Precision Regulation for Artificial Intelligence

By Ryan Hagemann, IBM Policy Lab co-Director (Washington, DC) & Jean-Marc Leclerc, IBM Policy Lab co-Director (Brussels)

Among companies building and deploying artificial intelligence, and the consumers making use of this technology, trust is of paramount importance. Companies want the comfort of knowing how their AI systems are making determinations, and that they are in compliance with any relevant regulations, and consumers want to know when the technology is being used and how (or whether) it will impact their lives.

62% of Americans and 70% Europeans prefer a precision regulation approach for technology, with less than 10% in either region supporting broad regulation of tech. 85% of Europeans and 81% of Americans support consumer data protection in some form, and 70% of Europeans and 60% of Americans support AI regulation.

As outlined in our Principles for Trust and Transparency, IBM has long argued that AI systems need to be transparent and explainable. That's one reason why we supported the EU and the OECD AI Principles, and in particular the focus on transparency and trustworthiness in both.

Principles are admirable and can help communicate a company's commitments to citizens and consumers. But it's past time to move from principles to policy. Requiring disclosure — as appropriate based on use-case and end-user — should be the default expectation for many companies creating, distributing, or commercializing AI systems. In an earlier Policy Lab essay, we articulated a disclosure requirement for law enforcement use-cases of facial recognition technology. Something similar should be required of AI more generally in order to provide the public with appropriate assurances that they are being treated fairly and equitably by AI-based determinations in sensitive use-cases.

That is why today we are calling for precision regulation of AI. We support targeted policies that would increase the responsibilities for companies to develop and operate trustworthy AI. Given the ubiquity of AI — it touches all of us in our daily lives and work — there will be no one-size-fits-all rules that can properly accommodate the many unique characteristics of every industry making use of this technology and its impact on individuals. But we can define an appropriate risk-based AI governance policy framework based on three pillars:

- **Accountability** proportionate to the risk profile of the application and the role of the entity providing, developing, or operating an AI system to control and mitigate unintended or harmful outcomes for consumers.
- **Transparency** in where the technology is deployed, how it is used, and why it provides certain determinations.
- **Fairness and security** validated by testing for bias before AI is deployed and re-tested as appropriate throughout its use, especially in automated determinations and high-risk applications.

Wisely, the OECD AI Principles suggest a solid accountability bedrock for this framework, arguing that "[g]overnments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems." This implicit recognition

of the fundamental difference in accountability between stages of AI development can help appropriately assign responsibility for providing transparency and ensuring fairness and security, based on who has better control over the protection of privacy, civil liberties, and harm-prevention activities in a given context.

In the lifecycle of AI capabilities in the marketplace, organizations may contribute research, the creation of tooling, and APIs; in later stages of operation, organizations will train, manage, and control, operate, or own the AI models that are put to real-world commercial use. These different functions may allow for a distinction between "providers" and "owners," with expectations of responsibilities based on how an organization's role falls into one or both categories.

Differentiating accountability can help to better mitigate potential harm by directing resources and oversight to specific applications of AI based on the severity and likelihood of potential harms arising from the end-use and user of such systems. Risk-based regulatory approaches like this — which also allow for more manageable and incremental changes to existing rules — are ideal means to protect consumers, build public trust in AI, and provide innovators with needed flexibility and adaptability.

Building from these pillars, we propose a precision regulation framework that incorporates 5 policy imperatives for companies, based on whether they are a provider or owner (or both) of an AI system. These policies would vary in robustness according to the level of risk presented by a particular AI system, which would be determined by conducting an initial risk assessment based on potential for harm associated with the intended use, the level of automation (and human involvement), and whether an end-user is substantially reliant on the AI system based on end-user and use-case.

1. Designate a lead AI ethics official. To ensure compliance with these expectations, providers and owners should designate a person responsible for trustworthy AI, such as a lead AI ethics official. This person would be accountable for internal guidance and compliance mechanisms, such as an AI Ethics Board, that oversee risk assessments and harm mitigation strategies. As the complexity and potential impact of AI systems increases, so too must the accountability embraced by different organizations providing various functions in the AI lifecycle. A market environment that prioritizes the adoption of lead AI ethics officials, or other designated individuals, to oversee and manage this increasing complexity could help to mitigate risks and improve public acceptance and trust of these systems, while also driving firms' commitment to the responsible development, deployment, and overall stewardship of this important technology.

2. Different rules for different risks. All entities providing or owning an AI system should conduct an initial high-level assessment of the technology's potential for harm. As noted previously, such assessments should be based on the intended use-case application(s), end-user(s), how reliant the end-user would be on the technology, and the level of automation. Once initial risk is determined, a more in-depth and detailed assessment should be undertaken for higher-risk applications. In certain low-risk situations, a more cursory appraisal would likely suffice. For those high-risk use-cases, the assessment processes should be documented in detail, be auditable, and retained for a minimum period of time.

3. Don't hide your AI. Transparency breeds trust; and the best way to promote transparency is through disclosure. Unlike other transparency proposals, this approach does not entail companies revealing source code or other forms of trade secrets or IP. Instead it focuses on making the purpose of an AI system clear to consumers and businesses. Such disclosures, like other policy imperatives here, should be reasonably linked to the potential risk and harm to individuals. As such, low-risk and benign applications of AI may not require the type of disclosure that higher-risk use-cases might require.

4. Explain your AI. Any AI system on the market that is making determinations or recommendations with potentially significant implications for individuals should be able to explain and contextualize how and why it arrived at a particular conclusion. To achieve that, it is necessary for organizations to maintain audit trails surrounding their input and training data. Owners and operators of these systems should also make available — as appropriate and in a context that the relevant end-user

can understand — documentation that detail essential information for consumers to be aware of, such as confidence measures, levels of procedural regularity, and error analysis.

74% of American and 85% of EU respondents are in agreement that artificial intelligence systems should be transparent and explainable, and strong pluralities in both countries believe that disclosure should be required for companies creating or distributing AI systems. Nearly 3 in 4 Europeans and two-thirds of Americans support regulations such as conducting risk assessments, doing pre-deployment testing for bias and fairness, and reporting to consumers and businesses that an AI system is being used in decision making.

5. Test your AI for bias. All organizations in the AI developmental lifecycle have some level of shared responsibility in ensuring the AI systems they design and deploy are fair and secure. This requires testing for fairness, bias, robustness and security, and taking remedial actions as needed, both before sale or deployment and after it is operationalized. Owners should also be responsible for ensuring use of their AI systems is aligned with anti-discrimination laws, as well as statutes addressing safety, privacy, financial disclosure, consumer protection, employment, and other sensitive contexts. For many use-cases, owners should continually monitor, or retest, the AI models after the product is released to identify and mitigate against any machine-learning resulting in unintended outcomes. Policies should create an environment that incentivizes both providers and owners to do such testing well. This can be done without creating new and potentially cumbersome AI-specific regulatory requirements, but rather by adhering to a set of agreed-upon definitions, best practices, and global standards.

To achieve this, governments should:

- **Designate, or recognize, existing effective co-regulatory mechanisms (e.g. CENELEC in Europe or NIST in the U.S.) to convene stakeholders and identify, accelerate, and promote efforts to create definitions, benchmarks, frameworks and standards for AI systems.** Ideally, standards that are globally recognized would help create consistency and certainty for consumers, communicating to end-users that the AI is trustworthy;

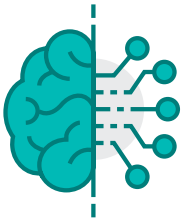
- **Support the financing and creation of AI testbeds with a diverse array of multi-disciplinary stakeholders working together in controlled environments.** In particular, minority-serving organizations and impacted communities should be supported in their efforts to engage with academia, government, and industry. Working together, these stakeholders can accelerate the development and evaluation criteria of AI accuracy, fairness, explainability, robustness, transparency, ethics, privacy, and security; and

- **Incentivize providers and owners to voluntarily embrace globally recognized standards, certification, and validation regimes.** One such potential mechanism is by providing various levels of liability safe harbor protections, based on whether and how an organization adheres and certifies to globally recognized best practices and standards.

Finally, any action or practice prohibited by anti-discrimination laws should continue to be prohibited when it involves an automated decision-making system. Whether a decision is fully rendered by a human or a determination is assisted by an automated AI system, impermissibly biased or discriminatory outcomes should never be considered acceptable. But whereas correcting the bias of humans is a daunting and difficult task, in AI systems it may be a matter of addressing historical bias in some training data by testing for, and correcting, statistical failures in the model. While this will take time, AI offers us the promise of a world where bias and discrimination may one day fade away. With precision regulations helping to promote trustworthy AI, that future could be sooner than we think.

IBM **Policy Lab**

Bold Ideas for a Digital Society



Since day one, IBM has pushed the boundaries of technology to address the challenges of tomorrow. We've done this while earning our clients' trust to innovate responsibly and carefully stewarding their data. We'll continue to drive forward new technological advances with the values of accountability, transparency, and trust that our clients and government partners have relied on since 1911.

The world — and IBM — has changed a lot over the past century. We've seen the march of progress move humanity from an analog era to the digital age and explosive innovation in both bits and atoms contribute to a wave of disruptive change. At IBM, we're optimistic about what the future holds, and the crucial role technological advancement will play in driving economic growth and societal well-being. Already, cloud computing has changed how work gets done and how connections are made, artificial intelligence has revolutionized our daily routines, and we can find information on practically anything at the touch of a button. Technology will fundamentally change society, bring us closer together, improve lives around the world and help us tackle some of our greatest challenges.

But no journey comes without challenges. We have already seen concerns materialize across emerging technologies on the implications of opaque AI systems making safety- and life-critical decisions; the growing pains of new digital platforms leading to the spread of illegal and harmful content online; and fears that a fully-automated future will displace more jobs than it creates. All of this comes amid a wave of global challenges to modern society, from the spread of protectionist impulses to the failure to address climate change.

But at IBM, we've seen how technological progress has improved the human condition over the past 100 years. We were optimistic about the future then, and we remain optimistic about the future to come. While there are challenges ahead, we believe there are clear and practical ways through them.

As businesses and governments break new ground and deploy technologies that are positively transforming our world, we want to work collaboratively to make sure public policy adapts to meet the challenges of tomorrow. That's why we've created the **IBM Policy Lab, a new forum that provides a vision with actionable recommendations to harness the benefits of innovation while ensuring trust in a world reshaped by data.**

Led by co-directors Ryan Hagemann and Jean-Marc Leclerc — two long-standing experts in tech and public policy — IBM Policy Lab convenes leading thinkers in public policy, academia, and technology to develop the concrete, common-sense policy ideas leveraging technology to tackle some of the most pressing issues facing our world. Our approach is grounded in the belief that tech can continue to disrupt and improve civil society while protecting individual privacy.

What We Do

- **Develop Industry-Leading Policy Positions** that don't just respond to the spot issues of today but look forward to the opportunities of tomorrow and the ways public-private cooperation can pave the way for an even brighter future. With the full benefits of artificial intelligence, blockchain, quantum computing and more still untapped, we'll put forward bold visions for public policy that harnesses innovation.

- **Collaborate with Global Thinkers, Stakeholders and Leaders** to collect input and share perspective from the diverse voices that must inform public policy.

- **Produce Data-Driven Studies and Research** to guide policymaking with specific, common-sense recommendations, and help industry leaders make critical decisions on policies impacting our future.

As technological innovation races ahead, our mission to raise the bar for a trustworthy digital future could not be more urgent. IBM Policy Lab is committed to developing and advocating the right policies that meet the demands of the moment and harness the power of technology as a force for good in the world.

How We're Different

- While some traffic in grandiose policy recommendations that stand little chance of becoming reality, IBM has always believed that big challenges require practical solutions. That's precisely what IBM Policy Lab has been chartered to create.

- Our policy recommendations will be concrete. Specific. Actionable. We will have big ideas, but they will be ideas that policymakers can implement on day one.

- We will also convene government, industry and civil society experts to think big about upcoming challenges and make space for collaborative solutions.

- Serious times call for serious solutions, and that's precisely what leaders in government, business and civil society can expect from IBM Policy Lab.

Jean-Marc Leclerc joined IBM's Government and Regulatory Affairs team in 2015, where he leads the EU Affairs team. Jean-Marc is the Chair of the EMEA Policy Committee at the Business Software Association (BSA), and he is a Vice-Chair of the Digital Economy Committee at the American Chamber of Commerce to the EU. Before joining IBM, he was a Policy Director at Digitaleurope 2013-15. He has also managed an association representing the music industry in Brussels 2006-13. Jean-Marc is a graduate from the universities of Paris III, Sciences Po, the Catholic Institute of Paris, and the College of Europe in Bruges.

Ryan Hagemann is the Co-Director of the IBM Policy Lab and a Technology Policy Executive on IBM's Government and Regulatory Affairs team. He was previously a senior policy fellow at the International Center for Law & Economics. Before joining ICLE, he was a senior fellow at the Niskanen Center, where he also served as the senior director for policy and director of technology policy. His policy expertise focuses on regulatory governance of emerging technologies, as well as a broader research portfolio that includes genetic modification and regenerative medicine, bioengineering and healthcare IT, artificial intelligence, autonomous vehicles, commercial drones, the Internet of Things, and other issues at the intersection of technology, regulation, and the digital economy. His work on "soft law" governance systems, autonomous vehicles, and commercial drones has been featured in numerous academic journals, and his research and comments have been cited by The New York Times, MIT Technology Review, and The Atlantic, among other outlets. He has been published in The Wall Street Journal, Wired, National Review, The Washington Examiner, U.S. News & World Report, The Hill, and elsewhere. Ryan graduated from Boston University with a B.A. in international relations, foreign policy, and security studies and holds a Master of Public Policy in science and technology policy from George Mason University.