



Advancing Technology-Enabled Audit Methods to Reduce Procurement Fraud, Vendor Risks, Contract Noncompliance, and Inefficient Spending in U.S. Public- and Private-Sector Organizations

Prosper Fuadedzi

Georgia State University Atlanta, GA, USA

* Corresponding Author: **Prosper Fuadedzi**

Article Info

ISSN (online): 2583-6641

Volume: 05

Issue: 06

Received: 22-08-2026

Accepted: 24-09-2026

Published: 26-10-2026

Page No: 11-17

Abstract

Procurement is one of the largest and most fraud-prone areas of organizational spending, yet auditors and inspectors general can review only a small share of contracts. Traditional red-flag checklists are easy to apply but rarely validated against outcomes. This study proposes and evaluates X-PRA, an explainable machine learning framework that ranks contract awards by the risk of cost growth beyond their planned ceiling, a measurable symptom of weak contract control and inefficient spending, and explains each ranking in terms of reviewer-facing risk signals. Using 30,088 U.S. Department of the Interior contract awards from public USAspending.gov records, the model is trained on fiscal year 2019 awards and tested on fiscal year 2020 awards using only information available at award time. X-PRA achieves an AUC of 0.798, compared with 0.769 for logistic regression, 0.602 for a six-rule red-flag checklist, and 0.599 for an unsupervised isolation forest. Reviewing the riskiest 10% of awards captures 42.1% of contracts that later overran and 73.8% of overrun dollars, versus 12.6% and 43.0% for the checklist. Classic competition flags such as single offers carried little predictive value, whereas contract design and the category of goods or services dominated. An exploratory link to SAM.gov exclusions shows that vendor-integrity risk needs separate signals. The framework uses only data agencies already publish, supporting scalable, transparent procurement oversight.

Keywords: procurement fraud, audit analytics, contract risk, cost overrun, vendor risk, red flags, explainable artificial intelligence, Shapley values, USAspending, public sector auditing

1. Introduction

The U.S. federal government obligates hundreds of billions of dollars a year through contracts, and private-sector organizations route a comparable share of their spending through purchasing and vendor management. Procurement is therefore a primary channel for fraud, waste, and abuse. The Government Accountability Office (GAO) has for decades kept federal contract management on its High-Risk List, citing persistent weaknesses in oversight, cost control, and vendor management^[1]. Typical failures range from deliberate fraud, such as bid rigging, kickbacks, and false claims, to noncompliance and poor contract control, in which contracts grow far beyond their planned

value through modifications, scope changes, and loosely controlled pricing arrangements.

Audit and oversight bodies cannot review every contract. Federal inspectors general, internal audit functions, and corporate procurement-compliance teams generally rely on *red-flag checklists*: rules such as “the contract was not competed,” “only one offer was received,” or “the award was made in the last month of the fiscal year”^[2, 3]. These rules are intuitive and easy to document, but they are seldom validated against outcomes, and a checklist cannot say how much each flag matters for a particular contract.

The Digital Accountability and Transparency Act of 2014 (DATA Act) made award-level federal spending data public through USAspending.gov^[4]. This creates an opportunity to move from untested checklists to *validated, data-driven* risk scoring that remains transparent enough for auditors to act on.

This paper proposes and evaluates X-PRA (Explainable Procurement Risk Analytics), a technology-enabled audit method that

- a. scores each new contract award at the moment of award for the risk that its value will grow far beyond its planned ceiling;
- b. evaluates the scores as an oversight team would use them, as a ranked list reviewed under a fixed budget; and

- c. explains each score with Shapley attributions grouped into reviewer-facing risk signals. We use cost growth beyond the planned ceiling as the primary outcome because it is objective, observable in public data, and directly tied to contract noncompliance and inefficient spending. It is not, by itself, evidence of fraud, but it marks the contracts where control has most often failed, and where fraud and waste are most likely to be found.

The contributions are as follows:

- We build a reproducible contract-level dataset of 30,088 U.S. Department of the Interior awards (FY2019–2020) from public USAspending.gov records, with award-time features and outcomes observed six or more years after award.
- We show that a common six-rule red-flag checklist performs only modestly better than random review in predicting cost growth (AUC 0.602), and that several classic flags, including single offers and year-end awards, carry little or even inverse signal.
- We propose X-PRA and show, under a strict train-on-FY2019, test-on-FY2020 design, that it significantly outperforms the checklist, logistic regression, and unsupervised anomaly detection.
- We provide global and contract-level explanations, and an exploratory analysis linking awards to vendors later excluded in SAM.gov, which clarifies what outcome-based scoring can and cannot detect about vendor risk.

2. Related work

Red flags and corruption-risk indicators. International guidance from the OECD catalogs procurement red flags across the tendering cycle [2]. Fazekas *et al.* [3] built a corruption-risk index from contract-level indicators such as single bidding, short advertisement periods, and non-open procedures. Decarolis *et al.* [5] linked discretion in procurement to corruption investigations in Italian public works.

Contract performance and cost growth. Bajari *et al.* [6] showed that incomplete contracts lead to substantial adaptation costs through change orders, and Decarolis [7] showed that aggressive low bids are associated with worse contract performance and cost overruns. Flyvbjerg *et al.* [8] documented systematic cost underestimation in public works. For U.S. federal procurement specifically, Liebman and Mahoney [9] found that information technology contracts awarded in the last week of the fiscal year received significantly lower quality ratings, supporting year-end spending as a red flag.

Machine learning for procurement fraud. Westerski *et al.* [10] deployed explainable anomaly detection for procurement fraud in a government setting and emphasized that auditors need explanations they can act on. Unsupervised methods avoid the need for labels but, as our results show, statistical unusualness and actual risk can diverge sharply. We contribute a supervised, outcome-validated approach on U.S. public data, with an explanation layer based on Shapley values [11, 12] and a direct comparison against a red-flag checklist.

3. Data

3.1. Contract Awards

We use the USAspending.gov Award Data Archive files for contracts awarded by the U.S. Department of the Interior in fiscal years 2019 and 2020 [13]. The files contain every contract transaction reported to the Federal Procurement Data System, together with award-level values as of the extract date (September 2026). The unit of analysis is a new award (modification number 0). We keep awards with a planned value of at least

\$10,000, the micro-purchase threshold in these years, and a reported final value. Table I summarizes the sample, which spans the department's bureaus, including the National Park Service, Bureau of Land Management, Fish and Wildlife Service, and

U.S. Geological Survey.

Table 1: contract award sample (u.s. Department of the interior)

	FY2019 (train)	FY2020 (test)
New contract awards	14,883	15,205
Distinct vendors	6,243	6,039
Total planned value (\$B)	5.77	5.33
Awards with cost growth >25% of ceiling	916 (6.2%)	928 (6.1%)
Value above 1.25× ceiling (\$M)	249.7	282.1

B. Outcome: Cost Growth Beyond the Planned Ceiling

For each award we compare its current total value, observed six to seven years after award, with its *planned ceiling*, the base value plus all options at award. The outcome $y_i = 1$ if the current value exceeds 1.25 times the planned ceiling, that is, the

contract grew by more than a quarter *beyond everything that was planned, including options*. Growth within the planned options is excluded by construction. We also record the overrun amount above the threshold, which allows a dollar-weighted evaluation.

C. Award-Time Features

All 23 features are known when the contract is signed, so the model can be used before any money is spent. They fall into six groups: *competition* (extent competed, not-competed flag, number of offers, single-offer flag, solicitation procedure,

set-aside type, commercial-item procedures); *award timing* (days to fiscal year-end, September award); *contract design* (pricing type, award type, task/delivery order, share of value in options, planned duration, performance-based acquisition); *contract scale* (log planned value); *vendor profile* (number and value of the vendor's earlier awards in the same year, parent company, small-business status); and *bureau and category* (awarding bureau, product/service code group, NAICS sector). Vendor ownership-demographic fields were deliberately excluded to avoid encoding protected characteristics into risk scores.

D. Vendor Exclusions

To examine vendor risk, we matched vendors by Unique Entity ID to the SAM.gov public exclusions extract (active records as of October 2026), obtained through OpenSanctions [14, 15].

4. Methodology

4.1. The X-PRA Framework

Fig. 1 shows the framework. The *learning layer* is a gradient-boosted decision tree ensemble [16, 17] with native handling of categorical features, trained with a class-weighted logistic loss that gives the rare overrun class the same total weight as the majority class. We use at most 400 trees with learning rate

0.03, at most 15 leaves per tree, a minimum of 40 samples per leaf, an ℓ_2 penalty of 1.0, and early stopping on a random 10% validation split of the training data. These settings were fixed in advance and not tuned on the test year.

The *explanation layer* computes, for each flagged award i , Shapley values ϕ_i of the model's log-odds output for each

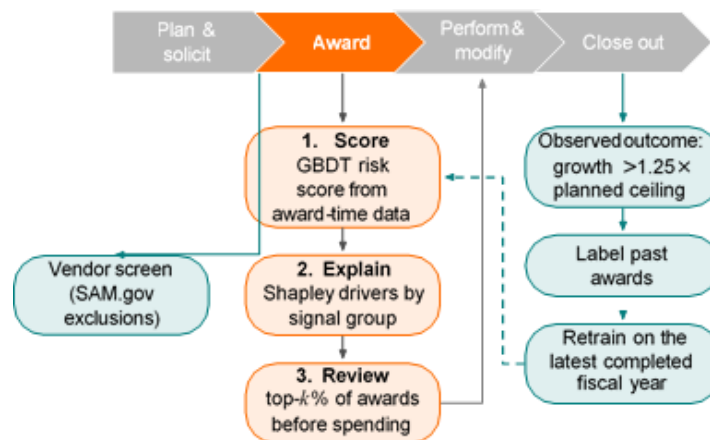


Fig 1: X-PRA embedded in the contract lifecycle. At award (orange), each new contract is scored from award-time data, explained by signal group, and, if it falls in the top-k%, reviewed before money is spent; reviewed contracts are then monitored as modifications arrive. Vendors are screened against SAM.gov exclusions at award. After close-out (teal), observed cost growth labels past awards, and the model is retrained each year (dashed feedback loop).

feature j [11, 12], so that the score decomposes additively as $F(x_i) = E[F] + \sum_j \phi_i j$. We estimate them by Monte Carlo sampling over 100 feature permutations against a background of 300 training awards, which gives unbiased estimates [11], and sum them within the six feature groups so a reviewer sees, for example, how much of the risk comes from contract design ver-sus competition.

4.2. Baselines

- **Red-flag checklist:** the number of six common flags triggered: not competed, single offer, September award, cost-reimbursement or time-and-materials/labor-hour pricing, planned value in the top 10% of the training year, and planned duration over one year [2, 3, 9]. Ties are broken by planned value.
- **Logistic regression:** class-weighted, with standardized numeric features and one-hot categorical features.

- **Isolation forest:** an unsupervised anomaly detector [18] trained on the same features without labels.

All models use scikit-learn [19].

4.3. Evaluation

Models are trained on FY2019 awards and tested on FY2020 awards, mimicking deployment on a new year of contracts. We report AUC, average precision (AP), precision and capture (the share of all overrun awards found) at review budgets of 5% and 10% of awards, and dollar capture, the share of total overrun dollars found within the budget. Uncertainty is quantified with a 500-replication bootstrap over test awards.

5. Results

5.1. How Well Do Classic Red Flags Work?

Table II compares the overrun rate of training-year awards with and without each flag. Cost-reimbursement, time-and-

Table 2: Overrun Rate With and Without Each Red Flag (FY2019)

Red Flag	Share Flagged	Rate if Flagged	Rate if Not Flagged
Cost-type, T&M, or labor-hour pricing	5.5%	24.0%	5.1%
Planned duration > 1 year	26.1%	11.3%	4.3%
Planned value in top 10%	10.0%	8.2%	5.9%
Not competed	17.0%	6.5%	6.1%
Single offer received	36.4%	6.3%	6.1%
Awarded in September	25.4%	4.2%	6.8%

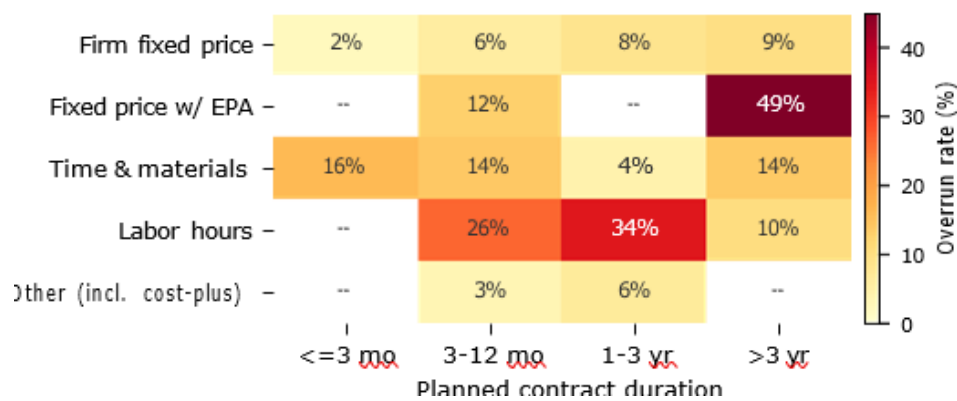


Fig 2: FY2019 overrun rate by pricing type and planned duration. Cells with fewer than 15 awards are not shown (--). EPA: economic price adjustment.

materials, and labor-hour pricing is by far the strongest flag, with an overrun rate of 24.0% versus 5.1%, and contracts planned to last more than a year overrun 2.6 times as often. In contrast, the two classic competition flags barely matter: not-competed and single-offer awards overrun at almost the same rate as others. Awards made in September actually overrun less often (4.2% versus 6.8%). This does not contradict evidence that year-end contracts are of lower quality [9]; rather, it shows that a single flag can point toward one kind of problem and away from another, which is why flags need validation against the outcome of interest. Flags also interact, which a checklist that simply counts them cannot capture. Fig. 2 crosses pricing type with planned duration: short firm-fixed-price awards almost never overrun (1.9%), while labor-hour contracts of one to three years overrun 34.2% of the time and fixed-price contracts with economic price adjustment lasting more than three years overrun 48.9% of the time.

5.2. Predictive Performance

Table III reports out-of-sample performance on FY2020. X-PRA achieves the highest AUC (0.798) and AP (0.277). Within a 5% review budget, 36.6% of the awards it flags later overrun, six times the base rate of 6.1%. A 10% budget captures 42.1% of all overrun awards and 73.8% of all overrun dollars. The checklist and the isolation forest perform only modestly better than random review in capturing overrun awards, although the checklist's high-value flag helps it capture a larger share of overrun dollars. Table IV confirms that X-PRA's advantages over all three baselines in AUC and award capture are statistically significant. Its dollar-capture advantage is significant against the checklist and the isolation forest, but not against logistic regression, be-

Table 3: Out-of-sample performance on fy2020 awards

Model	AUC	AP	Budget 5% Precision	Budget 5% Capture	Budget 10% Precision	Budget 10% Capture	\$ Capture
Red-flag checklist	0.602	0.078	7.8	6.4	7.7	12.6	43.0
Isolation forest	0.599	0.088	12.4	10.1	12.0	19.7	44.2
Logistic regression	0.769	0.216	32.0	26.2	22.8	37.3	69.9
X-PRA (proposed)	0.798	0.277	36.6	30.0	25.7	42.1	73.8
Random review	0.500	0.061	6.1	5.0	6.1	10.0	10.0

Precision (Prec.), capture (Capt.), and dollar capture (\$ Capt.) in percent. Capture is the share of all FY2020 overrun awards found within the budget; dollar capture is the share of all overrun dollars.

Table 5: X-pra minus baseline, bootstrap 95% intervals

Baseline	Δ AUC	Δ Capture@10%	Δ \$ Capture@10%
Checklist	+0.196 [0.174, 0.216]	+29.5 [25.9, 33.0]	+30.7 [17.0, 45.6]
Isolation forest	+0.199 [0.179, 0.219]	+22.4 [19.0, 25.6]	+29.6 [16.0, 46.9]
Logistic reg.	+0.029 [0.020, 0.036]	+4.8 [2.8, 6.8]	+3.8 [-4.8, 13.8]

Capture differences in percentage points. Bold: interval excludes zero.

cause a few very large overruns dominate the dollar totals. Fig. 3 compares the models at three review budgets. X-PRA leads at every budget, both in the number of overrun awards found and in overrun dollars. At a 20% budget it captures 59.6% of overrun awards (logistic regression 53.8%, checklist 27.4%) and 87.3% of overrun dollars (79.5% and 61.9%).

5.3. What Drives the Risk Scores

Fig. 4 shows the ten most influential signals for the 760 awards in the top 5% of FY2020 risk scores, colored by signal group. Aggregated by group, bureau and category (mean

absolute contribution 1.39 log-odds) and contract design (1.21) dominate, far ahead of vendor profile (0.11), award timing (0.11), competition (0.09), and contract scale (0.06). The product or service category is the single most influential signal. Services carry much higher risk than goods: in the training year, overrun rates reached 21.7% for equipment leases and rentals and 13.2% for professional and support services, compared with 1.2% or less for most product categories. Planned duration and pricing type follow, consistent with Fig. 2.

Fig. 5 explains one correctly flagged award, ranked 4th of 15,205: a firm-fixed-price delivery order planned at \$489,275

over about four years, awarded through competition with 13 of-fers. Its value later grew to 5.9 times the planned ceiling. It triggered only two of the six checklist rules (high value and long duration), placing it 2,278th in the checklist ranking, out-side a 10% review budget: it was competed, received many of-fers, used fixed pricing, and was not a September award. Fig. 5 traces how X-PRA reached its score. Starting from the aver-age award (-1.14 log-odds), bureau and category add $+2.08$, driven by its category (an equipment lease or rental), and con-tract design adds $+0.92$, mainly from its long planned duration, while the remaining groups add little, giving a final score of

2.17. For a reviewer, this points to concrete steps: confirm that the planned ceiling reflects realistic rental terms and duration, and monitor modifications against that ceiling.

5.4. Exploratory Analysis: Vendor Exclusions

Only 27 FY2020 awards, to 9 vendors, went to vendors that were later excluded from federal contracting in SAM.gov. These awards ranked, on average, at the 62nd percentile of X-PRA scores, and 26% fell in the top quarter, about what chance alone would produce. The sample is far too small for firm conclusions. The finding is nonetheless informative: a model trained to detect cost growth does not, by itself, identify vendors who will later be found to lack integrity. Vendor risk requires separate signals, such as exclusion screening at award, owner-ship and address links between vendors, and past-performance records.

6. Discussion

6.1. Implications for Audit and Oversight

For federal inspectors general and agency internal auditors, X-PRA provides a validated way to select contracts for review at the time of award, before money is spent, rather than af-ter an overrun has occurred. The explanations translate scores into specific review steps, such as verifying

ceilings for long-duration service contracts and increasing surveillance of labor-hour and time-and-materials arrangements, consistent with the extra oversight that the Federal Acquisition Regulation already requires for these contract types [20]. The weak performance of the checklist suggests that agencies should validate their red flags against outcomes before relying on them. The same method applies to private-sector organizations, whose purchase-order and vendor-master data contain equivalent fields.

6.2. Limitations

First, cost growth beyond the ceiling is a proxy for contract-control failure and inefficient spending, not proof of fraud; some growth reflects legitimate scope changes. Second, the study covers one department and two fiscal years; results should be confirmed for other agencies and later years. Third, the exclu-sion analysis uses only currently active exclusions, which omits expired records, and involves very few vendors. Fourth, Shap-ley values explain model behavior, not causes, and bureau and category effects may reflect differences in mission rather than in control quality.

7. Conclusion

This paper proposed X-PRA, an explainable, technology-enabled audit method that scores new contract awards for the risk of cost growth beyond their planned ceiling using only public, award-time data. On 30,088 U.S. Department of the Interior awards, trained on FY2019 and tested on FY2020, X-PRA achieved an AUC of 0.798 and captured 42.1% of over-run awards and 73.8% of overrun dollars within a 10% review budget, significantly outperforming a red-flag checklist, logis-tic regression, and unsupervised anomaly detection. The re-sults show that classic competition flags are weak predictors of cost growth, that contract design and the category of goods or services matter most, and that vendor-integrity risk requires complementary screening. Future work will extend the frame-

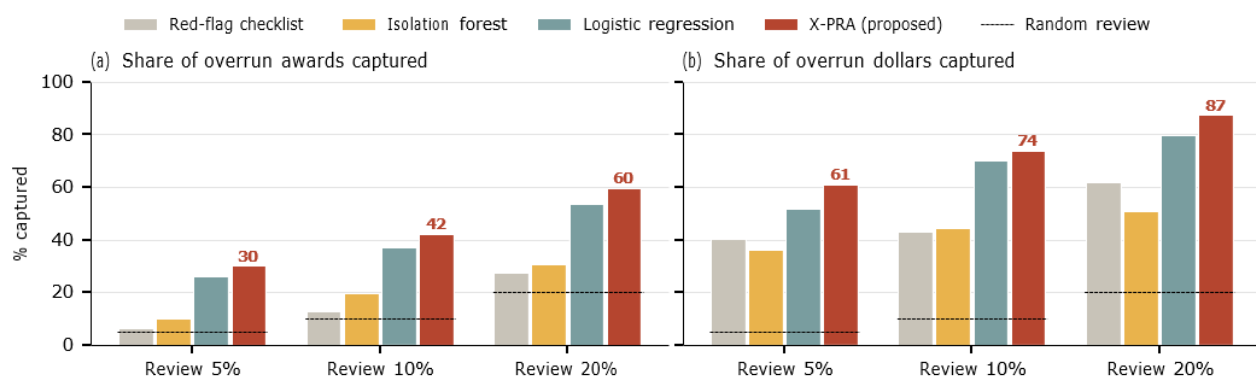


Fig 3: Capture of FY2020 overruns at review budgets of 5%, 10%, and 20% of awards. (a) Share of overrun awards; (b) share of overrun dollars. Dashed lines mark what random review would capture. Labels give X-PRA values.

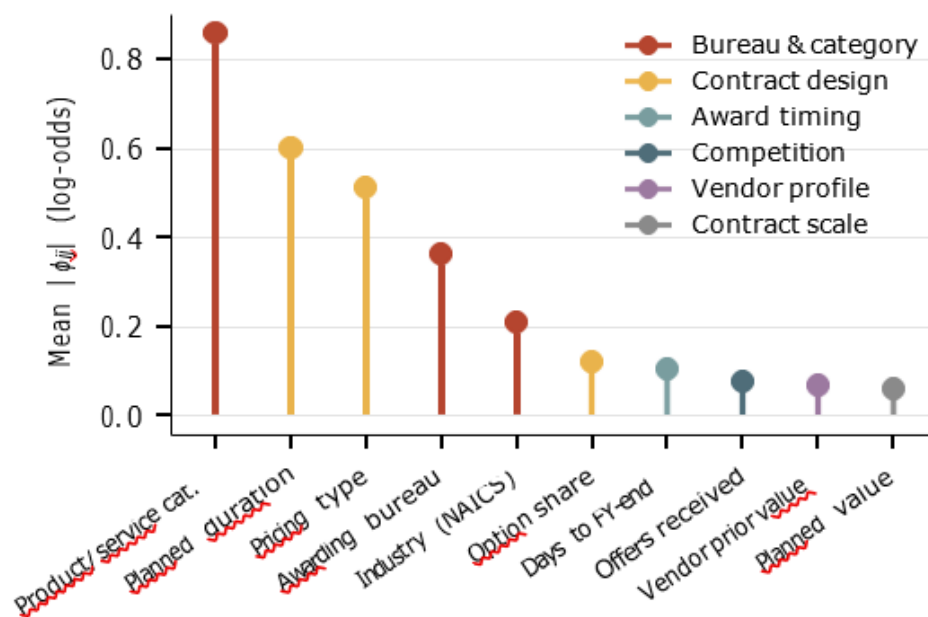


Fig 4: The ten most influential signals for the 760 FY2020 awards in the top 5% of X-PRA risk scores (mean absolute Shapley value), colored by signal group.

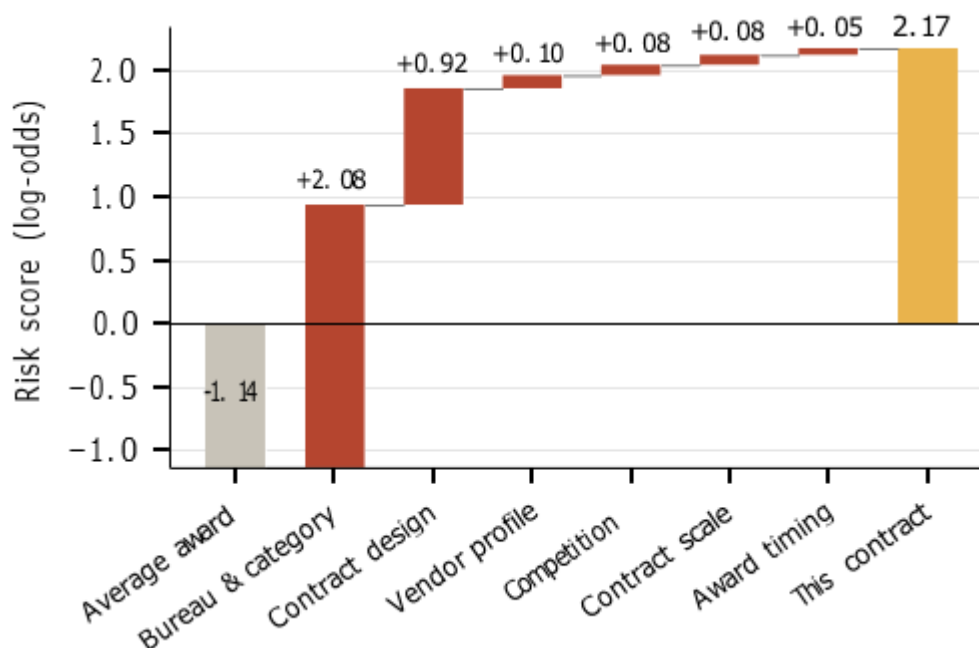


Fig 5: Waterfall explanation for a correctly flagged FY2020 award (rank 4 of 15,205): how each signal group moves the score from the average award to this contract.

work to more agencies and years, add vendor-network and past-performance signals, and evaluate it prospectively with over-sight teams.

References

1. U.S. Government Accountability Office. High-risk series: heightened attention to high-risk areas could yield billions in savings and a more efficient and effective government. GAO-25-107743. 2025 Feb.
2. Organisation for Economic Co-operation and Development (OECD). Preventing corruption in public procurement. Paris: OECD Publishing; 2016.
3. Fazekas M, Tóth IJ, King LP. An objective corruption risk index using public procurement data. Eur J Crim Policy Res. 2016;22(3):369-397.
4. Digital Accountability and Transparency Act of 2014. Pub L No. 113-101, 128 Stat 1146; 2014.
5. Decarolis F, Fisman R, Pinotti P, Vannutelli S. Rules, discretion, and corruption in procurement: evidence from Italian government contracting. National Bureau of Economic Research. Working Paper No. 28209; 2020.
6. Bajari P, Houghton S, Tadelis S. Bidding for incomplete contracts: an empirical analysis of adaptation costs. Am Econ Rev. 2014;104(4):1288-1319.

7. Decarolis F. Awarding price, contract performance, and bids screening: evidence from procurement auctions. *Am Econ J Appl Econ*. 2014;6(1):108-132.
8. Flyvbjerg B, Holm MS, Buhl S. Underestimating costs in public works projects: error or lie? *J Am Plann Assoc*. 2002;68(3):279-295.
9. Liebman JB, Mahoney N. Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. *Am Econ Rev*. 2017;107(11):3510-3549.
10. Westerski A, Kanagasabai R, Shaham E, Narayanan A, Wong J, Singh M. Explainable anomaly detection for procurement fraud identification: lessons from practical deployments. *Int Trans Oper Res*. 2021;28(6).
11. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. *Knowl Inf Syst*. 2014;41(3):647-665.
12. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30*; 2017. p. 4765-4774.
13. U.S. Department of the Treasury. USAspending.gov award data archive [Internet]. [cited 2026 Oct]. Available from: https://www.usaspending.gov/download_center/award_data_archive
14. U.S. General Services Administration. System for Award Management (SAM) public extract: exclusions [Internet]. [cited 2026 Oct].
15. OpenSanctions. US SAM procurement exclusions [Internet]. [cited 2026 Oct]. Available from: https://www.opensanctions.org/datasets/us_sam_exclusions/
16. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232.
17. Ke G, *et al*. LightGBM: a highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems 30*; 2017. p. 3146-3154.
18. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*; 2008. p. 413-422.
19. Pedregosa F, *et al*. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
20. Federal Acquisition Regulation. Types of contracts. 48 CFR Part 16, Subpart 16.6: Time-and-materials, labor-hour, and letter contracts.

How to Cite This Article

Fuadedzi P. Advancing technology-enabled audit methods to reduce procurement fraud, vendor risks, contract noncompliance, and inefficient spending in U.S. public- and private-sector organizations. *International Journal of Management and Organizational Research*. 2026;5(6):11-17

Creative Commons (CC) License

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms