Regulators Should Treat AI Like Employees to Avoid Stifling Innovation

November 29, 2022

We recently wrote about how rights-based regulatory regimes for artificial intelligence (as opposed to risk-based frameworks) can lead to a misallocation of resources because compliance will require too much effort on low-risk AI (*e.g.*, spam filters, graphics generation for games, inventory management, etc.) and not enough effort on AI that can actually pose a high risk of harm to consumers or the public (*e.g.*, hiring, lending, underwriting, etc.). In this follow-up blog post, we discuss why regulators should view AI risk the same way as employee risk for large companies, and accordingly adopt risk-based regulatory frameworks for AI.

The deployment of AI has resulted in increased efficiencies and entirely new business opportunities across sectors and industries. But it has also raised concerns about privacy, data security, bias, transparency and the quality of automated decision-making. Regulators are understandably trying to ensure that employees and consumers get the benefits of AI innovations, but are protected from these and other risks.

There are two different regulatory approaches that are developing to address these risks. The first is a rights-based approach, which treats all instances of AI that meet a certain definition equally, and therefore subjects all AI to the same compliance obligations. The second is a risk-based approach, which treats AI applications differently depending on the likelihood or severity of the potential harm they might cause, and therefore subjects different AI systems to different compliance obligations, if any. For example, the White House's non-binding Blueprint for an AI Bill of Rights proposes primarily a rights-based regime that argues in favor of regulatory oversight of all covered automated systems, largely regardless of their risks. New York City's AI hiring law is also largely rights based. By contrast, the draft EU AI Act is largely risk based, with a small number of high-risk categories of AI being subject to the more onerous compliance obligations.

We have noted that rights-based regimes, depending on their requirements, can be challenging or unworkable to implement in practice. Many organizations that have adopted AI are currently running hundreds, if not thousands, of models making decisions that range from consequential to insignificant. Requiring those organizations to put every single AI application through a complicated and time-consuming

compliance process is not an effective way to reduce the risks associated with automated systems. Rather, compliance efforts should be primarily focused on the small number of models that actually pose a high risk of causing harm, such as those that significantly influence decisions involving hiring, promotions, lending, detecting fraud, insurance underwriting, law enforcement, and education admissions.

One additional drawback of rights-based AI regulatory regimes is that they focus almost exclusively on the potential drawbacks of AI systems, and therefore leave little room to balance those drawbacks against the systems' countervailing benefits. They also often fail to acknowledge that many of the potential problems associated with AI systems are equally present in human decision-making, which can also be flawed, opaque, and biased.

To illustrate the point, we urge regulators to think of the risk posed by AI in the same way we all think about the risk posed by employees:

Employees	AI Applications
Nearly every employee theoretically has the potential to cause a significant amount of harm to, and thus meaningfully impact, an organization.	Nearly every AI application theoretically has the potential to cause a significant amount of harm to, and thus meaningfully impact, an organization.
Even the most junior employees have the potential to cause significant damage by stealing sensitive information, alienating customers, destroying valuable property, and undermining core company objectives.	Nearly every AI application has the potential to cause significant damage by malfunctioning and bringing down the operation they are associated with or making arbitrary or random decisions.
The decision to hire any particular employee is optional. With few exceptions, any individual hire is not essential to the company, although the hire may bring benefits to the company.	The decision to implement any particular AI application is optional. With few exceptions, any AI system is not essential to the company, although the AI system may bring benefits to the company.
Most large organizations cannot function without having hundreds or thousands of employees, which requires the constant hiring of new employees.	An increasing number of organizations cannot function without having hundreds or thousands of AI models running, which requires the constant deployment of new AI applications.
It is both unworkable and a waste of resources for large companies to require a lengthy and robust vetting process before each new employee at every level of the company is allowed to start their job.	It can be unworkable and a waste of resources for companies operating thousands of models to require a lengthy and robust vetting process

Employees	AI Applications
	before each new AI application is allowed to operate.
The hiring process for employees who are not employed in high-risk or sensitive domains is often limited to a resume review, one or two interviews, a basic background check, and their signature of relevant codes of conduct and other employment policies before they begin working.	The vetting process for most AI applications that are not deployed in high-risk or sensitive domains can be limited to a limited risk/impact assessment, a review of the input, training, and output data; some limited testing, and the implementation of relevant oversight processes at an organizational level.
For most employees, the risk of significant harm that they can cause is largely theoretical. To address this risk, however, after they are hired, these employees are given additional training, monitored, tested, measured against performance metrics, supervised, and reviewed on a regular basis.	For most AI applications, the risk of significant harm that they can cause is largely theoretical. To address this risk, however, after they are deployed, these applications are monitored, tested, measured against performance metrics, supervised, and reviewed or retrained on a regular basis to ensure that they are performing as expected.
Most companies have a small number of employees who hold higher profile positions or who work in high-risk or sensitive domains, whose mistakes or malfeasance are likely to cause significant financial harm, reputational damage, or legal liability to the company or the public.	Most companies have a small number of AI applications that pose a higher risk of causing significant financial harm, reputational damage, or legal liability to the company or the public.
The vetting process for these high-risk jobs is therefore more involved, and often includes a detailed submission from the candidate, a more thorough background check, informal and formal reference checks, and multiple rounds of substantive interviews, which can take several months.	The vetting process for these high-risk AI applications is therefore more involved, and often includes a detailed impact assessment, robust data review and testing, approval by a cross-functional internal committee, and the implementation of significant mitigation and oversight measures to ensure that these models do not operate in an unforeseen or incorrect manner.

Aside from the misallocation of resources, if large companies were required to take a rights-based, one-size-fits-all approach to employee vetting, that would benefit the market participants with the most resources. Many smaller companies or recent entrants would not have the resources to subject all of their new employees to onerous

vetting requirements, which could lead to greater market consolidation and reduced incentives to grow, further benefiting the largest companies.

By contrast, by taking a risk-based approach to employee vetting, companies can effectively balance the actual downside risk posed by the employees against the upside of having enough people on staff to carry out the essential operations of the company in a timely manner and at a reasonable cost. Similarly, a risk-based approach to AI vetting will allow companies to focus their resources on the automated applications that are most likely to cause harm, and better balance the downside risks of AI deployment against the upsides in efficiencies, product delivery, and improvements over human decision-making.

The authors would like to thank Debevoise Law Clerk Jackie Dorward for her contribution to this blog post.

To subscribe to the Data Blog, please click here.

The Debevoise Artificial Intelligence Regulatory Tracker (DART) is now available for clients to help them quickly assess and comply with their current and anticipated AI-related legal obligations, including municipal, state, federal, and international requirements.

* * *

Please do not hesitate to contact us with any questions.

NEW YORK

Avi Gesser
agesser@debevoise.com

Tricia Bozyk Sherno
tbsherno@debevoise.com

Frank Colleluori
facolleluori@debevoise.com

WASHINGTON, D.C.

Anna R. Gressel
argressel@debevoise.com

Jehan A. Patterson
jpatterson@debevoise.com