Journal of the International Society for the Systems Sciences | 70th Meeting of the International Society for the Systems Sciences, Cyprus.
Wilby, J., Makar, J. and Laouris, Y. eds.

JUDGMENT PRESERVATION IN AUGMENTED INTELLIGENCE SYSTEMS: A GOVERNANCE MODEL FOR DELEGATED ARTIFICIAL INTELLIGENCE ACTION

Ken Howarth
Mercer County Community College
howarthk@mccc.edu

Abstract

As Artificial Intelligence systems increasingly participate in recommendation, tool use, and delegated organizational action, institutions face a governance problem that is often treated too narrowly as a performance problem: under what conditions may system capability be translated into actionable authority without eroding human responsibility, contestability, and learning? This problem is important because socio-technical failures frequently arise not only from model error, but also from weak authorization design, ambiguous accountability, poor escalation pathways, and the thinning of meaningful human oversight. It is well-suited to systems science because it involves interacting human and technical actors, feedback loops, boundary conditions, incentive structures, and cross-level effects that cannot be adequately understood in isolation. This paper uses a systems-governance methodology that combines conceptual analysis, socio-technical systems modeling, boundary analysis, and failure-mode-oriented institutional design to distinguish capability from authority and to identify conditions under which delegated Artificial Intelligence action remains governable. The analysis develops a four-part governance model centered on bounded capability envelopes, execution-boundary revalidation at the point of action, competence-based human oversight, and observability structures that support escalation, review, and post-incident learning. The result is a framework for preserving meaningful human interpretive, supervisory, and justificatory agency under conditions of increasing machine mediation. The paper contributes to systems science by offering a transferable governance architecture for complex institutional settings and contributes to practice by clarifying how augmented intelligence can support action without surrendering responsibility in sectors such as public administration, education, health, and other regulated decision environments.

Keywords

augmented intelligence, delegated Artificial Intelligence action, socio-technical governance, human oversight, systemic leadership

Introduction

Artificial Intelligence governance is often narrated as a problem of safety, fairness, accuracy, or explainability. Those matters remain indispensable, but they do not exhaust the governance problem that emerges when systems move from generating outputs to participating in action. A recommendation engine may remain advisory. A workflow system may remain clerical. A model may remain one evidential input among others. Yet once systems trigger tool use, reorder tasks, draft communications for release, route cases, allocate attention, or otherwise participate in delegated organizational action, the question changes. The issue is no longer only whether a system performs well enough in a technical sense. The issue becomes whether institutional authority is being translated into machine-mediated action under conditions that remain answerable, interruptible, reviewable, and corrigible.

This paper argues that the crucial requirement at that point is judgment preservation. By judgment preservation I mean the maintenance of meaningful human interpretive, supervisory, and justificatory agency under conditions of increasing machine mediation. The claim is not that human beings must manually perform every step, nor that automation is suspect by definition. The claim is narrower and more demanding: delegated action remains governable only when human agents and institutions retain real capacities to set action bounds, revalidate execution at consequential thresholds, detect the need for escalation, contest outcomes, and learn from breakdown.

The paper is deliberately positioned to complement, rather than repeat, several adjacent lines of recent work. Recent work on judgment literacy treats formed human judgment as institutional infrastructure. Recent human-before-the-loop arguments stress the temporal and normative priority of human judgment in the design, authorization, and oversight of Artificial Intelligence systems. Recent noosphere or judgment-ecology work examines how Artificial Intelligence reshapes collective thought at a broader social scale. The present paper operates at a more middle level. It focuses on delegated action architecture inside organizations once the question is no longer whether a system can inform a human, but whether it can help initiate, shape, or carry forward action under institutional authority. Its central concern is therefore not judgment in general, but the system conditions under which judgment remains operationally real.

That makes the topic especially fitting for systems science. The relevant failures are rarely reducible to one bad model or one negligent operator. They arise from interacting human and technical actors, feedback loops, boundary conditions, incentive structures, weakly designed handoffs, and institutional learning failures. The unit of analysis is therefore not the model alone, but the socio-technical arrangement through which capability becomes or fails to become authority.

The timing matters because official guidance now tracks a move from discrete assistive tools toward workflow-shaping and agentic uses. The Organisation for Economic Co-operation and Development's 2025 report on governing with Artificial Intelligence, drawing on 200 government use cases, reports that 57 percent support automating, streamlining, or tailoring services and 45 percent enhance decision making, sense making, or forecasting. The Centers for Disease Control and Prevention's March 2026 guidance for agentic research in public health addresses tools that autonomously plan and execute multi-step research tasks while insisting on human oversight. The National Institute of Standards and Technology's Center for AI Standards and Innovation has likewise warned that models can cheat on agentic evaluations. The problem, then, is no longer only how to appraise outputs. It is how to govern delegation across a workflow when evidential reliability, interruption authority, and supervisory competence all matter at once (OECD 2025; Centers for Disease Control and Prevention 2026; National Institute of Standards and Technology 2026).

From Capability to Authority

A recurring mistake in Artificial Intelligence governance is to treat capability as if it already carried its own license for authority. If a system predicts well, classifies efficiently, drafts competently, or executes a chain of tasks with low apparent error, institutions begin to absorb its outputs into operational authority. That absorption may be explicit, as when systems are authorized to take bounded actions automatically. It may also be tacit, as when staff increasingly defer to system outputs because the surrounding organization has been redesigned for speed, scale, compliance, or liability management.

Capability and authority answer to different questions. Capability concerns what a system can do under specified conditions. Authority concerns when, where, how, and under whose answerable responsibility those capacities may rightly shape action. A system may be highly capable in one register while being unfit to bear authority in another. It may be useful for triage and unfit for final disposition. It may be useful for drafting and unfit for autonomous issuance. It may be useful for signal detection and unfit for sanction, denial, exclusion, or coercive intervention. To collapse these questions is to confuse technical performance with institutional warrant.

The distinction matters because institutional life is full of translation layers between output and action. Managers decide where automation fits. Designers decide what default options are visible. Procurement and compliance teams define what counts as acceptable risk. Staff adapt their own behavior to workflow pressures. Over time, advisory tools often become benchmarks, and benchmarks become quiet governors of decision pathways. Responsibility remains formally human while action logic becomes substantively machine-mediated. This gradual authority migration is one of the defining governance problems of augmented intelligence systems.

Official governance materials increasingly recognize pieces of this problem. The National Institute of Standards and Technology frames Artificial Intelligence risk as affecting individuals, organizations, and society, and it organizes risk management across design, development, deployment, and use rather than treating evaluation as a one-time technical checkpoint. Its Generative Artificial Intelligence profile extends that concern to dynamic, open-ended, and sociotechnically entangled deployments. The European Union Artificial Intelligence Act likewise ties obligations to use context and risk category rather than to raw model power alone. These moves are important, but institutions still often smuggle authority in through convenience, interface design, throughput pressure, or staffing assumptions that hollow out meaningful review (NIST 2023; NIST 2024; European Union 2024).

Why Delegated Artificial Intelligence Action Is a Systems Problem

Delegated Artificial Intelligence action is a systems problem for at least four reasons. First, action chains are distributed. The system generating a recommendation is not the whole operative unit. There are upstream data practices, model choices, interface decisions, institutional policies, escalation rules, staffing ratios, procurement terms, and documentation norms. The relevant object is therefore the socio-technical assemblage through which action is made possible, attractive, routinized, and insulated from challenge.

Second, the harms are interactional. Breakdowns emerge not only from model inaccuracy but from bad handoffs, misplaced trust, alert fatigue, missing exception pathways, undocumented overrides, or oversight roles that have become too thin to function. A technically decent model can still participate in an institutionally bad system.

Third, the feedback loops are dynamic. Human overseers learn from systems, but systems also reshape what human agents notice, remember, defer to, and feel authorized to challenge. Over time, organizational adaptation can erode judgment capacity even where policy language still presumes its presence.

Fourth, cross-level effects matter. Micro-level design choices can generate meso-level workflow distortions and macro-level legitimacy losses. A poorly designed override interface may look trivial in isolation, yet it can contribute to underreporting of uncertainty, which weakens auditability, degrades trust, and makes post-incident repair harder. This is the terrain on which systems science has distinctive value: it can show how local design, human roles, institutional incentives, and learning loops interact to determine whether authority remains governable.

Methodological Approach

The paper uses a systems-governance methodology composed of four interlocking moves. The first is conceptual analysis. This clarifies distinctions that are frequently blurred in current discourse: capability versus authority, assistance versus delegation, oversight versus responsibility theater, and automation of tasks versus migration of institutional judgment. Without such distinctions, governance debates slide too quickly from technical performance to institutional permission.

The second move is socio-technical systems modeling. The relevant object is not the model in isolation but the action architecture in which it participates. This includes upstream inputs, downstream users, authority nodes, review chokepoints, exception pathways, and learning loops.

The third move is boundary analysis. Institutions often fail not because they lack rules, but because they have weakly specified thresholds for when systems may act, when humans must intervene, when authority must be withdrawn, and when escalation is mandatory. Boundary analysis asks where those thresholds sit, how they are recognized, and who retains the power to redraw them.

The fourth move is failure-mode-oriented design. Instead of assuming good operation and then adding compliance checks, this paper works backward from predictable breakdowns: authority creep, rubber-stamp review, silent uncertainty, override suppression, procedural opacity, and post-incident amnesia. The aim is not only to diagnose failure but to identify a governance architecture that remains robust under ordinary institutional pressures rather than only ideal conditions. This orientation resonates

with longstanding systems concerns about regulation, complexity, hierarchy, and requisite variety (Ashby 1956; Simon 1962; Bertalanffy 1968).

A Four-Part Governance Model for Delegated Action

The proposed model centers on four interacting elements: bounded capability envelopes, execution-boundary revalidation, competence-based human oversight, and observability structures that support escalation, review, and post-incident learning. The point is not to add four independent controls to an otherwise unchanged system. It is to make governability a property of the action architecture itself.

Bounded Capability Envelopes

The first condition of governable delegation is a bounded capability envelope. Institutions must specify what the system is being relied on to do, under what conditions, at what stakes, over what classes of cases, and with what exclusions. The point is not only technical scope control. It is normative and institutional clarity. Without a bounded envelope, the system's role expands through convenience, habit, vendor ambition, or managerial pressure.

A bounded capability envelope should state at least five things: the task class, the context of use, the stakes profile, the allowed action range, and the explicitly non-authorized uses. It should also identify what evidence supports deployment in that envelope and what conditions trigger re-evaluation. This moves governance upstream. Instead of saying only that a tool is available or helpful, the institution must say what it is actually authorizing.

Execution-Boundary Revalidation

The second condition is execution-boundary revalidation. Even when a system has been placed inside a bounded envelope, real-world cases drift. Contexts change. Edge cases appear. Stakes escalate. New information arrives. A system fit for action yesterday may be unfit for this case today. Revalidation means that consequential action points require a renewed check of whether the present case still falls within the authorized envelope.

This matters especially when systems move from advisory output to initiating or materially shaping action. Revalidation may be partly automated, but it cannot be reduced to a box-ticking ritual. Its function is to interrupt routine authority migration and force recognition of salient change. In practice, revalidation can be tied to thresholds such as elevated uncertainty, rights impact, irreversible effects, contestation by an affected party, novelty of case profile, or mismatch between system output and contextual human knowledge.

Competence-Based Human Oversight

The third condition is competence-based human oversight. Human oversight is not real merely because a human being appears somewhere in the process. Oversight counts only when the person or role-holder has the authority, competence, time, information access, and institutional backing needed to recognize problems and intervene meaningfully. Organizations often speak as if the human layer automatically supplies judgment. In reality, oversight fails when staff are under-trained, over-burdened, procedurally boxed in, or conditioned to treat system outputs as presumptively correct.

Competence-based oversight requires at least three things: role clarity, domain literacy, and intervention reality. Role clarity concerns who is responsible for what. Domain literacy concerns whether the overseer can interpret outputs in context, recognize mismatch, and understand the stakes of acting or not acting. Intervention reality concerns whether the overseer can actually pause, override, escalate, document concerns, and trigger review without being punished for slowing the workflow. Here the paper complements judgment-literacy work while staying narrower than it. The focus is not institution-wide

formation as such, but the specific form of human competence that delegated action systems must preserve if oversight is to be more than ceremonial.

Observability, Escalation, and Post-Incident Learning

The fourth condition is observability structures that support escalation, review, and post-incident learning. Governable systems must make it possible to see not only what outputs were produced, but how they were used, when they were overridden, where uncertainty clustered, which cases escalated, and what kinds of breakdowns recurred. Observability is not mere logging. It is the design of visibility in service of answerability.

Some logs are technically rich yet institutionally poor because they do not illuminate authority transitions or failure modes. What matters is whether the system allows organizations to reconstruct action pathways, identify where delegation went wrong, and learn without defaulting to scapegoating. Escalation must also be real. If staff cannot move difficult or dubious cases upward, or if escalation is formally allowed but practically discouraged, then the system becomes brittle and self-sealing. Post-incident learning should feed back into envelope revision, threshold revision, training revision, and, where necessary, withdrawal of delegated authority.

Failure Modes and Governance Controls

Table 1 summarizes recurrent failure modes in delegated action systems. The table is deliberately practical. It treats failure not as exceptional misfortune but as a predictable expression of weakly governed socio-technical architecture. Each row links a typical mechanism of failure to the governance control most directly aimed at it.

Failure Mode	Typical Systemic Mechanism	Primary Governance Control
Authority creep	Advisory outputs become de facto action rules through routine deference and workflow redesign	Bounded capability envelope
Boundary drift	Novel, high-stakes, or rights-sensitive cases are treated as if they still fit the original deployment scope	Execution-boundary revalidation
Rubber-stamp oversight	Humans remain nominally present but lack time, authority, or competence to intervene	Competence-based human oversight
Silent uncertainty	System doubt, mismatch, or anomalous conditions are not made legible to users or managers	Observability for escalation and review
Override suppression	Formal override exists, but organizational pressure discourages interruption or escalation	Intervention reality plus escalation protection
Post-incident amnesia	Breakdowns are handled case by case without revising envelopes, thresholds, or training	Post-incident learning loop

Table 1. Recurrent failure modes in delegated action systems and the controls most directly aimed at them.

Illustrative Applications

Current application landscapes are broader and more action-adjacent than the early chatbot or co-pilot frame suggested. In government, education, health, and public health, institutions are experimenting not only with summarization and drafting, but with routing, advising, tutoring, research synthesis, anomaly detection, eligibility support, documentation, and other workflow-shaping uses. That spread makes it

more important to ask not simply where Artificial Intelligence appears, but where it begins to carry or condition institutional authority.

In public administration, the model applies wherever systems rank, route, flag, or prioritize cases that affect benefits, enforcement, inspection, or eligibility. The Organisation for Economic Co-operation and Development's 2025 survey of 200 government use cases finds that public-sector adoption is concentrated in public services, civic participation, and justice; 57 percent of cases support automating, streamlining, or tailoring services, while 45 percent enhance decision making, sense making, or forecasting. The Organisation for Economic Co-operation and Development's 2026 brief on an Artificial Intelligence-ready public workforce adds that public institutions need internal capability, proactive governance, and workforce upskilling if Artificial Intelligence is to improve service quality without weakening accountability. The core risk, then, is not only inaccurate output but the quiet transformation of administrative judgment into workflow obedience. Bounded envelopes and revalidation points help distinguish supportable triage from unfit de facto adjudication, and competence-based oversight matters because administrative staff need real authority to challenge system outputs rather than merely record compliance (OECD 2025; OECD 2026).

In education, systems increasingly assist with advising, retention prediction, writing evaluation, academic integrity screening, tutoring, and career or pathway navigation. The United States Department of Education's July 2025 guidance on the use of federal grant funds to improve education outcomes through Artificial Intelligence explicitly treats Artificial Intelligence as usable across key educational functions, including adaptive instructional materials, tutoring, advising, and navigation, while insisting that these tools should support educators rather than replace their critical role. UNESCO's 2025 report AI and education: Protecting the rights of learners likewise frames the issue in human-centred, rights-based terms rather than as simple tool uptake. Here the salient risk is authority expansion through convenience. Systems introduced as advisory often harden into default decision channels. Oversight must therefore remain pedagogically informed rather than dashboard-driven, and educational actors need enough contextual knowledge and institutional backing to interpret outputs in light of student history, mission, and educational purpose rather than raw system signal alone (United States Department of Education 2025; UNESCO 2025).

In healthcare and public health, system outputs may aid triage, documentation, imaging review, patient messaging, research synthesis, or risk scoring. The World Health Organization's 2025 guidance on large multi-modal models anticipates wide use in health care, scientific research, public health, and drug development. World Health Organization/Europe's November 2025 regional readiness report and April 2026 European Union snapshot both examine governance models, legal and ethical frameworks, workforce readiness, and uptake across health systems. The Centers for Disease Control and Prevention's 2026 considerations for agentic research in public health likewise describe deep research tools that autonomously plan and execute multi-step tasks while stressing human oversight and clear expectations. High stakes and heterogeneous contexts therefore make execution-boundary revalidation especially important. A system can be useful at one stage and dangerous if its action-shaping role expands unnoticed across the care pathway or from evidence gathering into operational response. Across these settings, the recurring lesson is the same: the question is not whether the system contributes usefully, but whether the institution retains governable control over how contribution becomes action (World Health Organization 2025; World Health Organization Regional Office for Europe 2025; World Health Organization Regional Office for Europe 2026; Centers for Disease Control and Prevention 2026).

Relation to Existing Governance Frameworks

The proposed model is not meant to compete with contemporary governance frameworks so much as to sharpen one of their underdeveloped questions: how institutional authority is actually translated into system-mediated action. The National Institute of Standards and Technology's Artificial Intelligence Risk Management Framework provides an important lifecycle and governance orientation. It asks organizations to treat risk as context-sensitive, to map functions, to measure, and to manage. The Generative Artificial Intelligence profile extends that logic into open-ended deployments where system behavior, downstream

misuse, and interactional harms are harder to stabilize. The present paper complements those frameworks by focusing more tightly on the authority transition itself. It asks what must be true at the moment a capable system is permitted to shape action under institutional warrant, and what must remain true if that permission is to stay governable.

The same is true of the European Union Artificial Intelligence Act. The Act rightly links obligations to risk category, use context, and governance responsibilities rather than treating all systems as interchangeable. Yet even where the legal architecture is robust, institutions still face a practical design problem inside the organization: who may rely on what system output, under what conditions, with what thresholds for interruption, and with what capacity to contest or withdraw delegated authority? A legal framework can require human oversight, documentation, and risk management, but it does not by itself make oversight competent, intervention-real, or learning-oriented. The governance model advanced here therefore sits one level below broad regulatory classification and one level above local workflow detail. It offers a way to operationalize governability inside the organization.

This middle-level contribution matters because organizations frequently mistake compliance artifacts for living controls. A policy, model card, or impact assessment may exist on paper while authority continues to drift operationally through interface defaults, staffing patterns, or pressure for speed. By centering bounded capability envelopes, execution-boundary revalidation, competence-based oversight, and observability for escalation and learning, the model helps specify what it would mean for compliance expectations to take institutional form. In that sense, the paper does not reject prevailing frameworks. It supplies a more explicit account of the socio-technical control points through which their governance ambitions either become real or collapse into ceremony.

Design Implications for Institutions

The model also yields a set of design implications for organizations adopting augmented intelligence systems. First, institutions should maintain an authorization register that states, in ordinary operational language, the specific task classes, stakes conditions, exclusions, and escalation triggers associated with each deployment. Many organizations document system purpose at too high a level of abstraction. They say that a tool supports case review, advising, risk identification, or service optimization. What they need instead is a bounded statement of what kinds of action the system may materially shape and what remains outside its warrant. Such registers create a practical reference point against which authority drift can be detected.

Second, organizations should design interruption as a normal feature of delegated systems rather than as a reluctant exception. This means building visible routes for pause, override, and escalation at points of uncertainty, novelty, rights impact, or downstream irreversibility. It also means protecting those who use such routes. An escalation path that exists formally but is treated as a sign of inefficiency or poor team alignment will soon become inert. Governability requires not only that interruption be technically possible, but that it be institutionally legitimate.

Third, institutions should treat oversight roles as formed and resourced positions rather than as residual duties. Competence-based oversight requires more than a checkbox that a human reviewed the output. It requires time, contextual knowledge, clarity about responsibility, and some protected space for reason-giving. Especially in public, educational, and health settings, the human layer must remain thick enough to notice when the system's local output is at odds with the case context, institutional mission, or normative stakes. This is where delegated-action governance reconnects with broader work on judgment formation and judgment literacy. A thin oversight role cannot preserve judgment because it no longer houses enough judgment to preserve.

Fourth, observability should be designed for institutional learning rather than only retrospective blame assignment. Organizations need to know which classes of cases trigger override, where uncertainty clusters, which units escalate frequently, and what kinds of mismatch recur between output and outcome. That information should not simply feed dashboards for productivity management. It should inform envelope revision, additional training, redesign of interfaces, and, when necessary, de-authorization of a

workflow that has become ungovernable. In this respect, the governance of delegated action is inseparable from the governance of organizational learning itself.

Seen through a Bettering Bearings lens, several recent lines sharpen the paper's practical stakes without requiring Bettering Bearings terminology to dominate the venue-facing surface. Delegation-with-Answerability keeps attention on who still owns action when a system is allowed to shape it. Human Before the Loop presses the design question upstream: oversight that appears only after deployment is already late. The Finding-Conditions Audit asks whether staff, authority, time, and escalation routes are present enough for oversight to be real. The AI Ethics Capture Audit guards against mistaking policy talk, dashboards, or ethics theatre for working controls. Fidelitying Technical Literacy marks a further requirement: enough technical and domain literacy to notice salient failure modes without yielding judgment to technical prestige. These lines do not replace the paper's four-part model. They specify what each element must still be able to do in practice.

Discussion: Systemic Leadership and Governability

One virtue of the model is that it does not treat human oversight as a magical residue left over after automation. Oversight itself must be designed, authorized, and supported. This is a leadership problem as much as a technical one. Systemic leadership in augmented intelligence settings requires more than compliance statements or aspiration language about keeping humans in the loop. It requires institutions to decide where delegation is warranted, what forms of uncertainty or rights impact trigger interruption, how staff are trained to intervene, and how breakdowns feed institutional learning rather than blame deflection. That leadership task is difficult because many organizations face strong incentives to blur the difference between assistance and authority. Speed, scale, cost pressure, and the symbolic prestige of automation all push in that direction. Yet precisely for that reason, a systems-governance approach is needed. It makes visible the points at which authority migrates, the conditions under which oversight becomes thin or fictive, and the controls required to keep augmentation answerable. The paper's broader contribution to systems science is therefore a governance vocabulary for delegated action: capability envelope, execution-boundary revalidation, competence-based oversight, and observability for escalation and learning. These concepts are meant to travel across sectors without pretending that one implementation fits all cases.

A final implication follows for institutions tempted to solve governance problems mainly through more detailed prediction, monitoring, or automation. Better measurement can be useful, but delegated-action failures are not always measurement failures. They are often failures of authorization design, role design, and answerability design. An organization may know a great deal about system performance and still know too little about when authority has shifted, when oversight has become ceremonial, or when escalation has been socially suppressed. Systems science is valuable here because it resists the reduction of governability to any single metric. It directs attention to the relationships among controls, actors, thresholds, and learning loops that determine whether an institution remains able to govern the systems it has chosen to rely on.

Conclusion

The most serious governance problem in augmented intelligence is not that machines produce outputs. It is that institutions increasingly allow those outputs to migrate into action without adequately governing the authority transition. This migration rarely occurs all at once. It proceeds through workflow redesign, staff adaptation, default settings, throughput pressure, and the quiet erosion of human capacities to question, interrupt, and justify.

Judgment preservation names the counter-requirement. If institutions want augmentation without surrender, speed without blind deference, and support without responsibility laundering, they must design delegated action systems that preserve meaningful human judgment. That requires bounded capability envelopes, execution-boundary revalidation, competence-based human oversight, and observability structures oriented toward escalation and learning. The systems-science lesson is equally clear:

governability cannot be read off a benchmark score or model card. It must be built into the socio-technical architecture through which capability becomes authority. Where that architecture is weak, nominal human oversight will not save the system. Where it is strong, augmentation can remain genuinely answerable.

References

Ashby, W. Ross. 1956. An Introduction to Cybernetics. London: Chapman & Hall.

Bertalanffy, Ludwig von. 1968. General System Theory: Foundations, Development, Applications. New York: George Braziller.

European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, 12 July 2024.

National Institute of Standards and Technology. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce.

National Institute of Standards and Technology. 2024. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. Gaithersburg, MD: U.S. Department of Commerce.

Centers for Disease Control and Prevention. 2026. Considerations for Agentic Research in Public Health. Atlanta, GA: Centers for Disease Control and Prevention.

National Institute of Standards and Technology. 2026. Analyzing Transcripts from AI Agent Evaluations. Gaithersburg, MD: Center for AI Standards and Innovation.

Organisation for Economic Co-operation and Development. 2025. Governing with Artificial Intelligence. Paris: OECD.

Organisation for Economic Co-operation and Development. 2026. Building an AI-ready Public Workforce. Paris: OECD.

UNESCO. 2025. AI and education: Protecting the rights of learners. Paris: UNESCO.

United States Department of Education. 2025. Guidance on the Use of Federal Grant Funds to Improve Education Outcomes through Artificial Intelligence. Washington, DC: U.S. Department of Education.

World Health Organization. 2025. Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-modal Models. Geneva: World Health Organization.

World Health Organization Regional Office for Europe. 2025. Artificial intelligence is reshaping health systems: state of readiness across the WHO European Region. Copenhagen: WHO Regional Office for Europe.

World Health Organization Regional Office for Europe. 2026. Artificial intelligence is reshaping health systems: state of readiness across the European Union. Copenhagen: WHO Regional Office for Europe.

Parasuraman, Raja, and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors 39 (2): 230-253.

Simon, Herbert A. 1962. The Architecture of Complexity. Proceedings of the American Philosophical Society 106 (6): 467-482.