New Automated Decision-Making Laws: Four Tips for Compliance

June 27, 2022

With the widespread adoption of artificial intelligence ("AI") and other complex algorithms across industries, many business decisions that used to be made by humans are now being made (either solely or primarily) by algorithms or models. Examples of automated decision-making ("ADM") include determining:

- Who gets an interview, a job, a promotion, or employment discipline;
- Which ads get displayed for a user on a website or a social media feed;
- Whether someone's credit application should be approved, and at what interest rate;
- Which investments should be made;
- When a car should break or swerve to stay in a lane;
- Which emails are spam and should not be read; and
- Which transactions should be flagged or blocked as possibly fraudulent, money laundering, or in violation of sanctions regulations.

Depending on the potential consequences of the decision, ADM may involve different levels of human involvement and oversight. For example, an AI model that predicts whether someone might have a disease by analyzing their symptoms and their x-rays is likely to have a human doctor confirm the predicted diagnosis before any information is shared with the patient. This is referred to as partial ADM with a human in the loop. A music app that determines which song to play next based on a person's past listening habits is often fully automated, but, because the person can easily reject the choice and pick their own song instead, it is considered "human over the loop." A fully automated decision that cannot be easily changed or appealed to a human (*e.g.*, a resume-screening tool that puts forward only 10% of candidates for interviews) is referred to as "human out of the loop."

These and other ADM technologies promise greater efficiency and lower costs for businesses, but they also carry significant risks for employees, customers, and investors including:

- **Operational Risks** – The AI does not work properly or as intended, resulting in harm.
- **Transparency Risks** – Persons impacted by the ADM are entitled to know that the decision is being made by a machine and what data the machine is relying on, but they are not provided with that information.
- **Explainability Risks** – Persons impacted by the ADM are entitled to know what information the model relied on in making its decision and which inputs most influenced the decision, but that information is unavailable.
- **Legal Process Risks –** Persons impacted by the ADM have the right to opt out of the ADM or appeal the decision to a human, but no such options are provided.
- **Discrimination Risks –** The model is trained on, or uses, data that is somehow biased, resulting in decisions that are discriminatory.

Across the globe, regulators and lawmakers have passed laws aimed at reducing these risks. In this Debevoise Data Blog post, we discuss several new laws focused on ADM that are either in effect today or will go into effect in 2023, as well as circumstances in which litigants have used these laws to challenge companies' uses of ADM tools. In light of these trends, we have also included four tips for companies seeking to establish practical compliance and governance programs related to their ADM systems.

What Laws Apply to Automated Decision-Making?

Because ADM technologies often depend on vast sets of personal data, privacy laws are becoming a common means by which lawmakers address their risks, including:

- **EU General Data Protection Regulation (GDPR) (*in effect*)**: GDPR Article 22 gives data subjects the right not to be subject to "solely automated" decisions if they produce "legal effects" for that person or "significantly affect" them. Although still subject to final interpretation by the EU Court of Justice, the European Data Protection Board views Article 22 as *prohibiting* ADM processes, unless one of three exceptions applies: (1) the ADM is authorized by applicable law; (2) the ADM is necessary for entry into or performance of a contract; or (3) the ADM is based on the

data subject's "explicit consent." 22(1)-(2). Where ADM is permitted under Article 22 based on a contractual relationship or explicit consent, the company must also implement measures to safeguard the data subject's rights, including rights to obtain human intervention, express their point of view, and contest the decision. Art. 22(3). The GDPR places certain restrictions on the use of special categories of personal data, such as race or ethnic origin, for ADM (Art. 22(4)); it also requires that data subjects be notified of the existence of ADM, including "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." Arts. 13(2)(f), 15(1)(h).

- **Brazil Lei Geral de Proteção de Dados ("LGPD") (*in effect*)**: Under the LGPD, data subjects have the right to request review of decisions made solely based on "automated processing of personal data affecting [their] interests," including any decisions intended to define their personal, professional, consumer and credit profile, or aspects of their personality.

- **California Privacy Rights Act ("CPRA") (*effective Jan. 1, 2023*)**: The CPRA establishes a new California Privacy Protection Agency ("CPPA"), charged with adopting regulations "governing access and opt-out rights with respect to businesses' use of automated decision-making technology," including providing meaningful information about the logic of the decision and the likely outcome with respect to the consumer. Notably, the CPPA's mandate to issue ADM regulations is not currently limited to "solely" automated decisions or those with legal effects.

- **Colorado**, **Virginia**, and **Connecticut** State Privacy Laws **(*effective in 2023*)**: Colorado and Virginia's privacy laws will enable individuals to opt out of "profiling in furtherance of decisions that produce legal or similarly significant effects" concerning the consumer, which is generally defined as the denial and/or provision of financial and lending services, housing, insurance, education enrollment or opportunities, criminal justice, employment opportunities, healthcare services, or access to basic necessities. Connecticut provides an opt-out right that is similar to Colorado's and Virginia's, but only for "solely automated decisions," aligning it more closely with the GDPR.

- **Québec Bill 64 (*effective Sept. 2023*)**: Québec, Canada's Bill 64 regulates a broad variety of ADM systems, applying to any "decision based exclusively on [the] automated processing" of personal data. It requires companies to provide notice to the individual of the ADM; a channel for individuals to submit "observations" to a company representative who can review the decision; and, upon request, information about the personal information used to reach the decision, the principal factors and parameters that led to the decision, and their right to correct the personal information used for the ADM.

Additionally, ADM technologies may often be regulated through laws that protect individuals from potentially discriminatory practices. For example, the U.S. Equal Credit Opportunity Act prohibits discrimination in the context of credit transactions and requires companies to explain to applicants the specific reasons for denying an application for credit or taking other adverse actions, including where the decision is based on AI or complex algorithms. At the municipal level, New York City also recently passed an "Automated Employment Decision Tool" law, effective Jan. 1, 2023, requiring employers to conduct independent bias audits of their automated employment tools, provide detailed notices to employees or candidates about the use of these tools, and provide an alternative selection process or accommodation upon request.

Contesting Automated Decisions: Initial Legal Challenges

Recent cases brought under the GDPR against "gig economy" companies illustrate where ADM applications carry significant legal and reputational risk. For example:

- **Uber**: In 2020, Uber was sued in two actions by current or former drivers concerning its automated driver dispatch system and its so-called "robo-firing" algorithm, which was purportedly used to automatically terminate workers' contracts based on alleged fraudulent activity, without providing the drivers with sufficient information about the decision. On March 11, 2021, the Amsterdam District Court found in favor of Uber in both suits (*see* here and here), holding that Article 22 did not apply because the decisions reached by these systems did not have "legal or similarly significant effects." Notably, with respect to the "robo-firing" algorithm, the Court found that the decisions involved a human "Operational Risk Team" that reviewed the data to confirm a pattern of fraud and deactivate drivers' accounts accordingly; thus, the automated decision had "no legal consequences," as it was merely a temporary deactivation while human review occurred.

- **Ola**: Much like the *Uber* suit, this case focused on the Indian ride-sharing company Ola's automated work allocation program, including automated penalties against drivers for invalid rides. In contrast to its *Uber* decisions, the Amsterdam District Court (*see* here) held that Ola's automated penalties were subject to Article 22 for three reasons: (1) the penalties had a legal effect on drivers because they were used to impose fines and sanctions; (2) the penalties were "solely" automated because no human intervention took place prior to issuing the decisions; and (3) no exception to Article 22 applied, as these penalties were not necessary to the driver's performance under their contracts, and Ola had not obtained the drivers' explicit consent. The Court thus required Ola to provide its former drivers with the criteria used in the

penalty assessment, as well as the ability to verify and correct the data used in their decisions.

- **Foodinho**: On July 5, 2021, the Italian DPA, the "Garante," fined food delivery company Foodinho €6 million for GDPR breaches related to its algorithm for ranking and assigning drivers to delivery slots. Specifically, the Garante held that this system had significant effects on the riders because it afforded them access to job opportunities, and Article 22 thus applied (*see* here). The Garante further determined that Foodinho failed to implement suitable safeguards under Article 22(3) to protect its riders' rights, including (i) disclosing how the rating system functioned and what data would be collected, (ii) ensuring the accuracy and fairness of the ratings, and (iii) allowing riders to dispute the app's decisions.

Taken together, cases indicate that, for certain types of ADM, companies will need a combination of notice, consent, human oversight, and an appeal process in order to avoid significant legal risks.

Four Tips for Reducing Risks

In light of these emerging trends, companies deploying ADM should consider implementing the following measures to reduce their legal and reputational risks:

- **Identify Models and Algorithms That May Be Subject to ADM Laws.** Companies could start by making an inventory of their machine-assisted decision processes. Although some laws (like the GDPR) only apply to ADM processes that are "solely" or "exclusively" automated, other laws (like the New York City AI employment law) apply to technical processes that "substantially assist" human decisions.

- **Risk-Assess ADM Systems.** In order to determine whether mitigation measures are appropriate, applications that may be subject to ADM laws could be risk-assessed based on (a) scope of deployment, (b) anticipated use case and context, (c) locations of use, (d) extent and form of human involvement, if any, (e) use of personal data (including biometric or other sensitive data), and (f) impact of the decision, including whether it likely involves "legal effects."

- **Implement Mitigations for Applications Likely Subject to ADM Laws.** For applications that are likely to be subject to ADM laws based on a risk assessment, determine whether one or more of the following mitigation measures may be appropriate:

- **Human Oversight**, including having a human review some or all of the decisions before they are implemented, or conducting a periodic review of a sample of decisions already made, especially for important decisions that are solely made by machines.

- **Notice**, including informing individuals impacted by a particular decision that it is being made (partially or exclusively) by a machine, as well as types of data that the machine is relying upon in making that decision.

- **Explanation**, including providing individuals with additional information about the most important factors that led to the particular decision, the consequences of the decision, and what the individual can do to obtain a better result in the future.

- **Complaints and Appeals,** including providing a channel for individuals to submit questions, comments, or complaints about the AMD to a human, as well as opportunities to correct any inaccurate data associated with the decision and have a human review the decision.

- **Opt-Outs,** including providing an option for some or all individuals to opt out of the ADM process. This may be appropriate where the ADM may result in unfair decisions in light of a person's particular circumstances. For example, persons with visual impairments may be disadvantaged by certain AI hiring tools that rely on visual games to test personality types or aptitudes.

- **Bias Testing**, which may be appropriate for ADM that is subject to federal civil rights laws, state human rights laws, and industry-specific laws prohibiting certain forms of unfair discrimination. Such testing may involve a qualitative evaluation of inputs to the model, as well as an examination of relevant policies, procedures, training, and governance. It may also, in certain circumstances, involve a quantitative analysis of the model outputs to see if there is a disparate impact on any protected classes.

- **Ensure Appropriate Documentation.** Companies should consider what data should be preserved relating to the ADM process, including (a) data sources used to train or operate the model, (b) the data inputs used to reach particular decisions, (c) the decisions made by the tool, (d) the factors that were most important for reaching the decision, and (e) data relating to any of the testing or validation of the model.

The authors would like to thank Summer Associate Annabella Waszkiewicz for her contributions to this article.

* * *

Please do not hesitate to contact us with any questions.

NEW YORK

Avi Gesser
agesser@debevoise.com

Samuel J. Allaman
sjallaman@debevoise.com

Anna R. Gressel
argressel@debevoise.com

Andres S. Gutierrez
asgutierrez@debevoise.com

Mengyi Xu
mxu@debevoise.com

LONDON

Robert Maddox
rmaddox@debevoise.com