How Policymakers Can Foster Algorithmic Accountability

Instead of overregulating, governments should adopt a policy framework based on the principle of “algorithmic accountability” to promote innovation while holding human operators responsible for harmful outcomes.

Increased automation with algorithms, particularly through the use of artificial intelligence (AI), offers opportunities for the public and private sectors to complete increasingly complex tasks with a level of productivity and effectiveness far beyond that of humans, generating substantial social and economic benefits in the process. However, many believe an increased use of algorithms will lead to a host of harms, including exacerbating existing biases and inequalities, and have therefore called for new public policies, such as establishing an independent commission to regulate algorithms or requiring companies to explain publicly how their algorithms make decisions. Unfortunately, all of these proposals would lead to less AI use, thereby hindering social and economic progress.

Policymakers should reject these proposals and instead support algorithmic decision-making by promoting policies that ensure its robust development and widespread adoption. Like any new technology, there are strong incentives among both developers and adopters to improve algorithmic decision-making and ensure its applications do not contain flaws, such as bias, that reduce their effectiveness. Thus, rather than establish a master regulatory framework for all algorithms, policymakers should do what they have always done with regard to technology regulation: enact regulation only where it is required, targeting specific harms in particular application areas through dedicated regulatory bodies that are already charged with oversight of that particular sector. To accomplish this, regulators should pursue algorithmic accountability—the principle that an algorithmic system should employ a variety of controls to ensure the operator (i.e., the party responsible for deploying the algorithm) can verify it acts in accordance with its intentions, as well as identify and rectify harmful outcomes. Adopting this framework would both
promote the vast benefits of algorithmic decision-making and minimize harmful outcomes, while also ensuring laws that apply to human decisions can be effectively applied to algorithmic decisions.