The Future of AI Regulation (Part 4): 24 Ways That Companies Can Reduce Their Regulatory and Reputational AI Risks

May 6, 2021

Our three previous articles in this series on the future of AI regulation have discussed the RFI on AI issued by U.S. banking regulators, the draft EU AI regulation, and the FTC's recent guidance on AI bias and fairness. In this fourth post, we have taken those important developments in AI regulation, along with some other recently issued guidance, and prepared a list of 24 measures that companies can adopt now to prepare for the coming AI regulatory landscape, which is an update to a post we wrote last year on this same topic.

Although new AI regulations are probably not going to be effective for a few years, there are several reasons to begin future-proofing AI programs now, including:

- As the both the FTC and the U.S. banking regulators made clear in their recent AI releases, existing laws such as Section 5 of the FTC Act, the Fair Credit Reporting Act, and the Equal Credit Opportunity Act already prohibit unfair, deceptive, and discriminatory uses of AI.

- Companies are building AI models now that they hope will last for many years. To the extent that those models turn out to have been designed, trained or operated in a way that is contrary to future regulatory requirements (or future interpretations of existing laws), it is possible that some models will need to be substantially modified or decommissioned, which may be very costly and disruptive.

- For many companies, the governance structure that will be needed for AI regulatory compliance will include (1) identifying high-risk uses of AI and treating those as enterprise-level risks, (2) establishing sufficient oversight of AI by senior management and the Board, and (3) ensuring that the appropriate representatives from Legal, Compliance, Risk, Information Security, Privacy, Internal Audit, etc. are consulted in the design, implementation, and continued oversight of high-risk AI. For some companies, creating these new and complicated governance structures will take time, and many decisions will need to be made along the way in terms of how much to rely on existing frameworks (e.g., model risk management, vendor management, data management, incident response plans, etc.) and what additional

governance structures may be needed. Many companies are starting this work now, knowing that it could take months or years to fully build out their AI governance framework.

- As the illustrated by the recent report issued by the New York Department of Financial Services ("DFS") on allegations of gender bias relating to the Apple credit card's credit limits, companies face significant reputational risk in adopting AI. The same measures that reduce potential regulatory risk can also reduce reputational risk. While the DFS report absolved Apple (the card's offeror) and Goldman Sachs Bank USA (the underwriter) of any fair lending violations, DFS observed that consumers' concerns about being affected by a "black box model" could have been mitigated through clearer disclosures and increased responsiveness to customer appeals. By adopting some of the measures discussed below, companies may better position themselves to be able to quickly and effectively defend their AI from allegations of bias or unfairness from customers and in the press.

To help companies meet their evolving AI legal obligations, regulators have started to provide detailed guidance on practices for managing AI risks. For example, the UK Information Commissioner's Office has recently published a toolkit containing an extensive set of suggested risk mitigation measures that significantly overlap with the suggestions made in other AI regulatory guidance we're tracking worldwide.

Below we have set out (in alphabetical order) 24 measures for companies to consider as possible ways to reduce the regulatory and reputational risks associated with their AI programs:

1. **Accountability**—Place overall responsibility for AI regulatory compliance and AI risk with a senior individual or committee to ensure meaningful oversight of AI development, implementation, and ongoing monitoring. Require certain high-risk AI applications to obtain approval from that person or committee before deployment.

2. **Appeal Rights**—Provide persons who may be negatively affected by AI decisions with the ability to challenge those decisions, as well as the right to receive and review the information necessary to conduct such a challenge or correct any erroneous personal data.

3. **Bias Testing**—Ensure that appropriate bias testing is conducted before deployment of models that are expected to process sensitive personal data and that might negatively affect consumers, job applicants, or employees.

4. **Board Reporting**—Management should periodically report to the Board on AI use and risk within the company. The Board should have information that is sufficient for it to assess whether the company's risk appetite around its use of AI is appropriate.

5. **Business Continuity**—Ensure that the business can continue operating without significant interruption if a particular AI program fails, is hacked, or cannot be used for some other reason.

6. **Cybersecurity**—Protect models, inputs, testing data, and outputs from unauthorized access by both insiders and third parties, including through anonymization, deletion of data not being used, and other measures to prevent hacking, ransomware, data poisoning, adversarial example attacks, and other malicious training or use.

7. **Dark Patterns**—Avoid AI use cases that could be viewed as manipulating human decisions through subliminal techniques or exploiting vulnerabilities due to age or mental condition in a manner that harms individuals or groups (e.g., targeting advertisements to individuals who have bipolar disorders during their manic phases), which carry enormous reputational and regulatory risk.

8. **Data and Privacy Rights**—Ensure that privacy obligations and other data rights have been respected for any data used for testing, validating, and operating an AI system, including complying with applicable disclosure obligations, consents, IP, or contractual rights related to the data, as well as any limitations on sharing or using personal or third-party information.

9. **Disclosures**—Ensure that company statements about its AI, including risk factors in financial statements, investor communications, privacy policies and public codes of conduct, are accurate and, as appropriate, backed by evidence and documentation.

10. **Documentation**—Maintain accurate records and proper documentation on algorithms used for decision-making, including risk assessments, training, data testing, and output logs to ensure compliance with applicable regulatory record-keeping obligations.

11. **Escalation**—Ensure that high-risk AI incidents (such as a model significantly deviating from expected behavior, findings that a model's input or training data is fundamentally flawed, or credible claims of bias) are promptly reported to the appropriate executives within the company. Timely escalation

facilitates important risk-mitigation decisions regarding the continued use of the model and whether to notify potentially affected persons or regulators.

12. **Explainability**—For AI decisions that may have negative effects on individuals, ensure the appropriate level of explainability, which may include the basics of the model's functioning, how results were reached, grounds upon which the results were based, and how someone may make changes to improve their results.

13. **Guardrails**—Create automatic circuit-breakers or guardrails that prevent a model from significantly departing from expected performance ("model drift") or alert the company when drift is occurring.

14. **Human Oversight**—Consider establishing human oversight of certain high-risk AI decisions. For example, some automated decisions may not become effective until reviewed by a human ("Human in the Loop"), or those decisions may generate alerts so that human intervention is possible shortly after the decision is made ("Human over the Loop").

15. **Inventory**—Maintain a list of the AI models in use and in development, along with a risk rating for each AI model based on its intended or reasonably foreseeable uses.

16. **Model Validation**—Test the data and the models for accuracy and integrity. Document the testing procedures. Ensure that data being used from different sources is being properly harmonized. Stress test or re-validate the model before deployment or following any significant change in use, context, or scope.

17. **Ongoing Monitoring**—For high-risk AI applications, ensure that ongoing monitoring will alert management when the AI significantly departs from expected performance, especially for dynamically updating models.

18. **Opt-Out Rights**—Consider providing consumers with the option to decline or opt out of automated decision-making under certain circumstances.

19. **Policies**—Create a set of written principles, policies and procedures for the development, deployment and use of AI that are aligned with the organization's risk appetite and ethical or corporate principles. Conduct periodic audits of the company's compliance with its AI policies.

20. **Regulatory Compliance**—Conduct a legal review of AI systems to confirm compliance with applicable laws and regulations, particularly for AI systems that may impact individuals in protected or vulnerable groups, that leverage sensitive data, or that might otherwise be considered high risk.

21. **Risk Management**—Integrate AI adoption and use into the overall risk-management framework of the company, and create a risk-assessment framework for AI programs to identify those applications that are considered high-risk along with the basis for those determinations, as well as any appropriate risk mitigation.

22. **Training**—Provide training for employees who develop, approve, or use AI applications, including boards and management, as well as special training for certain employees relating to bias and model validation.

23. **Transparency**—Ensure accurate disclosure to those who may be negatively affected by AI decision-making or interact with AI systems or content, such as chatbots, roboadvisors, deepfakes, sentiment monitoring, or biometric profiling.

24. **Vendor Management**—Establish an AI vendor risk framework for third parties involved in developing and implementing AI for the company, or providing external data used in the company's AI systems, which may include questionnaires, risk assessments, due diligence, and appropriate risk mitigation through contractual provisions, insurance, or other risk allocation.

* * *

Debevoise has developed the Debevoise AI Regulatory Tracker ("DART"), an online tool to help our clients keep track of AI regulatory developments across the globe. For a demonstration of the DART, please contact us at agesser@debevoise.com and argressel@debevoise.com.

To subscribe to the Data Blog, please click here. Please do not hesitate to contact us with any questions.

NEW YORK

Avi Gesser
agesser@debevoise.com

Anna R. Gressel
argressel@debevoise.com

Tara Raam
traam@debevoise.com