INSIGHT ARTICLE NO. 6 · INNOVATION

Human-AI Collaboration in Government

The right model for government AI is augmentation with accountable humans in command. Getting the doctrine right early prevents both automation without accountability and resistance without reason.

CENTER FOR AI & PUBLIC SECTOR INNOVATION

JULY 2026

Somewhere in a federal office this morning, a claims examiner opened a disability file that would once have consumed her entire day. A software system had already read the medical records, assembled the chronology, flagged the evidence bearing on each rating criterion, and drafted a summary. Her day's work now begins where it used to end: at judgment. Does the evidence actually support the claim? Did the machine miss the buried nerve-conduction study? Is there something in this veteran's file that no pattern model would weight correctly? She decides, she signs, and if she is wrong, there is a name on the decision and an appeal against it. That office, multiplied across government, is the whole argument of this article.

The question facing federal AI has quietly changed from whether to on what terms, and the terms on offer reduce to three. Replacement: the machine decides, at machine speed and machine cost. Refusal: the humans decide alone, as they always have, backlog and all. Or the examiner's office: augmentation with an accountable human in command. The claim here is that the third model outperforms both extremes on the merits, and that establishing it as doctrine now, while the systems are young, is the cheapest governance decision the government will ever make.

The doctrine has a moral root

Begin with why replacement is wrong even where it is accurate, because the reason is older than the technology. In the American constitutional order, public power is exercised by officials who answer for it: they are appointed under law, they act under law, and the citizen aggrieved by their decisions can demand an accounting, in an appeal, a hearing, a courtroom, or a congressional office. Responsibility is the license for power. A consequential decision that cannot be traced to a responsible official has slipped that license, whatever its statistical accuracy, because the citizen it

wrongs has no one to answer him. The tradition of subsidiarity sharpens the point: decisions belong close to accountable persons, and delegating judgment upward into an unaccountable abstraction is the same error whether the abstraction is a distant bureaucracy or a model. The conservative suspicion of any arrangement in which no one can be blamed is a suspicion the founders wrote into the structure of the government itself. Diffusing responsibility into the machine does not eliminate blame. It orphans it.

> *Diffusing responsibility into the machine does not eliminate blame. It orphans it.*

The doctrine has an evidentiary case

Federal policy has, to its credit, already chosen the collaboration model on paper. OMB Memorandum M-25-21 requires that high-impact AI, systems serving as a principal basis for decisions about people, operate under minimum risk-management practices before deployment, with human oversight and governance attached.[1] And the government's own inventories show where the model is being tested at scale. The Department of Veterans Affairs reported 367 AI use cases in 2025, more than 200 of them designated high-impact, the deepest concentration in government, aimed overwhelmingly at claims and clinical workflows where the volume is crushing and the stakes are individual.[2]

Claims processing is the perfect proving ground because it exposes both extremes' failure modes. Pure automation fails on the tails: the atypical file, the incomplete record, the veteran whose case resembles no training distribution, precisely the cases where an erroneous denial does the most harm and where due process exists for a reason. Pure human processing fails on the volume: the backlog is itself a harm, measured in months of a claimant's life, and no hiring surge clears it at the rate the reading load grows. The collaboration model attacks each failure with the other's strength. The machine compresses the reading, the assembly, the retrieval, the consistency checking, work where it is tireless and fast; the human supplies the judgment on the tails, the skepticism about the model's blind spots, and the signature that keeps the decision inside the constitutional order. Casework triage across the benefit agencies follows the same logic: let the model sort the queue so that trained judgment lands first on the cases that need it most, which is a better use of scarce humans than the chronological grind, and a better outcome for the citizens at the front of the sorted line.

The objection worth taking seriously

The strongest objection to the collaboration doctrine says that the human in command is a comforting fiction. Automation bias is among the most replicated findings in human-factors research: give a person a machine recommendation and a caseload, and the person converges toward approving the recommendation, especially under time pressure, which is the only pressure government knows. The signature stays human while the judgment quietly becomes the machine's, and the accountability the doctrine promises decays into ceremony. This objection deserves respect

because it is describing a real failure mode, and any honest doctrine must be built against it. But notice what the objection assumes: an untrained reviewer, an unmeasured process, and an institution indifferent to whether its humans are actually deciding. Those are design defects, and they have design answers. Override rates can be monitored, and a reviewer who never disagrees with the model is a signal, visible in the data, that judgment has gone to sleep. Sampling regimes can route a share of machine-recommended cases to independent blind review. Training can teach examiners the specific failure modes of the specific system, which converts oversight from ritual to skill. The objection, followed to its end, is an argument for building the command function seriously rather than nominally. It indicts the cheap version of the doctrine, and the cheap version deserves the indictment.

Set the terms early

Doctrines harden with infrastructure. Systems procured, workflows designed, and habits formed in the next few years will carry their assumptions for decades, which is why the accountability question must be settled at the design stage, while it costs a requirements document, rather than at the litigation stage, when it costs a scandal and a citizen's rights. Settled correctly, the terms are short enough to state in a sentence each. Every consequential decision traceable to a named, responsible official. Every high-impact system deployed with trained reviewers, measured override behavior, and a preserved avenue of appeal. Every efficiency gain from the machine banked as faster, better service, with the judgment work retained by people who answer for it. The examiner's office is the model: the machine did the reading, the human did the deciding, and the veteran got both speed and a signature. On those terms, AI makes government more accountable, because the routine work that once buried judgment is lifted off of it. The question was never whether AI enters government. It is whether responsibility stays, and that is a decision, and it is being made now.

NOTES

1. OMB Memorandum M-25-21, Accelerating Federal Use of AI Through Innovation, Governance, and Public Trust (Apr. 3, 2025), §§ 5-6 (high-impact AI definition and minimum risk management practices).
2. Office of Management and Budget, 2025 Federal Agency Artificial Intelligence Use Case Inventory (released Apr. 2026): Department of Veterans Affairs, 367 reported use cases, including more than 200 high-impact designations.