Herbert Simon, Generative AI, and CUTIE STREET Walk into a Bar…

Is Being Smart Not Enough?

MEOW as a Simonian Decision Architecture for Generative AI in Organizations

Aoki, Kenichi

Independent Researcher

Correspondence: aokikenichi@gmail.com

Position paper (preprint), November 2, 2026

Abstract

Generative AI lowers the cost of generating decision-relevant summaries, alternatives, and arguments, but expanded model capability does not by itself yield more rational organizational decisions. This conceptual position paper argues that generative AI can redistribute bounded rationality: when generated outputs expand faster than the capacity to assess and use them, binding constraints shift toward verification, attention, decision-premise governance, and learning. Building on Herbert A. Simon and research on human–AI collaboration, AI governance, and organizational routines, the paper proposes MEOW: Model Intelligence, Epistemic Verification, Organizational Decision Premises, and Work Routines. These dimensions describe complementary capabilities enacted within an AI-mediated decision episode, rather than stages in a linear pipeline. Organizational rationality is defined here procedurally as evidence-grounded, premise-explicit, accountable, and revisable decision-making; it does not imply optimal outcomes or ethical legitimacy by managerial authorization alone. Five propositions address generation–verification imbalance, premature satisficing, capability complementarity, risk-calibrated verification, and trace-enabled learning. The framework translates these relationships into roles, evidence thresholds, decision gates, records, and feedback loops. Its proposed contribution is a Simonian integration centered on how AI outputs become accepted decision premises, with observable conditions under which that passage may succeed or fail. It does not claim priority for verification, oversight, or governance gates. The propositions and accompanying decision canvas, checklist, and log remain unvalidated and are offered for empirical testing and context-sensitive organizational design.

Keywords: generative AI; bounded rationality; Herbert Simon; organizational decision-making; decision premises; epistemic verification; AI governance; organizational routines; human–AI collaboration

1. Introduction

Generative AI is rapidly entering knowledge work and managerial decision-making. It can search and synthesize information, produce alternative courses of action, draft analyses, and evaluate proposals at a speed and scale that were previously expensive or impractical. Experiments and field studies report productivity and quality gains in writing, customer support, and consulting tasks, while also showing that performance depends on the task, the user, and whether the task lies within the model's uneven capability frontier (Brynjolfsson et al., 2025; Dell'Acqua et al., 2026; Noy & Zhang, 2023). In strategic decision-making, empirical studies show that large language models can generate and evaluate strategies; implications for search, representation, and aggregation are examined theoretically (Csaszar et al., 2024).

It is tempting to infer that smarter models will produce more rational organizations. That inference confuses the intelligence of an artifact with the rationality of a decision system. Organizations must still determine what goals matter, which evidence is admissible, what assumptions are acceptable, who is authorized to judge an output, and who remains accountable for action. A fluent answer can reduce the cost of generating a candidate judgment while increasing the difficulty of recognizing whether the judgment is grounded, contextually appropriate, or legitimate.

This paper develops a Simonian response to that problem. It asks a simple research question:

> **If generative AI substantially expands the cognitive capacity available to organizations, why does organizational bounded rationality persist?**

Its central thesis is:

> **Generative AI does not eliminate bounded rationality; it can redistribute it.** It can relax bounds on search, articulation, and computation while increasing the relative importance of verification, attention allocation, premise governance, coordination, and learning.

The implication is that **model intelligence does not by itself establish organizational rationality**. Intelligence is a resource available to a decision system; organizational rationality is a system-level achievement. Here, organizational rationality means a procedural capacity to make decisions that are evidence-grounded, consistent with explicit and contestable premises, authorized through accountable roles, and revisable in light of outcomes. This is a working definition for the framework, not an exhaustive interpretation of Simon or a claim to optimality. Internal authorization is also insufficient to establish ethical or societal legitimacy: an authorized decision can still violate justified stakeholder claims. Decision quality, procedural accountability, and legitimacy must therefore be assessed separately.

Research also examines how generated outputs become admissible within organizational decision systems (Saup et al., 2026). This paper builds on that research. Its question is how a Simonian account of decision premises can connect such arrangements to conditional predictions about verification, search termination, capability complementarity, and learning.

The paper proposes the **MEOW framework**: **Model Intelligence (M), Epistemic Verification (E), Organizational Decision Premises (O), and Work Routines (W)**. MEOW is not a maturity slogan or a technical stack. It is a conceptual configuration for an **AI-mediated decision architecture**. The framework treats organizational rationality as an achievement of linked capabilities, not as a property transferred from a model to an organization.

The paper develops three linked contributions. First, it specifies conditions under which generation–verification asymmetry may relocate a binding constraint. Second, it uses the AI-mediated decision episode as a unit connecting human–AI interaction to premises, authority, and routines, without claiming to originate episode-based analysis. Third, it derives five propositions and corresponding design artifacts that make the proposed relationships open to empirical challenge. The contribution is an integrative argument and research agenda, not evidence that MEOW improves decisions.

The paper proceeds as follows. Section 2 explains the conceptual approach and scope. Section 3 connects Simon's theory to adjacent literatures. Section 4 develops the redistribution argument. Section 5 presents MEOW as a configuration of complementary capabilities. Section 6 turns the framework into an organizational decision architecture. Sections 7 and 8 discuss contributions, implications, boundary conditions, and a research agenda; Section 9 concludes.

Figure 1 summarizes the distinction between greater model intelligence and procedural organizational rationality; the relationship is not presumed to be automatic.



Figure 1: Why greater model intelligence does not, by itself, establish procedural organizational rationality.

2. Conceptual Approach and Scope

This is a position and conceptual paper. It does not report an empirical test, a systematic literature review, or a validated maturity model. Its research design uses theory synthesis to develop an integrative conceptual model, drawing on Jaakkola's (2020) distinction among conceptual research designs. Simon's account of bounded rationality, decision premises, and organizational design provides the focal theoretical lens because the problem is not only whether AI improves individual cognition, but how organizations structure the premises, authority, and routines through which bounded actors decide. That lens is applied to a setting in which generative AI changes the cost of producing decision-relevant outputs; the paper does not claim to revise Simon's theory in its entirety.

The synthesis integrates four adjacent streams selected for their explanatory roles: generative AI at work informs changes in task performance; human–AI decision research informs reliance and judgment allocation; AI governance informs risk and accountability; and organizational routines research informs repeated practice and learning. Source selection is purposive rather than exhaustive. A targeted pre-submission update checked closely related publications available by September 21, 2026, including direct challenges to the proposed contribution. No systematic search protocol or claim of literature completeness is implied. The framework asks what is needed for generated content to become sufficiently supported, acceptable under stated premises, authorized, and subject to outcome review. These are analytically distinct requirements that may recur in any order. The five propositions are deductions from this synthesis, not findings established by the cited studies.

The analysis is bounded in four ways. First, the primary setting is organizational decision support rather than fully autonomous machine action. Second, the focal technology is generative AI, including large language models used to produce summaries, options, evaluations, recommendations, and explanations. Third, the focal outcomes are decision quality, legitimacy, accountability, and organizational learning rather than model accuracy alone. Fourth, the proposed artifacts are diagnostic and design aids; they do not automatically authorize a decision or replace domain expertise.

The unit of analysis is the AI-mediated decision episode: a bounded sequence in which AI output influences problem framing, alternatives, evaluation, authorization, or action. For empirical work, an episode begins when a focal decision problem is opened and ends with commitment, rejection, or deferral; any later outcome review is linked by its decision identifier. A meeting may contain several episodes, and one episode may span several meetings. The capability-bearing entity is the organization, team, or decision system that supplies M, E, O, and W. Observations are therefore episodes nested within teams or organizations, while learning claims require multiple episodes over time. Organizational capability should not be inferred from a single completed checklist.

3. Theoretical Foundations and the Research Gap

3.1 Simon: rationality is organized

Simon rejected the image of decision-makers who know every alternative, consequence, and preference and then optimize. Human decision-makers operate under limits of information and computation. They therefore search selectively and often **satisfice**, accepting an alternative that crosses an aspiration threshold rather than proving it globally optimal (Simon, 1955).

For the present argument, Simon's concept of **decision premises** is especially important. Organizations coordinate decisions not only through direct commands but by shaping the factual and value premises from which members decide: goals, classifications, rules, assumptions, criteria, channels of information, and structures of authority (Simon, 1947/1997, Chapter I and its commentary). Authority, communication, specialization, training, identification, and routines can all be understood as mechanisms through which an organization structures the premises available to bounded actors. In this sense, an organization can be read as a **decision-premise architecture**: it does not remove boundedness, but channels it.

Organizational rationality is therefore not the sum of individually intelligent choices. It is produced through designed environments that direct attention, communication, authority, and action (Simon, 1973). Boundedness resides not only inside an individual decision-maker but also in the interaction between limited actors and the organizational architecture that shapes what they notice, know, value, and are authorized to do.

Generative AI matters because its outputs can function as decision premises. A summary influences what is noticed. A list of options shapes the feasible set. A recommendation embeds criteria. A confident explanation can make one causal story appear more reasonable than another. Generative AI is thus not only an information tool; it is a potential **premise-producing infrastructure**.

3.2 Generative AI changes the cost structure of deciding

Research on generative AI demonstrates meaningful gains in output speed and, in many settings, output quality (Brynjolfsson et al., 2025; Noy & Zhang, 2023). At the same time, those gains are not uniform. The "jagged technological frontier" describes how an apparently capable model can perform very well on one task and poorly on a neighboring task that is difficult for users to distinguish in advance (Dell'Acqua et al., 2026). Strategic applications may similarly expand search and evaluation while leaving open questions about judgment, competition, and organizational integration (Csaszar et al., 2024).

Research on AI-assisted decisions suggests that explanations do not reliably produce complementary performance: in the tasks studied by Bansal et al. (2021), explanations increased acceptance of recommendations without selectively improving reliance on correct advice. Buçinca et al. (2021) found that cognitive forcing interventions could reduce overreliance, with costs to user experience. These experiments studied particular AI-assisted tasks, rather than establishing effects across contemporary generative AI deployments. A broader meta-analysis likewise found that human–AI combinations did not, on average, outperform the better standalone performer, with substantial variation across tasks (Vaccaro et al., 2024). These findings motivate, but do not validate, MEOW's propositions about premature satisficing and organizational complementarity.

3.3 Adjacent literatures and the proposed integration

Existing literatures illuminate important parts of the problem. Human-AI collaboration examines the allocation and combination of human and machine capabilities (Jarrahi, 2018; Raisch & Fomina, 2025; Raisch & Krakowski, 2021). Research on algorithms at work examines how computational systems can reorder knowledge, control, roles, and power (Faraj et al., 2018; Kellogg et al., 2020). Responsible AI governance identifies structural, relational, and procedural practices across the AI lifecycle (Papagiannidis et al., 2025), and the NIST Generative AI Profile provides a practical risk-management reference (Autio et al., 2024). Organizational routines research explains how repeated action patterns store knowledge while enabling both stability and change (Feldman & Pentland, 2003; Levitt & March, 1988; Nelson & Winter, 1982).

The closest work already connects several of these functions. Saup et al. (2026) use a qualitative organizational case to develop an adoption pathway with admission gates for AI outputs. Shrestha et al. (2019) distinguish delegated, sequential, and aggregated decision structures, while Raisch and Fomina (2025) theorize hybrid problem-solving. Applying Simon to AI is also established territory (Schwarz et al., 2022). Neither institutional embedding nor the transition from generated output to decision input can therefore serve as an unrestricted novelty claim.

MEOW offers a narrower integration: it connects verification capacity to the acceptance and revision of factual and value premises within a decision episode. Its proposed explanatory value lies in the five conditional propositions and in distinguishing failure locations that otherwise appear as a generic failure of human oversight. The question is not whether these neighboring literatures lack governance or organizational analysis, but whether the premise-centered account adds explanatory value beyond their existing constructs. That incremental value remains to be demonstrated.

Closely related work	Established contribution	MEOW's proposed emphasis
Saup et al. (2026)	Organizational adoption pathway and admission gates	Examines capability relationships within and across decision episodes
Shrestha et al. (2019)	Allocation of human and AI decision authority	Examines the grounds on which outputs are accepted within those structures
Raisch and Fomina (2025)	Hybrid problem-solving arrangements	Specifies conditional interactions among M, E, O, and W
Schwarz et al. (2022)	Simonian rationality and AI in public organizations	Adapts that lens to generative output, verification workload, and premise governance

These emphases overlap. MEOW is not offered as a replacement for the cited frameworks or as a comprehensive taxonomy of organizational capabilities.

4. Redistributing Bounded Rationality: Where the Bounds Move

4.1 From information scarcity to verification scarcity

Before widespread generative AI, many knowledge-work decisions were constrained by the cost of finding, reading, comparing, and articulating information. Generative AI sharply reduces some of those costs. It can produce more summaries, alternatives, arguments, and drafts within the same time budget. In this limited but important sense, some aspects of cognition become less bounded: search can be broader, representation faster, and alternative generation cheaper.

Yet boundedness does not disappear. Human and organizational attention remains finite; goals and evaluation criteria still have to be specified; outputs still have to be verified; conflicts among values and stakeholders still have to be governed; and action still requires authority and coordination. The relevant question therefore changes from whether AI reduces bounded rationality in the abstract to where the binding constraints move when model capability expands.

Under these conditions, a **generation–verification asymmetry** may arise. The supply of plausible outputs can grow faster than the organization's capacity to validate sources, inspect assumptions, test contextual fit, adjudicate disagreements, allocate attention, and assign responsibility. The phrase **verification scarcity** is used here descriptively for a possible organizational bottleneck when review demand exceeds available verification capacity. This capacity is treated as a scarce and allocable organizational resource rather than an automatic consequence of stronger models.

Redistribution of bounded rationality refers here to a change in the relative importance of constraints, not to a conserved quantity moving through the organization. "Relocated boundedness" and "verification scarcity" are descriptive shorthand for resulting bottlenecks, not additional independent constructs. Verification is not assumed to be the binding constraint in every episode. A more accurate model, better evidence infrastructure, or automated testing may reduce the verification burden, and organizations may restrict which outputs enter review.

Generative AI can therefore shift the dominant constraint from producing candidate answers toward judging them, from information access toward attention allocation, from cognitive production toward goal and premise specification, and from individual reasoning toward organizational governance. Relocation does not imply that all organizations experience the same bottleneck. The binding constraint is contingent and may move over time as models, expertise, evidence infrastructure, and routines change.

The conditional mechanism can be summarized without assuming that every capability improvement increases verification demand:

Change	Necessary condition for a verification bottleneck	Observable implication
Generation becomes cheaper	More material outputs are submitted for review	Review arrivals increase
Submitted outputs require assessment	Triage and automated checks do not offset the added workload	Required review time rises relative to available capacity
Required workload exceeds capacity	Review standards are maintained	Waiting time or backlog grows
Required workload exceeds capacity	Decisions proceed without added capacity	More material outputs receive incomplete verification

Improved accuracy, reduced checking effort, or stronger triage can interrupt this mechanism.

The claim is not that information ceases to be scarce, nor that every AI output is unreliable. In many domains, source access, proprietary knowledge, and current data remain scarce. The claim is comparative: when generation expands faster than verification, the binding constraint of a decision episode can move downstream.

Proposition 1 (generation–verification imbalance). *When lower generation costs increase the volume of material AI outputs submitted for review, while review capacity, task mix, and per-output verification requirements remain unchanged, the additional demand will increase review waiting time or the share of outputs used without completing required verification where spare review capacity is limited.*

The operative mechanism is induced review demand, not cheaper generation alone. The proposition predicts observable workload consequences; it does not establish that the predicted demand increase occurs in every deployment.

4.2 Premature satisficing and premise capture

Generative AI may also change the stopping problem. Fluent and apparently complete output may encourage acceptance before further search would reveal a material defect. Here, premature satisficing means accepting a recommendation that fails a decision-relevant evidential or performance criterion established independently of the observed acceptance decision, when a feasible check within the available time and resource budget would have exposed that failure. Ordinary satisficing can be rational; neither short search nor acceptance of AI advice is, by itself, evidence of premature stopping.

A related risk is unexamined adoption of AI-suggested premises, termed "premise capture" here only as descriptive shorthand. AI may shape problem definitions, categories, evidence selection, or evaluation criteria without making the final decision. An organization can thus retain formal approval authority while leaving substantive framing unexamined. Explanation acceptance in Bansal et al. (2021) offers a reason to investigate this possibility, not direct evidence for premise capture in organizational generative AI use. The proposed mechanism is that making premises and stopping conditions explicit creates an opportunity to discover a mismatch before commitment.

Proposition 2 (premature satisficing). *For comparable decision episodes with feasible diagnostic checks, requiring decision-makers to state the governing premises and stopping thresholds and to examine an alternative and counterargument will reduce acceptance of AI recommendations that fail independently specified decision criteria, relative to ordinary AI use without those requirements, provided that review time and expertise are sufficient.*

An equal-time comparison is needed to distinguish the effect of this structure from the effect of simply spending longer on a task. Reduced acceptance of defective advice should also be assessed alongside rejection of sound advice; indiscriminate skepticism is not improved judgment.

Figure 2 summarizes the conditional pathways through which generation and verification burdens can change.

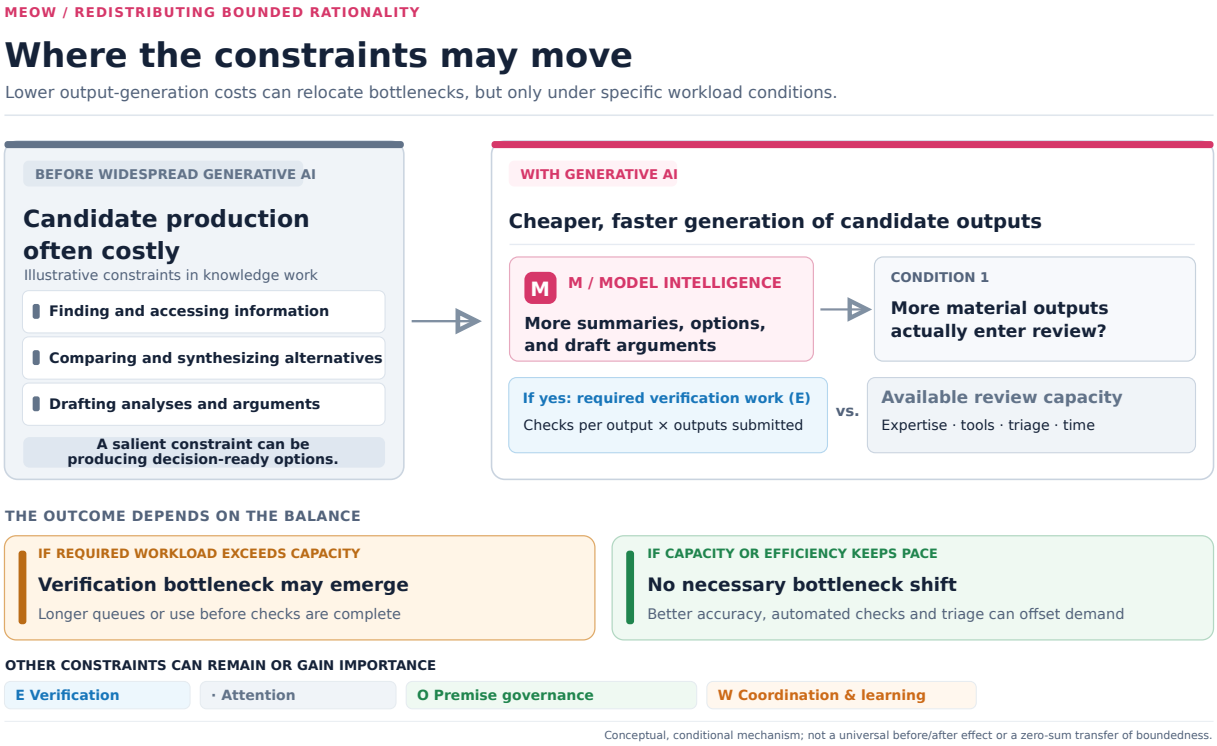


Figure 2: Conditional redistribution of bounded rationality as generation and verification demands change.

5. The MEOW Framework

MEOW organizes four dimensions of capability relevant to converting model outputs into accountable action. The labels denote different objects: M concerns the deployed model configuration, E concerns assessment, O names the premises being governed, and W names recurring practices. When O is described as a capability, it means the capacity to specify, authorize, contest, and revise those premises; their mere existence is insufficient. These dimensions are analytically separable but may share people and artifacts. O often precedes generation

by specifying goals and authority, and W surrounds the episode through coordination and review. MEOW is a configuration, not a causal pipeline or a claim that the four dimensions exhaust organizational rationality.

Capability	Definition	Core diagnostic question	Failure mode when weak	Observable evidence
M - Model Intelligence	The situated ability of a model, tools, data, and prompts to generate decision-relevant options, representations, and analyses	What task is the model expected to support, and where is its capability boundary?	Capability mismatch; confident output outside the model's effective frontier	Task specification, model/tool record, input provenance, baseline comparison
E - Epistemic Verification	The capacity to establish whether claims and reasoning are sufficiently grounded for the decision's risk	What must be checked, against which evidence, by whom, and to what threshold?	Hallucination, weak evidence, false confidence, uncalibrated reliance	Sources, tests, counteranalysis, expert review, uncertainty statement
O - Organizational Decision Premises	The capacity to specify, authorize, contest, and revise goals, values, criteria, constraints, classifications, and authority governing a decision	Which premises govern the decision, and who has authority to set or contest them?	Hidden criteria, unauthorized premise formation, value conflict, accountability gap	Decision criteria, risk appetite, authority map, rejected premises, dissent record
W - Work Routines	The capacity to enact and revise recurrent practices for preparation, review, authorization, action, records, and learning	How does the episode become repeatable, traceable, and improvable?	Shadow use, inconsistent review, missing provenance, no organizational learning	Workflow, checklist, decision log, escalation rule, post-decision review

5.1 From Model Intelligence to Organizational Rationality

Adding a more capable model does not necessarily improve the surrounding decision system. Its contribution depends on goals, evidence, authority, and opportunities for correction. This is consistent with the conditional nature of human–AI performance in prior research (Vaccaro et al., 2024), rather than a newly identified fallacy. MEOW asks which organizational conditions account for that dependency.

Model Intelligence is therefore a conditional organizational resource. Its value depends on the architecture into which it is inserted. The relevant unit is not the model in isolation but the system that converts generated cognition into authorized action. Organizational rationality is a system property: it emerges from the alignment of model capability with epistemic verification, legitimate premises, human and institutional judgment, and routines for action and learning.

This yields a conditional bottleneck implication: improvements in M may deliver smaller gains when verification, premise governance, or routines constrain the use of better outputs. The effect is not necessarily a diminishing-returns curve. A model improvement might also reduce error rates or make verification easier, thereby relieving an E bottleneck. Proposition 3 concerns moderation by independently measured organizational capabilities, not a universal weakest-link law.

5.2 M: Model Intelligence

Model Intelligence is not an abstract benchmark score. It is the task-relevant performance of a deployed configuration of model, retrieval, tools, data, prompts, and interface. For measurement, M concerns what that configuration can produce under specified use conditions; E concerns assessment of its outputs, O concerns the

governing premises, and W concerns how work is coordinated and repeated. The same model can be useful for divergent option generation but unsuitable for factual adjudication. Treating model intelligence as situated prevents the organization from transferring broad reputational claims about a model into a specific decision episode.

The M question is therefore not "Is this model smart?" but "What can this configuration reliably contribute to this decision, relative to available alternatives?" This includes identifying where AI is exploratory, where it is analytical, and where its output could become a formal decision basis.

5.3 E: Epistemic Verification

Epistemic Verification is the capacity to assess whether output is sufficiently supported and relevant for its intended use. It may involve source tracing, recalculation, replication, internal-data comparison, expert review, sensitivity analysis, and independent judgment. The NIST Generative AI Profile recommends checking sources and citations in generated outputs (Autio et al., 2024, action MS-2.5-003). In MEOW, adversarial prompting may identify questions to investigate, but another generated response is not treated as independent evidence without checking its basis. Where models share sources or failure modes, agreement need not provide independent corroboration. The evidential basis, rather than the number of approving outputs, determines what has been checked. The appropriate threshold depends on consequence and reversibility; verification is not synonymous with exhaustive proof.

E also requires calibration. Organizations must distinguish outputs used to stimulate thought from outputs used to justify action. A brainstorming suggestion may tolerate uncertainty that would be unacceptable in a safety, legal, financial, or public-impact decision.

5.4 O: Organizational Decision Premises

Organizational Decision Premises include factual assumptions and value judgments governing action: problem definitions, goals, success criteria, constraints, risk tolerance, and stakeholder obligations. Governance of those premises includes the authority to accept, challenge, and revise them. This is the framework's focal Simonian construct, not a new category of premise. E assesses the support for factual claims; O governs whether supported claims are relevant and whether the goals and constraints under which they are used are defensible. Evidence can inform value judgments without settling them.

The O capability makes premise formation visible and contestable. It asks which criteria were specified by the organization, which were introduced by the model, and which changed during deliberation. It also restores the distinction between advice and authority: AI may generate or evaluate a proposal, but the organization must define who is entitled to accept, reject, or revise its premises.

Some premises can be encoded in prompts, retrieval configurations, policy rules, and evaluation rubrics. Computational mediation of organizational knowledge and control predates generative AI (Faraj et al., 2018; Kellogg et al., 2020). The present concern is that generative systems can reproduce and reshape such premises across many outputs, including frames that were never explicitly specified. Encoding does not make values fully programmable: the organization still needs to examine whose criteria are represented and how revisions are authorized.

5.5 W: Work Routines

Work Routines turn episodic AI use into organizational capability. They specify when AI may be used, what records must be kept, who performs verification, how exceptions are escalated, and when outcomes are reviewed. Routines are not merely rigid controls. Their ostensive form provides a recognizable pattern, while their performance can adapt to local circumstances (Feldman & Pentland, 2003).

The W capability can support organizational memory. Research on routines links experience to organizational memory and the revision of practice (Feldman & Pentland, 2003; Levitt & March, 1988). Building on that account, MEOW proposes that a trace linking AI contributions, verification, premises, authorization, and outcomes can make later reconstruction more feasible and inform routine review. Such a trace is neither necessary for all informal learning nor sufficient to establish that AI caused a better outcome. Records should be proportionate, with access and retention limits appropriate to their sensitivity.

5.6 Capability complementarity and distributed constraints

MEOW hypothesizes complementarity among its capability dimensions. Increasing M while E, O, and W remain weak can make failure faster and more persuasive. Strong verification without explicit premises can establish

that a claim is factually supported while leaving its relevance or legitimacy unresolved. Clear premises without routines may depend on individual diligence and disappear under time pressure.

The relevant bounds may be distributed across people, models, evidence systems, authority, and routines. This system-level interpretation does not replace human bounded rationality or imply a new autonomous decision-maker. It directs empirical attention to interactions: whether better model output is used more effectively when the surrounding capabilities are stronger.

Proposition 3 (capability complementarity and bottlenecks). *For comparable tasks, the improvement in independently assessed decision quality produced by a demonstrated increase in M will be larger in teams with stronger E, O, and W than in teams with weaker complementary capabilities, when using the improved output requires verification, premise adjudication, and coordinated action.*

M must be measured through task-relevant performance outside the focal decision outcome, and E, O, and W through practices or resources observed before that outcome. This avoids defining stronger capabilities as whatever produced a better decision. The claim predicts positive moderation within its stated scope; it does not require multiplicative scores, a strict minimum function, or beneficial effects from every additional control.

Figure 3 presents MEOW as a configuration of complementary capabilities, rather than a mandatory M–E–O–W sequence.



Figure 3: MEOW as a configuration of complementary capabilities, not a fixed linear sequence.

6. From Framework to AI-Mediated Decision Architecture

An **AI-mediated decision architecture** is the arrangement of roles, evidence thresholds, authority, artifacts, gates, and feedback through which AI outputs enter organizational decisions. The relationship map below summarizes the architecture. Premises and authority set the boundary conditions. Model intelligence generates candidate outputs. Verification determines epistemic admissibility. Work routines carry the episode through authorization, action, recording, and review. Outcomes then modify future premises, verification rules, model configurations, and routines.

The term "conversion architecture" describes this arrangement rather than a separate central construct. Premises shape generation; verification can challenge factual premises or trigger reconsideration of criteria; authorization determines permitted action; and outcome review can revise all four dimensions. "Conversion" need not involve deliberate endorsement: unnoticed adoption of an AI-suggested frame is itself a failure that the architecture is intended to expose.

Relationship	Function in the decision episode	Return or revision path
O shapes M and E	Establish the problem, criteria, authority, and evidence standards	New evidence or disagreement can reopen premises
M supplies candidates for E	Generate representations, options, and analyses	Failed checks can lead to regeneration, new evidence, or abandonment
E informs authorization under O	Distinguish supported claims from unresolved assumptions	A decision may proceed within accepted limits, be deferred, or be rejected
W coordinates the episode	Allocate work and record authorization, action, and exceptions	Implementation problems trigger escalation
Outcome review revises M, E, O, and W	Compare expectations with observed consequences	Revise configurations, checks, premises, and practices for later episodes

These relationships can recur; the rows do not prescribe a fixed sequence.

6.1 Conversion relationships and allocation of judgment

The four MEOW capabilities should be read as coupled relationships. **M–E** asks whether increased generation is matched by sufficient verification. **E–O** asks whether factual support is relevant to the organization's stated goals, values, and authority. **O–W** asks whether legitimate premises are actually embedded in repeated practice. **W-to-M/E/O** captures organizational learning: outcomes can change model configurations, verification rules, premises, and future routines.

The architecture also allocates judgment within the structures examined by Shrestha et al. (2019) and the organizational arrangements examined by Saup et al. (2026). It asks what evidence and authority support the acceptance of a judgment under a given allocation. Some assessments may be delegated to models; others require domain experts, managers, or formal authorities. The design problem is to allocate judgment according to demonstrated capability, consequence, contestability, and accountability. Naming a human approver does not ensure meaningful oversight.

6.2 Five design elements

The framework proposes five design elements for evaluation in practice.

1. Decision classification. The organization identifies the decision, its consequence, reversibility, uncertainty, affected stakeholders, and the role assigned to AI.

2. Premise specification. Goals, criteria, constraints, evidence standards, and authority are stated before an AI-generated recommendation is treated as a decision basis. Premises can be revised, but revisions are made visible.

3. Verification design. The organization assigns checks, reviewers, and escalation thresholds in proportion to risk. Review competence and independence should match the risk; role separation is one option, while a separately performed evidence check may be feasible in a small team.

4. Decision gate and trace. The final decision distinguishes the AI contribution from the human or institutional judgment. It records accepted and rejected recommendations, unresolved uncertainty, the accountable owner, and the rationale for action.

5. Outcome review. The organization compares expected and actual outcomes and updates its prompts, sources, model selection, evidence thresholds, premises, and routines.

6.3 Risk-calibrated verification

Uniform verification can waste scarce review capacity or leave high-risk outputs insufficiently checked. Risk-based frameworks motivate context-sensitive controls, but they do not establish the optimal intensity for every decision (Autio et al., 2024; Papagiannidis et al., 2025). The table below proposes starting defaults, not legal requirements or validated thresholds. Risk classification should reflect the intended downstream use, including foreseeable effects on affected parties; an exploratory output must be reassessed before consequential use.

Decision tier	Typical characteristics	Proposed verification design	Record and authority
Exploratory	Low consequence, reversible, internal, no external commitment	Check material facts; label uncertainty; consider an alternative	Lightweight log; task owner decides
Consequential	Material budget, customer, staffing, operational, or reputational impact	Source trace; suitably independent check; explicit counterargument; premise check; sensitivity analysis where relevant	Decision record; named verifier and accountable manager
High stakes	Safety, legal rights, regulated activity, major financial exposure, public impact, or low reversibility	Primary evidence; qualified domain review; adversarial or independent analysis; formal escalation; documented residual risk	Formal approval gate; separation of proposer, verifier, and authorizer; scheduled outcome audit

Proposition 4 (risk-calibrated verification). *Across a portfolio of AI-mediated decisions with heterogeneous risks, allocating a fixed review budget using ex ante estimates of preventable error harm will reduce harm-weighted decision errors more than uniform review of equal total cost, when the risk estimates are informative and additional review can detect and correct material errors.*

Consequence, reversibility, uncertainty, external effects, and opacity inform allocation; none alone guarantees that more review will help. Review independence is useful when it supplies additional evidence or reduces shared blind spots. Harm weights and review costs, including delay, must be set before outcomes are assessed. Mandatory safeguards and non-negotiable rights remain constraints on allocation, rather than harms to be traded away. This formulation tests the benefit of risk calibration without assuming that unlimited verification is desirable.

6.4 Illustrative decision episode

Consider a hypothetical decision about launching a new service; this illustration is not an empirical case or evidence of effectiveness. A team uses generative AI to synthesize interviews, generate market segments, and draft a business case. Under a tool-use approach, the output may move directly into a presentation. Under MEOW, the organization first records the decision owner, success criteria, budget limit, risk tolerance, and evidence threshold (O). It specifies that AI may synthesize and generate options but may not establish market size without traceable evidence (M). Analysts check quotations against interview records, recompute market estimates from primary data, and formulate a counter-case (E). The proposal then passes through a meeting and approval routine that records accepted assumptions, rejected AI suggestions, unresolved uncertainty, and a six-month review date (W).

The same model may be used in both cases. The difference is not model intelligence; it is the architecture through which model output acquires organizational consequence.

6.5 From decision traces to learning

A useful decision trace links the AI contribution, verification, accepted premises, authorization, and observed outcome. Reconstructing these links may help teams distinguish recurring failure locations and decide what to revise. However, documentation alone cannot separate the causal contributions of AI, human judgment, implementation, and environmental change. That inference requires an appropriate comparison or additional evidence. The learning argument draws on the role of experience in routines (Levitt & March, 1988) and assumes that reviewers can report failures and implement revisions.

Proposition 5 (trace-enabled learning). *Across repeated, comparable decision episodes, teams that routinely review linked records of AI contributions, verification, accepted premises, authorization, and outcomes will identify recurring failure modes sooner and show fewer recurrences after revision than teams that review outcomes with output-only records, holding review time, expertise, and opportunity to revise practices comparable.*

The prediction applies when outcome feedback is sufficiently informative, records are usable, and failures can be reported without incentives for concealment. A factorial comparison of record linkage and review practice can distinguish the contribution of provenance from the effect of review itself.

Figure 4 maps the decision episode, including checks, authorization gates, return paths, and outcome review.



Figure 4: An AI-mediated decision episode with verification, premise review, authorization, and outcome learning.

7. Contributions and Implications

7.1 Theoretical contributions

To bounded rationality. MEOW specifies a conditional account of how cheaper generation may change which constraint binds. Its focus is the interaction among output demand, checking effort, review capacity, and premise governance, rather than a claim that information abundance or limited attention is newly discovered.

To organizational rationality. The paper separates model-level intelligence from organization-level rationality and treats the latter as a system property. This avoids additive accounts in which smarter components are assumed to produce a smarter organization automatically, and it directs attention to conversion bottlenecks among generation, verification, premises, authority, and routines.

To human–AI collaboration. The framework shifts the analytical focus from immediate user-model performance to the organizational passage through which an output becomes a decision premise. Complementarity is treated as institutional as well as cognitive.

To AI governance. MEOW uses the decision episode to connect governance practices to the frames, factual assumptions, and criteria actually accepted in a decision. This is an interpretive integration with existing governance work, not the first account of admission gates or accountability.

To organizational routines and learning. The framework makes AI decision provenance an input to routine dynamics. Logs, checklists, and review meetings are not administrative residue; they are artifacts through which organizations preserve, contest, and revise knowledge.

7.2 Managerial implications

The first managerial implication is to evaluate generative AI with two scorecards. A **productivity scorecard** asks about speed, cost, and throughput. A **decision-quality scorecard** asks whether AI broadened alternatives, improved evidence, made uncertainty and premises visible, clarified authority, and generated learning. Productivity is valuable, but it is not a proxy for organizational rationality.

The second implication is that organizations should invest in verification capacity as deliberately as they invest in model access. This includes source infrastructure, domain review, independent challenge, and escalation paths. If submitted verification workload grows beyond capacity, backlogs or incomplete checks may result unless outputs are triaged, demand is constrained, or checking becomes more efficient.

The third implication is to govern premise formation. Policies should distinguish exploratory AI output, advisory

output, and authorized decision input. Important decisions should reveal which frames or criteria came from the organization and which were suggested by AI.

The fourth implication is to prefer lightweight, risk-calibrated routines over universal bureaucracy. A two-minute pre-mortem or source spot check may be appropriate for a reversible decision; a decision affecting safety or legal rights may call for stronger evidence and role separation under the applicable requirements.

7.3 Practical contribution of MEOW

MEOW's practical contribution is to organize established design concerns around a common diagnostic question, not to claim that verification, oversight, provenance, or admission gates are new. The proposed value of the canvas and log is their explicit linkage of generated claims, accepted premises, authority, and later outcomes. Whether these artifacts outperform existing governance practices is an empirical question:

> *Through what evidential, authorized, and contestable process did this AI output become a premise of organizational action?*

This question can be applied to a strategy deck, approval memorandum, risk assessment, hiring recommendation, policy option, or supervised agentic workflow. The decision canvas, checklist, and log in the appendices operationalize the question and create material for future empirical research.

Figure 5 locates the five propositions within the framework as testable claims, not validated effects.

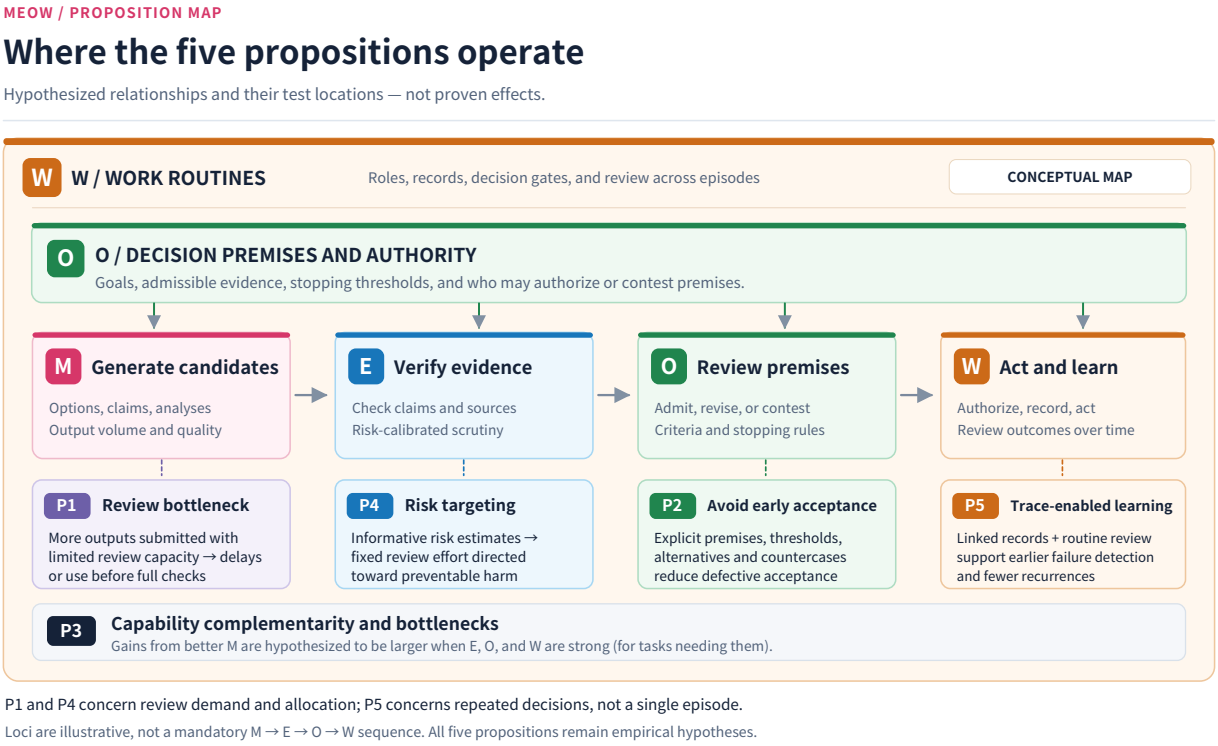


Figure 5: Illustrative locations of the five theoretical propositions within the MEOW architecture.

8. Boundary Conditions, Limitations, and Research Agenda

MEOW is conceptual and requires empirical validation. The propositions are conditional predictions, and the artifacts are design hypotheses. Existing studies motivate components of the argument; they do not validate the combined framework. Its incremental contribution should be tested against existing decision structures and governance practices, rather than only against an absence of review. Model capabilities and organizational practices are changing, so the framework remains revisable.

The framework also has important boundary conditions. In highly structured domains with formal verification or deterministic tools, technical controls may reduce the organizational verification burden. In creative exploration, heavy premise specification may suppress valuable surprise. In small teams, role separation may be impractical. In political organizations, explicit premises do not remove conflict; they may reveal it. Moreover, decision traces can support learning but also create surveillance, defensive documentation, or power asymmetries.

MEOW currently gives limited treatment to emotions, identity, professional jurisdiction, and the distributional consequences of AI adoption. Algorithmic systems can reconfigure control and power, not merely improve decisions (Kellogg et al., 2020). Future versions should examine who is entitled to define premises, whose knowledge counts as evidence, who bears verification labor, and who can contest an AI-mediated decision.

Empirical tests should specify the task, sampling frame, comparison, time horizon, and smallest practically meaningful effect in advance. Decision quality should be assessed independently of MEOW compliance, using task-appropriate accuracy, calibration, or attainment of predefined objectives. Accountability, stakeholder legitimacy, review burden, and adverse effects should be reported separately rather than collapsed into a favorable composite. Episodes are nested within teams, and nonrandom adoption can confound comparisons through expertise, resources, leadership, or task selection.

Proposition	Illustrative test and comparison	Observable outcomes	Evidence against the prediction within its stated scope
P1: generation–verification imbalance	Staged generation-cost reduction; track actual review arrivals, with limited spare capacity, task mix, and required checking effort held comparable	Review waiting time, backlog, and share used before required checks finish	Material review demand rises under limited spare capacity without greater delay or incomplete checking, beyond what added capacity or triage explains
P2: premature satisficing	Randomize structured premise/threshold/alternative review versus an equal-time comparison; use independently specified criteria and feasible checks	Acceptance of defective advice; rejection of sound advice; criterion attainment	Structured review yields no practically meaningful reduction in defective acceptance despite adequate time, expertise, and implementation
P3: capability complementarity	Factorial comparison of demonstrated M improvements and organizational supports; measure E, O, and W before outcomes	Task-specific decision quality and interaction effects; review burden	M gains are no larger with stronger complementary capabilities, or the interaction reverses, for tasks requiring those capabilities
P4: risk-calibrated verification	Assign comparable decision portfolios to ex ante risk-based or uniform review at equal total cost	Preweighted error harm; review and delay costs; safeguard compliance	Risk-based allocation fails to reduce weighted errors despite informative risk estimates and effective corrective checks
P5: trace-enabled learning	Longitudinal or factorial comparison of linked versus output-only records and routine review, with equal review resources	Time to independently confirmed failure-mode detection; recurrence per comparable opportunity after revision	Linked-record review yields no practically meaningful detection or recurrence advantage despite usable feedback and authority to revise

A null result is informative only when measurement, implementation, and statistical precision support it; post hoc changes to boundary conditions should not be used to shield a proposition from failure. Observational studies can assess associations and mechanisms but cannot establish causality without a defensible identification strategy. Further process research should examine judgment allocation and who performs, controls, or can contest verification work.

The practical artifacts also require validation. They should be treated as **design hypotheses and boundary objects for empirical refinement**, not as mature governance instruments. The ordinal checklist categories are not interval measures and should not be summed into an approval score. The readiness rules in Appendix B are design defaults, not empirically proven cutoffs. Case studies should begin with real decision episodes and revise the artifacts based on observed work rather than forcing work to fit the acronym.

9. Conclusion

Generative AI can make organizations faster at producing answers. It cannot by itself determine which answers deserve belief, which premises are legitimate, who may authorize action, or how the organization should learn from consequences. Those are problems of organizational design.

This paper has argued that generative AI can redistribute bounded rationality. When increased generation creates more review demand than the organization can absorb, constraints may shift toward verification and premise governance. That is a conditional mechanism, not an inevitable consequence of better models; capability improvements can also reduce the effort needed to check outputs.

MEOW—Model Intelligence, Epistemic Verification, Organizational Decision Premises, and Work Routines—organizes this argument around the AI-mediated decision episode. Its proposed contribution is a Simonian integration of how outputs become accepted premises and how those premises are authorized, enacted, and revised. The five propositions and accompanying artifacts make this integration open to empirical testing; their benefits remain to be demonstrated. The central message is that being smart is not enough: model intelligence alone cannot establish evidence-grounded, accountable, and revisable organizational decision-making, nor can internal authorization alone establish ethical legitimacy.

Declarations

Data availability. No datasets were generated or analyzed for this conceptual paper.

Conflict of interest. The author declares no conflict of interest.

Generative AI assistance. Generative AI tools assisted with literature discovery and organization, source checking, conceptual critique, language revision, and manuscript preparation. The manuscript was drafted using ChatGPT (OpenAI) under the direction and review of Kenichi Aoki, who takes full responsibility for its content, citations, and submission. AI-assisted source checking does not substitute for author review.

References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. S. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3411764.3445717

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. *The Quarterly Journal of Economics, 140*(2), 889–942. https://doi.org/10.1093/qje/qjae044

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction, 5*(CSCW1), Article 188. https://doi.org/10.1145/3449287

Csaszar, F. A., Ketkar, H., & Kim, H. (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science, 9*(4), 322–345. https://doi.org/10.1287/stsc.2024.0190

Dell'Acqua, F., McFowland, E., III, Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science, 37*(2), 403–423. https://doi.org/10.1287/orsc.2025.21838

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*(1), 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94–118. https://doi.org/10.2307/3556620

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review, 10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Levitt, B., & March, J. G. (1988). Organizational learning. *Annual Review of Sociology, 14*(1), 319–338. https://doi.org/10.1146/annurev.so.14.080188.001535

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Harvard University Press.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science, 381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems, 34*(2), Article 101885. https://doi.org/10.1016/j.jsis.2024.101885

Raisch, S., & Fomina, K. (2025). Combining human and artificial intelligence: Hybrid problem-solving in organizations. *Academy of Management Review, 50*(2), 441–464. https://doi.org/10.5465/amr.2021.0421

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Saup, T.-O., Asghar, J., Kanbach, D. K., & Kraus, S. (2026). From pilots to decision systems: Embedding generative AI into strategic decision-making through a socio-technical and governance lens. *Journal of Decision Systems, 35*(1), Article 2597835. https://doi.org/10.1080/12460125.2025.2597835

Schwarz, G., Christensen, T., & Zhu, X. (2022). Bounded rationality, satisficing, artificial intelligence, and decision-making in public organizations: The contributions of Herbert Simon. *Public Administration Review, 82*(5), 902–904. https://doi.org/10.1111/puar.13540

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83. https://doi.org/10.1177/0008125619862257

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1973). Applying information technology to organization design. *Public Administration Review, 33*(3), 268–278. https://doi.org/10.2307/974804

Simon, H. A. (1997). *Administrative behavior: A study of decision-making processes in administrative organizations* (4th ed.). Free Press. (Original work published 1947)

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour, 8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Appendix A. MEOW Decision Canvas

Purpose. The MEOW Decision Canvas provides a structured record for one AI-mediated decision episode. It is intended to make explicit the decision being considered, the role of generative AI, the evidence required, the premises governing judgment, the responsible actors, and the arrangements for subsequent review. The following page contains the printable canvas in A4 landscape format.

How to use the canvas

1. **Frame the decision.** State the choice or commitment, its consequences and reversibility, the decision owner, and the goals, criteria, values, and constraints that should govern it.
2. **Document the AI contribution (M).** Specify whether AI is used for exploration, synthesis, analysis, or recommendation. Record the model, tools, material sources and inputs, and known capability limits.
3. **Establish evidence and authority (E and O).** Define what must be checked, against which evidence, by whom, and to what standard. Identify who may set, challenge, revise, approve, or reject a decision premise, and make unresolved uncertainty visible.
4. **Embed review in work routines (W).** Identify the points for review, escalation, authorization, recording, and implementation. Specify when and how actual outcomes will be compared with expectations.

Interpretation and limits

The canvas is a *documentation and reflection aid*, not an automatic approval instrument or a validated scoring scale. Completing its fields does not establish factual accuracy, ethical legitimacy, or authorization to act. The depth and independence of verification should be appropriate to the decision's consequences, uncertainty, and reversibility. For consequential decisions, use the risk-calibrated checks in Appendix B and, where appropriate, maintain the linked decision record in Appendix C.

Worksheet on the next page: MEOW Decision Canvas (A4 landscape).

MEOW Decision Canvas

One canvas per decision episode | A record for review, not an automatic approval instrument

- M Model Intelligence
- E Epistemic Verification
- O Decision Premises
- W Work Routines

M AI role

Is AI exploring, summarizing, analyzing, recommending, or supplying a formal decision basis?

M Model configuration

Which model, tools, sources, inputs, and prompt conditions were used?

M Capability boundary

Which parts of the task are inside or outside the model's demonstrated competence?

Decision

What choice or commitment is being made?

E Evidence threshold

What must be verified before the output may influence action?

E Verification plan

Which claims, sources, assumptions, calculations, and counterarguments will be checked, by whom?

O Goals and criteria

What outcomes, criteria, values, and constraints govern the decision?

O Authority and contestability

Who may set, challenge, revise, approve, or reject the premises?

Consequence and reversibility

Who or what may be affected, and how easily can the decision be reversed?

Decision owner

Who is accountable for the final judgment and outcome?

W Workflow and gate

Where will review, escalation, approval, documentation, and action occur?

Residual uncertainty

What remains unknown or disputed at the time of decision?

W Outcome review

When and how will expected and actual outcomes be compared?

Appendix B. Risk-Calibrated MEOW Checklist

Use the ordinal categories **0 = absent, 1 = planned or incomplete, 2 = implemented for this episode with supporting evidence, 3 = implemented with supporting evidence and reviewed across prior episodes**. A first-time decision can satisfy an item at level 2. Do not sum the categories or use them as a validated measurement scale. Select applicable critical items before reviewing the output; record reasons for any non-applicable item. A material unresolved item blocks progression to consequential use under these proposed defaults, irrespective of ratings elsewhere.

B.1 Critical items for every consequential decision

Capability	Critical check
M	The role of AI and the boundary between generated output and final judgment are explicit.
M	The model, tools, material inputs, and relevant source provenance are recorded.
E	Decision-critical factual claims and calculations are supported by appropriate checks, or are explicitly identified as assumptions with their consequences assessed.
E	Uncertainty, counterarguments, and known capability limits are stated.
O	Goals, criteria, constraints, and risk tolerance are explicit.
O	The accountable decision owner and authority to approve are explicit.
W	Verification and authorization are recorded in the decision workflow.
W	A review trigger or outcome-review date is defined.

B.2 Readiness judgment

- **Green:** Every applicable critical item reaches at least category 2; verification matches the risk; the authorized owner accepts the documented residual uncertainty within applicable constraints. This is a readiness judgment, not a guarantee of correctness or legitimacy.
- **Yellow:** A material check, premise, role, or record is incomplete but can be resolved; use remains exploratory until the gap is closed or the decision basis is revised.
- **Red:** The proposed use depends on a material claim that cannot be adequately supported, lacks an accountable authorizer, or exceeds accepted risk or applicable constraints; the output must not serve as that decision basis.

B.3 Additional checks by risk

"Required" denotes a proposed MEOW default once adopted for a use case, not a claim about universal regulatory requirements. Document any equivalent control and its rationale. The risk tiers and control choices should be calibrated before consequential use.

Check	Exploratory	Consequential	High stakes
At least one non-AI or independently generated alternative	Recommended	Required	Required
Direct source or primary-data verification	When material	Required for material claims	Required
Independent check or reviewer	Optional	Required; evidence-based check with suitable independence	Required with role separation

Check	Exploratory	Consequential	High stakes
Domain-qualified expert	As needed	As needed	Required
Explicit counter-case or pre-mortem	Recommended	Required	Required and documented
Sensitivity or robustness analysis	Optional	When decision depends on estimates	Required where applicable
Legal, security, ethics, or compliance review	As applicable	As applicable	Formal review where applicable
Residual-risk acceptance	Informal	Named owner	Formal authorizer
Decision provenance retained with access and retention limits	Lightweight	Required	Required and auditable
Scheduled outcome review	Recommended	Required	Required with escalation criteria

Appendix C. AI-Mediated Decision Log

Field	Record
Decision ID and date	
Decision and accountable owner	
Risk tier and rationale	
AI role in the episode	
Model, tools, inputs, and sources	
Material AI outputs used	
Material AI outputs rejected	
Verification performed and verifier	
Goals, criteria, constraints, and authority	
Accepted and revised premises	
Alternatives and counterarguments considered	
Final decision and rationale	
Residual uncertainty and accepted risk	
Approval and implementation owner	
Expected outcome and indicators	
Review date or trigger	
Actual outcome	
Lessons and changes to M, E, O, or W	
Record access, retention, and redaction rules	