

Which Decisions Should AI Make—and Which Must Stay Human?

AI is moving from analysis into decision-making and action. That changes the executive question. The issue is no longer simply what AI can do, but which decision rights the organization should transfer, under what conditions, and where human authority still creates more value than machine speed.

Executive Operations Insights | Saša Cibula

An operations team deploys an AI system that can monitor demand, inventory, labor availability, transportation constraints, equipment conditions, and customer priorities across a regional network. At first, the system produces recommendations. Planners review them and act. As confidence grows, leadership begins asking the obvious question: if the system is already identifying the best response in thousands of routine situations, why should a person approve every one?

The answer is not as simple as “keep a human in the loop.” Human review can reduce risk, but it also adds latency, cost, and capacity limits. A manager who approves hundreds of machine-generated decisions may eventually become a ceremonial checkpoint rather than a meaningful control. Conversely, removing people simply because the model is accurate can transfer authority faster than the organization has transferred governance, judgment, or the ability to recover when conditions change.

The executive challenge is therefore not to choose between human decisions and AI decisions. It is to allocate **different levels of authority to different types of decisions**. Some decisions should remain human. Some should be machine-recommended. Some should be machine-executed within clear boundaries. The difficult work is deciding which is which.

The Decision Is Not Binary

Human-automation research has treated this as a design problem for decades. Sheridan and Verplank’s classic work described a continuum ranging from fully manual control to systems that select and execute actions autonomously. Parasuraman, Sheridan, and Wickens later made the framework more useful by separating automation into four information-processing functions: information acquisition, information analysis, decision and action selection, and action implementation.

That distinction matters enormously in the AI era. An organization does not have to make a decision either “human” or “automated.” AI can operate at a high level of autonomy in one stage while human authority remains decisive in another. A system might continuously acquire data and analyze risk, recommend an action, and still require a human to authorize execution. In another process, AI might select and execute routine actions automatically while escalating only exceptions.

The implication for executives is straightforward: **decision rights can be designed at the level of the decision stage, not merely at the level of the job or process**. That is a much more precise way to think about human-AI work than asking whether a role should be automated.

TYPES AND LEVELS OF HUMAN-AUTOMATION INTERACTION

Automation can be allocated separately across stages of a decision—not simply switched “on” or “off.”

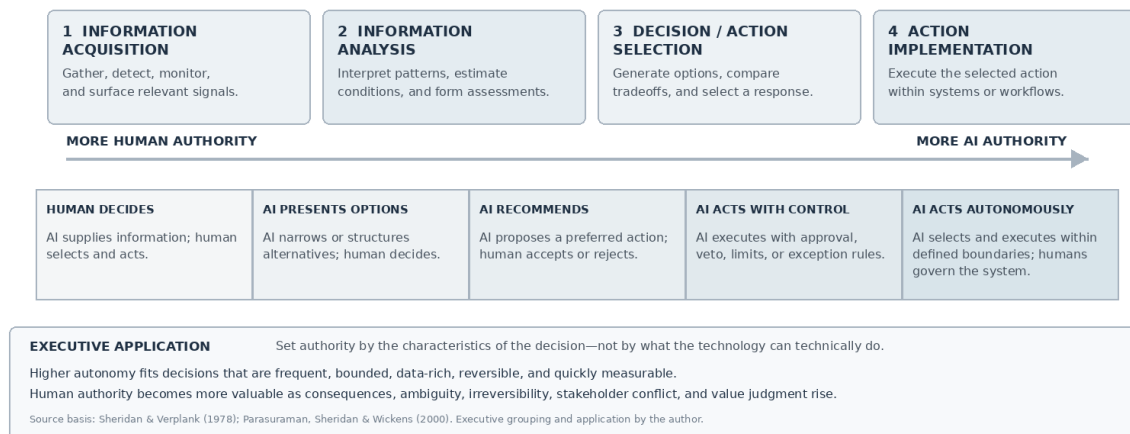


Figure 1. Types and levels of human–automation interaction, adapted for executive decision authority.

Start With the Decision, Not the Technology

The most common mistake in AI governance is to begin with capability: the model can do this, therefore the organization should allow it to do this. Technical capability is necessary, but it is not sufficient. Decision authority should be determined by the economics and risk characteristics of the decision itself.

Five characteristics are especially useful when executives consider how much authority to transfer. None creates an automatic answer, but together they reveal why two decisions that use the same AI system may require very different governance.

Frequency and boundedness

High-volume, repetitive decisions with clearly defined objectives are generally stronger candidates for greater automation. Replenishment within approved inventory bands, routine scheduling adjustments, low-value purchase execution, or standardized routing changes can occur too frequently for human review to add proportional value. The more bounded the decision space and the clearer the permitted actions, the easier it is to delegate authority without delegating unlimited discretion.

Consequence of error

The relevant question is not whether AI will make mistakes. Humans make mistakes as well. The more important question is what happens when the decision is wrong. A poor recommendation that shifts a small replenishment quantity can often be corrected quickly. A decision that stops production, releases a regulated product, terminates a supplier, changes a safety control, or materially affects a customer may have consequences that justify a different level of human authority even if the model is highly accurate.

Reversibility

Reversibility changes the economics of autonomy. Decisions that can be detected and reversed at low cost can tolerate more machine authority than decisions whose effects become difficult or impossible to unwind. This is why a system may be allowed to rebalance inventory autonomously within a network but not to close a facility, permanently alter a customer commitment, or approve a major capital project.

Ambiguity and value judgment

AI performs best when objectives and constraints can be expressed with sufficient clarity. Human judgment becomes more valuable when the decision requires reconciling competing values, incomplete context, political or reputational consequences, or goals that cannot be reduced cleanly to a single optimization function. A model may determine

which customer allocation maximizes margin and service probability. Leadership may still decide that a strategically important relationship warrants a different answer.

Quality and speed of feedback

Autonomy is easier to govern when the organization can observe outcomes quickly and determine whether the decision was good. A system making thousands of decisions without timely feedback can accumulate error before anyone understands the pattern. Fast, reliable feedback allows organizations to monitor performance, tighten boundaries, and progressively expand authority. Weak feedback should make executives more conservative, regardless of how sophisticated the model appears.

The Hidden Cost of Keeping Humans in Every Loop

The phrase “human in the loop” often sounds like the safest answer, but it can conceal a poor operating design. If every machine recommendation requires human approval, the organization eventually caps automation at the speed and attention of the approvers. McKinsey’s work on agentic organizations makes this point directly: as AI systems scale, human oversight capacity can itself become a productivity bottleneck.

There is also a human-factors problem. When a system is correct most of the time, people can become less vigilant while continuing to carry nominal approval responsibility. The result is a dangerous combination: the machine effectively makes the decision, but the process retains the cost and appearance of human control. A signature or click is not meaningful oversight if the person lacks the time, information, or expectation to challenge the recommendation.

The better design is often **human governance around the decision rather than human approval of every decision**. Leaders define objectives, boundaries, prohibited actions, escalation triggers, monitoring requirements, and conditions under which authority must return to a person. Within low-risk and well-understood boundaries, the system can act at machine speed. Humans concentrate on exceptions, ambiguous cases, changing conditions, and decisions where judgment creates real value.

This is increasingly how organizations are approaching agentic systems. NIST’s AI Risk Management Framework emphasizes governance, context, risk measurement, monitoring, and defined human oversight rather than prescribing a single universal level of human intervention. Recent enterprise research similarly treats AI autonomy as a decision-rights problem, with greater autonomy appropriate inside explicit policy and risk boundaries.

A Practical Example: Network Inventory and Service

Consider an illustrative consumer-goods network with six distribution centers serving different regions. An AI planning agent can forecast near-term demand, detect inventory imbalance, assess transportation capacity, identify service risk, and initiate stock transfers. Historically, planners review every recommendation before a move is executed.

Leadership wants to increase speed and reduce planning workload. Giving the agent full autonomy across the entire process would be simple technologically, but poor management design. The decisions inside the process are not equally consequential.

Routine transfers between facilities, inside predetermined inventory and transportation limits, are frequent, reversible, measurable, and economically bounded. The agent can reasonably select and execute these moves automatically while logging the rationale and outcome. Transfers that would reduce a region below a strategic service floor can require human approval. Decisions involving allocation during a major shortage can remain human-led because they may involve customer relationships, commercial commitments, reputational consequences, and tradeoffs that are only partially represented in the data.

The same system can therefore operate at several levels of authority simultaneously. It may autonomously acquire information and analyze conditions across the entire network. It may automatically execute low-risk rebalancing. It may recommend higher-impact actions to planners. And it may be prohibited from executing certain decisions at all.

That is the value of function allocation. The question is no longer “Should the AI run inventory?” It becomes: **Which parts of the inventory decision should the AI perform, at what level of authority, and under what conditions should authority shift back to a person?**

Authority Should Be Dynamic, Not Permanent

Executives should also resist treating autonomy as a one-time design choice. Decision authority can expand or contract as conditions change. A system that operates autonomously during stable demand may need tighter human control during a product launch, labor disruption, cyber incident, regulatory change, or severe supply shortage. A new model may begin in recommendation mode, move to approval-based execution after validation, and later gain bounded autonomy once performance is demonstrated.

The reverse should also be possible. If model drift, data-quality problems, abnormal operating conditions, or repeated exceptions reduce confidence, the organization should be able to step down the level of autonomy without redesigning the entire process. In this sense, human-AI decision architecture should behave more like a control system than a permanent organizational chart.

This is where governance becomes an operating capability rather than a compliance activity. Leaders need visibility into what decisions AI is authorized to make, the boundaries of that authority, how performance is measured, what triggers escalation, and who can change the level of autonomy. As agentic AI increasingly moves from recommendation to action, McKinsey has described that shift accurately as a transfer of decision rights—not simply another software feature.

The Counterargument: Human Judgment Is Not Automatically Safer

There is an important counterargument to preserving human authority: people are inconsistent, biased, fatigued, politically influenced, and often slow. In many bounded decisions, forcing a person to remain the final decision maker can reduce quality rather than improve it. Human discretion is not inherently superior merely because the consequences are important.

The right comparison is therefore not “AI risk” versus an idealized human. It is the performance of a governed AI system versus the actual human decision process it would replace or augment. If human decisions are highly variable, poorly documented, slow, or driven by incomplete information, automation may improve both control and consistency. Conversely, a high-performing human process with rich context and experienced judgment may justify a much higher threshold before authority is transferred.

This is why the objective should never be maximum autonomy or maximum human control. Both can be expensive. The objective is the **lowest-cost, highest-quality allocation of decision authority consistent with the organization’s risk appetite and strategic requirements.**

The Executive Question Is Where Judgment Creates Value

AI will continue to improve at acquiring information, recognizing patterns, generating alternatives, selecting actions, and executing work. As capability rises, the temptation will be to treat increasing autonomy as evidence of progress. That is the wrong measure.

Progress is not giving AI more decisions. Progress is assigning each stage of a decision to the party that can perform it best while preserving the speed, control, adaptability, and judgment the enterprise actually needs.

For frequent, bounded, reversible, data-rich decisions with fast feedback, the economic case for machine authority can become compelling. As consequences become larger, outcomes harder to reverse, objectives more ambiguous, or stakeholder tradeoffs more consequential, human judgment becomes more valuable—even when AI supplies most of the analysis.

The executive task is therefore not to keep humans in every loop. It is to decide **which loops still need them.**

Selected Sources

Sheridan, T. B., and Verplank, W. L.. *Human and Computer Control of Undersea Teleoperators*. 1978. [Source](#)

Parasuraman, R., Sheridan, T. B., and Wickens, C. D.. *A Model for Types and Levels of Human Interaction with Automation*. IEEE Transactions on Systems, Man, and Cybernetics - Part A, 2000. [Source](#)

National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. January 2023. [Source](#)

McKinsey & Company. *Trust in the Age of Agents*. March 2026. [Source](#)

McKinsey & Company. *The Agentic Organization: Contours of the Next Paradigm for the AI Era*. 2025. [Source](#)

Boston Consulting Group. *Enterprise AI Control Plane: The CIO's Guide to Governing and Accelerating AI Agents*. August 2026. [Source](#)