Stanford University
Human-Centered
Artificial Intelligence

JUNE 2021

The European Commission's Artificial Intelligence Act

Marietje Schaake

RECENT ADVANCES IN ARTIFICIAL INTELLIGENCE (AI) have led to excitement about important scientific discoveries and technological innovations. Increasingly however, researchers in AI safety, ethics, and other disciplines are identifying risks in how AI technologies are developed, deployed, and governed. Academics, policymakers, and technologists have called for more proactive measures to tackle risks associated with AI and its applications. These range from voluntary frameworks to supranational legislation. Legislative action is on the rise. The world's first legal framework for AI was unveiled on April 21, 2021, when the European Commission published a comprehensive proposal to regulate "high-risk" AI use cases.

Introduction

From the stock market "flash crash" of 2010, to the 2019 wrongful arrest of an innocent African-American man living in Michigan, policymakers around the world are becoming increasingly familiar with the harms that poorly functioning or under-regulated AI systems cause. While there have been various attempts to address AI accidents, risks, or misuses, governments everywhere are realizing

KEY TAKEAWAYS

- To address the variety of risks associated with societal adoption of AI, the European Commission has proposed a set of regulations that promote the uptake of AI and try to mitigate or prevent harms associated with certain uses of the technology.
- Under the proposal, developers of high-risk AI systems will need to perform both pre-deployment conformity assessments and post-market monitoring analysis to demonstrate that their systems meet all requirements in the AIA's risk framework.
- The AIA expressly prohibits the use of AI for subliminal distortion of a person's behavior that may cause physical or mental harm; exploiting vulnerabilities of specific groups of people like the young, the elderly, or persons with disabilities; social scoring that may lead to unjustified or disproportionate detrimental treatment; and real-time remote biometric identification in publicly accessible spaces by law enforcement (except for specific actions like searching for missing persons or counterterrorism operations).

more must be done. Today's AI consists of complex algorithms that learn from constantly expanding and changing datasets. How, exactly, an AI system generates its outputs is often unknown to end-users. This lack of insight and transparency makes it difficult for people to anticipate the risks, harms, or rights-violations they incur. As AI evolves, future risks or harms should not be inevitable.

To address risks associated with the many possible applications of AI, the European Commission has put forward a set of regulations in the AI Act (AIA). The legislative proposal aims to promote the uptake of AI while mitigating or preventing harms associated with certain uses of the technology. The AIA does not cover most existing AI systems but imposes regulatory requirements specifically when an AI system is likely to pose high risks to the rights or safety of what the proposal terms *natural persons*—i.e. an individual human being, rather than a legal entity. The AIA also prohibits a limited number of specific use cases that are deemed to carry unacceptable risk as described further below. The proposal builds upon the EU's fundamental rights framework.

As with the landmark General Data Protection Regulation (GDPR) privacy law enacted in 2018, the European Commission seeks to *de facto* externalize its laws to apply outside its borders when companies do business with consumers in the EU. The AIA will likely have a standard setting impact known as the 'Brussels Effect.' Understanding the new AI framework should therefore be of interest to companies operating internationally as well as governments and civil society organizations seeking to understand the impact of the proposal as it progresses. This brief explores the proposed regulation and briefly discusses its implications.

> *Reactions to the AIA proposal have predictably been mixed: some organizations disapprove of the carve-outs for public security uses, while others lament that too little is being done to incentivize and support EU innovation and entrepreneurship in this space.*

Defining AI

Defining AI for the purpose of a legal text presents a number of challenges since definitions need to be both specific and future-proof. For example: section 238(g) of the FY 2019 National Defense Authorization Act in the United States offers a five-part definition that ranges from "[a]ny artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight," to "[a] set of techniques, including machine learning, that is designed to approximate a cognitive task," and even to "[a]n artificial system designed to act rationally." Jonas Schuett of the Legal Priorities Project recommends defining designs, use cases, and capabilities according to a risk-based approach. The European Commission opted for a hybrid approach, using both a broad definition for AI as well as defined categories and use cases.

Article 3 of the AIA defines an *artificial intelligence system* as "software that is developed with one or more of the techniques that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." The techniques identified include:

(a) Machine learning approaches, including supervised, unsupervised, and reinforcement learning, using a wide variety of methods including deep learning

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference, and deductive engines, (symbolic) reasoning and expert systems

(c) Statistical approaches, Bayesian estimation, search- and optimization methods

There are concerns that as drafted, the AIA's AI definition is too broad and may present loopholes for companies that seek to evade regulation. However, to ensure the regulation can be adjusted to emerging uses and applications, the AIA empowers the European Commission to amend these definitions as well as the list of high-risk AI systems referenced below.

> *Given the urgency of shaping global norms around AI, cooperation between the United States and the European Union will be revived the coming months.*

Four Prohibitions on Unacceptable AI

Beyond codifying an EU wide definition of AI, one element of the European Commission's proposal that has received significant attention is the prohibition of four specific use cases:

1. Subliminal techniques to distort a person's behavior that may cause physical or mental harm
2. Exploiting vulnerabilities of specific groups of persons such as the young or elderly, and persons with disabilities
3. Social scoring leading to unjustified and disproportionate detrimental treatment
4. Real-time remote biometric identification in publicly accessible spaces for law enforcement purposes

Article 5 describes how member states can bypass the EU's ban on real-time remote biometric identification in public spaces—sometimes termed 'mass surveillance'—under two conditions. First, if provisions in the states' own national law permit them to do so, provided certain conditions are met; and second, if the technologies performing these activities are strictly necessary and covered by any of the following three specific exceptions: the targeted search for specific victims of crime (including missing children), the prevention of a specific, substantial, and imminent threat to public safety (including terrorist attacks), or the detection, identification, and prosecution of specific criminal suspects facing significant jail time. These exceptions have led to criticism, including from the European Data Protection Supervisor.

High-Risk AI

Article 6 of the AIA proposes a definition of what constitutes a "high-risk" AI system. This section states that high-risk AI systems are those that can be used as a safety component of a product, covered by any of the nineteen EU regulations designed to harmonize standards for certain products across the market, or deployed in any of the following high-risk verticals:

- Biometric identification and categorization of natural persons
- Critical infrastructure where AI could put people's life and health at risk
- Educational and vocational settings where the system could determine access to education or professional training
- Employment, worker management, and self-employment
- Access to essential private and public services (including access to financial services such as credit scoring systems)
- Law enforcement
- Migration, asylum, and border control (including verifying the authenticity of travel documents)
- The administration of justice and democratic processes

Article 7(2) delineates the factors considered for assessing whether an AI system is high-risk. These include the intended purpose of the AI system; extent the AI system has been used or is likely to be used; whether the system has already caused harm or created concern for harm; the potential extent of such harm; how dependent those who are harmed by an AI system are on the technology; how vulnerable those who are potentially impacted are to imbalances of power, knowledge, economic or social circumstances, or age; how easily the outcome of the system can be reversed; and the extent to which existing EU legislation provides for effective redress or prevention of harm.

> *While the AIA does not cover low-risk AI systems, certain use cases such as deepfakes, chatbots, and other AI systems made for human interaction would need to follow transparency rules ensuring that consumers know they are interacting with manipulated content.*

Before these high-risk systems become available for consumer use, they are subject to strict reviews known in the AIA as 'conformity assessments', which determine whether the system meets all requirements in the AIA's risk framework. The developer or provider of the AI system will conduct these assessments and report their results to independent oversight authorities in each member state known as *notified bodies*. There are, however, some exceptions. For remote biometric identification systems, such as facial recognition technologies, the notified bodies themselves will administer the conformity assessment. Similarly, systems covered by existing product safety legislation and conformity assessments in the EU will stay covered by existing procedures.

Articles 9 through 15 detail how providers can be compliant with the AIA. Specifically, they should:

- Establish and maintain a risk management system
- Ensure training, validation, and testing data sets are subject to appropriate data governance and management practices
- Publish and update technical documentation of a high-risk AI system before it is placed on the market or put into service
- Incorporate logging capabilities to ensure a level of traceability of the AI system's functioning throughout its lifecycle
- Guarantee a certain level of transparency and provide users with relevant information (for example the characteristics, capabilities, and limitations of performance of the high-risk AI system)
- Put in place measures to guarantee human oversight and ensure high-risk AI systems can be overseen by natural persons during the period in which they are in use
- Design and develop systems in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity while performing consistently in those respects throughout their lifecycle

Since these algorithms are specifically designed to evolve over time as they learn from ever-growing data, the systems covered by the AIA will also need to comply with mandatory post-market monitoring obligations. These obligations require developers or providers to analyze relevant data about the performance and continuous compliance of high-risk systems with special attention paid to how the programs have changed throughout their lifetime. The proposal further establishes a European Artificial Intelligence Board composed of representatives from the member states and the European Commission. While the AIA does not cover low-risk AI systems, certain use cases such as deepfakes, chatbots, and other AI systems made for human interaction would need to follow transparency rules ensuring that consumers know they are interacting with manipulated content.

Since these algorithms are specifically designed to evolve over time as they learn from ever-growing data, the systems covered by the AIA will also need to comply with mandatory post-market monitoring obligations.

Next Steps

The European Commission's proposal is a first step toward devising a more responsible framework to govern risky AI use cases. Over the next few years, the European Commission will need to negotiate the proposal with the European Parliament and European Council. Each of the EU institutions will adopt its own position, then negotiate a compromise text that will be voted on before entering into law. By way of comparison, the GDPR was first conceived in 2012 and only entered into force in 2018, but a more rapid process is anticipated for the AIA.

Reactions to the AIA proposal have predictably been mixed: some organizations disapprove of the carve-outs for public security uses, while others lament that too little is being done to incentivize and support EU innovation and entrepreneurship in this space. Intense lobbying is expected from private companies, civil society, and governmental representatives alike.

It is worth noting that under the proposal, regulators will presume that high-risk AI systems currently in conformity with harmonized EU standards meet AIA requirements. Standard-setting organizations are now important conduits of economic power (because they often consist of patented technologies), legal dominance (because of their inclusion in trade law), and political choices (because of geographical lock-in and bifurcation).

Experts have noted that while the issuance of the EU's draft regulation may close a few doors for collaboration with the United States, it potentially opens up others. Further, in the United States, a number of states and cities have begun exploring restrictions on government use of facial recognition technologies. Given the urgency of shaping global norms around AI, cooperation between the United States and the European Union will be revived the coming months. This provides a golden opportunity to align both ecosystems on cybersecurity requirements, market access conditions, and standard setting through a possible EU-US Trade and Technology Council. Additionally, topics including data protection and digital taxation are on the table in Brussels and Washington.

The AIA is likely to evolve throughout the remainder of the drafting process with the European Parliament and the Council of the European Union. As the United States and other like-minded partners further engage with the law, policymakers and the broader public should consider the opportunities and responsibilities the proposal offers to ensure future technologies reflect our highest aspirations and contribute to realizing our shared dream of a better future for all of humanity.

Stanford University's Institute on Human-Centered Artificial Intelligence (HAI), applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu.**

Marietje Schaake is a former Member of the European Parliament; International Policy Fellow at the Stanford Institute for Human-Centered Artificial Intelligence (HAI); International Policy Director, Stanford Cyber Policy Center; President, Cyber Peace Institute.

Stanford University
Human-Centered
Artificial Intelligence

Stanford HAI: Cordura Hall, 210 Panama Street, Stanford, CA 94305-4101
T 650.725.4537 **F** 650.123.4567 **E** HAI-Policy@stanford.edu **hai.stanford.edu**