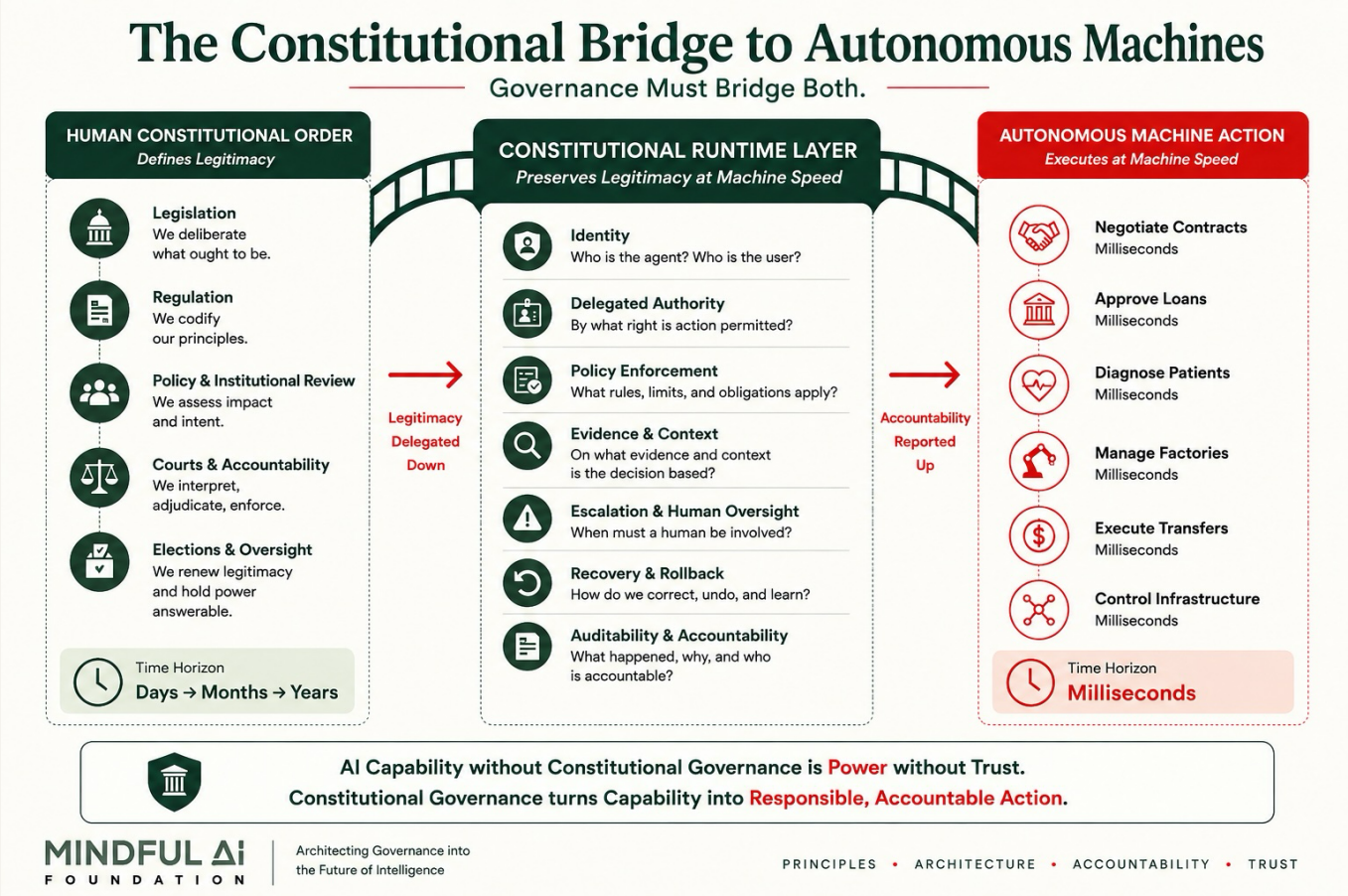
When Machines Outrun Institutions

Why Autonomous Machines Need Constitutions

Max Michaels, Mindful AI Foundation

The Stanford Critical AI Manyfesto is a wakeup call to address one of the defining questions of our age: ***How should democratic societies govern artificial intelligence?*** It begins from a conviction we readily share. AI is a new expression of power, and every expression of power eventually confronts the oldest constitutional questions: ***Who decides? By what authority? For whose benefit? Who remains accountable?***

We begin where that conversation naturally leads. The Manyfesto invites us to think constitutionally about AI in society. We ask the complementary question: ***What would it mean for autonomous machines themselves to operate under a constitutional order?***



Yet constitutions have always existed in two worlds. There is the world of *ideals*, where principles are articulated and rights are declared. And there is the world of *institutions*, where those principles become procedures, constraints, records, and habits that shape everyday action. The distance between the two is often where history unfolds. AI now occupies that same distance.

Policies, regulations, and democratic institutions tell us what *ought* to happen. But an autonomous machine lives in another domain, one measured in milliseconds rather than legislative sessions. Before it acts, it must somehow answer questions that constitutions have always asked of those who wield authority: *By what right? Under whose authority? On what evidence? Who remains answerable for the consequence?* These are no longer merely legal or philosophical questions. Or even technical questions on clock speed. They have become architectural ones.

Every transformative technology eventually discovered that capability alone was insufficient. Railroads required signaling systems before they could become national infrastructure. Aviation required air traffic control before flight became routine. Electricity required the grid before it became civilization's invisible utility. These systems did not restrain technology. They made trust possible. They transformed isolated acts of capability into enduring public institutions.

Computing now stands at a similar threshold. ***It was not architected to govern***. Its architecture was designed to compute, to optimize, and to communicate. As autonomous systems begin to negotiate contracts, approve loans, diagnose patients, manage factories, and increasingly act on behalf of people and institutions, governance can no longer remain something imposed from outside through policy documents, audits, or after-the-fact review. It must become part of the architecture.

This is where the work of the Mindful AI Foundation seeks to extend the conversation. We share the Manyfesto's constitutional ambition, but propose that constitutions must also exist within autonomous systems themselves. We describe this as **Three Constitutions.** The first is the ***constitution of society***, which expresses the values, rights, and institutions that technology exists to serve. The second is the ***constitution of cognition***, which governs how intelligent systems reason, align, and interact. The third, still largely absent from contemporary AI, is the ***constitution of action***: the architecture that determines what a machine may do, at what speed, under what authority, with what evidence, and with what capacity for explanation, recovery, and accountability.

The defining achievement of the AI age will not be that machines learned to think. It will be that societies learned how to govern them. The age of AI will be no different. The Manyfesto provides a *constitutional vision* for governing AI in society. Mindful AI proposes the *constitutional architecture* that enables autonomous systems to honor those same democratic principles at machine speed. ***Together they suggest that the future of AI will be shaped not only by intelligence, but by the constitutions that ensure intelligence remains answerable to humanity.***

MINDFUL AI
FOUNDATION