The alignment gap: Overcoming the structural bottlenecks in public sector AI

Erli Wang,[1,*] Belinda Pfeffer,[1] Dakshi Kapugama Geeganage,[1] and Opec Kemp[1,*]

[1]Queensland Government Procurement, Department of Housing and Public Works, The Queensland Government, Brisbane QLD 4000, Australia

*Correspondence: Erli.Wang@hpw.qld.gov.au (E.W.); Opec.Kemp@hpw.qld.gov.au (O.K.)

Received: February 13, 2026; Accepted: July 5, 2026; Published Online: July 15, 2026; https://doi.org/10.59717/j.xinn-inform.2026.100056

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Wang E., Pfeffer B., Geeganage D. K., et al. (2026). The alignment gap: Overcoming the structural bottlenecks in public sector AI. *The Innovation Informatics* 2:100056.

Public procurement represents roughly 13% of GDP across OECD countries,[1] shaping markets at a scale comparable to monetary or industrial policy. While artificial intelligence (AI) is increasingly used to support public-sector decision-making, the principal challenge is no longer predictive performance but the *alignment gap*—the mismatch between AI optimisation objectives and the institutional conditions required for outputs to inform real-world decisions. As a result, systems that perform well on technical metrics may still fail to generate actionable public value. Using public procurement as a case study, we examine how this alignment gap manifests through automated categorisation, transaction-level confidence, and the interpretation of analytical outputs within institutional decision processes. We argue that shared semantic structures are essential for transforming procurement data into actionable information, supporting benchmarking, supplier analysis, forecasting, and governance outcomes. To bridge the gap between analytics and action, we propose a closed-loop data value chain linking data, analytics, and decision-making. The central implication is that public-sector AI should be evaluated not solely by predictive accuracy, but by its ability to align information and uncertainty with institutional authority, budget cycles, and accountability requirements. Addressing this alignment gap is essential for moving from isolated AI applications towards more effective governance with and by AI.

PUBLIC PROCUREMENT AS A STRUCTURAL LENS ON THE ALIGNMENT GAP

Recent studies have examined artificial intelligence (AI) adoption across three core public-sector functions: policymaking, service delivery, and internal management.[2] While AI has the potential to improve efficiency, effectiveness, and service quality, governments remain in a transitional phase toward governance "of, with and by AI". Public procurement, a key area of internal management where AI is increasingly used for contract analysis, risk detection, and administrative automation, provides a valuable setting for examining this transition.

Public procurement

Public procurement is one of the largest levers of government action. More than a transactional function, procurement supports economic resilience, public services, and broader policy objectives. Procurement data can be viewed through buyers, suppliers, and procurement categories, providing a rich representation of government demand, market activity, and institutional priorities. As a result, procurement offers an ideal lens through which broader structural challenges in public-sector AI become visible. It sits at the intersection of policy intent, administrative practice, market response, legal obligations, transparency requirements, and historically siloed data systems.

Using automated procurement categorisation and its downstream analytics as a case study, we argue that the principal bottleneck in public-sector AI is no longer computational capability or modelling sophistication. Instead, it lies in a misalignment between AI optimisation objectives and the informational structure of public institutions. Governments do not simply process data; they interpret and act on information that is context-dependent, legally constrained, and embedded within regulatory frameworks. Consequently, AI systems optimised for predictive accuracy alone may fail to align with the institutional contexts in which their outputs must be interpreted and acted upon.

Understanding this alignment gap is therefore essential not only for procurement, but for the design and governance of AI systems across public administration more broadly.

Prediction accuracy is no longer sufficient

Most AI systems are optimised under a simple objective: minimise expected prediction error. Formally, this is often expressed as

$$\min_{f};E_{(x,y)}\left[L\left(y,f(x)\right)\right]$$

where x denotes inputs, y a ground-truth label, and $L(\cdot)$ a loss function such as 0–1 loss, squared error, or absolute percentage error. This formulation has been highly successful in domains where labels are stable and improved predictions directly translate into improved actions. Public procurement differs in a fundamental way.

•Labels are shaped by dynamic policy, incentives, and administrative practice.

•Feedbacks are limited and usually the loops are long and indirect.

•Decisions are mediated by human actors operating under accountability constraints.

•Data volume is extremely large, siloed, often of poor quality, and must be analyzed within practical time limits.

For these reasons, predictive accuracy remains necessary but is no longer sufficient as the primary objective for AI systems in public procurement. The key challenge is to align representations, uncertainties, and predictions with institutional authority and decision processes. We refer to this mismatch between predictive optimisation and institutional actionability as the alignment gap. These examples illustrate why predictive performance alone is insufficient for effective public-sector AI.

SOME PRACTICES OF AI IN PUBLIC PROCUREMENT

Automated procurement categorizations

Automated procurement categorisation is one of the most mature AI applications in public procurement. Standardised taxonomies such as the United Nations Standard Products and Services Code (UNSPSC) provide a hierarchical structure for representing procurement activity and support downstream applications including benchmarking, forecasting, supplier analysis, and policy evaluation. The transition to automated categorisation is already underway. The eSCPRS Purchase Order Data repository provides a standardised procurement dataset supporting research into spending optimisation,[3] cost reduction, and supply-chain risk management. More recently, Queensland Government Procurement deployed an automated categorisation solution combining machine learning and Large Language Models (LLMs) to process millions of transactions.[4]

Categorisation illustrates a key manifestation of the alignment gap. Procurement transactions do not have a single universally correct label but can be represented using multiple taxonomies (e.g., GL codes, UNSPSC, ANZSIC, GS1/GTIN, and HS codes), each serving different stakeholders and decision contexts. The challenge is therefore not only predicting labels accurately, but selecting the representation that best supports the institutional decision. As such, the value of categorisation lies in providing a shared semantic foundation for downstream decision-making, not merely achieving high predictive accuracy.

Transaction-level confidence

Aggregate measures such as accuracy and F1-score provide useful summaries of model performance, but they often conceal substantial variation across individual transactions. This limitation is particularly important in procurement datasets, where transaction values range from a few dollars to many millions. A misclassified stationery purchase and a misclassified infrastructure contract contribute equally to conventional accuracy metrics despite having vastly different financial and governance implications.

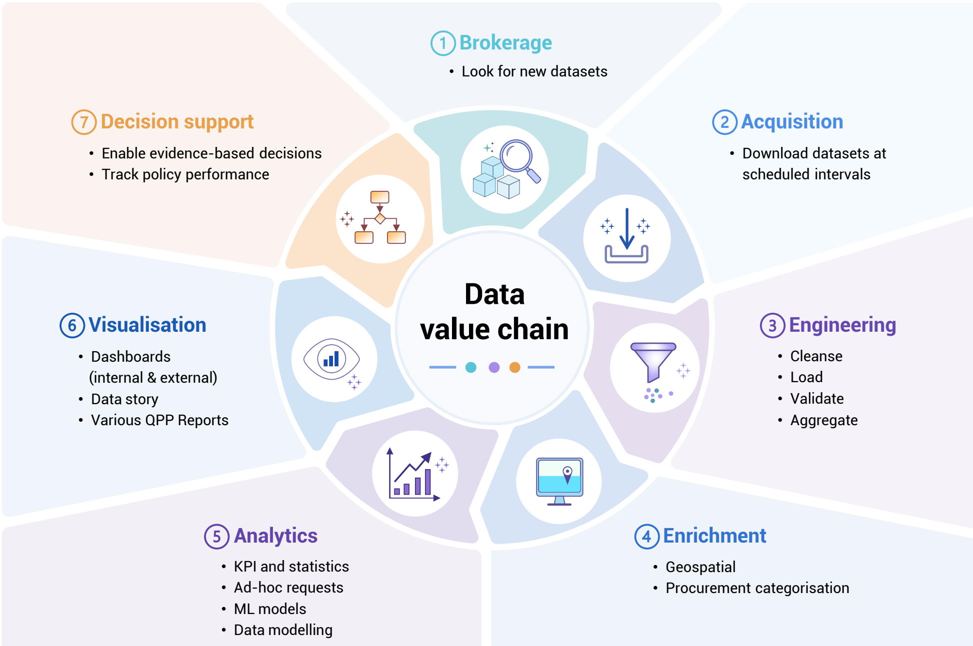

Figure 1. Data value chain for public procurement.

Although expert validation remains the gold standard, manual review cannot scale to the millions of procurement transactions processed each year. Recent research suggests that LLMs can support evaluation and annotation,[5] enabling confidence-aware review frameworks that combine automated predictions with selective human oversight. In practice, organisations are less concerned with overall accuracy than with identifying which transactions can be trusted automatically and which require review. Confidence estimation therefore becomes a mechanism for aligning uncertainty with institutional accountability and governance.

Institutional semantics for action

Public procurement semantics create value only when they support institutional action. Shared procurement categories enable expenditure comparisons across agencies, support supplier benchmarking, and provide a common language for procurement analysis. At the operational level, they reduce fragmentation caused by agency-specific coding schemes and enable analysis across organisational boundaries. At the market level, they provide consistent representations of supplier capabilities, supporting assessments of supplier concentration and supply-chain resilience.

Institutional semantics are particularly important in forecasting and decision support, where procurement officers and budget owners operate at specific levels of organizational authority. Higher predictive accuracy does not necessarily improve decisions: distinguishing between pothole repair, kerb restoration, and drainage maintenance may add little value when planning and budgeting occur at the broader road-maintenance category. The usefulness of procurement analytics therefore depends not only on statistical precision but also on alignment with institutional responsibilities, budget structures, and decision-making needs.

Close-loop data value chain

A closed-loop data value chain (Figure 1) describes how procurement data are transformed into decision-ready insights and continuously improved through feedback from policy and operational outcomes. The process begins with governed data acquisition, followed by enrichment through cleansing, categorisation, supplier alignment, and contextual tagging. Analytics and visualization then generate actionable insights for monitoring, forecasting, benchmarking, and risk assessment, informing procurement strategies and operational decisions. These decisions create new procurement activity that feeds back into data collection, enrichment, and modelling, completing the cycle. By linking data, analytics, and decision-making, the framework shifts AI evaluation beyond predictive accuracy towards institutional learning, decision effectiveness, and improved governance outcomes.

INSIGHT EXAMPLE: SME PARTICIPATION

The share of procurement spend awarded to small and medium-sized enterprises (SMEs) is often used as an indicator of supplier diversity and procurement policy outcomes. Yet interpreting changes in SME participation is not straightforward. Procurement transactions must first be transformed into a consistent representation of suppliers, industries, and spending categories before meaningful comparisons can be made. The resulting evidence must then be evaluated in terms of reliability and interpreted within the context of procurement policy objectives.

For example, a decline in SME share accompanied by sustained growth in large-business spending, industry-specific expenditure shifts, and increasing supplier concentration provides stronger evidence of a structural change in procurement allocation than any individual indicator alone. Such insights enable agencies to distinguish temporary fluctuations from emerging market trends and to consider whether policy intervention is warranted. This example illustrates how procurement value is created not from prediction alone, but from converting transaction data into actionable insights that support monitoring, policy evaluation, and continuous improvement.

SOCIAL IMPACTS

This analysis suggests that public-sector AI should be designed, evaluated, and governed according to how well it aligns information, uncertainty, and institutional capacity for action, rather than predictive accuracy alone. Taxonomies, confidence measures, and forecasting horizons should reflect decision authority, budget cycles, and accountability mechanisms. Such alignment improves governance through more consistent data and decision support, enables fairer and more transparent procurement, and strengthens budget transparency and resource allocation. Public trust ultimately depends on AI systems being explainable, accountable, and aligned with public objectives.

The central challenge of public-sector AI is therefore not automation, but alignment: aligning statistical objectives with institutional semantics, analytical outputs with feasible interventions, and technical performance with public value. Although procurement provides a clear illustration, these principles apply broadly wherever AI supports policy-driven, human-mediated decision-making across government.

REFERENCES

1. OECD (2024). Governing with Artificial Intelligence: Are governments ready? *OECD Artificial Intelligence Papers* No. 20. OECD Publishing. DOI:10.1787/26324bc2-en
2. Van Noordt C. and Misuraca G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of AI in government across the European Union. *Gov. Inf. Q.* 39:101714. DOI:10.1016/j.giq.2022.101714
3. Department of General Services, State of California (2019). Purchase order data. https://data.ca.gov/dataset/purchase-order-data (accessed Jul-2026)
4. Wang E., Geeganage D.K. and Kemp O. (2025). Categorization architecture with predictive reasoning and alignment for UNSPSC. *AI 2025: Advances in Artificial Intelligence* 16370. DOI:10.1007/978-981-95-4969-6_24
5. Gilardi F., Alizadeh M. and Kubli M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proc. Natl. Acad. Sci. USA* 120:e2305016120. DOI:10.1073/pnas.2305016120

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.