April 29th, 2021

AI-REGULATION.COM

Machines Learn That Brussels Writes the Rules: The EU’s New AI Regulation

Mark MacCarthy and Kenneth Propp

To cite this article:

Mark MacCarthy a,d Kenneth Propp,
Machines Learn That Brussels Writes the Rules: The EU’s New AI Regulation, AI-Regulation.com, April 29th, 2021, https://ai-regulation.com/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/

AI-Regulation.com
CHAIR LEGAL AND REGULATORY IMPLICATIONS OF ARTIFICIAL INTELLIGENCE

The European Union's proposed artificial intelligence (AI) regulation, released on April 21, is a direct challenge to Silicon Valley's common view that law should leave emerging technology alone. The proposal sets out a nuanced regulatory structure that bans some uses of AI, heavily regulates high-risk uses and lightly regulates less risky AI systems.

The proposal would require providers and users of high-risk AI systems to comply with rules on data and data governance; documentation and record-keeping; transparency and provision of information to users; human oversight; and robustness, accuracy and security. Its major innovation, telegraphed in last year's White Paper on Artificial Intelligence, is a requirement for ex-ante conformity assessments to establish that high-risk AI systems meet these requirements before they can be offered on the market or put into service. An additional important innovation is a mandate for a postmarket monitoring system to detect problems in use and to mitigate them.

Despite these innovations and a sound risk-based structure, the regulation appears to have some surprising gaps and omissions. It leaves Big Tech virtually unscathed. It lacks a focus on those affected by AI systems, apparently missing any general requirement to inform people who are subjected to algorithmic assessments. Little attention is paid to algorithmic fairness in the text of the regulation as opposed to its accompanying recitals. And the newly required conformity assessments turn out to be merely internal processes, not documents that could be reviewed by the public or a regulator.

Nevertheless, the proposal is a comprehensive and thoughtful start to the legislative process in Europe and might prove to be the basis for trans-Atlantic cooperation to throw a common regulatory net over a consequential emerging technology, as White House National Security Adviser Jake Sullivan noted in his statement welcoming the EU's new AI initiative.

Scope

The regulation's broad definition of artificial intelligence captures software embodying machine learning, the older rules-based AI approach, and also the traditional statistical techniques that have long been used in creating models for credit scoring or recidivism. Although users of AI systems are covered, the regulation focuses largely on providers, the entities that develop an AI system and either place it on the market or put it into service for their own use.

The jurisdiction of the regulation covers providers of AI systems in the EU irrespective of where the provider is located, as well as users of AI systems located within the EU, and providers and users located outside the EU "where the output produced by the system is used in the Union." This "effects" test potentially extends the law's reach to companies without a market presence in the EU that use AI systems to process data about EU citizens.

Prohibitions

The regulation bans AI systems that cause or are likely to cause "physical or psychological" harm through the use of "subliminal techniques" or by exploiting vulnerabilities of a "specific group of persons due to their age, physical or mental disability." It prohibits AI systems from providing social scoring for general purposes by public authorities. It also precludes the use of "real-time" remote biometric identification systems, such as facial recognition, in publicly accessible spaces for law enforcement purposes.

These prohibitions sound stern, but there may be less here than meets the eye. The bans focused on manipulative or exploitative AI systems to address the "dark patterns," which many groups and scholars have identified as aiming to trick users into making decisions against their own best interests—such as divulging more personal information than is needed to get an online service. But an enforcing agency will have to determine when a system is exploitative or manipulative, so the effect of the ban is dependent on future regulatory action. The ban on social scoring seems a matter of optics, declaring that European values reject systems such as those reportedly under development in China.

In addition, the proposed regulation would allow for the use of remote biometric identification systems to search for crime victims such as missing children; to prevent the specific threat of a terrorist attack; or to detect, identify, or prosecute serious crime with certain protections including prior judicial or administrative approval. Moreover, the prohibition does not appear to extend to law enforcement's use of facial recognition on databases of "pictures or video footage" gathered previously.

These compromises will disappoint civil liberties advocates but satisfy EU member states' significant concerns over internal security and terrorism. For example, French authorities may use facial recognition technology to search a train station for suspects after a North African radical Islamic group threatened to detonate a bomb. And the regulation might allow member state police to use facial recognition for after-the-fact identification of suspects, as the FBI did after the Capitol riot.

The Rules for High-Risk AI Systems

High-risk AI systems include those intended to be used as safety components of products that are already regulated under existing product safety law, including machinery, toys and medical devices. In addition, the legislation specifically defines certain stand-alone AI systems as high risk when they pose special risks to established fundamental rights. The high-risk list includes systems used for remote biometric identification systems, safety in critical infrastructure, educational or employment purposes, eligibility for public benefits, credit scoring, and dispatching emergency services. Some AI systems used for law enforcement, immigration control and the administration of justice are also deemed high risk.

The regulation mandates, in general terms, the kind of governance program that financial regulators in the United States have established for the financial analysis models used by companies they regulate. The regulation properly notes that these rules "are already state-of-the-art for many diligent operators'' and are derived from the Ethics Guidelines of the European Commission's High Level Expert's group.

Providers of AI systems must establish "appropriate data governance and management practices" and must use datasets that are "relevant, representative, free of errors and complete." They must draw up technical documentation to demonstrate that the system conforms with the rules. This documentation must contain a general description of the AI system; its main elements, including validation and testing data; and information about its operation, including metrics of accuracy.

High-risk AI systems must be "sufficiently transparent to enable users to understand and control how the high-risk AI system produces its output." Disclosures include "the level of accuracy, robustness and security," the circumstances of use where the system would create risks to safety and rights, the general logic and design choices of the system, and a description of the system's training data.

High-risk AI systems must be designed to allow users to "oversee" them in order to prevent or minimize "potential risks." Design features must enable human users to avoid overreliance on system outputs ("automation bias") and must allow a designated human overseer to override system outputs and to use a "stop" button.

The regulation requires high-risk AI systems to "meet a high level of accuracy that is appropriate for their intended purpose" and to continue to perform at that level of accuracy in use. It requires resilience against "errors, faults or inconsistencies" and also against "attempts to alter their use or performance by malicious third parties intending to exploit system vulnerabilities."

A key obligation of providers is to conduct conformity assessments demonstrating that the high-risk system complies with these rules. The providers of AI systems used as safety components of consumer products, who are already subject to third-party ex-ante conformity assessment under current product safety law, now must also demonstrate compliance with the AI rules. The providers of stand-alone high-risk systems are required to conduct internal assessments themselves, except that independent third parties have to conduct the assessments for uses of facial identification.

After a high-risk AI system is sold or put into use, the providers must establish a "proportionate" postmarket monitoring system to collect data on the system's operation to ensure its "continuous compliance" with the regulation and to take corrective action if needed. Systems that continue to learn after they have been put into use would need a new conformity assessment if the modifications from learning are substantial. If the user of such a system modifies it substantially, then the user, not the provider, is responsible for doing the new conformity assessment.

The regulation seeks to be future proof in the face of evolving technology by providing the European Commission with delegated authority to update the definition of artificial intelligence, to update the list of high-risk AI systems, and to amend the requirements for technical documentation of high-risk systems. Delegated acts are a type of administrative action initiated and controlled by the European Commission that is less onerous procedurally than legislative change

.

Governance and Enforcement

Overseeing and enforcing EU law is always complex, due to the EU's inherent division of responsibilities between Brussels regulators and member states, but the scheme devised for artificial intelligence is particularly baroque. National supervisory authorities are to take the lead in conducting "market surveillance" of AI products, following the usual EU practice of deferring to experts in capitals. Member states need not create new, specialized AI regulatory authorities—existing authorities already entrusted with related regulatory responsibilities under other EU legislation may take on AI tasks as well. These authorities' AI powers include investigating health and safety or fundamental rights risks, and ordering companies to take corrective action, such as by removing offending systems from the market. A member state, or the European Commission, may object to a decision made by another member state, triggering an EU-wide consultation process that could lead to its reversal.

The national supervisory agencies have access to all information, documentation and data necessary to enforce the law, including access to source code if needed, and must "protect intellectual property rights, and confidential business information or trade secrets … including source code."

A newly created European Artificial Intelligence Board would be chaired by the European Commission and would have one representative from each national supervisory authority and the European Data Protection Board (EDPB). This new board would have the power to issue opinions and interpretive guidance on implementing the legislation, share best practices, and develop harmonized technical standards. The commission thus would have a leading role in EU-level AI governance, unlike with the member state-led EDPB for data protection.

Last, the European Commission has presented potential penalties of eye-catching severity, even beyond those of the General Data Protection Regulation (GDPR). Member states may impose administrative fines of up to 30 million euros or 6 percent of global annual corporate turnover, whichever is higher, on companies that market AI systems for prohibited purposes (like social scoring) or that fail to comply with data training requirements. Violation of other requirements carries lesser penalties, on a graduated scale.

Surprising Omissions and Gaps

The recitals accompanying the regulation are full of references to the well-documented concerns about the risks of algorithmic bias. But the text of the regulation is surprisingly thin on the need for conducting and publishing disparate impact assessments.

The regulation hints at disparate impact assessment in several places. The data governance rule allows providers of AI systems to use data concerning sensitive properties such as race, gender and ethnicity in order to ensure "bias monitoring, detection and correction." The robustness rule requires some AI systems to guard against "possibly biased outputs[.]" A system's technical documentation must contain "metrics used to measure … potentially discriminatory impacts" and information about "the foreseeable unintended outcomes and sources of risks to … fundamental rights and discrimination[.]"

But these vague references do not specifically require impact assessments on protected classes. Moreover, the documentation that arguably must contain a bias assessment does not need to be provided to users, the public or those potentially affected by discriminatory algorithms. It is available only to regulators upon request. In contrast, the legislation clearly mandates assessments and disclosure of system accuracy.

Overall, Big Tech emerges virtually unscathed under the new AI legislation despite being the object of widespread and growing concern over the use of AI-driven algorithms and the focus of most of the cutting-edge applied AI research. The regulation does not treat the algorithms used in social media, search, online retailing, app stores, mobile apps or mobile operating systems as high risk. It is possible that some algorithms used in ad tracking or recommendation engines might be prohibited as manipulative or exploitative practices. But as noted above, this would take an assessment by a regulator to determine.
The regulation is light on information that must be disclosed to the people who are affected by AI systems. The regulation requires that people be informed when they "interact with" an AI system or when their emotions or gender, race, ethnicity or sexual orientation are "recognized" by an AI system. They must be told when "deepfake" systems artificially create or manipulate material. But not in other cases. For instance, people need not be told when they are algorithmically sorted to determine eligibility for public benefits, credit, education or employment.
The requirement for conformity assessment is significantly less protective and revealing than it appears. The conformity assessment is a procedure, not a document, and an *internal* check-off for most of the providers of high-risk AI systems. So, there is no audit report for the public or the regulator to review. Instead, the public gets a "mark" affixed to the AI system indicating compliance with the rules. AI system providers must draw up a "declaration of conformity" and "keep it at the disposal of" regulators, but even this statement of compliance with the rules can be withheld from the public.

These apparent shortcomings will almost certainly be discussed and revisited as the legislation moves through the complicated and lengthy European legislative process.

Trans-Atlantic Dimensions

The EU regulation contrasts with the piecemeal approach to AI taken in the United States. The Trump administration delegated AI responsibility to specific regulatory agencies, with general instructions not to overregulate—an extension of the Obama administration's treatment of AI regulation. The Biden administration is likely to maintain this decentralized approach, with perhaps a greater emphasis on the need to regulate to avoid potential AI risks. The EU's comprehensive AI structure of horizontal risk-based rules, premarket conformity assessments, and postmarket monitoring duties will not be copied in the United States, although some cities and states have taken action over certain uses of AI by banning or heavily regulating facial recognition.

Still, there are opportunities for trans-Atlantic cooperation on AI. The European Commission greeted the Biden administration with an ambitious blueprint for trans-Atlantic cooperation. High on its list was the creation of a Trade and Technology Council, and "work on an AI agreement[,]" although no further detail was provided.

The EU's decision to carve out its own bold regulatory direction on AI inevitably poses the question of what space is left for a trans-Atlantic agreement. Obviously, the United States is not going to formally adopt the EU's definition of "high risk" nor its precise system of conformity assessments or, for that matter, its approach of comprehensive regulation. The realities of differing regulatory philosophies—and varying constellations of business, consumer and civil society interests—will sharply limit trans-Atlantic ambition.

Nevertheless, both sides will find it politically attractive to cooperate on AI. And while the issuance of the EU's draft regulation may close a few doors for collaboration with the United States, it also potentially opens others. For example:

- Once EU-wide criteria for high-risk AI systems are in place, the United States might seek an arrangement with the EU that allows companies located in the U.S. to self-certify as meeting them, subject to U.S. government control, under a system similar in concept to the Privacy Shield. Mutual recognition of conformity assessments also could be considered, as suggested in a recent Center for Strategic and International Studies paper.
- There may be scope for government-structured public-private partnerships on AI research and development. Such cooperation could be based on the U.S.-EU Science and Technology Agreement, following the approach the United States has taken recently with the United Kingdom.
- Specialized U.S. government agencies pursuing their own AI regulations, such as in the areas of finance, consumer protection, or employment protection, could compare approaches with counterpart regulators in the European Union, perhaps eventually yielding selective mutual recognition arrangements. These specific conversations could be organized under the framework of a larger intergovernmental dialogue on AI.

Conclusion

The European Commission's AI regulatory proposal is but the latest addition to an ambitious digital legislative agenda that Brussels has unveiled gradually over the past two years. Its Digital Services Act and Digital Markets Act took aim at the behavior of U.S. platform giants, perhaps explaining why this AI legislative proposal seems focused elsewhere. The commission also self-consciously has portrayed the AI regulation as a defense of European values against less scrupulous AI developers in China.The European Union prides itself on developing regulatory frameworks that have an impact outside its borders—witness the example of the GDPR. But whether its AI regulation ends up becoming the dominant global set of rules in a competition with the United States and China is far from the whole story. The commission's proposal reflects wide-ranging thinking and concrete choices on difficult policy issues. For that alone it will prove valuable internationally, even as it evolves like the technologies it seeks to master.

This post has first been published with Lawfare and is republished here with the kind permission of the editors

These statements are attributable only to the author, and their publication here does not necessarily reflect the view of the other members of the AI-Regulation Chair or any partner organizations.