November 29, 2021

Discussion Paper: Identifying Common Transatlantic Principles for AI Regulation

By: Meredith Broadbent, CSIS

In Europe and the United States, future economic growth and societal improvement will be fueled by the advancements in artificial intelligence (AI) employed by the digital, ICT, and manufacturing sectors. Among many other benefits, AI technologies can identify and amplify solutions to fundamental challenges in healthcare, reduce carbon emissions, and increase sustainable crop yields. But AI also has the potential to be used as a tool for repression, surveillance, violation of privacy, and institutionalized bias.

Because of the opportunities and risks, both Europe and the United States are grappling with how best to monitor and regulate the development of AI technology and its applications. Policymakers and business leaders are looking to adopt appropriate safeguards to protect the public against harm without stifling innovation that will enable a myriad of future public benefits. The challenge—particularly for the European Union, which is home to only six of the top 100 AI startups worldwide—will be to develop lighter-touch, adaptable regulations that facilitate rather than impede positive innovation and the uptake of AI.

Regulatory Framework Initiatives in the United States

- Executive Orders: E.O. 13859 establishes the American AI Initiative to promote R&D investment and coordination, reduce barriers to the use of AI technologies, develop technical and international standards around AI innovation, and train the workforce to develop and use AI. E.O. 13960 promotes the use of trustworthy AI in federal government and improve public trust and confidence through federal government use.

- Office of Management and Budget – Guidance for Regulation of Artificial Intelligence Application 2020 Report: This document lays out 10 principles for the stewardship of AI applications, including topics such as risk assessment, fairness and non-discrimination, disclosure and transparency, and interagency coordination. It also supports engagement in international regulatory cooperation efforts.

- Legislation: The 2021 National Defense Authorization Act includes the National Artificial Intelligence Initiative Act of 2020 (NAIA). Pursuant to this law, in 2021, the Office of Science and Technology Policy launched the National AI Initiative to support and coordinate federal AI activities. The 2021 Consolidated Appropriations Act established within the General Services Administration (GSA) an AI Center of Excellence to facilitate the adoption of AI technologies in the federal government. The National Institute for Standards and Technology has requested public input that will inform a voluntary Risk Management Framework for the use and development of AI systems.

- The 2020 NAIA authorizes $4.79 billion in funding for AI research at the National Science Foundation over the next five years, $1.15 billion at the DoE, and $390 million at NIST. A

March 2021 report from the US National Security Commission on Artificial Intelligence (NSCAI), chaired by Eric Schmidt, calls for the allocation of at least $8 billion towards AI R&D annually.

Proposed Legislation in the European Union

In April 2021, the European Commission (EC) published a legislative proposal for a Coordinated Plan on Artificial Intelligence to address the human and ethical implications of AI. The draft legislation, the Artificial Intelligence Act (AIA) follows a horizontal and risk-based regulatory approach that differentiates between uses of AI that generate: i) minimal risk; ii) low risk; iii) high risk; and iv) unacceptable risk, for which the EC proposes a strict ban. The EC legislative proposal requires that high-risk AI systems abide by a risk management system, be continuously maintained and documented throughout their lifetime, and enable interpretability of their outcomes and human oversight. The proposal also encourages European countries to establish AI regulatory sandboxes to facilitate the development and testing of innovative AI systems under strict regulatory oversight.

The law applies to any company selling an AI product or service in the EU, so will be extraterritorial in nature, similar to the GDPR.

The EU's Coordinated Plan proposes an increase in public and private investments in AI to a total €20 billion per year over the course of the next decade. According to the Technical Report by the Joint Research Centre, the EU invested between €7.9 billion and €9 billion in AI in 2019.

In 2019, the G20 agreed to commit to a human-centered approach to AI, adopting the G20 AI Principles, which were drawn from the OECD AI Principles. In 2020, the G20 agreed to advance the G20 AI Principles in each country. The OECD contributed to the 2020 Declaration of G20 Digital Economy Ministers by providing a report on examples of policies to advance the AI Principles.

Key Points of the U.S.-EU Trade and Technology Council (TTC) Inaugural Joint Statement

September 19, 2021(Annex III):

- The U.S. and EU will develop and implement trustworthy AI as part of a commitment to a human-centered approach, as demonstrated by an endorsement of the OECD Recommendation on AI. It recognizes that the U.S. and EU are also the founding members of the Global Partnership on Artificial Intelligence.
- The U.S. and EU oppose use of AI that does not respect democratic values and universal human rights, such as rights-violating systems of social scoring. They share concerns that authoritarian governments aiming to implement social control systems at scale pose a broad threat to fundamental freedoms and the rule of law.
- Regulatory measures should be "proportionate to the risks posed by the different uses of AI."

- Areas of cooperation
    - The U.S. and EU will uphold and implement the OECD Recommendation on AI; develop mutual understanding on principles underlining trustworthy and responsible AI; and discuss measurement and evaluation tools and activities to assess the technical requirements for a trustworthy AI (for example, protections for accuracy and bias mitigation.)
    - The U.S. and EU agree AI technologies should be designed to enhance privacy protections and will jointly "undertake an economic study examining the impact of AI on the future of workforces with attention to outcomes in employment, wages, and the dispersion of labor market opportunities."

Potential Areas for Future Regulatory Harmonization, Cooperation, and Partnership

The U.S. and EU plan to cooperate on assessing and developing technical requirements for trustworthy AI. Entities engaged in research and development of AI define "Trustworthy AI" as a framework that addresses challenges related to AI ethics and governance by ensuring that it is transparent, respects privacy, safe, impartial, reliable and responsible. In other words, trustworthy AI is premised on the idea that trust builds the foundation for realizing the full potential of AI. The U.S. and EU can also coordinate on the risk assessments processes and implications, while exploring potential interactions with data privacy and cybersecurity. The list below outlines possible areas for U.S.-EU cooperation on AI frameworks.

- Build smart cities and promote data protection, particularly with autonomous vehicles (AVs)
    - Joint standards and regulation on liability for algorithms in AVs and elsewhere
- Create frameworks that protect human agency
- Adjust conditions for which AI is developed
    - Having diverse coders is important so that, for example, algorithms identify black women as human
- Implement equal carrots and sticks
    - Incentivize the development of "beneficial" AI
    - Create regulations that reign in harmful algorithms and applications of AI
- Focus on defining outcomes, which can range from basic legal compliance to enhanced standards and extend to more ambitious moonshots
- Create optimal conditions for the innovation processes required to achieve them, such as the availability of data, skills, infrastructure, competition, and capital
- Define objectives but stay out of the code
    - AI develops more quickly than policy, so prescribing how algorithms work could constrain the innovation ecosystem and soon become outdated
- Regulate innovatively to "protect human agency"
- "Future of Work"
    - Invest in re-skilling and apprenticeship programs to minimize job loss from AI deployment

Beyond U.S. and EU Initiatives on AI

Many countries are starting to consider experimental models or co-regulatory approaches. These approaches allow experimentation to better understand the effects of AI systems and provide controlled environments to facilitate the scale-up of new business models. These take place in parallel to regulatory approaches that help create a policy environment that supports the transition from research to deployment of trustworthy AI systems. The concept of sandboxes was formally introduced in the United States. Subsequently, experimentation with sandboxes was conducted by the United Kingdom's Financial Conduct Authority. The objective of these sandboxes was to test new fintech products and services before they officially enter the market. Since then, a number of sandboxes have emerged in a broad range of sectors.

The Global Partnership on AI (GPAI) is an international and multi-stakeholder initiative to undertake research and pilot projects on AI priorities to advance the responsible development and use of AI. The Partnership was launched in June 2020 with 15 founding members: Australia, Canada, France, Germany, India, Italy, Japan, Mexico, New Zealand, Korea, Singapore, Slovenia, the United Kingdom, the United States, and the European Union. Brazil and the Netherlands have since joined. The Quad (Australia, United States, Japan, and India) also has standards setting for AI on its agenda.

Issues for Discussion

So far, collaboration between the U.S. and EU, most recently in the Trade and Technology Council, has focused on broad principles that, seemingly, have not been particularly controversial or politically challenging. Whether the U.S. and EU can avoid costly divergence in the regulation of AI in the future will come into more focus as discussions move towards developing common views on legal definitions and metrics for risk management requirements. As Nigel Corey of ITIF warns, the U.S. and EU should seek common principles, norms and regulations "but they should not expect to achieve complete convergence."

A key problem for regulators is that artificial intelligence is not yet an established field, technology, or capability, and countries have not converged on basic definitions for AI. Under the AIA, distinctions between AI, deep learning, algorithms, automated processes and "traditional software" are vague, making it difficult to get a basic understanding of compliance. As such, artificial intelligence may not be ripe for comprehensive regulation as is proposed in the EU's Artificial Intelligence Act (AIA).

Some EU Member States and private sector commenters have indicated that the current description of high-risk applications in the AIA is "still insufficiently clear and too broad, which can lead to overregulation and an unnecessary implementation burden." The definitions of high-risk and prohibited systems in the AIA have been criticized as being infeasible to implement in practice. Allied for Startups, a European network of startup advocacy organizations, states that definitions of high-risk use and the responsibilities and roles of AI providers, operators, and users must be clearly defined, particularly for purposes of assessing liability. Many other commenters on the AIA emphasize that there is "uncertainty about roles and responsibilities of the different actors in the AI value chain, namely developers, providers, and users of AI systems."

The private sector has bristled at some *ex-ante* requirements envisioned by the Commission, such as the need to turn over training data, algorithms, and programming history for audit. Business groups view requirements to reveal source code to Commission regulators as running up against protections for intellectual property established by the Trade Secrets Directive. There are serious concerns about how data protection requirements under the GDPR will work in tandem with AI applications that require flexible access to a wide variety of data sets.

Perfecting promising uses of AI depends in large measure on the availability of sizable quantities of training data; the more data available, the more algorithms can learn. According to the Information Technology Industry Council "the absence of sufficiently large and heterogeneous European datasets may foreclose the viability of critical AI innovations and advancements in Europe without relying on third-party data." The use of third-party data sets, a common practice to address the lack of data availability, would seem to be inconsistent with the wide range of European data restrictions.

The AIA envisions many requirements for documentation to establish accuracy, replicability, reproducibility and "explainability" of AI models. Developers of AI and regulators will be challenged by the problem that many effective AI models are not well-understood. Much of the time it is not possible to explain how AI models make certain determinations, making it impossible for companies to produce an explanation as required under the current AIA. In many ways AI models can be black boxes difficult to understand, yet alone explain to regulators. The terms "explainable" and "accurate" can be ambiguous when the goal is to reach an undefined level of human understanding of the functioning of a company's AI model.

Many organizations who submitted comments said that the new legislation, by increasing the cost and legal difficulties of using AI at an early stage, will reduce the capacity of EU firms to innovate socially beneficial applications of AI in league with the United States and China.

There will be harms and costs to limiting the use and development of AI in Europe. It can be expected that well-meaning regulators will be aggressive in their request for algorithms, data sets and programming history, reinforcing disincentives for thinly staffed start-ups to develop and scale AI research in the EU, when more permissive, voluntary environments exist in the United States and elsewhere.

Discussions on these issues and many others will take place as the draft AIA is considered in the European Parliament and scrutinized by Members States in the Council of Ministers. After it is passed, the AIA will be subject to a two-year implementation period.