NYU | DRI
Development Research Institute

Harnessing Agentic AI: The Public Sector Can Gain as Well

November 13, 2026

Jerry Haar and Driss Temsamani

If ever there was a "gamechanger," it is artificial intelligence (AI). And while *generative* AI automates a range of functions, *agentic* AI acts and makes decisions in a way that a human might, according to MIT professor John Horton.

Recognizably, agentic AI is regarded as the domain of the private sector. For example, LinkedIn deploys a multi-agent recruiter workflow (using a supervisory orchestrator and specialized sub-agents) as its hiring assistant to automate candidate sourcing and multi-step evaluation. In another case, Google implements agentic commerce and checkout features that allow shoppers to set target prices, track real-time inventory updates, and have Google autonomously execute purchases on their behalf via Google Pay.

Within the public sector, agentic AI can improve performance when it is used as a *bounded digital operator:* a system that can interpret a goal, plan a sequence of steps, retrieve information, call approved tools, execute permitted actions, check results, and escalate exceptions. Unlike a conventional chatbot that mainly generates an answer, an agent can route a case, update a record, initiate a transaction, run a security test, or coordinate several specialized agents across a workflow.

The strongest near-term public-sector applications are not autonomous policy decisions. They are high-volume, rules-rich workflows in which the agent handles search, document extraction, validation, routing, follow-up, and case preparation while a public official retains authority over consequential decisions. Evidence from the United Kingdom, Singapore, Ukraine, and the United Arab Emirates indicates potentially large reductions in processing time and administrative burden; U.S. deployments are expanding, but many remain new, optional, or in pilot stages and therefore lack independently validated outcome data.

Cybersecurity is a particularly promising domain because agents can continuously inspect code and telemetry, correlate signals, investigate attack paths, prioritize vulnerabilities, and prepare remediation. In a 2026 UK government pilot, structured AI-agent pipelines helped identify 407 findings—including previously unknown weaknesses. All critical weaknesses were remediated, and human experts manually verified every important escalation.

Agentic AI should not be treated as synonymous with generative AI. A language model may draft a memo or summarize a file without acting; an agentic system uses one or more models as reasoning components within a workflow that can affect other systems. Multi-agent designs assign specialized functions—such as planner, researcher, validator, policy checker, and supervisor—to separate components, but this added capability also creates additional failure and security pathways.

Public-value opportunities

The highest-return initial candidates generally share five characteristics: high volume, substantial manual coordination, clear authoritative rules, measurable outcomes, and low cost of reversing an error. Poor initial candidates include criminal charging, coercive enforcement, final benefits denial, clinical disposition, use-of-force decisions, or any workflow in which the agent's reasoning cannot be reconstructed and an error cannot be quickly remedied.

Evidence from various countries

U.S. Department of State

In July 2026, the Technology Modernization Fund began supporting a Department of State agentic-AI program intended to connect fragmented help desks, assist foreign-policy research and scenario planning, and provide a secure environment in which trained employees can build agents. State operates across more than 270 locations, supports80,000 users and relies on more than 450 technology systems; routine tasks can require navigating 15 or more applications.

The project expects to save more than 100,000 staff hours annually when deployed across help desks, with estimated first-year time value of nearly $15 million and rising to about $20 million the following year. These are projected benefits from an active 2026 investment, not independently verified realized savings; the design requires a person to review every automated decision.

United Kingdom: consultation analysis

The UK government's Consult tool analyzed more than 50,000 responses to the Independent Water Commission in roughly two hours for £240, followed by 22 hours of expert checking. It agreed with at least one of two expert groups almost 83% of the time, while the two human groups agreed with each other 55% of the time; reported F1 scores were 0.79 and 0.82 versus 0.74 between human reviewers.

Consult is best classified as an AI-assisted analytical workflow rather than a highly autonomous transaction agent. Nevertheless, it demonstrates the public-value logic of agentic systems: machines perform high-volume classification and organization, experts validate outputs and focus on interpretation, and detailed stakeholder submissions still receive manual review. The UK estimated that scaled use could avoid 75,000 staff-days of manual analysis annually, though that is a forecast rather than a fully realized saving.

Singapore: multi-agent analytics

GovTech Singapore and SkillsFuture Singapore developed a multi-agent system that models specialized roles including business analyst, data analyst, code executor, graph analyst, and supervisory agents. The system decomposes questions, queries tabular and graph data, executes

analysis, consolidates results, and produces human-validated insights from customer-relationship-management records.

Singapore reports that the resulting agent-and-GraphRAG (https://microsoft.github.io/graphrag/) workflow processes approximately 30,000 quarterly cases 62.5 times faster than the previous approach, enabling analysis that had previously been feasible only quarterly and improving the timeliness of business-process redesign. This remains a proof of concept moving toward an official internal tool, so long-term reliability and cost evidence should continue to be monitored.

United Arab Emirates: federal HR

The UAE Federal Authority for Government Human Resources deployed an HR AI Agent serving more than 50,000 employees across over 40 federal entities. Its initial rollout offered 108 services, automated 80% of self-service HR transactions and delivered instant responses to over three-fourths of HR-related legal inquiries.

The government estimated annual savings of approximately 170,000 work hours in technical support and legal HR services. Because this estimate comes from the implementing authority, an independent evaluation should verify realized hours, accuracy, escalation rates, employee experience, and whether faster service changes substantive HR outcomes.

Limitations and risks

Agentic AI is not a panacea. It does have its limitations and risk. The principal ones are:

Technical reliability

Agents can hallucinate facts, misread ambiguous instructions, choose the wrong tool, stop too early, loop indefinitely, or optimize a proxy rather than the government's true objective. Errors can compound across a long workflow, and conventional single-answer benchmarks often fail to measure how an agent interacts with changing systems over time.

Legal authority and due process

An agent has no independent public authority; it acts only through powers delegated by law and by accountable officials. Automated actions affecting benefits, licensing, employment, enforcement, privacy, or safety require clear legal authority, notice where applicable, contestability, accessible human review, and a remedy for error. OECD guidance emphasizes that consequential actions need a legal basis, defined permissions, a named accountable sponsor, logs, and the ability to pause or reverse the action where feasible.

Bias and unequal access

Historical administrative data can encode unequal enforcement, service access, or documentation patterns. Language, disability, digital access, and data completeness can affect who benefits from agentic services, while fraud controls can disproportionately burden groups with atypical records. Agencies should test subgroup performance, retain non-digital service channels, and monitor whether the system changes denial, escalation, and appeal rates.

Bottom line

Agentic AI can make government faster and more effective by converting fragmented administrative processes into supervised, end-to-end workflows. The most credible evidence today comes from bounded applications: consultation analysis, CRM analytics, service completion, internal HR transactions, and multi-agent cyber vulnerability discovery.

Agentic-embodied technology should not be framed as an autonomous civil servant. It is better understood as a controlled operational layer connecting people, rules, data, and government systems. Public value will depend less on the sophistication of the underlying model than on process redesign, authoritative data, secure identity, legal clarity, measurable outcomes, human judgment, and the ability to stop or reverse the system when it fails.

****

Jerry Haar is a business professor at Florida International University and a fellow at both New York University's Development Research Institute and the Baratta Center for Global Business Education at Georgetown University. He is the co-author with Ricardo Ernst and Santiago Gutierrez of *Winning in the New Global Business Landscape.* Driss Temsamani is a managing director at Citigroup where he heads digital banking for the Americas. He is the author of *The Agentic Bank: How AI and Intelligent Systems Are Redefining Finance.*