Collaborative Contracting Can Help Combat Bias In AI

By **Boris Segalis, Joshua Fattal and Neal Dittersdorf** (June 1, 2021)

Boris Segalis

Joshua Fattal

Neal Dittersdorf

Last month, the European Union proposed rules to govern the use of artificial intelligence systems that affect EU citizens.[1]

The framework is intended to encourage the development and spread of AI systems that are secure, trustworthy and transparent. The regulations also emphasize the development of AI technology that is ethical and unbiased.

AI researchers are also increasingly discussing the effects that AI can have on issues of systemic racism, bias and fairness.

As AI technology has evolved, it has begun to influence many facets of individuals' daily lives — from financial services such as extensions of credit or insurance, to recruiting and employment, to targeted online advertising. The rapid proliferation of AI has led AI developers to recognize the dangers of explicit and implicit biases and to consider how to properly quantify bias and address it.

Because the use of automated decision systems is subject to anti-discrimination laws and regulatory and public scrutiny, customers of AI solutions are understandably seeking to limit their risk of liability when using these tools.

These concerns are starting to arise during contracting, where parties often find themselves at loggerheads: Customers demand representations or warranties from providers that their AI solutions will not result in bias, while providers resist those guarantees.

Given the rapid evolution of AI technology and its uses, along with the uncertain legal and regulatory environment, an open discussion between contracting parties will be necessary to find mutually agreeable contract provisions. A binary allocation of liability to one party or another may be less useful than an approach that takes into account the relative roles of each party in controlling for bias and looks to the evolving tools available to help parties mitigate bias.

The use of AI solutions that can adversely affect protected groups creates risks for companies.

While the sources of law, regulatory requirements and enforcement practices differ, both the U.S. and Europe have regulated AI-powered decision making. Specifically, the U.S. and Europe both recognize claims of disparate impact and indirect discrimination.

The possibility of these claims, along with reputational concerns and a desire to play a positive role in society, are motivating businesses to look for ways to minimize this type of discrimination.

AI solutions may adversely affect one group of people with a protected characteristic — such as gender, race, national origin or religion — even though the algorithm is facially neutral.

In the U.S., this type of discrimination is known as disparate impact. Individuals who experience this type of bias are entitled to relief. A number of federal laws, such as the Equal Credit Opportunity Act, Fair Housing Act and Fair Credit Reporting Act, as well as state and city laws, address automated decision making that may have a disparate impact on protected groups.

Over the last decade or so, U.S. regulatory agencies such as the Federal Trade Commission, the Consumer Financial Protection Bureau and the U.S. Equal Employment Opportunity Commission have applied disparate impact analysis under these anti-discrimination laws to automated decision making across a number of sectors, and that scrutiny has begun to extend to AI.

Most recently, on April 19, the FTC released a blog post noting that the FTC Act prohibits the use or sale of racially biased algorithms.[2]

One early example of regulators' focus on automated processing of consumer information is a 2010 case in the U.S. District Court for the Northern District of Ohio, EEOC v. Kaplan Higher Education Corp., in which the EEOC sued Kaplan on the grounds that Kaplan's algorithm's use of credit reports had a disparate impact on Black applicants. The EEOC argued that the algorithm resulted in the rejection of more Black applicants than white applicants.

While the EEOC lost the case,[3] a number of state laws now protect individuals against widespread use of credit reports in hiring and promotion.[4] In 2020, New York became one of the latest states to introduce a bill to regulate the use of automated employment decision tools.[5]

Lawmakers have also scrutinized bias in lending algorithms. In 2020, a group of five senators requested that Upstart Network Inc., a company that provides lending algorithms to banks and had previously agreed to share information about its lending decisions with the CFPB, describe the tests used to ensure it complies with fair lending laws.[6]

In March, the CFPB, along with other federal regulators, issued requests for information on the use of AI by financial institutions, including the use of AI in enhancing credit decisions.[7] In the requests, the regulators explained that they are specifically concerned with the perpetuation or amplification of bias due to inaccuracies inherent in an algorithm's training data as well as incorrect predictions due to the incompleteness or nonrepresentative nature of data sets.

Similarly, in 2018, in Assistant Secretary for Fair Housing and Equal Opportunity v. Facebook Inc., the U.S. Department of Housing and Urban Development alleged that Facebook's algorithm violated the prohibition on housing discrimination by excluding certain protected groups from some housing advertisements.[8]

European law also prohibits indirect discrimination based on protected grounds such as gender; race, ethnic or social origin; religion; and membership in a national minority.

In Europe, AI algorithms that involve the processing of personal data fall under the General Data Protection Regulation. The GDPR imposes limitations and enhanced transparency requirements on automated decision making and offers Europeans certain legal rights regarding automated decision making that may affect individuals in legally significant ways.

While the current regulatory guidance on the GDPR does not reference discrimination based on protected grounds, European legislators have been increasingly attentive to the problems of racism and other biases. In 2019, the European Commission published guidelines on AI ethics, emphasizing the need to ensure legal clarity in AI-based applications.[9]

This past September, the European Commission released a five-year anti-racism plan for the EU, acknowledging the problems of structural racism and outlining focus areas to address racism more effectively.[10]

As mentioned above, the European Commission published its draft AI regulations on April 21, which require high-risk AI systems — algorithms that affect material aspects of people's lives such as assessing someone's credit score or determining whether they can get a loan — to meet minimum standards regarding trustworthiness and safety and fine companies that fail to comply.[11]

Additionally, the Council of Europe will address nondiscrimination in its upcoming legal framework for AI at the end of 2021.[12]

At the same time, the use of AI continues to advance, as both providers and customers see substantial benefits in current and anticipated applications of this technology. The applications in which companies see the benefits of AI are some of the very areas that are sparking regulatory and public concern.

For example, AI may have benefits such as matching consumers with appropriate products that even better suit their financial profiles, helping job candidates be matched with available positions that meet their capabilities and aspirations, and even facilitating reductions in bias and increasing diversity and inclusion.

A collaborative approach to contracting may be necessary to address AI bias.

In this uncertain environment, customers of AI solutions may seek to shift liability for bias, seeking contractual representations and warranties from the providers that the technology will not produce biased results. At the same time, AI providers often have serious concerns about taking on any liability for potential bias in their AI technology.

Absolute positions on who should bear liability may not be tenable for some time, if ever. Even though providers and customers may not intend to discriminate, unintended bias may develop at many points along the continuum from initial development to output to direct or indirect decision making based on AI.

Efforts to reduce bias in or resulting from AI continue, but they will require time to develop.[13] In the meantime, the legal and regulatory response continues to evolve. Contracting parties may need to adopt a more nuanced approach that is better suited to this rapidly changing environment.

At the outset, contracting parties should understand their respective abilities to mitigate bias. What they may discover is that, depending on the use case, the parties may control or share different aspects of how AI may have an impact on bias.

In many cases, the AI developer selects and supplies the data that is used to train the AI. In other cases, the AI is trained using data selected or supplied by the customer, or the output of the AI may be influenced by the customers' input.

In some cases, the greatest risk of bias may be reflected in the results of the AI processing, while in others the risk may arise primarily from decisions made by the customer. A full discussion of these factors will assist the parties in designing an allocation of liability that tracks more closely to each party's ability to control for and mitigate bias.

In light of these factors, contract parties may determine that they are best served by focusing on bias controls and testing, rather than no-bias guarantees. For example, the contract may require the AI provider to use controls designed to mitigate and test for bias. The contract also could require the provider to disclose when test results indicate bias, which may then enable the customer to take steps to mitigate the bias in its use of the AI output.

Prospective customers also may want to use precontract diligence to understand the controls and testing employed by the provider, or in some cases even reference certain controls and testing in the contract itself.

We acknowledge that this approach raises a host of issues, such as the protection of intellectual property or proprietary information and the imposition of contractual constraints on the ability of an AI developer to continually adjust its own controls and tests. The parties may need to consider solutions to these problems, balancing the need to place some limits on the disclosure of information against the benefits of disclosure generally.

The evolution of contract provisions governing information security may, by analogy, offer some guidance here. Providers and customers certainly still wrangle about a binary allocation of representations, warranties and liability.

As the understanding of information security has matured, however, some of the focus has moved to an agreement on documented security controls that a provider will maintain, or a standard against which the provider's controls will be measured, together with a combination of assessments, audits and reporting of test results or security incidents.

This kind of nuanced approach may also be appropriate for AI, where absolute guarantees may be unrealistic, and responsibility for mitigating bias does not necessarily sit with just one party.

Even when a customer seeks to license an AI tool that has already been developed, it may not be too late for the parties to agree to a controls and testing regime for the algorithm's design. AI engines typically undergo continual training, fed by additional data. Although an AI provider may not be able to provide assurances about the past, the parties still may agree on forward-going testing and mitigation.

Finally, contracts may need built-in provisions that anticipate changing legal and regulatory requirements. For example, the EU's proposed AI rules point to requirements that may affect providers as well as customers and require contact provisions to ensure compliance and allocate liability.

Because it may be difficult at this point to predict the exact regulatory requirements that will come into effect, contracting parties may want to include provisions requiring the parties to confer and make adjustments when new requirements are enacted.

The use and development of AI is still in its early stages. Efforts on the part of developers, governments and the AI community to mitigate biases in automatic decision making are only beginning, and the commercial response will also continue to evolve.

In some cases, contracts will be ground zero for how the market works through these issues while AI solutions continue to grow and spread. Ultimately, a collaborative approach between parties may be the best way to both facilitate the benefits of AI and continue to mitigate bias.

Correction: A previous version of this article listed the incorrect job title for author Neal Dittersdorf. The error has been corrected.

Boris Segalis is a partner and Joshua R. Fattal is an associate at Goodwin Procter LLP.

Neal Dittersdorf was previously executive vice president, general counsel and secretary at iCIMS Inc.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[2] https://www.ftc.gov/news-events/blogs/business-blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

[3] https://www.employerslawyersblog.com/2014/04/eeoc-loses-kaplan-credit-check-appeal-race-discrimination-disparate-impact-background-checks-title-vii-brad-cave.html.

[4] https://www.nolo.com/legal-encyclopedia/can-prospective-employers-check-your-credit-report.html.

[5] https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=Advanced&Search.

[6] https://www.bankingdive.com/news/wells-fargo-upstart-student-loan-case-study/571849/.

[7] https://www.jdsupra.com/legalnews/cfpb-and-federal-banking-agencies-issue-4673784/.

[8] https://www.hud.gov/sites/dfiles/PIH/documents/HUD_01-18-0323_Complaint.pdf.

[9] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[10] https://ec.europa.eu/info/sites/info/files/a_union_of_equality_eu_action_plan_against_racism_2020_-2025_en.pdf.

[11] https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.

[12] https://www.politico.eu/article/europe-artificial-intelligence-blindspot-race-algorithmic-harm/.

[13] https://techcrunch.com/2019/07/25/bias-in-ai-a-problem-recognized-but-still-unresolved/; https://www.wired.com/story/ai-biased-how-scientists-trying-fix/; https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html.