Fifteen Ways to Reduce Regulatory and Reputational Risks for Your AI-Powered Applications – Lessons from Recent Court Decisions and Regulatory Activity

February 20, 2020

It is only February, but, so far, 2020 looks like it is going to be the year that courts and regulators look seriously at artificial intelligence ("AI").

Recent developments in both Europe and the United States provide some insight into where AI is likely to face tough scrutiny and ways to mitigate risks of using AI.

On February 5, 2020, a Dutch court halted that country's use of an automated system for detecting welfare benefits fraud (called "SyRI"), finding that it violated the human rights of Dutch citizens—even though the court recognized that welfare fraud was a very significant problem for the Dutch government. SyRI analyzed government data falling into 17 categories—such as tax records, land registries, and vehicle registrations—to identify individuals whose welfare benefits should receive added scrutiny. Noting that SyRI was primarily deployed in poorer neighborhoods, the Dutch district court criticized the program for lacking transparency and ruled that SyRI violated Dutch citizens' "right to private life" under Article 8 of the European Convention on Human Rights. The court was concerned that SyRI, as applied, would lead to discrimination against the poor, many of whom are immigrants. The court also cited concerns that SyRI may violate requirements of the EU General Data Protection Regulation.

This concern was recently echoed by the European Commission, which released a White Paper on February 19, 2020 outlining its proposed EU regulatory approach for AI. Noting the potential violation of citizens' rights that can result from AI decision systems, the White Paper states that "economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, [and] any algorithmic exploitation of consumer behavior in violation of existing rules shall be not permitted and violations shall be accordingly punished."

In the United States, state government uses of AI are also being actively challenged, although under different legal theories. On December 5, 2019, a Court of Appeals in Michigan allowed a class action to proceed in a civil suit against the Michigan Unemployment Insurance Agency. The plaintiffs alleged that the Agency violated their

constitutional rights by utilizing an automated fraud-detection system (called "MIDAS") to determine that individuals had engaged in unemployment benefits fraud, without providing them with notice or an opportunity to present evidence. Many of these plaintiffs were assessed with large fines—some over $50,000—and had their wages garnished and tax returns seized if they did not pay. Notably, Michigan's auditor general examined MIDAS and found that the model was wrong 93 percent of the time.

Although the state later removed the algorithm and issued some refunds, the plaintiffs argued that they were not made whole—particularly, as the tax refund seizures affected their background checks and credit scores when applying for new jobs. The court held that the "absolutely egregious nature" of the Agency's algorithm violated the due process rights of the many people who had suffered sanctions resulting from the system's automated decisions.[1]

U.S. regulators and lawmakers have also started scrutinizing the use of AI. For example, the New York Department of Financial Services (DFS) issued a Circular Letter last year that imposed two obligations on life insurers who use AI in their underwriting process. First, insurers using alternative data (such as zip codes, educational attainment, and credit information) to power their AI tools must independently confirm that any given source of data "does not use and is not based in any way on race, color, creed, national origin, status as a victim of domestic violence, past lawful travel, or sexual orientation in any manner, or any other protected class." Second, insurers using alternative data for the purposes of any adverse underwriting decision for any particular consumer must disclose to the consumer the details about all information on which the insurer relied to make the decision, including the specific sources of that alternative data.

In addition, the U.S. House Financial Services Committee Task Force on AI held a hearing on February 12, 2020 to probe witnesses about concerns with ensuring that AI is used fairly.

To be sure, these developments largely involve governments using AI to make decisions that have significant impacts on the benefits and services provided to their citizens, which understandably face substantial scrutiny, in part due to legal restrictions on such actions.

Moreover, many commercial AI applications pose very little legal risk to consumers, either because the inputs do not involve any information about individuals, or the decisions do not have a significant impact on anyone. For example, the Roomba vacuum's use of automated systems to "learn" the best route for cleaning houses, and

---

[1] *See Bauserman v. Unemployment Ins. Agency*, No. 333181, 2019 WL 6622945, at *12 (Mich. Ct. App. Dec. 5, 2019) (internal quotation marks and citation omitted).

the Intelligent Oven's ability over time to identify common foods and adjust temperatures for the best results, are the kinds of AI applications that carry little legal or reputational risk.

Nonetheless, understanding how and why AI is facing increased scrutiny can be helpful for companies that would like to manage their automatic decision programs in ways that might help avoid legal challenges and reputational damage.

And so, here are 15 considerations to reduce risks to the use of AI. These have been gleaned from recent legal developments, as well as our experience helping clients in developing their AI programs—recognizing that some AI applications may have such low risk as to only merit the implementation of one or two of these considerations, if any:

Fifteen Considerations for Reducing Legal and Reputational Risk to AI Programs

1. Governance – Having policies and procedures concerning when AI can be used, how the appropriate inputs to AI models should be determined, and who must be involved in final approvals of any applications.

2. Impact – Determining when to conduct AI impact assessments before specific applications of AI decision-making are put into production—including, where appropriate, consideration of ways to minimize potentially negative impacts.

3. Authorization – Making sure there is authorization to use the proposed input data and a process to ensure the data's use is consistent with the authorization received, including diligence and sampling to ensure that actual input data is consistent with what people think is (and is not) going into the model.

4. Bias – Evaluating model inputs to make sure that they are relevant for the particular decision, and that they do not introduce unintended biases or discrimination.

5. Responsibility – Establishing accountability for the AI systems, including who is responsible for ensuring that model inputs, training, updating and operations are carried out according to policy.

6. Ongoing Evaluation – Periodically testing the AI system's functioning to monitor for unintended drift and to ensure that the model is operating as intended. Such testing is not currently a routine part of the AI development process and there is no consensus as to what constitutes effective testing.

7. Transparency – Informing stakeholders that AI is being used for a particular decision or process, as well as potentially providing information on key variables or parameters.

8. Appeal – Providing people who have been negatively affected by the AI system's decisions with a right to appeal or correct the decision.

9. Explainability – Providing stakeholders—including consumers, regulators, or other users of an AI system—with meaningful and clear information about how the model reached its decision, and why that result matters. This can be difficult depending on the type and complexity of the model, but is increasingly an area of focus for regulators and lawmakers.

10. Training – Ensuring that operators of the AI system and those implementing its decisions receive appropriate training on how to spot improper design, functions, or uses of the model.

11. Supervision – Having the AI system provide only a recommendation or preliminary decision, which then is reviewed by a human to make sure the recommendation or decision makes sense and is consistent with general expectations or domain-specific expertise (such as medical treatment guidelines).

12. Security – Providing appropriate security controls on the data being used for the AI system to reduce the risk of unauthorized access, including anonymization, deleting data not being used, and implementing other reasonable cybersecurity measures.

13. Reassessment – Reassessing all of the above for any significant change to the inputs, model or its applications.

14. Documentation – Making sure there is a good record of the steps taken above, including any key decisions that might later need to be defended to regulators or other stakeholders.

15. Oversight – Informing senior management and the board as appropriate about anticipated uses of AI, including risks and benefits, and steps being taken to mitigate potential risks and problems.

Recent development show that regulators and plaintiffs are not waiting for new laws to challenge AI programs. Rather, they are using existing antidiscrimination and consumer protection laws to seek redress for injuries allegedly caused by flawed AI decisions.

And even in the absence of significant legal risk, companies whose AI tools are criticized have faced negative media attention—as was the case for Microsoft's chatbot, Tay, which famously learned racist language less than 24 hours after going live on Twitter.

Accordingly, companies implementing AI programs may wish to review the foregoing considerations.

We will continue to monitor and assess legal developments regarding the use of AI and provide updates on anything of significance.

* * *

Please do not hesitate to contact us with any questions.

New York

Avi Gesser
agesser@debevoise.com

Henry Lebowitz
hlebowitz@debevoise.com

Washington, D.C.

Jeffrey P. Cunard
jpcunard@debevoise.com

Jim Pastore
jjpastore@debevoise.com

Lisa Zornberg
lzornberg@debevoise.com

Luke Dembosky
ldembosky@debevoise.com

Anna R. Gressel
argressel@debevoise.com

Steve Tegrar
sgtegrar@debevoise.com