Journal Article
Journal of Generative Artificial Intelligence in Public Sector
www.aipost.com
Volume 1 July 2025 Issue 1

The Use Case Illusion: Why the Public Sector's Approach to AI Is Undermining Transformation

By ALI (AL) NAQVI*

Abstract

The public sector's prevailing approach to artificial intelligence (AI) emphasizes use cases and pilot projects as indicators of progress. While well-intentioned, this mindset is deeply flawed. Measuring AI maturity through the number of projects undertaken leads to fragmented, siloed automation efforts that lack systemic coherence and fail to deliver strategic transformation. This article argues that the "use case mindset" stems from legacy business process reengineering paradigms and remains fundamentally ill-suited to generative AI and other advanced systems. The goal of AI is not merely task-level automation but the reconfiguration of work itself – both cognitive and physical – across the organizational graph. Public institutions should move beyond linear workflows and embrace models that treat agencies, economies, and even governments as complex adaptive systems. Only through this systems-based lens can GenAI fulfill its potential to increase institutional productivity, responsiveness, and strategic capability. The article concludes with a call to redefine AI strategy away from pilot counting and toward full-system optimization, offering a framework for agencies to escape the use case trap.

Keywords: US Government, Use Cases, Artificial Intelligence, GenAI

* Ali (Al) Naqvi is the Chief Executive Officer of the American Institute of Artificial Intelligence

1. Introduction: The Illusion of Progress

Across the U.S. public sector, AI progress is routinely narrated through inventories of "use cases" and project counts. Agencies are required by OMB Memorandum M-24-10 and CIO Council guidance to compile annual, machine-readable catalogs that describe each AI application, indicate stage of deployment, and flag "rights- or safety-impacting" uses [1]. This reporting regime has produced large, public repositories – over two thousand entries drawn from more than forty agencies as of early 2025 – and has normalized a metric in which *more entries* appear to signal *more progress*. Yet such aggregation risks conflating administrative activity with progress and institutional transformation, particularly when the inventories themselves are silent on productivity, cross-workflow coherence, capability development, and mission outcomes.

In this context, the "use case" has become the de facto *unit of strategy*. Operationally, a *use case* denotes a bounded application of AI to a specific task or problem (e.g., document summarization for claims, a citizen-service chatbot, or a fraud-risk triage model). A *project* is the vehicle that funds and executes that bounded application along a lifecycle (exploration, pilot, deployment). Inventories therefore tally projects that instantiate use cases. The seductive simplicity of this framing is managerial: use cases are discrete, estimable, and easy to count; projects are procureable, schedulable, and easy to report. But a list of use cases is not a strategy – it is a collection of point solutions anchored to legacy processes, with no necessary guarantee that they interoperate, scale across silos, or compound into system-level gains. The very guidance that standardizes inventories emphasizes classification and compliance (e.g., rights, governance, ethics, or safety-impacting determinations) rather than systemic redesign, thereby reinforcing a project-centric optics of success [2].

This article interrogates that optics. It argues that counting use cases and projects systematically overstates progress while under-measuring transformation, and that a reliance on inventory metrics can entrench "islands of automation." As the consolidated federal catalogs expand – periodically updated with additional entries from new and existing agencies – the risk is that institutional attention tracks the growth of the spreadsheet rather than the growth of capability [3]. We therefore advance an alternative analytic lens for public-sector AI: shifting from a task- and project-bounded paradigm to a system-level, outcome-oriented view that evaluates whether AI reconfigures work (and its interdependencies) in ways that measurably improve mission results. The remainder of the paper develops this claim, situates it within federal policy context, and proposes evaluation criteria that privilege coherence, capability, and impact over inventory size.

We can capture the evolution of Use Case in four stages:

Stage 1 Early Stages: In the 1980s and 1990s we observed the rise of "Use Case" when the term began as a software requirements artifact [4] – a structured narrative of how an *actor* interacts with a *system* to achieve a goal (Jacobson's Objectory, then UML formalization). It is explicitly human–system, goal–flow oriented, built for clarity, testability, and traceability in functional requirements.

Stage 2 Consolidation & Practice (1990s–2000s): Cockburn's "actors & goals" formalized use cases and standardized templates and writing discipline [5]. That is when use cases became the lingua franca for scoping functionality and acceptance tests, and a backbone for stakeholder alignment and documentation.

Stage 3 Agile Adaptation (Use Case 2.0): As delivery shifted to agile, teams kept the narrative power of use cases but sliced them into incremental, releasable "use-case slices", often pairing with user stories for sprint-scale work. The concept retained rigor while gaining iteration speed.

Stage 4 Semantic Expansion in AI Era (2010s–today): In AI, "use case" broadened into a business-level label (e.g., "fraud detection," "document summarization"), and a unit for portfolio planning and governance (AI registries; risk triage). It ceased to be only a stepwise interaction script and became a strategic tag for applications – useful for visibility and oversight, but looser in precision.

2. The Use-Case Mindset: Origins and Shortcomings

2.1 Definition of the Mindset

By "use-case mindset" I refer to a planning and reporting posture that treats the use case – a bounded scenario describing how an *actor* (typically a user) interacts with a *system* to achieve a goal – as the *primary unit of strategy*. In software engineering and Human Computer Interaction (HCI), a use case classically captures a dialogue between an external actor and a system, enumerating steps, alternatives, and postconditions. This framing is explicit in the foundational literature (Jacobson's OOSE tradition; Cockburn's requirements guidance) and in UML's formalization of "actors" and "use cases." In short, the mindset assumes that *progress = more well-specified actor–system interactions implemented as projects*. [6]

2.2 Intellectual Lineage: BPR and Industrial Process Logic

The use-case mindset inherits much of its appeal from the Business Process Reengineering (BPR) era, which privileged decomposition of work into tasks and linear processes amenable to redesign and automation. BPR's promise – dramatic performance gains via radical process redesign – encouraged organizations to view technology as a means to streamline discrete workflows. In public administration, this translated into projectized interventions against specific processes, often evaluated by throughput and cycle-time metrics rather than system-level effects. The result is conceptually tidy portfolios of "use cases," each anchored to an extant process rather than to emergent, cross-boundary capability. [7]

2.3 Why This Template Misfits Generative AI

Generative AI (GenAI) exposes several limits of the use-case template:

a. Unit of analysis – Use cases privilege *user–system* interactions; GenAI routinely operates across *system–system* and *agent–agent* interactions (e.g., autonomous agents negotiating tasks), where no single "primary actor" or stable dialogue suffices. The UML/Cockburn framing is necessary for requirements capture but proves insufficient for modeling multi-actor, multi-modal, continuously learning systems [6].

b. Determinism vs. generativity – Use cases assume relatively deterministic scenarios with enumerated alternatives. GenAI produces probabilistic, context-compositional outputs whose value often lies in *reconfiguring* tasks and information flows, not merely executing a predefined script.

c. Local tasks vs. system behavior – The use-case lens optimizes local tasks; GenAI's highest leverage appears when it alters the topology of work (who does what, in what sequence, with which artifacts), i.e., when organizations are treated as systems of interdependent nodes rather than pipelines of isolated steps.

d. Static boundaries vs. permeable ecologies – Use cases typically presuppose a boundaryable "system under consideration." GenAI thrives in permeable data and capability ecologies (cross-silo retrieval, model ensembles, agent swarms), where value emerges from *interoperation* and *coordination*, not from atomized implementations.

e. Compliance optics vs. capability growth – Inventories of use cases, standardized for reporting and compliance, encourage counting and classification at the expense of coherence and compounding capability (e.g., shared models, shared data planes, shared assurance). The governance machinery that catalogues uses is valuable, but it can unintentionally entrench project-centric optics. [8]

2.4 Interim Conclusion

In sum, the use-case mindset accurately describes how a user and a system interact, and it remains a useful artifact for requirements elicitation and local delivery. But as a governing logic for GenAI strategy, it under-specifies (i) multi-actor agency, (ii) emergent, non-deterministic behavior, and (iii) the system-level reconfiguration that GenAI enables. Treating inventories of such use cases as proxies for transformation thus risks mistaking activity for capability and projects for progress.

3. Where This Mindset Comes From – and Why It Misfits GenAI

The contemporary use-case mindset inherits its appeal from several intertwined traditions. First, it aligns naturally with the logic of business process reengineering (BPR) and industrial efficiency: decompose work into tasks, optimize the flow between them, and measure cycle time or throughput. Within that paradigm, a use case is an ideal scoping device – clear about user goals, explicit about preconditions and postconditions, and readily testable – so it reliably delivers *local* improvements to a bounded workflow. Second, public-sector technology is typically procurement-driven and projectized. Budgets, schedules, and oversight mechanisms require discrete, auditable units, and the "AI use case" serves as a convenient procurement object and registry entry. This encourages the growth of catalogs of point solutions rather than the cultivation of shared capabilities that compound across missions. Third, classical use cases are rooted in human-centric task decomposition. They formalize an actor–system dialogue (main flows and alternates) and privilege bounded interfaces where a person initiates and supervises the interaction. That modeling choice, powerful for requirements engineering and HCI, underrepresents system-to-system, agent-to-agent, and other emergent patterns of coordination that increasingly characterize contemporary AI ecosystems.

These roots render the use-case template ill-fitted to generative AI. Use cases presuppose enumerable scripts; GenAI is probabilistic and generative, often delivering value precisely by reconfiguring tasks and the topology of information rather than executing a predefined pathway. Traditional actor–system narratives assume a primary human interlocutor; modern deployments increasingly involve agentic ecologies – multiple AI services negotiating, planning, and verifying one another's outputs – where value arises from coordination dynamics, not a single dialogic exchange. Catalogs of use cases are excellent for visibility and risk triage, but they privilege countable projects over the properties that determine institutional transformation: coherence across silos, reusability of models and data planes, and compounding capability through shared infrastructure and learned policy. Finally, as an artifact of requirements, the use case excels at describing how existing work should be automated; GenAI invites prior questions – whether the work should exist at all, where to relocate cognition along socio-technical boundaries, and how to re-architect the organizational graph to achieve mission outcomes.

In short, "use case" has evolved from a precise engineering device to a convenient portfolio and governance label, and it remains valuable for communication, traceability, and oversight. But when elevated to the governing logic of AI strategy, it binds institutions to project-centric, task-bounded thinking that systematically undermeasures systemic capability, interoperation, and mission impact – the very arenas where GenAI yields step-change value. Accordingly, this article argues for retaining use cases for cataloging and compliance, while replacing them as the unit of transformation with system-level, outcome-linked capability models that explicitly reward coherence, reuse, and compounded learning across the enterprise.

4. What AI Actually Enables

At its core, contemporary AI – especially large, generative, and agentic systems – is not merely an automation technology. It functions as a cognitive reconfiguration layer that can reorganize information flows, decision rights, and work topologies across an institution. Rather than optimizing a predefined sequence of steps, AI can surface alternative problem framings, synthesize multi-modal evidence, and continuously adapt outputs to shifting context – properties that move beyond classic, task-bound automation. Public guidance already recognizes this socio-technical, system-level character of AI and encourages organizations to evaluate AI not only at the component or application level, but across interactions, contexts, and organizational processes, underscoring that risk and value emerge from the system as a whole [9].

First, AI enables institutions to question the purpose of the task itself, not just how to execute it faster. By generating alternatives, counterfactuals, and synthesized rationales, generative systems can reveal when a task is duplicative, mis-scoped, or better relocated to a different point in the workflow (or eliminated altogether). Evidence from applied domains – such as clinical and administrative uses of LLMs – shows that the principal gains often arise from rethinking information work (summarization, triage, drafting, coordination), not merely automating a narrow step, suggesting a broader reframing of what the task should be [10].

Second, AI allows organizations to rethink institutional boundaries. When models can retrieve across silos, reason over heterogeneous data, and interface with other services via tools and APIs, the relevant unit of design shifts from the single process to interdependent systems. Defense research has articulated this as "mosaic" or system-of-systems thinking – composing capabilities dynamically across platforms and echelons – an idea that generalizes to civilian agencies as cross-unit assembly of data, models, and services in pursuit of mission outcomes [11].

Third, AI now permits machines to collaborate in cognition. Multi-agent systems (MAS) and emerging "multi-AI" collaboration frameworks demonstrate how specialized agents can plan, critique, verify, and negotiate with one another to complete complex tasks – behaviors that exceed the classic actor–system dyad of legacy use-case modeling. This agentic ecology foregrounds coordination, role assignment, and protocol design (who does what, when, with which information), making collaboration a first-class design variable rather than an afterthought [12, 13].

Finally, these properties introduce the practical possibility of system-level intelligence: organizations that learn, adapt, and self-reconfigure as complex adaptive systems (CAS). In such systems, value emerges from the interactions among many semi-autonomous components ("agents") that co-adapt over time; AI provides both the computational substrate (models, agents, tool-use) and the governance prompts (profiles, controls) to make this tractable within public institutions. Designing for CAS dynamics – rather than optimizing isolated tasks – aligns evaluation with coherence, compounding capability, and mission outcomes, which system-level frameworks like the NIST AI RMF explicitly encourage [14, 9].

Implication. If AI is treated as cognitive reconfiguration rather than point automation, the unit of strategy must shift accordingly: from counting use cases to engineering system behavior – how information, authority, and action propagate across the enterprise under algorithmic mediation.

5. From Linear Workflows to Complex Adaptive Systems

Public institutions are often managed and measured as if work proceeds along linear workflows – stable, decomposable processes with fixed roles and handoffs. A more accurate and useful lens for AI-era transformation is the complex adaptive system (CAS): a system composed of many interacting components ("agents") whose collective behavior emerges from local interactions and adapts over time through learning and feedback. In plain terms, a CAS is an organization that changes how it works as it works, because the parts influence one another and update their behavior in response to outcomes. Foundational accounts emphasize distributed control, rich interdependence, and adaptation as defining features of complexity in social and institutional systems [15].

A CAS view invites graph-based thinking about institutions: people, services, data stores, and algorithms are nodes; relationships, handoffs, and data flows are edges. Network science provides language and tools – paths, centrality, communities, and bottlenecks – to analyze how information and authority propagate, where failures concentrate, and which subgraphs form emergent "functions" even when no single process description exists. This perspective enables optimization not only of steps within a process, but of the topology of the organization – which nodes should connect, which bridges reduce distance, and which communities should be reconfigured to improve outcomes [16, 17].

Within this systems frame, generative AI is not merely a faster step in a fixed chain; it is a cognitive reconfiguration layer that alters the graph itself:

- Rerouting information flows. Retrieval-augmented generation, tool-use, and multi-agent orchestration allow models to pull from, write to, and coordinate across multiple nodes, dynamically re-wiring who informs whom and in what sequence. In practice, this looks like AI agents that plan, critique, and hand off tasks to one another – changing "who talks to whom" inside the enterprise without a human specifying every pathway [18, 19].
- Redesigning work clusters. Network methods identify tightly connected subgraphs ("communities") that function as de facto work clusters. GenAI can consolidate or redistribute their cognitive load (e.g., summarization, triage, drafting, adjudication), enabling new cluster boundaries that cut across legacy silos and shorten decision paths [17].
- Discovering new configurations of mission execution. By composing capabilities across heterogeneous services and teams – often in system-of-systems fashion – AI supports agile recombination of sensors, data, models, and human roles for a given objective. This is the institutional analogue of "mosaic" assembly in defense: building larger, adaptive effects from interoperable, disaggregated pieces. For civilian agencies, the same principle enables cross-program tasking, shared data planes, and reusable model services that assemble on demand around a mission [20].

Designing for CAS dynamics aligns with contemporary governance guidance that treats AI as socio-technical and system-level: risk and value emerge from interactions among models, data, people, and procedures, not from components in isolation. Evaluating and steering AI at this level means optimizing coherence, compounding capability, and mission outcomes – not merely counting automated tasks – so that the organization learns to reconfigure itself in response to evidence [9].

6. Strategic Consequences

A use-case/counting posture fragments AI effort – and with it, state capacity. When agencies optimize for inventories of discrete projects rather than for coherence of shared data, model services, and cross-workflow learning, the result is a patchwork of "islands of automation." The federal reporting regime formalizes this bias: OMB's M-24-10 requires agencies (with limited exceptions) to enumerate AI use cases annually and post public inventories, a practice that has produced large consolidated catalogs across dozens of agencies. Recent consolidations and compliance plans describe these inventories in detail and emphasize classification and disclosure – important for transparency, but not substitutes for system design [8, 21].

The risk is measurable: oversight bodies now document rapid growth in reported use cases – for example, GAO notes that counts roughly doubled from 2023 to 2024 at a set of large agencies and that generative-AI use cases increased sharply – yet also catalog persistent governance, workforce, and integration challenges that impede impact. Counting activity, in other words, can outpace alignment [21, 22].

This fragmentation carries national-level consequences. For national security, NSCAI's final report frames AI as a strategic, system-of-systems capability – warning that the United States must organize for integrated adoption to remain competitive. Fiscal sustainability is likewise implicated: duplicative point solutions and siloed procurements raise lifecycle costs while under-delivering shared capability. And for service equity and legitimacy, federal policy explicitly recognizes "rights- or safety-impacting" AI and calls for protections against algorithmic discrimination; a fragmented implementation landscape complicates consistent safeguards across programs and jurisdictions [23, 21, 8, 24].

Meanwhile, peer competitors are moving toward more integrated, systemic AI approaches. China's New Generation AI Development Plan (2017) articulates a top-level design to 2030, and current "AI+" policies stress whole-of-nation deployment across sectors – an explicitly coordinated posture that seeks compounding effects rather than isolated pilots. Independent analyses describe this as a state-directed, vertically integrated model across the AI stack. While the efficacy of such policies is debated, the strategic intent is clear: alignment, not merely activity [25, 26].

Finally, U.S. guidance already points beyond inventories. The NIST AI Risk Management Framework treats AI as a socio-technical, system-level phenomenon – placing emphasis on interactions, contexts, and organizational processes. If agencies adopt RMF-style lenses while continuing to report use cases for transparency, they can pivot from project counting to capability alignment: shared data planes; reusable model services; common assurance; and outcome-linked metrics. In short, the United States cannot afford to confuse activity with alignment. The policy scaffolding exists; the strategic task is to organize for coherent, compounded capacity rather than a larger spreadsheet [9].

7. Call to Action: Escaping the Use-Case Trap

Escaping the use-case/counting mindset requires replacing project-by-project optimization with system design. The practical path is diagnostic first, redesign second, and institutional alignment throughout.

Map cognitive workflows. Begin with a cognitive work map – a graph of how information, judgment, and authorization move through the institution. Go beyond swimlanes and SOPs: enumerate decision points, evidence requirements, latency tolerances, handoffs (human↔human, human↔system, system↔system), and failure modes. Treat people, services, data stores, and models as nodes, and their dependencies as edges. The artifact should make visible where cognition is duplicated, starved, or delayed.

Identify redundancies and chokepoints. Use the graph to locate (i) redundant judgments (multiple units re-interpreting the same evidence), (ii) serial bottlenecks (single nodes that gate many downstream actions), (iii) long paths (excessive hops between evidence and decision), and (iv) orphan outputs (work products generated but rarely consumed). These are the targets for consolidation, parallelization, or removal.

Use GenAI for synthetic redesign – not bolt-on automation. Treat GenAI as a cognitive reconfiguration layer:

- Reroute flows by inserting retrieval-augmented agents that deliver just-in-time evidence to the point of decision.
- Collapse steps by co-locating summarization, drafting, critique, and adjudication in a multi-agent pattern (planner, solver, verifier).
- Relocate cognition by shifting routine judgments from scarce expert nodes to supervised AI agents, reserving humans for exception handling and policy setting.
- Remove work that becomes unnecessary once upstream information is synthesized (design for “non-events,” not just faster events).

- Document these changes as capability patterns (reusable blueprints that specify inputs, guardrails, roles, and expected outcomes), not as isolated use cases.

Align funding, procurement, and governance to systemic outcomes.

- Funding. Budget for shared capabilities (data planes, model services, assurance tooling) rather than one-off pilots. Create line items for platform teams and cross-program enablement, with Service Level Objectives tied to mission outcomes (e.g., decision cycle time, error rates, equity measures), not project counts.
- Procurement. Specify interoperability and reuse as first-order requirements (APIs, model cards, evaluation protocols, lineage), and score offers on contribution to shared capability (not just local fit). Prefer modular contracts that allow composition and substitution of models/agents over time.
- Governance. Replace inventory-centric dashboards with system health dashboards: coherence across silos, reuse ratios, outcome deltas, assurance coverage, and incident learning. Institutionalize AI assurance (risk, testing, monitoring) as a continuous function embedded in the platform, not a one-time gate at project end.

Measure what matters. Retire "number of use cases" as a success metric. Track mission-linked outcomes (timeliness, accuracy, equity), topology metrics (average path length from evidence to decision; reduction in redundant nodes), and capability compounding (percentage of workloads using shared models/data; rate of pattern reuse). Publish deprecation plans for legacy steps that redesign makes obsolete.

Organize to sustain change. Stand up a cross-functional AI platform team (engineering, data, security, policy, evaluation) with a mandate to deliver reusable services and patterns. Pair it with mission design cells that apply those patterns to high-value workflows and run controlled trials with rigorous evaluation. Establish a policy-tech review cadence where doctrine, controls, and capabilities evolve together based on evidence.

Codify the portfolio. Maintain a capability portfolio (not a use-case list) that articulates: (i) shared services available, (ii) the patterns they enable, (iii) adoption and reuse metrics, and (iv) outcome impacts across programs. Use the portfolio to guide sequencing, investment, and sunset decisions.

Taken together, these steps shift the unit of strategy from projects to properties of the system – coherence, reuse, assurance, and measurable mission impact – so that GenAI is used to redesign how the institution thinks and acts, rather than to decorate existing processes with isolated automations.

8. Conclusion

The public sector's prevailing reliance on use-case inventories and project counts has produced an illusion of progress while entrenching structural fragmentation. The cost of this incoherence is tangible: duplicated effort across silos, brittle point solutions that do not interoperate, escalating lifecycle costs, uneven safeguards, and – most importantly – mission outcomes that fail to improve commensurately with investment. Counting implementations is administratively convenient; it is not analytically meaningful. A larger spreadsheet of isolated automations does not constitute a more capable state.

Generative AI sharpens this diagnosis and widens the opportunity. Its value does not lie primarily in accelerating predefined steps, but in reconfiguring cognition and coordination across the enterprise. That requires moving beyond the question "Which tasks can we automate?" to the prior and more consequential questions: What is the work now? Where should cognition live? How should information, authority, and action propagate? In other words, GenAI demands a redefinition of work, not merely faster execution of legacy workflows.

Accordingly, the unit of strategy must shift from the use case to system-level capability – shared data planes, reusable model services, multi-agent patterns, and embedded assurance that compound across programs. Evaluation must likewise pivot from activity metrics to outcome and topology measures: coherence across silos, reuse ratios, shortened evidence-to-decision paths, improved timeliness, accuracy, equity, and resilience. Institutions that organize around these properties will see GenAI translate into durable capacity; those that do not will continue to amass isolated projects and underperform at the mission edge.

The choice before the public sector is therefore clear: persist with a project-centric optics that mistakes activity for alignment, or design for complex, adaptive systems in which intelligence is a property of the whole. Only the latter approach is proportionate to the promise – and the stakes – of the present moment.

References

1 CIO.gov Report Guidance for 2024 agency artificial intelligence reporting per EO 14110 (2024) https://www.cio.gov/assets/resources/2024-Guidance-for-AI-Use-Case-Inventories.pdf

2 https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

3 https://flatgithub.com/ombegov/2024-Federal-AI-Use-Case-Inventory?filename=data%2F2023_consolidated_ai_inventory_raw.csv&sha=2a91f52c539771a57e5f8db19f1560f197a12a6f

4 Jacobson, I. (1987, December). Object-oriented development in an industrial environment. In *Conference proceedings on Object-oriented programming systems, languages and applications* (pp. 183-191).

5. Cockburn, A. (1997). Structuring use cases with goals. *Journal of object-oriented programming*, *10*(5), 56-62.

6. Cockburn, A. (2000). Writing Effective Use Cases. https://people.inf.elte.hu/molnarba/Informaciorendszerek_ELTE/Writing_effective_Use_cases_Cockburn.pdf

7. https://www.sciencedirect.com/topics/computer-science/reengineering-process

8. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

9. Artificial Intelligence Risk Management https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

10. Meng, X., Yan, X., Zhang, K., Liu, D., Cui, X., Yang, Y., Zhang, M., Cao, C., Wang, J., Wang, X., Gao, J., Wang, Y. G., Ji, J. M., Qiu, Z., Li, M., Qian, C., Guo, T., Ma, S., Wang, Z., Guo, Z., … Tang, Y. D. (2024). The application of large language models in medicine: A scoping review. *iScience*, *27*(5), 109713. https://doi.org/10.1016/j.isci.2024.109713

11. Serbu, J. (2020) *DARPA's contribution to JADC2: 'Mosaic' warfare* *https*://federalnewsnetwork.com/defense-main/2020/12/darpas-contribution-to-jadc2-mosaic-warfare/

12. Zhang, X, Dong, X., Wang, Y., Zhang, D., Cao, F A Survey of Multi-AI Agent Collaboration: Theories, Technologies and Applications. https://arxiv.org/html/2501.06322v1

13. Tran, K., Dao, D., Nguyen, M., Pham, Q,, O'Sullivan, B. Nguyen, H. Multi-Agent Collaboration Mechanisms: A Survey of LLMs https://arxiv.org/pdf/2501.06322v1

14. Holland, John H., 'Complex adaptive systems (CAS)', *Complexity: A Very Short Introduction*, Very Short Introductions (Oxford, 2014; online edn, Oxford Academic, 24 July 2014)

15. Mitchell, M. Complexity: A Guided Tour (2009) Oxford University Press

16. Newman, M. E. J. (2010), Networks: An Introduction, Oxford University Press

17. Barabasi, A. L. (2015), Network Science, Cambridge University Press

18. Wu, Q et al (2023) AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation https://arxiv.org/abs/2308.08155v2 https://doi.org/10.48550/arXiv.2308.08155

19. Multi-agent Conversation Framework https://microsoft.github.io/autogen/0.2/docs/Use-Cases/agent_chat/

20. https://www.darpa.mil/news/mosaic-warfare

21. https://files.gao.gov/reports/GAO-25-107653/index.html

22. https://www.nextgov.com/artificial-intelligence/2025/07/agency-ai-use-doubled-2024-gao-finds/407067/

23. https://reports.nscai.gov/final-report/

24. https://www.gao.gov/assets/gao-24-107332.pdf

25. https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

26. https://www.geopolitechs.org/p/china-releases-ai-plus-policy-a-brief