

Programmable Incentives: Aligning Behavior in AI Strategy

S. Dutta and Y. Doz

*Former Founding Dean, SC Johnson College of Business, Cornell University, USA
Solvay Chaired Professor of Technology Innovation Emeritus, INSEAD, France*

An agentic supply-chain system detects a disruption. It sees rising demand in one region, delayed shipments from a supplier, tightening logistics capacity, and early signs of inventory pressure. It reasons through alternatives and recommends a response: shift inventory from lower-priority markets, ask suppliers for updated production data, reroute shipments, and temporarily reward partners that can add capacity.

The system may know what action to initiate. It may even act on part of the response. But the outcome still depends on behavior. Suppliers must share accurate data. Logistics partners must reprioritize capacity. Store managers must accept substitutions. Customer-facing teams must communicate changes. Employees must trust the system enough to act. External partners must see a reason to cooperate. Increasingly, some of these decisions will be made not by people directly, but by AI agents acting on their behalf.

This is the alignment problem at the heart of AI strategy. As firms deploy agentic systems, strategy is no longer only about making better decisions. It is about aligning behavior across thousands of human, organizational, and machine agents interacting continuously, often in real time and at scale.

This points to a distinction that sits at the heart of AI strategy. In the age of agentic AI, competitive advantage depends on two capabilities that are related but distinct. The first is *intelligence generation*: building systems that can perceive, predict, and decide — the data, models, decisions, workflows, and interfaces that make up the AI Intelligence Stack. The second is *behavioural orchestration*: ensuring that all the actors around those systems — employees, partners, suppliers, customers, developers, and AI agents — actually behave in ways that turn that intelligence into outcomes at scale. Most of the current conversation about AI strategy has focused on the first. This paper argues that the second is equally important, and that programmable incentives are the mechanism through which behavioural orchestration is achieved.

Traditional incentives were not designed for this world. Bonuses, contracts, service-level agreements, loyalty programs, and platform rankings are often too slow, too coarse, and too static. They operate periodically, while agentic systems operate continuously. They reward broad outcomes, while AI systems can detect granular contributions.

A new capability is becoming necessary: programmable incentives. Programmable incentives are adaptive, software-based mechanisms that dynamically reward, recognize, prioritize, penalize, or

grant access based on real-time behavior and contribution. They may use money, tokens, credits, reputation, ranking, access rights, governance rights, preferred status, or other digital units of value.

If programmable strategy embeds strategic intent into systems, programmable incentives embed behavioral alignment into systems.

Why Intelligence Alone Is Not Enough

The first wave of AI strategy focused on intelligence: better data, better models, better predictions, better automation. That was necessary, but incomplete.

An AI system may identify the optimal action, but the organization still depends on actors whose incentives may point elsewhere. A model may know which supplier behavior improves resilience, but suppliers may still withhold data. A platform may know which content improves long-term community health, but creators may still chase short-term visibility. A bank may know which employees improve risk signals, but internal rewards may still prioritize speed or volume. This gap between what the system can reason about and what participants actually do becomes more consequential as agentic AI proliferates.

In earlier organizations, alignment was challenging but relatively bounded. Managers aligned employees through hierarchy, compensation, culture, and process. Firms aligned partners through contracts, pricing, and relationship management. Platforms aligned users and contributors through rules, rankings, and reputation.

In agentic organizations, alignment becomes harder. Decisions are more distributed. Interactions happen faster. Participants include employees, suppliers, customers, creators, developers, communities, devices, and AI agents acting on behalf of people or firms. Behavior must be shaped continuously, at the level of granular actions.

The strategic question therefore changes. It is no longer enough to ask: *Can our AI system identify or initiate the right action?* Leaders must also ask: *Will the relevant human and agentic actors behave in ways that make the desired outcome happen, continuously and at scale?* This is where programmable incentives matter.

Reframed at the strategic level: *behavioural orchestration* is the complement to *intelligence generation*. A firm that generates intelligence without behavioural orchestration will find its AI systems identifying optimal actions that the organisation cannot execute — because the employees, partners, and agents on whom the outcome depends are not aligned. A firm that invests in behavioural orchestration without intelligence will coordinate its actors efficiently around purposes that its data, models, and decisions have made irrelevant or obsolete. In the AI era, both are required. Programmable incentives are how behavioural orchestration is built.

From Static Incentives to Programmable Incentives

Traditional incentives are blunt instruments. Annual bonuses reward broad performance. Supplier contracts specify service levels but are hard to adjust quickly. Loyalty programs reward transactions but not always valuable behavior. Procurement agreements reward cost and delivery, but may fail to reward resilience, data sharing, or flexibility.

AI changes the economics of measurement. Firms can now observe behavior with much greater granularity: who shares useful data, who improves system resilience, who contributes secure code, who strengthens community trust, who reduces operational friction, and who helps an AI system learn.

Programmable incentives translate these observations into behavioral signals. They can reward a supplier for sharing accurate real-time inventory data. They can grant visibility to a developer whose contribution improves platform reliability. They can give status to users who strengthen community trust. They can allocate access to partners who improve ecosystem resilience. They can adjust incentives when conditions change.

In this context, token economics is especially important because it shows what becomes possible when incentives are encoded as digital units. Tokens can represent value, access, reputation, participation rights, or governance claims. They can be issued, tracked, adjusted, redeemed, or, in some cases, transferred.

But token economics is not the whole story. Not every programmable incentive needs to be a crypto token, tradable asset, or blockchain-based instrument. The broader strategic shift is that incentives can now become part of the operating system of the firm. Incentives are moving from static management tools to adaptive behavioral infrastructure.

The problem of aligning behavior is not new. Strategy and organization scholars have long studied how firms guide action through incentives, resource allocation, routines, and governance. Joseph Bower and Robert Burgelman showed how strategy is shaped by the processes through which initiatives are selected, funded, and reinforced. Herbert Simon, Richard Cyert, and James March emphasized that organizations are systems of boundedly rational actors responding to rules, routines, and incentives. Research on platforms, ecosystems, open innovation, and collective action extended the problem beyond the firm, showing that value increasingly depends on contributors who cannot be managed through hierarchy alone. What is new is that these incentive systems can now become programmable: measured at granular levels, adjusted in real time, and applied across human, organizational, and machine agents.

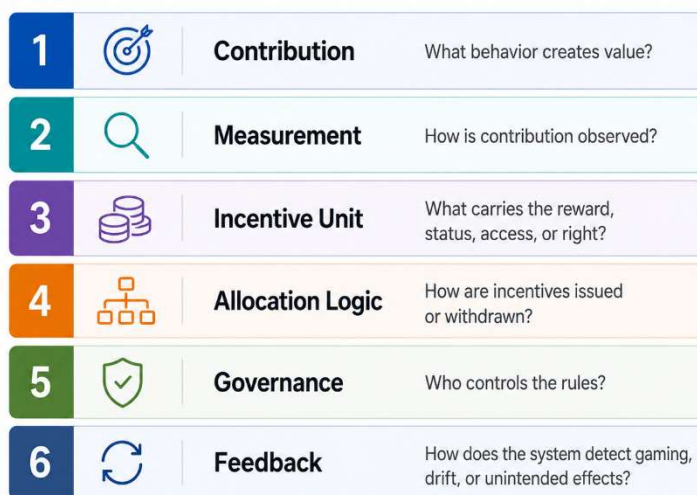
The Programmable Incentives Stack

Programmable incentives are not simply better rewards. They are a new architecture for behavioral alignment. To work, six layers must operate together.

The first is the **contribution layer**: what behavior creates value? In a supply chain, valuable behavior may include accurate data sharing, rapid response to disruption, resilience investment, lower emissions, or flexible capacity. In an open-source ecosystem, it may include secure code, documentation, bug fixing, mentoring, or community moderation. This layer is strategic because not all activity creates value. Clicks are easier to measure than trust. Code volume is easier to measure than maintainability. Delivery speed is easier to measure than resilience.

The second is the **measurement layer**: how is contribution observed and evaluated? This is where AI changes the incentive problem. Firms can now observe contribution at a level of granularity that was previously impossible. A supplier is not judged only by whether it delivered on time, but by whether it shared useful early-warning data. A developer is not judged only by code submitted, but by whether the code is secure, reusable, and maintained. The key is that contribution becomes visible.

The Programmable Incentives Stack



The third is the **incentive unit layer**: what carries the incentive? Sometimes the carrier is money. Sometimes it is a token, credit, badge, ranking, reputation score, access right, preferred status, revenue share, compute credit, governance vote, or priority privilege. This is where token economics becomes central. Tokens show how incentives can become digital units that are issued automatically, linked to contribution, tracked over time, and embedded into rules. They can represent not only payment, but also recognition, status, access, or influence.

Google Local Guides uses points, levels, badges, early access, and invitations to encourage users to contribute reviews, photos, and local knowledge. GitHub Sponsors allows communities to financially support open-source contributors whose work includes code, documentation, maintenance, mentoring, and project management. Neither is a crypto-token system. But both

illustrate token-like incentive units: digital markers that translate contribution into recognition, access, visibility, or reward.

The most powerful programmable incentives will not always be monetary. In many communities, status, belonging, identity, access, recognition, and voice may matter more than payment. Token economics is useful not because everything should be financialized, but because it shows how different forms of value — economic, reputational, social, or governance-related — can be encoded and allocated through digital systems.

The fourth is the **allocation logic layer**: how are incentives issued, adjusted, or withdrawn? This is where incentives become executable. Ride-sharing platforms use dynamic pricing when demand and supply are out of balance. Energy demand-response programs compensate customers for reducing or shifting electricity use when the grid is under stress. In both cases, a system detects imbalance and changes the incentive structure. Participants respond. The system learns. In agentic settings, the responding participant may be a person, a company, a device, or an AI agent acting within pre-set rules.

The fifth is the **governance layer**: who controls the rules? Programmable incentives shape behavior directly, so leaders must decide who can change the rules, how disputes are resolved, how gaming is detected, how unfairness is corrected, and when human judgment should override system logic. Governance also determines which incentives should not be programmed. Not all valuable behavior should be monetized, scored, or ranked.

The sixth is the **feedback layer**: how does the system learn and adapt? If rewards produce gaming, they must be redesigned. If incentives crowd out intrinsic motivation, they must be recalibrated. If agents optimize locally but damage system-wide performance, the rules must change.

Together, these six layers form the Programmable Incentives Stack. In the age of agentic AI, incentives are no longer merely compensation, contracts, or rewards. They become programmable infrastructure for aligning behavior across human, organizational, and machine agents.

Inside the Firm and Across Ecosystems

Programmable incentives matter first inside the organization. But in agentic organizations, “inside the firm” no longer means only employees, teams, and managers. It increasingly includes AI agents that recommend, route, escalate, prioritize, and sometimes act autonomously.

Consider sales. A company may deploy AI agents to improve forecasting, identify opportunities, recommend next-best actions, and prioritize accounts. If salespeople are rewarded only for short-term revenue, they may underinvest in accurate customer data or weak market signals. If the AI agent is optimized only for conversion probability, it may steer attention toward easy wins and away from strategically important relationships.

A programmable incentive approach would align both. It would recognize not only closed deals, but also data quality, forecast accuracy, useful customer insight, and long-term customer value. It could reward employees for improving the intelligence base while recalibrating AI agents when their recommendations produce short-term gains but weaken strategic priorities.

The greater novelty lies beyond the firm. Modern firms depend on contributors they do not employ and cannot command: suppliers, developers, creators, users, data providers, logistics partners, communities, complementors, and agentic systems acting on their behalf. The issue here is not simply that firms depend on ecosystems; it is that AI makes it possible to observe, evaluate, and influence ecosystem behavior at a new level of granularity.

Open-source software is one example. Platforms such as GitHub depend on coders, maintainers, reviewers, documenters, security experts, community moderators, and now AI coding agents. Human contributors respond to reputation, status, visibility, access, belonging, and influence. AI agents need objectives that prioritize secure, reusable, and maintainable code rather than simply producing more output.

Supply chains provide another example. AI can identify suppliers that share accurate data, respond quickly to disruptions, maintain buffer capacity, reduce emissions, or improve resilience. But supply chains increasingly involve both human decision-makers and autonomous planning agents operating across firms. Unless both are aligned, the system can fragment.

Energy systems show the same logic in physical infrastructure. As grids become more distributed, value depends on households, businesses, storage providers, renewable producers, electric vehicle owners, and automated energy-management systems. AI can forecast grid stress and identify when demand should shift. Programmable incentives can encourage both people and devices to reduce consumption, store energy, or release capacity during peak periods.

Token economics represents the frontier case. Helium, a decentralized wireless-network project, uses small devices called Hotspots to expand wireless coverage. Businesses and individuals host the devices and can earn HNT, Helium's cryptocurrency, when their Hotspots provide connectivity. In this model, token rewards are tied directly to network contribution.

Whether or not every firm adopts tokenized systems, the underlying logic matters: distributed infrastructure and ecosystem participation can be shaped when contributions are measured and incentives are programmable. In this sense, programmable incentives are a digital extension of an old ecosystem problem: how to encourage collective action when participants bring different contributions, face different costs, and cannot be directed through hierarchy.

The Risks of Over-Programming Behavior

Programmable incentives carry real risks. One is surveillance. If every action is measured, scored, and rewarded, participants may feel controlled rather than empowered. A second is narrowing motivation. People do not act only for financial rewards. They act for purpose,

identity, pride, reputation, belonging, and commitment. A third is gaming. If a platform rewards activity, users may generate noise. If it rewards engagement, creators may produce outrage. If it rewards data sharing, suppliers may provide quantity rather than quality. A fourth risk is legitimacy. Participants must believe the system is fair. If incentives are adjusted by opaque algorithms, firms may generate mistrust rather than alignment.

These risks do not argue against programmable incentives. They argue for better design. The most effective systems will not monetize every behavior. They will combine economic incentives with social, reputational, cultural, and governance mechanisms. They will recognize that alignment is not only transactional. It is also emotional, institutional, and strategic.

Strategy as Behavioral Infrastructure

Programmable incentives change the strategist's task. Strategy has long been concerned with resources, capabilities, positioning, and ecosystems. Those questions remain. But in AI-enabled organizations, leaders must also ask a more behavioral question: What behaviors are we encoding into the system, and how do we ensure they scale? This is not a minor extension of compensation or contracting. It is a new form of strategic design.

Agentic AI systems can reason, act, and adapt. But their effectiveness depends on the behavior of the broader socio-economic system around them. Employees must contribute useful data. Partners must respond to signals. Suppliers must share information. Developers must strengthen platforms. Users must participate constructively. Communities must trust the system. AI agents must act within objectives that reflect the firm's intent.

Programmable incentives are the mechanisms through which these behaviors can be shaped dynamically and at scale. That is why token economics matters. Not because every firm will issue tradable digital assets, but because token economics reveals a deeper possibility: incentives can be encoded as programmable units that assign value, access, reputation, or rights based on contribution.

This paper highlights the two complementary foundational competitive capabilities in the age of agentic AI: *intelligence generation* and *behavioural orchestration*. Neither is sufficient without the other. Most firms competing in the AI era have invested heavily in the first and underinvested in the second. Programmable incentives are the architecture through which the second is built.

Programmable strategy embeds intent into systems. Programmable incentives embed alignment into behavior. Together, they define how firms compete in the age of agentic AI.