A FEDERAL WORKFORCE STUDY

Complementarity, Augmentation, or Substitutivity?

The Impact of Generative AI on the U.S. Federal Workforce

William G. Resh Yi Ming Xinyao Xia Michael Overton Gul Nisa Gürbüz Brandon De Bruhl

Working Forward Forum | April 23, 2026 | Under review at Public Administration Review

WHERE WE ARE

Federal workforce decisions are being made on a category error

The current rhetoric treats three different things as one thing. They are not the same thing.

"AI can help a federal worker do X."

is not the same claim as

"AI can replace a federal worker doing X."

Proposals for broad workforce reductions rest on the assumption that they are.

The analysis I'll show you tells you what happens across 435 federal occupations when you stop assuming it.

WHAT YOU GET OUT OF THIS

One framework, three operational roads

C

HIGH COMPLEMENTARITY

Tool access and adoption

The competency is intact. Investment: licensing, infrastructure, light training on how to use AI outputs effectively.

A

HIGH AUGMENTATION

Structured reskilling

How the worker exercises the competency has to change. Investment: AI literacy, output validation, collaborative workflows.

S

ELEVATED SUBSTITUTIVITY

Reallocation, not elimination

Routine subtasks shift to the system. The worker moves to oversight, exception handling, compliance, final determinations.

Every finding in the rest of this talk is calibrating which road applies to which occupation, and why.

ROADMAP

Three propositions I want to convince you of

01 Competencies, not tasks.

Tasks get reshuffled every time tools or policy change. Competencies (KSAs) are the durable unit on which federal HR is already organized. That's where the analysis should sit.

02 Three dimensions, not one.

A single automation-risk number collapses three distinct situations that imply three different workforce investments.

03 Reconfiguration, not replacement.

Across 435 federal occupational series, the dominant pattern is AI complementing and augmenting federal work, not substituting for it. Even the highest-substitutability roles retain essential human functions.

Durable units survive successive technology cycles

Tasks are what workers do. Competencies are what enables them to do it. Strategy belongs at the second level.

TASKS

What workers do

- Transient composites, reshuffled whenever tools, regulations, or strategy change
- Task-level forecasts age fast once workflows are reorganized
- Dominant unit of analysis in existing AI-and-labor work

COMPETENCIES (KSAs)

What lets workers do it

- Knowledge, skills, and abilities: the capacities workers use to produce, adapt, and recombine tasks
- Durable across successive technology cycles
- Already the foundation of federal HR through OPM classification standards and FWCI

THE FRAMEWORK

Three dimensions of AI impact on competencies

C — Complementarity

AI ENHANCES

AI enhances the competency without changing what the worker does.

FEDERAL EXAMPLE

Census Bureau BEACON. AI suggests NAICS codes in the Economic Census; respondents and analysts still make the classification call.

A — Augmentation

AI TRANSFORMS

AI requires the worker to change how the competency is exercised.

FEDERAL EXAMPLE

DOE DAMaL tools. Analysts integrate AI-generated diagnostics into their judgment and retain authority over final decisions.

S — Substitutivity

AI REPLACES

AI replaces human performance of a discrete subtask.

FEDERAL EXAMPLE

DHS counter-UAS detection. AI fuses sensor feeds that used to require continuous human monitoring; humans still adjudicate.

HOW WE BUILT IT

A research-grade reading assistant

Grounded in the actual federal documentation, not the language model's guess about what it says.

1

Read the actual federal documentation

OPM classification standards, Handbook of Occupational Groups and Families, O*NET competency taxonomy, Federal Workforce Competency Initiative ratings.

2

Extract the 9 core competencies per occupation

3 predominant knowledge bases, 3 skills, 3 abilities. Grounded in what the job descriptions actually say, not in what the model happens to recall.

3

Score each competency on C, A, and S

Each dimension is scored in an independent model call against a curated corpus of labor-market research and federal policy. Every score carries a written justification.

WHY IT'S CREDIBLE

Grounded in evidence.
Not the model's free-association.

Auditable.
Every score has a written justification you can inspect.

Independent.
C, A, and S scored separately to prevent one judgment from anchoring the next.

WHAT'S NEW SINCE THE IBM REPORT

Findings are consistent. The sharpness is new.

The substantive conclusions converge with the earlier model. The update makes them more nuanced, more accurate, and more informative.

CONSISTENT WITH PRIOR WORK

Substantive findings converge

Same ordering: **C > A > S**

Same distributional shape across occupations

Rank ordering largely preserved

*The update is a **refinement**, not a reversal. Findings are more nuanced, more accurate, more informative.*

WHAT'S SHARPER

01 **Front-end extraction triangulates OPM to O*NET.**

*Competency identification now semantically matches OPM classification text to the O*NET KSA taxonomy, then pulls the best-fit O*NET descriptions. More standardized, less drift.*

02 **Cross-model replication.**

Full pipeline run independently on GPT-4o and GPT-5.2. Same ordering on both. Structural, not model-specific.

03 **Richer narrative reasoning.**

GPT-5.2 justifications name specific subtasks and caveats. More defensible when someone pushes on a score.

04 **More rigorous evidence base.**

Stage 2 retrieval corpus mapped to a published survey of AI-labor forecasting approaches, not an ad hoc reading list.

05 **Validation pipeline is now scoped.**

Entailment testing, reliability testing across additional open-source LLMs, position-level extension from 3M+ USAJobs postings.

SCOPE

435 occupational series, two foundation models, same answer

SCOPE OF ANALYSIS

435

federal occupational series

308 white-collar

127 trade, craft, and labor

Same finding on two different foundation models

We ran the full pipeline on GPT-4o and GPT-5.2 separately. If the two models disagreed, that would tell us we're measuring an artifact of one model. They agree.

- Same rank ordering of occupations on C, A, and S
- Same distributional shape on every dimension
- Same fundamental ordering across both models

C > A > S

The framework is capturing structure in the occupations, not a quirk of one model.

THE HEADLINE FINDING

Reconfiguration, not replacement.

Mean CAS scores across 435 occupational series, all competencies pooled

COMPLEMENTARITY

HIGHEST

3.56

AI most often supports federal work

AUGMENTATION

MODERATE

3.31

Many roles will require workers to adapt how they deploy existing competencies

SUBSTITUTIVITY

LOWEST

2.38

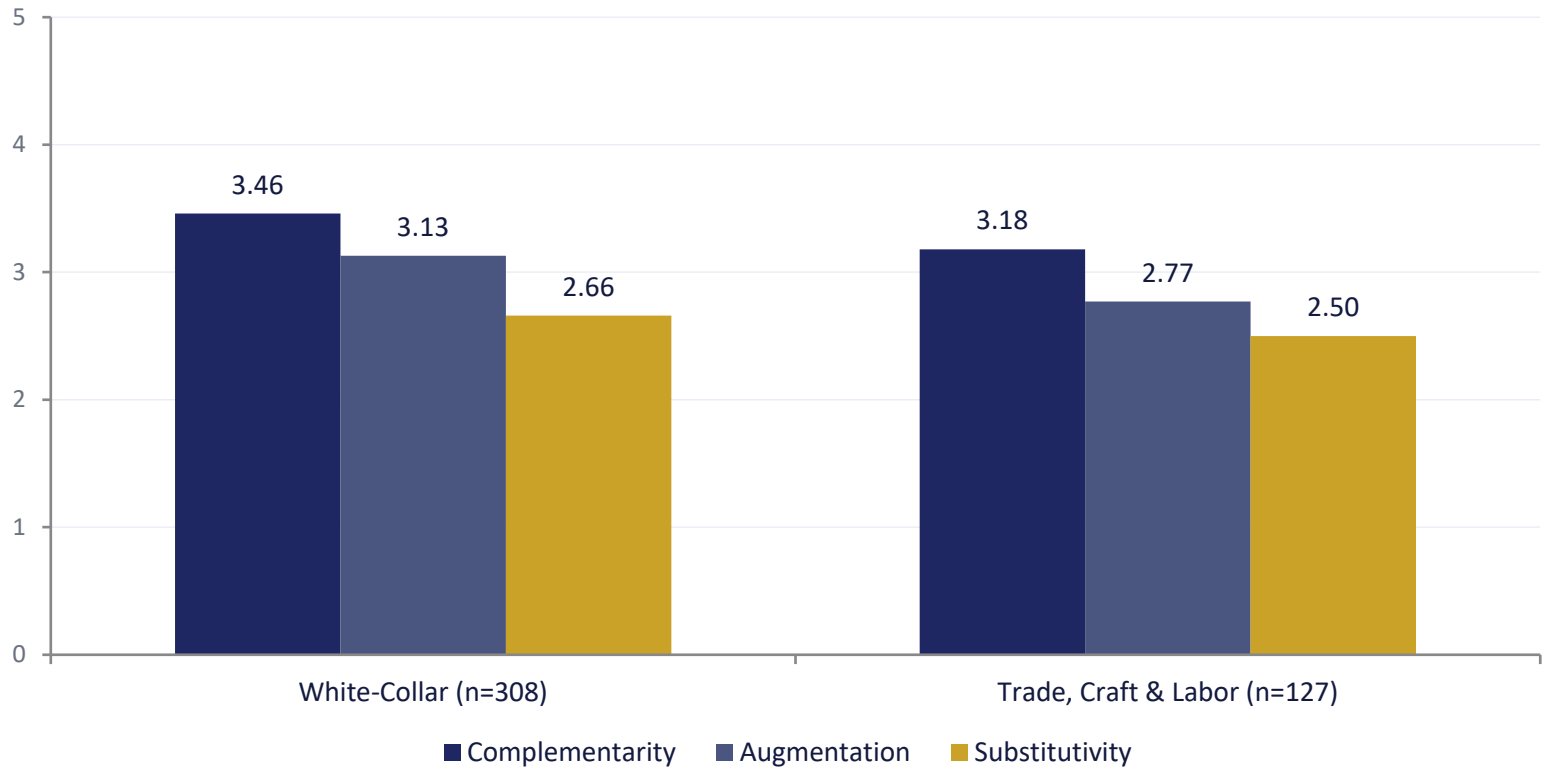
Even the most substitutable roles reflect AI replacing subtasks, not positions

n = 435 occupational series | GPT-5.2, cross-checked against GPT-4o | scale: 1 (no impact) to 5 (very high impact)

HOLDS ACROSS OCCUPATIONAL FAMILIES

Same ordering in white-collar and blue-collar federal work

White-collar roles are more exposed on complementarity and augmentation. Substitutivity stays low in both.



TAKEAWAY

White-collar roles are **more exposed** to AI on complementarity and augmentation.

TCL roles retain **physical and contextual** elements AI cannot replicate.

Substitutivity stays low in both. The C > A > S ordering is robust.

Three roads, three investment logics

Different dimensions call for different workforce investments. One-size-fits-all is inconsistent with the evidence.

HIGH COMPLEMENTARITY

C — Tool access, not retraining

The competency is intact. Workers do what they did before, faster or at larger scale. Invest in licensing, infrastructure, and light training in using AI outputs effectively.

HIGH AUGMENTATION

A — Structured reskilling

The worker has to change how the competency is exercised. Invest in AI literacy, output validation, human-AI collaborative workflows, and professional judgment about when to trust AI.

ELEVATED SUBSTITUTIVITY

S — Reallocation, not elimination

Routine subtasks move to the system. Workers move to oversight, exception handling, compliance verification, and final determinations. Cutting the position eliminates the residual work.

Even the most substitutable jobs are not going away

What AI actually does in the highest-substitutivity roles, and what the human role shifts to.

HIGHEST SUBSTITUTIVITY (white-collar)

1. Voucher Examining
2. Secretary
3. Miscellaneous Clerk and Assistant
4. Language Clerical
5. Insurance Accounts
6. Financial Clerical and Technician
7. Communications Clerical
8. Clerk-Typist
9. Data Transcribing
10. Civilian Pay

WHAT AI ACTUALLY DOES HERE

- Drafts routine correspondence
- Extracts and classifies documents
- Compares records across systems at scale
- Flags exceptions for human review

WHAT HUMANS STILL DO

Exception handling **Compliance verification**
Final determinations and accountability

Where substitution potential is near zero

Real-time judgment, interpersonal engagement, and physical accountability consistently floor the substitutivity score.

LOWEST SUBSTITUTIVITY (white-collar)

1. Aircraft Operation
2. Wildlife Refuge Management
3. Recreation Aid and Assistant
4. Range Technician
5. Nuclear Materials Courier
6. Medical Technician
7. Medical Instrument Technician
8. Diagnostic Radiologic Technologist
9. United States Marshal
10. Customs and Border Protection Interdiction

The common thread

- **Real-time physical control**

  Embodied execution in dynamic environments

- **On-site coordination**

  Reading a scene the AI cannot observe

- **Interpersonal engagement**

  Credibility, trust, de-escalation

- **Accountable human judgment**

  Someone who can be held responsible

INSIDE A SINGLE OCCUPATION

Federal Economist (0110) across nine competencies

Same occupation. Nine competencies. Three dimensions. Radically different investment implications.

Competency	Complementarity	Augmentation	Substitutivity
Economic Foundations (k1)	4	4	2
Statistical Methodology (k2)	4	4	3
Economic Data Systems (k3)	4	4	3
Data Compilation (s1)	4	4	3
Statistical Modeling (s2)	5	5	4
Interpretive Reporting (s3)	5	5	4
Research Leadership (a1)	4	4	3
Advanced Problem Solving (a2)	3	4	2
Policy Advisory Role (a3)	3	4	2

READING THE HEATMAP

Statistical Modeling & Interpretive Reporting
Score 5 on C and A. This is where AI leverage is highest.

Policy Advisory Role
Substitutivity 2. Judgment and context remain human.

A single index would average these to 3.5 and miss every useful distinction.

WHY THIS MATTERS FOR WORKFORCE DECISIONS

Conflating C, A, and S costs you exactly the capacity you meant to gain

THE MOVE THAT BREAKS STRATEGY

Read C and A evidence as S evidence.

The rhetoric that cites rising AI productivity as justification for workforce reductions is doing exactly this. Evidence that AI helps workers becomes evidence that the workers are replaceable.

The first kind of evidence does not support the second kind of conclusion.

WHAT YOU LOSE

The competency the AI was supposed to enhance.
Cut the position and you lose the human capacity the tool was leveraging. The AI without the worker is worth a fraction of the AI with the worker.

Exception-handling capacity.
In every high-substitutivity occupation, the residual work is where compliance and audit defensibility sit. That work is not optional.

The institutional memory that makes the tool credible.
Someone has to know whether the AI output is right. That someone is the experienced worker you just laid off.

ON THE HORIZON

Two extensions that turn this from prototype into operational tool

01 | VALIDATION

Stress-testing the measurements

- Replicate the pipeline across additional open-source language models
- Formal reliability and validity testing of the scores and the narrative justifications
- Ground-truth validation with agencies willing to check the outputs against actual workforce experience

02 | POSITION-LEVEL ANALYSIS

From occupation to actual position

3M+

federal job postings,
USAJobs, 2018 to present

Resolves the biggest weakness of occupation-level analysis: within-series heterogeneity. An agricultural economist is not a labor economist. A GS-14 at HQ is not a GS-9 in a field office.

Lets the framework speak to the level at which agency workforce planning actually happens.

What transfers, and what doesn't

FRAMEWORK IS PORTABLE

The three-dimensional CAS taxonomy

Any organization with a **standardized competency model** and structured job documentation can apply the same analysis to its own workforce.

State and local government, OECD civil services, shared-services organizations, large private sector employers.

INFRASTRUCTURE IS NOT

The research apparatus that makes it credible. What does is…

- Curated scholarly and policy corpus
- Vetted competency data for the target organization
- Occupation- and position-level documentation
- Model architecture, prompts, validation procedures

This is not a turnkey tool. Adopting the framework means investing in analogous infrastructure, ideally in partnership with research teams.

THE ASK

What I want to carry out of this room.

1 A sharper argument against the category error.

When proposals cite AI as justification for broad workforce reductions, the better counter-argument is methodological, not political. The CAS framework is that argument.

2 Differentiated workforce planning in your own shop.

Three roads. Tool access for complementarity. Reskilling for augmentation. Reallocation for substitutivity. Not one-size-fits-all.

3 Ground-truth partnerships.

Agencies willing to let us test the CAS outputs against what your workforce is actually doing. This is the constraint on the project right now, and this is the room that can unlock it.

William G. Resh, PhD *Georgia State University* **wresh@gsu.edu**