Old Laws for New Machines

Biagio Bossone

SCIENCE FICTION WRITER ISAAC ASIMOV'S ROBOT PRINCIPLES OFFER AN ETHICS GUIDE FOR AI IN FINANCE

It's 7:45 a.m. in a central bank's operations room. Screens glow with color-coded maps of the financial system. Overnight data from thousands of banks, brokers, and payment processors have been analyzed by an AI model that monitors liquidity conditions. A small red dot flashes beside a bank whose payment activity has suddenly dropped. The analysts glance up. The machine has detected something unusual—perhaps a transient error, or perhaps the first signal of stress.

Moments like this are now routine in financial centers around the world. AI is now the silent partner of global finance—embedded in credit scoring, fraud detection, algorithmic trading, and even monetary analysis. It sits in the background of every transaction, scanning vast oceans of data for patterns invisible to the human eye.

The promise is dazzling: faster insights, earlier warnings, and more efficient markets. Yet the perils are equally large. AI's rapid and coordinated responses can produce system-wide dynamics that have no human equivalent, including feedback loops that no one controls. The same systems that make finance more efficient can also make it more fragile.

How can we ensure that this new intelligence serves the system rather than undermines it? Curiously, the answer may lie in an old idea from science fiction.

Asimov's laws

In the 1940s, Isaac Asimov imagined intelligent robots governed by three simple principles: They should avoid harming humans, obey legitimate commands unless doing so causes harm, and protect their own functioning while respecting those two higher duties. He later added the "Zeroth Law," requiring robots to safeguard humanity as a whole.

Finance now faces a similar challenge. Machine learning models increasingly influence who receives credit, how portfolios adjust, and how supervisors detect risks. They affect how information flows and prices adjust across the economy.

Reinterpreted for financial oversight, Asimov's hierarchy becomes an ethic for AI in finance: It should first do no harm to financial stability or consumer integrity; it should obey institutional mandates without compromising fairness or safety; and it should preserve its own resilience without escaping accountability. Above all, it should serve the higher goal of maintaining the trust on which finance rests.

Yet any analogy to Asimov's laws requires cau-

Illustration by Matthieu Bourel

tion. AI systems have no moral compass. Their generative nature makes it easy to circumvent broadly accepted principles. Real-world failures show that hard-coded constraints are fragile.

Doing no harm

A large commercial bank never truly closes. Its systems work continuously across time zones, processing transactions, updating exposures, and scanning for anomalies that might signal fraud or stress. Yet errors happen. AI can misinterpret patterns, inherit historical biases, or reinforce market signals that accelerate volatility.

Asimov's injunction to do no harm has operational meaning: The obligation to prevent damage applies not only to consumers but to the stability of the system itself. That requires rigorous testing, sound data governance, and, crucially, explainability—because when an algorithm cannot justify its decision, neither can the institution that depends on it. Current AI models cannot explain their reasoning; they can generate post hoc narratives. This makes explainability an institutional—not a technical—requirement.

AI's macro-financial effects can be large. When algorithms react to new information at machine speed, price adjustments and liquidity shifts become sharper and more synchronized than in human-driven markets. Episodes such as the 2010 flash crash show how automated systems can impact markets long before policymakers can intervene.

Today's environment is even more complex. AI does not merely react faster; it reacts *differently*. Its internal logic not only accelerates the feedback loop between information and action, it *reshapes* it. AI-driven models influence credit pricing, asset allocation, and risk premiums in ways that reinforce each other. When they collectively adjust strategies—because they interpret signals similarly or their optimization routines converge—they can affect the transmission of economic shocks and the policy response of central banks.

Obedience and judgment

Obedience is built into AI. It optimizes whatever goal it is given—profit, accuracy, compliance. But blind obedience is dangerous in finance, where tradeoffs are moral as well as technical. A credit model designed to minimize default may respond by excluding entire high-risk groups, deepening inequality rather than managing it. A trading algorithm tasked with maximizing returns may exploit micro-price signals so efficiently that it destabilizes markets.

AI uses may be legitimate. However, they must remain tools, not decision-makers. They should inform judgment, not replace it. Monetary policy, financial supervision, and crisis management require human discretion. Models cannot drive these functions, especially if they are trained on past data and cannot predict new shocks. And because AI alters expectations, it may weaken the traditional signaling channels through which policy operates. Humans fail too; as the saying goes, policymakers are always best prepared for the previous crisis. The lesson is not that machines fail or people do, but that both must learn together—each aware of the other's blind spots.

Despite all the code and computation, finance remains a human enterprise. Judgment, empathy, and responsibility cannot be automated. Central banks, regulators, and financial institutions must therefore cultivate not only data literacy but ethical literacy. AI can augment oversight, but governance remains a moral task.

Resilience and accountability

The law of self-preservation translates into resilience. AI systems must function reliably under stress, and institutions must be accountable for their algorithms.

Technical resilience means redundancy, monitoring, and testing under extreme scenarios. Institutional resilience means openness: Regulators should be able to audit AI decisions, even when proprietary code is involved. This requires the skills and tools to validate and challenge companies' AI models.

The Bank for International Settlements (BIS) Innovation Hub has developed prototype tools to help supervisors analyze large datasets and detect anomalies. These efforts are promising, and their underlying principle is simple: If an algorithm affects financial stability, it should be open to supervisory scrutiny.

Secrecy breeds fragility. When models are black boxes, errors accumulate unseen. The global financial crisis of 2008 is a reminder that complexity without transparency leads to collapse. AI raises the same warning in digital form.

Accountability also extends to governance. Financial institutions should have AI risk officers, parallel to chief risk or compliance officers, ensuring that algorithms are explainable and auditable. Regulators, in turn, must develop AI literacy to interpret and challenge the outputs they receive. The goal is not to slow innovation but to make it safe, fair, and comprehensible.

A higher law

The Zeroth Law—no harm to humanity—finds its real-world equivalent in the preservation of trust. Trust is the invisible infrastructure of finance.

If AI undermines that trust—by being biased, unstable, or unaccountable—it threatens the foundation of the system. But when aligned with the public interest, AI can enhance trust. It can detect fraud faster, make supervision more proactive, and extend financial access to those long excluded.

AI's higher law, then, is to serve the social contract of money—to reinforce confidence, fairness, and stability. Every institution that uses AI should be judged by whether it strengthens or weakens that contract.

AI's impact is particularly visible in emerging markets, where digital finance is evolving rapidly and data scarcity has long constrained access to credit and public services. Rather than replacing existing digital tools, AI magnifies what these systems can do by extracting patterns from large, unstructured datasets that traditional models cannot interpret.

In Kenya, for example, mobile-money ecosystems such as M-Pesa generate rich transaction footprints increasingly analyzed by AI-based scoring models. The behavioral patterns and cash-flow regularities that emerge allow lenders to assess risk for borrowers with no formal credit record. This has expanded credit access for small entrepreneurs and previously unbanked populations. In India, digital identity systems and real-time platforms are paired with machine learning tools that aim to better target government transfers and expand microloan access.

But AI can also entrench exclusion. Data poverty—limited or biased data—means entire communities remain invisible to algorithms. If women, rural populations, or informal workers are underrepresented in datasets, they will be underrepresented in outcomes.

International organizations are stepping in. The IMF and the World Bank are increasingly integrating digital finance and AI governance issues into their capacity-building programs.

Global coordination

AI moves faster than regulation and across borders faster than money. Yet there is a lot policymakers can do within existing mandates and legal frameworks while the law catches up. Cross-border coordination is essential to prevent fragmentation and to ensure that a global regulatory approach to AI in finance emerges, buiding on evolving international best practices.

The Financial Stability Board, the BIS, and the IMF are exploring frameworks for responsible AI. A global set of principles—analogous to the Basel Core Principles for banking—could ensure consistency while allowing flexibility. Such a framework would emphasize fairness, explainability, accountability, and proportionality.

The IMF, through its surveillance and technical assistance, could help countries identify AI-related financial risks, share best practices, and avoid a digital divide in supervision. To this end, it should attract skilled professionals from the research and fintech communities. The BIS could host a repository of supervisory algorithms, allowing regulators to collaborate on open-source models.

The World Bank and regional development institutions can complement these efforts by building AI capacity and digital infrastructure in emerging markets. Through their technical assistance, policy dialogue, and financing instruments, they can help countries design responsible AI frameworks for financial inclusion, strengthen data governance, and integrate ethical AI standards into digital finance ecosystems.

Together, these institutions can ensure that the benefits of AI extend beyond advanced economies. The goal is digital multilateralism: ensuring that AI serves all economies. No country can manage these dynamics alone.

AI financial laws

Asimov's laws distill moral complexity into clear priorities: Protect people, obey within limits, preserve responsibly, and serve humanity. In an age when technology outpaces law, such simplicity is priceless.

The choice is not between progress and prudence, but between intelligent governance and blind automation, remembering that even as machines learn, humans remain responsible. The future of finance will increasingly be written in code. Yet the principles behind it must remain human. A system governed by safety before obedience, transparency before secrecy, trust before profit would not eliminate risk, but it would make it manageable and moral.

If central banks, regulators, and financial institutions embrace these principles, AI could become a stabilizing force rather than a source of fragility. It could extend financial inclusion, enhance oversight, and strengthen the legitimacy of monetary systems.

These same principles must be reinforced through international cooperation—ensuring that AI supports a financial system that is not only safer and fairer, but also more coherent at the global level. The challenge, for supervision and policy alike, is to ensure that as intelligence becomes artificial, judgment remains real.

The machines are learning. So must we. F&D

BIAGIO BOSSONE *is a senior advisor to international organizations, government agencies, and financial institutions.*