

How Many AI Regulators Has the Pope?

di Alex Friedman

Artificial intelligence may be humanity's greatest opportunity and risk. Where moral leaders have too little influence and states cannot align, actors with system-wide exposure and cross-border reach—in other words, large asset owners—become the only entities capable of imposing needed constraints at scale.

PRINCETON—Pope Leo XIV has now weighed in on the most pressing issue of our time: the advent of artificial intelligence and the recognition that humanity's technological capability is advancing faster than the institutions designed to govern it. In his first encyclical, [Magnifica Humanitas](#), Leo warns that AI without guardrails risks subjugating humankind; with safety measures, it can ensure that humanity thrives. But as morally powerful as Leo's words are, it seems doubtful that the Catholic Church will set the global technology agenda.

The lever of government looks unpromising as well. AI is the embodiment of a transnational issue that demands a multilateral governance framework. But this is unlikely at a time when the United States is abandoning global leadership, the US and China are competing to dominate frontier technologies, India is pursuing strategic autonomy, and most of the developing world is focused on growth, not constraint. The political conditions for an “AI Bretton Woods” simply do not exist.

There is, however, one powerful lever left to pull, and it does not depend on the cooperation of rivals or the bandwidth of struggling governments. The lever of capital, and the institutions that control it at scale, may be our era's most important and least understood governance actors.

Sovereign [wealth funds](#), public [pension systems](#), and large endowments collectively manage more than \$75 trillion, or around 70% of world GDP. They have a fiduciary

obligation to seek long-term returns, which requires a stable economic system from which those returns can be generated.

For individual technology companies, increased risk can be rational if it yields a competitive advantage. For nation-states, accelerating technological capability at all costs can be rational if it shifts the balance of power (think of America's Manhattan Project to develop the atomic bomb). But for the investment institutions that own a majority share of the world economy, a calamitous failure anywhere leads to loss everywhere—a direct threat to their mandate.

If states cannot coordinate effectively, and firms cannot self-regulate under competitive pressure, what is needed is a global compact whereby a critical mass of major asset owners treats existential risk as a binding category in investment governance. Accordingly, access to their capital would be contingent on adherence to defined oversight standards in AI. These requirements would apply not only to their portfolio investments but to the asset managers they use and through which their capital is deployed.

This idea has precedent. The Sullivan Principles, articulated by Reverend Leon Sullivan in 1977 and embraced by major institutional investors, used capital allocation to influence labor and civil-rights standards on US firms operating in apartheid South Africa. (Sullivan renounced the Principles in 1987 and called for a total boycott and corporate divestment, though he later worked with UN Secretary-General [Kofi A. Annan](#) on an expanded [Global Sullivan Principles](#) to advance corporate social responsibility around the world.) The Carbon Disclosure Project, today a reporting standard for thousands of companies, originated from a coalition of institutional investors looking for comparable climate data. In both cases, enough investors joined to drive market-wide conditions.

The capital compact would begin with a founding group of sovereign funds, public pensions, and foundation endowments. While these institutions differ politically, they all face a unique combination of long-duration liabilities, system-wide exposure, and governance frameworks that have already begun considering factors beyond pure return maximization.

Operationally, the compact would focus on standardized disclosure, benchmarking, and enforcement. The members would agree on a core set of metrics to be tracked (such as independent safety audits and human oversight of AI systems' roles in critical decisions) and establish reporting requirements for every asset manager receiving their capital. Compact members would then aggregate that data across their full portfolio and submit results to a neutral, independent nonprofit that they establish.

The nonprofit would anonymize and aggregate submissions, publishing benchmarks that allow each member to assess its exposure relative to peers. Members would publicly disclose their own benchmark results and how their exposures have changed over time. Above certain thresholds, disclosure would require independent verification. Firms in high-risk domains would undergo independent safety audits and need to demonstrate compliance with established risk-management protocols as a condition of capital access.

The data infrastructure to support this also has a precedent. In 1971, a group of US insurance companies created the non-profit Insurance Services Office (ISO) by combining state, regional, and national ratings bureaus for property and casualty insurance. This established a neutral intermediary for anonymized, contributed-loss data that later became the for-profit company Verisk Analytics.

The premise behind ISO was that no single insurer had sufficient data to price complex, correlated risks accurately, but that the industry collectively did. ISO enabled competitors to benchmark their own exposure against industry-wide patterns and identify systemic risks that no individual vantage point could discern. Governance rested on impartial aggregation by a mutually established non-profit, shared standards, and peer benchmarking, a model the capital compact would replicate for existential technological risk.

This proposal will inevitably invite comparison to Environmental, Social, and Governance investing, which deserves credit for the insight that investment decisions have externalities and that firms have a financial interest in managing them. But ESG tackled too broad a range of objectives and suffered from insufficient standardization, extraneous metrics, and the vulnerability of being perceived as values-driven advocacy cloaked in fiduciary language—"woke" capitalism to critics.

But the proposed capital compact is not a values initiative. It asks institutions to manage a specific, quantifiable category of financial risk that threatens the long-term value of every asset they hold. An institution that refuses to measure its exposure to the technologies most likely to undermine its core objectives is not avoiding ideology; it is avoiding its job.

Unlike failed climate-focused investor coalitions such as the Net Zero Asset Owner Alliance and the Glasgow Financial Alliance for Net Zero, where pledges appeared to be at the expense of financial returns, the capital compact's fiduciary logic is grounded in risk management. Moreover, there is no antitrust risk: the compact's structure of standardized disclosure and benchmarking through a neutral nonprofit (not coordinated portfolio decisions among competitors) is analogous to the ISO model, not a cartel.

AI may be humanity's greatest opportunity and risk. Where states cannot align, actors with system-wide exposure and cross-border reach become the only entities capable of imposing needed constraints at scale. In today's global economy, they are large, diversified asset owners. It falls to them to do what religion and government cannot.