competence an important leverage for enforcement. She drew from the Federal Trade Commission's example, with its competence over both competition and consumer protection. In addition, she highlighted the advantage of an outcome-focused enforcement rather than a process-focused one.

In a constructive dialogue on advancing solutions for a stronger enforcement, European Union expert **Paul Nemitz** suggested that the EU should develop an enforcement book that allows for a seamless compliance by companies, with coherence among different jurisdictions and different sectors. Furthermore, he advocated for a broader international cooperation based on a common approach of "democracy's primacy over business models" to curb the risks posed by AI.

TAKEAWAYS

- Regulatory enforcement goes beyond lawmaking and is not dependent on it: different institutions can leverage their competence and jurisdictional mandate to shape behaviors and provide clarity to the market around AI regulatory compliance.
- The work of the Transatlantic Reflection Group on Democracy and the Rule of Law in the Age of AI with the manifesto highlights the importance of civil society thought leadership in guiding the enforcement of AI governance.

Jack Clark, *Co-Founder, Anthropic*

Marek Havrda, *Deputy Minister of European Affairs, Czech Republic*

Sasha Rubel, *Public Policy Lead, Amazon Web Services*

Lucilla Sioli, *Director, Artificial Intelligence and Digital Industry, European Commission*

**Mark Brakel**, *Director of Policy, Future of Life Institute*

In an action-oriented and dynamic panel moderated by **Mark Brakel**, panelists and audience brainstormed what future-proof AI governance responses should look like to uphold the rule of law while fostering innovation. What should regulators and companies, across the value chain, do to respond to the rapid advancements in AI such as ChatGPT? How can governance be designed to overcome the challenges of AI's dual-use nature? The audience participated in discussions and expressed a favorable view of research and development (R&D) regulation through a hypothetical question (see Figure 1).

Lucilla Sioli stressed the importance of international organizations in contributing for a stronger enforcement of AI governance in the EU and beyond. The OECD's repository for metrics and tools could be used to check compliance in cases of high-risk systems and guide smaller companies on how to comply; and UNESCO's Recommendation on the Ethics of Artificial Intelligence, their readiness index, and their capacity-building tools could foster alignment in AI governance worldwide[2].

In a response to an audience question about the private sector leadership in general-purpose

POLL

Should AI research and development be regulated?

No 10%

Maybe 18.3%

Yes 71.7%
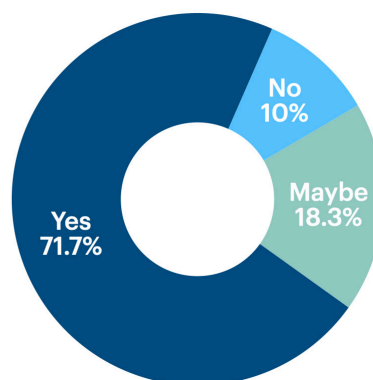
AI systems (GPAIS) breakthroughs, **Jack Clark** acknowledged the urgency in addressing power asymmetries in research and development (R&D). He suggested that governments invest in building capacity and shared experimental infrastructure to level the competition for the development of AI breakthroughs.

Marek Havrda explored actionable steps for stronger AI governance with a focus on AI research and development. He highlighted the importance of "regulatory learning" – through tools such as sandboxes — to produce use cases for evidence-based decision-making. The

proposed regulatory learning entails: adopting a human-centric approach, developing metrics that are fit for an agile approach based on scientific methods, and focusing on regulating the use of AI rather than research at large.

Sasha Rubel echoed Havrda's call for a use-case approach to AI governance in order to achieve policies that are tech-enabled but human-centric. In addition, she laid out the priorities for future-proofing AI governance at a global level. First, she called for the development of a shared understanding of what AI technology is, how to responsibly deploy it, and how to scale and innovate for impact. Second, she asked for a collective response to market fragmentation through interoperability and international standards.

Third, Rubel suggested following the [European Commission's digital compass](#) approach to digital transformation globally. Specifically, she

advocated for more participatory public policy making processes, cooperation between the public and private sectors and academia, working with start-ups on the ground to understand their limitations and guarantee implementation of policies and regulations, and building trust through capacity-building for different levels of stakeholders.

Furthermore, Jack Clark highlighted two main requirements to regulate AI R&D globally: an international set of standards and metrics to evaluate AI systems; and the need for every nation to invest in technical capacity so we can reach a harmonization and reduce power asymmetries worldwide.

Finally, Marek Havrda pointed out the need for Europe to work with China on R&D governance. Developing shared metrics, as suggested by other panelists, will be key in such efforts.

2. For more information on OECD's and UNESCO's work for international coordination and implementation of AI governance, see "Panel: Operationalizing International Agreements on AI".

TAKEAWAYS

- Priority actions to future-proof AI governance:
  - develop metrics and international standards to support "use case"-based regulatory learning (e.g. sandboxes)
  - invest in capacity-building within governments (i.e. upskilling and investment in technical capacity) and civil society (i.e. AI education focused on critical thinking about AI's potential and impact)
  - strengthen cooperation between the public and private sector, especially for GPAIS R&D governance (e.g. shared experimental infrastructure and constant dialogue regarding regulatory enforcement)

CLOSING REMARKS

Clara Neppel

Senior Director, IEEE Europe

In her closing remarks, Clara Neppel echoed the previous panel's call for metrics and risk definitions. Neppel underscored **the need for socio-technical standards and conformity assessments by independent bodies**. She stressed that measures and standards must be aligned with international AI value chains under a robust framework in order **to enable net-positive impact for citizens and legal certainty for providers in the international market**.

Gabriela Ramos

Assistant Director-General for the Social and Human Sciences, UNESCO

Gabriela Ramos alerted the audience to the critical ethical challenges posed by AI. **Ramos pointed out the need for more gender equality**, since women are by and large underrepresented in training datasets and in AI industry employment and entrepreneurship. She also stressed the need for changing current AI business models, which are characterized by high market concentration and a diversity deficit.

The UNESCO Recommendation on the Ethics of Artificial Intelligence seeks to **provide concrete policy actions and practical tools** to address these challenges. Ramos shared that UNESCO is also developing tools to help Member States with implementing the recommendation: a readiness assessment methodology and an ethical impact assessment to be deployed by states and companies.

Brando Benifei

Member of the European Parliament

Brando Benifei presented his remarks with a focus on next steps for the EU AI Act, **balancing innovation and security** of products in the European market. Echoing Clara Neppel, Benifei underscored the importance of **transparency, accountability, and risk assessment** by third parties throughout AI systems' lifecycle. He pointed out that special conditions should be applied to SMEs to reduce the costs of compliance. Benifei defended an **output-oriented**, rather than a technology-oriented approach to regulation and laid out three priorities for a responsible AI governance:

- fundamental rights impact assessment, with particular attention to vulnerable people
- stakeholder engagement
- effective enforceability

Jean-Noël Barrot

Minister for Digital Transition and Telecommunications, France

The last speaker of the day, French Minister Jean-Noël Barrot, opened his remarks laying out the conditions for AI to be a real opportunity for economic competitiveness: **gaining users' trust and developing and deploying AI in a safe, explainable, and ethical manner**.

To illustrate national efforts and capabilities driving change in AI governance, Barrot talked about France's efforts to advance **trustworthy AI**, the industrial project Confiance AI to advance trustworthy and interoperable systems, and France's National AI strategy for AI adoption, energy efficiency, and capacity-building. Finally, in line with The Athens Roundtable's mission, he mentioned the need to **establish a clear communication with businesses, administrations, and citizens regarding regulatory developments.**

OPENING REMARKS

[Watch video]

Kyriakos Pierrakakis
Minister of State and Digital Governance, Hellenic Republic

The Greek Minister Kyriakos Pierrakakis welcomed the audience to the second day of the fourth edition of The Athens Roundtable. He shared his first-hand experience with the regulatory and legal challenges posed by AI, advocating for a **values-based approach to AI** and stressed the significant role of **discussion fora** such as The Athens Roundtable in shaping AI governance.

Through the example of a recently launched Greek law[3], Pierrakakis shared that Greece has decided all corporations using AI algorithms on labor decisions must publish them online or notify those affected by the decision. The legal implications of AI in the workplace were further explored in the sessions that followed.

Marija Pejčinović Burić
Secretary General, Council of Europe

Council of Europe's Secretary General Marija Burić urged participants to collectively address AI challenges such as algorithmic discrimination and threats against election systems. Burić

provided insights into the treaty under negotiation at the [Council of Europe's Committee on Artificial Intelligence](). This effort aims at complementing national legislations to guarantee the protection of rights throughout the design, development and deployment of AI.

In an inspiring account of the potential of international efforts to advance AI governance, Burić stated that "**despite the pace of change, there is still a window of opportunity for AI to develop in a responsible, human-centered way**".

Tawfik Jelassi
Assistant Director-General for Communication and Information, UNESCO; Professor of Strategy and Technology Management, IMD Lausanne

Echoing Ms. Burić's remarks, Tawfik Jelassi underscored the importance of bringing the international community together around frameworks and implementation strategies for AI governance and risk assessment. The UNESCO [Recommendation on the Ethics of Artificial Intelligence]() was the first normative instrument of its kind, with ratification from 193 Member States in November 2021. However, **to establish a human-centered approach and an AI landscape that's open and accessible to all, it's urgent to invest in training all stakeholders about AI's impact and the protection of fundamental rights**.

In this sense, UNESCO has been implementing **capacity-building** programs with judicial

---

3. For more information on the recent legal efforts in Greece, see e.g. https://kglawfirm.gr/wp-content/uploads/2022/11/Emerging-Technologies-Law-4961-2022-c-6.pdf. For more on the legal implications of AI in the workplace, see "Fireside Chat: The U.S. Regulatory Landscape".

operators, civil servants, and other professionals globally. Notably, UNESCO and The Future Society launched the [MOOC on AI and The Rule of Law](#) earlier in 2022, which thus far has trained over 4,500 judicial operators worldwide.

Furthermore, Jelassi presented UNESCO's broad stream of work to address the ripple effects of AI breakthroughs in society through inclusive multi-stakeholder consultations and convenings such as the upcoming [Regulating Digital Platforms for Information as a Public Good](#) conference.

Anthony Gooch
Director of Public Affairs and Communications, OECD

Anthony Gooch discussed OECD's latest initiatives, focusing on three main challenges for AI governance that require immediate action: **accountability, interoperability, and anticipation**.

The [OECD AI Principles for a Responsible Stewardship of AI](#) published in 2019 provided practical and flexible standards for a swift adoption of AI governance measures. However, we are still grappling with the uncertainties around legal responsibility over the outcome of AI systems.

Regarding risk-based approaches in AI regulations such as the EU AI Act, this keynote underscored the need to build a basis of evidence for policies and mitigation strategies. The OECD is working on a framework for AI incident reporting and monitoring. In addition, **broader concerted efforts are key to reaching global consistency and interoperability in incident reporting**.

In terms of enforcement, **regulatory sandboxes are an opportunity for policymakers to improve their understanding of new technologies, and to anticipate standardization, interoperability, and regulatory needs**. Sandboxes must be coupled with spaces where policymakers and legislators can discuss the challenges of AI implications and share initiatives. According to Gooch, the OECD aims to fill that gap with the Global Parliamentary Network Group on Artificial Intelligence.

Michael Punke
Vice President of Global Public Policy, Amazon Web Services

Michael Punke centered his remarks around AWS' work in helping customers leverage the power of the cloud in AI and reap the benefits of digital transformation. For instance, he mentioned how AI can help double the output of the agriculture sector and meet the needs of a growing population while preserving the environment.

Punke stressed AWS cloud's role in **democratizing the use of AI among small businesses and start-ups**. By leveraging the cloud, stakeholders can also facilitate the green transition and reduce the carbon footprint of innovation.

Finally, Punke highlighted the importance of **capacity-building** in democratizing AI: AWS has launched educational initiatives through its Machine Learning University with the mission of training the next generation of AI developers.

Nikos Loutas, *Head Data and AI Policy, NATO*

Marietje Schaake, *International Policy Director at Stanford University Cyber Policy Center and International Policy Fellow at Stanford HAI*

The fireside chat on AI for defense kickstarted the second day's sessions and set the tone for action-oriented discussions with experts from different backgrounds. NATO's Nikos Loutas and Stanford's Marietje Schaake addressed a major blind spot to current AI regulation discussions: the lack of provisions regarding AI development and deployment in defense.

Drawing from his experience in the defense sector, Nikos Loutas positioned AI as a dual-use technology that can be leveraged to strengthen countries' defenses but also to amplify threats to national security and international stability. Technological breakthroughs in defense have a track-record of being repurposed and applied to a myriad of industries (e.g. from the Internet to urban vehicles and home appliances) with trickle-down effects in all sectors of society. Hence the concern with the lack of regulation in the development of AI systems for the defense sector.

Following NATO's recent adoption of responsible AI principles, Loutas noted that its operationalization is a pressing strategic need.

The organization is investing in multi-stakeholder coordination, capacity-building for their workforce, and creating a secure and resilient cyber ecosystem. In 2023, NATO's data and AI review board will develop a first version of a responsible AI certification standard focused on the defense sector. The organization will also focus on use cases to better understand AI's impact.

Marietje Schaake expressed concern over the threat of instrumentalizing AI for warfare. AI has challenged well-established notions that are fundamental to international law and international humanitarian law, such as that of sovereignty, territoriality, and attribution. Democratic states should take a stance on an updated interpretation of international law for the challenges posed by emerging technologies.

Schaake stressed that states should bear the prerogative of defending the Homeland to foster a free and open society. However, there is a gap in democratic governance and accountability among actors in this field, with the private sector leading the development of offensive and defensive capabilities.

TAKEAWAYS

- It's urgent to move beyond principles towards a rule of law-based system to strengthen accountability of private actors developing AI for defense.
- Further state action is needed to rebalance power relations between private and public actors regarding the collection and processing of data for AI use and the development of systems for the defense sector.

Andrea Renda - *Senior Research Fellow, CEPS (Brussels)*

Ariel Conn - *Head of the IEEE-SA Research Group on Issues of Autonomy and AI for Defense Systems*

Renée Cummings - *Data Science Professor, Data Activist in Residence, Criminologist at the School of Data Science, University of Virginia*

The second panel of the day built upon the previous discussion on AI for defense and drilled into the impact of AI breakthroughs – commonly originated in the defense sector – on human rights. The moderator, IEEE expert **Ariel Conn**, pointed out that technologies such as autonomous weapons can be used dangerously beyond military contexts, with little to no regulatory oversight.

Taking stock of recent AI developments and their risks to human rights, Professor **Renée Cummings** alerted the audience about how opaque decision systems can impact people's lives without their input, undermining autonomy, agency, and self-determination. In addition, Professor Cummings provided a U.S. perspective on the Blueprint for an AI Bill of Rights' approach to AI auditing and deployment. She highlighted the challenge of defining an adequate amount of due diligence and rigorous ethical guardrails against potentially violating uses of AI.

Measurement and evaluation are critical elements to ensure effective ethical guardrails. They can be operationalized through standards-setting coupled with civil society initiatives such as the Working Group on Interoperable Benchmarking of AI Systems established by The Athens Roundtable in 2021, which aims to provide empirical data needed for the objective assessment and adherence to the normative instruments.

Andrea Renda analyzed how normative guardrails were included in the EU AI Act through prohibited uses of AI such as mass surveillance by governments. However, the legislation doesn't consider the web of technological development that will likely happen as AI systems interact with each other in a given ecosystem. Such blind spots can create challenges to human agency if regulators lose sight of how individuals interact with each system.

Furthermore, Professor Cummings urged policymakers to look into the kind of datasets used to build AI systems: "we haven't had the critical conversation around data and its power dynamics, with privileges and prejudices that it creates". So far, data governance decisions have been left to the private entities developing AI systems.

Moving the conversation to enforcement, Renda drew upon his experience at the Trade and Technology Council (TTC) and analyzed how standards can facilitate compliance and reduce regulatory costs. Effective enforcement depends on regulators: they must define which governing entity or oversight body will ensure compliance to AI regulations and how standards can be updated in real time.

Regarding global-level action, Renda commented on TTC's work and the EU-US agreement to converge on certain definitions regarding AI and

to advance digital infrastructure and connectivity in third countries such as Jamaica and Kenya[4]. Finally, the panelists discussed how to advance international coordination and cooperation with the Global South: through AI literacy initiatives, data stewardship, and value retention at the local level throughout global supply chains. It is critical to establish standardized auditing practices at a global level, while also respecting cultural diversity and investing in AI-focused initiatives in developing countries.

4. For more details on the EU-US work at the Trade and Technology Council, see: https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_22_7516

TAKEAWAYS

- The quality and the democratic handling of data are key to advancing trustworthy AI and preventing large-scale human rights violations, especially in light of growing capabilities and pervasiveness of large-language models.
- Internationally recognized due diligence directives and best practices that pertain to human rights protection should be leveraged to increase responsible practices in AI (e.g. the 2011 United Nations' Guiding Principles on Business and Human Rights).
- International standards, civil society groups focused on evidence-based projects for measurement and evaluation, and investment in the Global South are critical to ensuring a global uptake of responsible AI.

KEYNOTE

Alexander Bornyakov
Deputy Minister, Ministry of Digital Transformation of Ukraine

[▷] Watch video

Ukrainian Deputy Minister **Alexander Bornyakov** provided an overview of Ukraine's plan of becoming a high-tech hub for AI development. He presented the country's path towards digital transformation, highlighting examples of the use of AI in Ukraine's public sector, ranging from an automated analysis model applied in the court of justice to forest plantation monitoring with the Deep Green Ukraine application.

Furthermore, he focused on how Ukraine has been using **AI for defense** to protect its territory and people from Russia's illegal armed attack. For instance, the army has deployed softwares that analyze images and identify war crimes evidence in Ukrainian cities; and drones that can spot and destroy enemy equipment. These use cases provide a realistic account of how AI plays a role in conflict and the urgency in discussing AI's implications for geopolitics and for international humanitarian law.

KEYNOTE

Mike Rounds
United States Senator

[▷] Watch video

Following the previous keynote, U.S. Senator Rounds stressed the urgency in addressing gaps in international and domestic laws regarding the use of AI in armed conflicts. He urged the international community to **strike the right balance between public safety, security, and privacy**.

Moving to the economics of AI, Senator Rounds mentioned the competitive advantage countries will have by investing in AI across the public and the private sectors with a complementary approach. In this sense, the United States Congress passed the National Artificial Intelligence Initiative Act in 2020 for a coordinated effort across the federal government to accelerate AI research and application in support of economic prosperity and national security. As a result of the Act, the US launched a National Artificial Intelligence Advisory Committee (NAIAC), with broad and interdisciplinary expertise from across sectors, to advise the President in matters pertaining to AI.

Regarding his co-chairmanship of the Senate AI Caucus in 2023, Senator Rounds stressed the importance of international coordination and adequate cybersecurity-focused laws for AI to be leveraged in a safe and secure manner.

PANEL — Operationalizing International Agreements on AI

[▷] Watch video

Marielza Oliveira - *Director for Partnerships and Operational Programme Monitoring Communications and Information, UNESCO*

Jan Kleijssen - *Director of Information Society - Action against Crime, Directorate General Human Rights and Rule of Law, Council of Europe*

Karine Perset - *Head of the AI Unit, OECD*

Marc Rotenberg - *President and Founder, Center for AI and Digital Policy*

The panel moderated by **Marc Rotenberg** discussed recent advances in international coordination for AI governance and explored synergies across key institutions and international normative agreements pertaining to AI. Opening the floor, Rotenberg introduced the latest developments in AI international policy frameworks and highlighted convergence in implementation as the next step for AI governance.

Jan Kleijssen presented the Council of Europe's (CoE) work in fostering human rights and democracy through legal cooperation and law-making. He also analyzed the prospects of a global convention on AI through the CoE's open ratification instrument by non-Member States. Pushing forward basic principles such as having a human in the loop and a remedy to redress harms, the convention would be complementary

to the EU AI Act. "While the EU efforts relate to the products, the Council of Europe's efforts relate to the process". Notably, CoE's global convention can confer rights – to be informed, to a remedy, to access stored systems – that citizens can invoke before courts in their respective jurisdictions.

Karine Perset focused on OECD's efforts towards international interoperability and risk management throughout the AI system's life cycle. Regarding the challenge of increasing convergence among stakeholders, the OECD is working on identifying common guideposts to decrease the burden of those that will have to implement all existing frameworks, particularly smaller organizations that are operating internationally. Mindful of the context-based challenges of AI applications, Perset further explored OECD's three actions for evidence-based policies:

- Monitor AI risks through incidents and case law across jurisdictions;
- Build a database of national AI policies through the OECD.AI Policy Observatory;
- Build a catalog of tools and metrics for trustworthy AI with the U.S. National Institute of Standards and Technology (NIST).

Beyond the significance of UNESCO's Recommendation on the Ethics of AI, **Marielza Oliveira** stressed the need to guarantee that such principles do not remain solely word on paper. UNESCO has been raising awareness about the impact of AI, developing guidelines and instruments that facilitate the deployment of responsible and trustworthy AI, and building capacity among stakeholders. According to Oliveira, the latter is critical to operationalizing AI governance at the international level.

Drawing from previous sessions' calls for stronger international coordination and shared values and standards, this panel drilled into the work of the main international institutions advancing AI governance. The Council of Europe's upcoming convention has the potential to complement existing normative instruments with a rights-based approach that can be enforced in multiple jurisdictions. Furthermore, international institutions' operationalization efforts must aim to facilitate a seamless adoption of standards and frameworks worldwide, taking into account different company sizes.

TAKEAWAYS

- In addition to risk-based, product-focused frameworks such as the EU AI Act, international organizations should work on rights-based normative frameworks specifically designed for the challenges posed by AI.
- The Council of Europe is currently drafting an international convention that can be incorporated by national jurisdictions worldwide, thus complementing regional efforts in regulating AI.
- Multi-stakeholderism at the center: from OECD's evidence-based strategy to UNESCO's education and training projects, enforcement frameworks are taking shape and require buy-in from multiple actors for effective AI governance.

Sebastian Hallensleben, *Head of Digitalisation and AI, VDE Association for Electrical, Electronic & Information Technologies*

Elham Tabassi, *Chief of Staff, Information Technology Laboratory, US NIST*

Gry Hasselbalch, *Senior Expert, International Outreach for a Human-Centric Approach to AI (InTouchAI.eu)*

Eileen M. Lach, *Former General Counsel and Chief Compliance Officer, IEEE*

Eileen Lach moderated a discussion on how measurement and evaluation can be used to reduce the risks posed by increasingly powerful and diffuse AI systems. The panelists provided recommendations to operationalize international, coordinated measurement and evaluation initiatives.

POLL

In the future, what applications of AI are in most need of measurement and assessment for trustworthy, reliable usage in the legal, compliance and/or judicial contexts?

none   ai in education

ai in the workplace   ai for social assistance

ai in health   ai in finance

ai use in judiciary

ai for accessibility   ai in policing

public service   ai for elections

Gry Hasselbalch pointed out the challenges posed by power dynamics in AI governance. Just like the Internet has been since its inception, AI is also a space of competing interests where democracy is not the primary guiding interest. She advocated for the need to actively steer a democratic socio-technical infrastructure for AI, through laws, standards, and literacy.

Drawing from previous panels and the challenge of bridging the EU and the US approaches to AI governance, **Elham Tabassi** mentioned that standards can be developed regardless of the

policy and regulatory landscape. She advocated for high-quality technical standards that are clear, implementable, and that consider the ambiguity around definitions and metrics. The ultimate objective for standards must be to foster a culture of critical understanding, communication, and management of risks.

Tabassi also highlighted the importance of evaluations in strengthening research communities. Ultimately, evaluations can create or jeopardize the conditions for an informed trust in AI systems. As a next step, NIST will publish a

voluntary risk management framework in January 2023, providing a structured and measurable way to map, measure, manage, and govern AI risks. Furthermore, context matters for evaluations: accuracy measures alone don't provide enough information on the AI system's impact. We must adopt rigorous methodology for transparent, repeatable, interoperable evaluations of real-world cases with real-world data.

Sebastian Hallensleben defended taking advantage of already established legally binding instruments and bringing the myriad of frameworks developed for AI systems into these instruments. As an example, the frameworks could be brought into the OECD due diligence guidelines to responsible business conduct for multinational enterprises when it comes to AI applications.

The panelists recognized that, even though there are substantial convergences in frameworks being developed by different institutions and across jurisdictions, the devil lies on the details. We still need to grapple with a wide range of definitions of risk, AI, transparency, and fairness.

TAKEAWAYS

- Standards-setting bodies and other institutions engaged with measurement and evaluation must coordinate efforts towards interoperable AI system evaluations.
- AI governance can gain from analyzing other domains' experience in creating measurement, such as with software performance assessment methodologies and comparison performance benchmarks.

FIRESIDE CHAT The U.S. Regulatory Landscape

Keith Sonderling - *Commissioner, US Equal Employment Opportunity Commission*

Anna Gressel - *Senior Associate, Debevoise & Plimpton LLP.*

Avi Gesser - *Partner, Debevoise & Plimpton LLP.*

Watch video

The fireside chat with Commissioner Keith Sonderling and legal experts Avi Gesser and Anna Gressel focused on the governance of AI pertaining to the workplace. The moderator **Avi Gesser** provided an overview of how the application of AI to hiring has been a widespread use-case and raises particular concerns about algorithmic discrimination. Emerging laws in the US – such as the NYC Law 144 –

signal the relevance in analyzing how this issue is regulated in the US and the EU so that companies are compliant across jurisdictions.

Commissioner **Keith Sonderling** commented on the Equal Employment Opportunity Commission's (EEOC) work in providing more clarity to employers around how discrimination laws that pre-date the use of AI in the workplace are being applied to this new reality. Ultimately, U.S. laws are outcome-oriented and the responsibility lies on employers, not on vendors of AI systems. Hence the importance of making such systems more transparent and reassessing liability in AI governance.

Anna Gressel touched upon the differences between the EU and the US regulations for employment discrimination. While hiring is considered a high-risk application in the EU AI Act, with both providers and users having to provide information on the AI system, regulations in the US focus on the result, which is borne by employers. In this sense, the vendor marketplace for AI tools is a key space for AI governance, since it's rare that employers will create their own AI tools in-house.

TAKEAWAYS

The fireside chat unveiled a pressing need to reassess the premises behind the liability regime in U.S. legislation.

- Liability must be adapted to the rise of third-party AI tools being procured and contracted by employers without them having full visibility over the functioning of the system.
- The US can learn from the EU AI Act, which proposes a broader liability regime and "AI life cycle" approach.
- Beyond regulation, best practices and standards for the use of AI in the workplace are crucial to update and strengthen EEOC's jurisdictional mandate and help employers in overcoming transparency and accountability gaps.

KEYNOTE

Dr. Alondra Nelson
Deputy Assistant to the President and Deputy Director for Science and Society, White House Office of Science and Technology Policy

▶ Watch video

Dr. Alondra Nelson provided valuable insights on how the U.S. Blueprint for an AI Bill of Rights can help protect democratic values. The document was preceded by consultations with different stakeholders to understand society's needs when it comes to AI governance. More than principles, the blueprint sets out the premises for AI:

- AI systems should only be used if they're **safe and effective**.
- People should be protected from **algorithmic discrimination** and algorithmic systems should be designed and used in an equitable way.
- Abusive data practices should be forbidden through **built-in protections**.
- Every individual should have **agency** over how data about them is used.
- People should know when automated systems are used and why and how those systems contribute to **outcomes** that impact their lives.
- Anyone should be able to **opt out** of AI-automated services and have access to a human being that can quickly help them solve the problem.

Dr. Nelson mentioned that **collaboration across institutions and jurisdictions** will be critical for the work ahead. She highlighted that the OECD's set of principles for the responsible stewardship of trustworthy AI was an important first step, and that the US is working on translating them into practice.

PANEL   Keeping Ahead of the AI Regulatory Compliance Curve

Peter Schildkraut, *Partner and Technology, Media & Telecommunications Industry Group Co-Leader, Arnold & Porter*

Navrina Singh, *Founder & CEO, Credo AI*

Miriam Vogel, *President & CEO, EqualAI; Chair, NAIAC*

▶ Watch video

In this panel, legal expert **Peter Schildkraut** moderated a compliance-focused discussion with business executives **Navrina Singh** and **Miriam Vogel**. The panelists examined what regulatory compliance will entail regarding forthcoming AI regulations and standards and how business can prepare for emerging technological trends.

Miriam Vogel alerted the audience that companies not thinking of compliance and regulatory standards might already be in breach of existing ones. For instance, the human in the loop requirement is already foreseen in the GDPR, and other regulations enforced in the US by the FTC, the EEOC, and the Department of Justice are being applied to AI-related cases as well, as presented in the previous fireside chat.
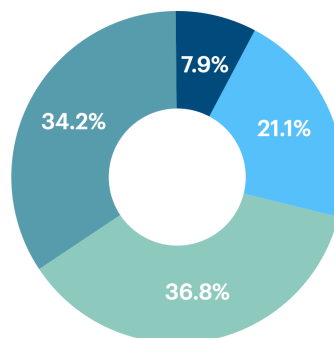
Drilling into risk management challenges, **Navrina Singh** mentioned three pain points common to most companies:

- A misalignment in incentives: technical stakeholders have different incentives than C-suite level actors, generating a gap in oversight.
- Access to the right datasets and tools: even though we have increasingly sophisticated assessment requirements, there is still a lack of tool infrastructure to put the oversight functions into practice
- Lack of clear organizational accountability structures: companies need to establish who is responsible for looking at the unintended consequences of AI, with the right accountable structure.

Reflecting on companies' challenges before the evolving AI regulatory landscape, Vogel evoked "good AI hygiene" as a tool for compliance. This entails mapping where AI is being used, planning where the company wants to apply AI, documenting the process and the systems' tests, choosing the adequate framework from the myriad of available options, and securing buy-in from the senior leadership.

The panelists discussed how complying with AI regulations is a competitive advantage in building trust with customers, building loyalty with employees, and standing out in the market for consumers. Furthermore, they provided guidance for risk management beyond regulations.

TAKEAWAYS

- Companies should address the internal misalignment in incentives and invest in preemptive compliance strategies.
- Regardless of headquarters' jurisdiction, companies should adopt internationally-oriented best practices and soft-law governance mechanisms – principles, guidelines, codes, standards – to increase responsible AI governance in global value chains.

Isabela Ferrari - *Federal Judge, Federal Regional Court of the 2nd Region, Brazil*

Prateek Sibal - *Programme Specialist, UNESCO*

Lilian Itemba - *Judge, High Court, Tanzania*

Watch video

A survey launched by UNESCO in 2021 identified judicial operators' priorities in terms of capacity-building for AI: leveraging AI to increase access to justice and addressing the challenges of virtual courts in light of the Covid-19 pandemic. Drawing from UNESCO's engagement with the topic, **Prateek Sibal** moderated a thought-provoking chat with Judges Isabela Ferrari (Brazil) and Lilian Itemba (Tanzania).

Reflecting on the challenges pertaining to AI in the Judiciary and how to further strengthen AI training and education, Judge and instructor of the MOOC on AI and the Rule of Law **Isabela Ferrari** commented on the growing trend of algorithmic decision systems in Brazil's courts. She stressed the need to educate stakeholders about how such systems are prone to biases in two ways: through biased datasets, and through biased hyperparameters, which are a set of rules the system follows to achieve the desired outcome.

Judge Lilian Itemba reflected on the use of a natural language processing software in Tanzania to interpret and transcribe court proceedings. Drawing from Tanzania's challenges with early-stage deployment, she stressed the importance of learning from countries in advanced stages of AI uptake.

TAKEAWAYS

- An effective enforcement of AI governance demands that judicial operators have a solid legal understanding of AI and its implications. Initiatives such as the MOOC on AI and the Rule of Law must be multiplied with constant updates on state-of-the-art AI capabilities.
- For regulations to truly uphold rights and liberties, we must expand the debate around AI and invest in **capacity building** – especially in countries that are not yet advanced in their adoption of AI and regulation.
- Countries must build ecosystems of international cooperation for AI governance and development, integrating industry players, regulators, and judicial actors in exchanges about AI development and deployment, so that regulation and legal enforcement reflect appropriate judicial remedies for the reality on the ground.

KEYNOTE

Lee Tiedrich
Distinguished Faculty Fellow for Ethical Technology, Duke University

Watch video

Professor Lee Tiedrich laid out the emerging trends and challenges in AI development and advocated that enforcement be consistent with OECD AI Principles, such as **promoting inclusive growth, sustainable development, and well-being**. She suggested 3 actionable steps to balance responsible AI innovation and the protection of fundamental rights:

- Build appropriate AI enforcement capacity.
- Develop AI standards and related tools such that either applying existing laws to the AI context or interpreting upcoming AI legislation can be adopted and enforced.
- Decide AI cases based on the relevant facts, in a **context-based approach**.

CLOSING REMARKS

Watch video

Christos Dimas
Deputy Minister of Development and Investments, Research, Innovation and Technology, Government of the Hellenic Republic

Greek Deputy Minister Christos Dimas presented **Greece's steps toward digital innovation**. Notably, the country launched a **start-up registry**, "Elevate Greece", to gather data about the start-up ecosystem, which will facilitate evidence-based policies for innovation and economic development. In addition, he commented on the work of Greece's four **research centers** that are paving the way for the country's rise in AI and digital innovation leadership. Representing the Hellenic Republic and underscoring their partnership with the Roundtable, Dimas commended the value-add of its multistakeholder and action-oriented convening.

Yannis Mastrogeorgiou
Special Secretary for Strategic Foresight at the Presidency of the Government of the Hellenic Republic

Expanding on the previous statement, the Special Secretary for Strategic Foresight at the Presidency of the Government of the Hellenic Republic Yannis Mastrogeorgiou shared his foresight of challenges societies will face with the potential negative outcomes of unregulated generative AI.

29

Drawing from Meta's Galactica example, Mastrogeorgiou called for all stakeholders to perceive the potential **toxic outputs of generative AI as a shared challenge to be overcome**. Although frameworks and norms are urgently needed, he stressed that AI will continue to have impacts we can't possibly yet understand. Thus, in the meantime, organizations in the forefront of AI development should adopt responsible behaviors and refrain from deploying systems that can cause harm.

Alexandros Spyridonos

Lawyer and Co-Founder of the European Law Observatory on New Technologies

Alexandros Spyridonos closed the second day of the fourth edition, highlighting the need to address AI governance with a worldwide approach and close transatlantic collaboration in leading such efforts. He further suggested exploring the application of regulations such as the EU AI Act and the need for a new draft bill of rights for the AI space. Spyridonos urged participants to **leverage the great potential of AI by building trust and prioritizing safety above all.**

CLOSED-DOOR DIALOGUE ON THE GOVERNANCE OF GENERATIVE AI

Jack Clark, *Co-Founder, Anthropic*

Anna Gressel, *Senior Associate, Debevoise & Plimpton*

Anna Makanju, *Head of Public Policy, OpenAI*

Dr. Emmanuel Kahembwe, *CEO - UK&I, VDE*

Connor Leahy - *CEO, Conjecture*

Kilian Gross - *Head of Unit, DG CNECT A/2 Coordination and Development of AI Policy*

Over the past year, a new wave of generative AI systems has taken the world by storm. ChatGPT, GitHub Copilot, DALL-E 2, Stable Diffusion, AudioLM, and Make-a-Video are only a handful of new tools redefining the manner in which content is produced and distributed — transforming our information networks and revolutionizing many business models. Due to the rapid emergence of generative AI, policymakers are unequipped to understand the risks it poses to individuals and society and are facing barriers towards designing or implementing regulatory safeguards.

In light of these circumstances, The Future Society organized a closed-door dialogue in the European Parliament on the morning of December 2nd (outside of the public agenda of the Athens Roundtable) to discuss the governance of generative AI. Speakers and participants included EU regulators, policymakers, civil society, standards associations, and companies developing generative AI systems.

During the event, one speaker remarked that the past year's rapid advancements of large and generative AI models signals the beginning of a wave of even more powerful AI systems and capabilities in the next few years. However, society and policymakers are currently unequipped and unprepared to govern them. To this end, policymakers drafting application-oriented regulation, including the EU AI Act, should consider proactive measures to govern this evolving area of AI research, development, and deployment. At the same time, policymakers have a responsibility to not regulate too heavy-handedly, so that companies within their jurisdictions are capable of competing in a global market. In this regard, regulatory sandboxes have emerged as an innovative tool for facilitating the production of laws and policies that promote safe technological development while fostering innovation.

The discussion touched frequently on the importance of understanding AI value chains for risk mitigation. AI development does not occur in a linear fashion, but rather in complex value chains with many actors and stages.

By examining the different stages of the AI value chain, from data collection, labeling and processing to algorithmic training and application deployment, decision-makers can identify potential ethical, safety and social risks, as well as opportunities for more responsible AI development. Additionally, understanding the actors' roles and their underlying incentives throughout the AI value chain – including technology providers, data owners, and users – can help to inform stakeholder engagement and the development of governance frameworks that promote accountability, transparency, safety, and fairness in the long term.

Moreover, speakers and audience discussed the importance of developing better tools to interpret and understand the mechanisms underpinning general-purpose AI systems ("GPAIS", used to describe the large, pretrained models that serve as the foundation for many of today's most popular generative AI systems). Greater interpretability would not only facilitate research and precautions for trust and safety, but would also inform policymakers about emerging AI risks and impacts. Several participants expressed an interest in better industry-wide benchmarks capable of meaningfully assessing the emergence and degree of new capabilities in general-purpose AI systems. In this regard, and to ensure that such

capabilities are tested against reliable references and processes are standardized, further work to develop interoperable benchmarks and metrology for GPAIS remains crucial.

The session also considered many new and unresolved concerns raised by generative AI. For example, issues regarding copyright law and remuneration arise for training data inputs and the outputs of generative AI systems. There are also new issues concerning responsibility, accountability, and civil liability with respect to products, such as code-generating AI. The European Commission's Draft Liability Rules for Artificial Intelligence (or Liability Directive for AI) comes into play towards ensuring accountability and liability across complex value chains for generative AI. To this end, corporate governance mechanisms such as licenses and terms of service may be valuable levers for governing generative AI across the value chain.

This session brought to light a number of significant, unresolved issues and concerns regarding generative AI. In the coming years, it will be necessary for regulators, policymakers, developers, and civil society to collaborate on policy approaches that preempt the risks posed by these novel transformative technologies.

THE ATHENS ROUNDTABLE
ARTIFICIAL INTELLIGENCE AND THE RULE OF LAW

2022 4TH EDITION REPORT

www.aiathens.org