AI-Powered Vendor Evaluation Lightens the Load in Government Procurement

GEORGIA TECHNOLOGY AUTHORITY

AWARD CATEGORY: Operational Efficiency

STATE: Georgia

PROJECT DATES: January 2025 – May 2025

CONTACT: Shawnzia Thomas (State CIO & GTA Executive Director)
Shawnzia.Thomas@gta.ga.gov

EXECUTIVE SUMMARY

As artificial intelligence (AI) becomes embedded in the operation of public agencies, the procurement process must evolve to ensure responsible acquisition of AI solutions. Public sector entities, including state agencies, frequently evaluate multi-million-dollar vendor bids that require rigorous scrutiny across technical, financial, and ethical dimensions.

Traditional vendor evaluation methods though are often manual and time-intensive, subject to human bias, and challenged to maintain transparency and consistency.

These hurdles create risks of delays, subjective decisions, and oversight gaps in selecting the right vendor partners for critical projects.

To help minimize such risks, the Georgia Technology Authority (GTA) has introduced a practical framework using AI to automate and enhance the public procurement evaluation process.

The framework leverages AI multimodal capabilities and advanced semantic reasoning to systematically evaluate vendor responses against predefined technical, financial, and ethical criteria.

IDEA

WHAT PROBLEM DOES THE PROJECT ADDRESS?

To address longstanding challenges in public procurement evaluations, the Georgia Technology Authority (GTA) developed an AI-powered vendor evaluation framework. This scalable approach leverages advanced automation to ensure consistent, traceable, and defensible decision-making across high-value, multi-vendor procurements.

WHY DOES IT MATTER?

Procurement teams are increasingly tasked with evaluating multi-million-dollar, high-value proposals from numerous vendors, often with closely balanced scores. This pressure to make quick, fair, and accurate decisions highlights the need for a more efficient and transparent evaluation process.

In line with GTA's goal of integrating AI in a policy-compliant way, this project addresses inefficiencies in traditional evaluation methods.

An AI tool is used to support and enhance human judgment in procurements by providing consistent, data-driven insights while also ensuring final decisions remain squarely in the hands of experienced professionals.

WHAT MAKES IT DIFFERENT?

The innovative framework automates scoring, aligns evaluation criteria with justifications, and extracts key insights received from vendors in response to state government procurement requests and extensive proposal documents. By leveraging real-time semantic analysis, it ensures evaluations are objective, transparent, and defensible. Future enhancements include the development of domain-specific templates and empirical validation to support responsible, AI-driven procurement practices.

WHAT MAKES IT UNIVERSAL?

Key use cases common across government include:

Public Sector Procurement Teams:

State agencies, government procurement offices, and public entities that need to evaluate multiple vendor proposals, especially for large-scale technology projects.

Evaluation Committees for High-Value Bids:

Teams responsible for scoring, ranking, and justifying vendor selections for multi-million-dollar contracts where decisions must be objective, transparent, and defensible.

Technology Project Management Offices (PMOs):

Departments tasked with ensuring that new AI systems being acquired are technically robust, ethically aligned, and financially sound.

Auditors and Oversight Bodies:

Groups that need clear, auditable trails of how vendor evaluation decisions were made ensuring accountability in public procurement processes.

IMPLEMENTATION

WHAT WAS THE ROADMAP?

Figure 1 Design of experiment framework

WHO WAS INVOLVED?

#	TEAMS	RESPONSIBILITY
1	**Enterprise PMO (EPMO)**	Provided oversight and aligned AI use with state procurement protocols.
2	**DevOps / Digital Services Team**	Supported infrastructure, security, and API integration.
3	**GTA Innovation lab and Vendor Partners**	Assisted in prompt tuning and AI deployment.
4	**Procurement Evaluators & SMEs**	Participated in training and validation of scoring outputs.

HOW DID YOU DO IT?

Integration with Local and Cloud-Based Systems:

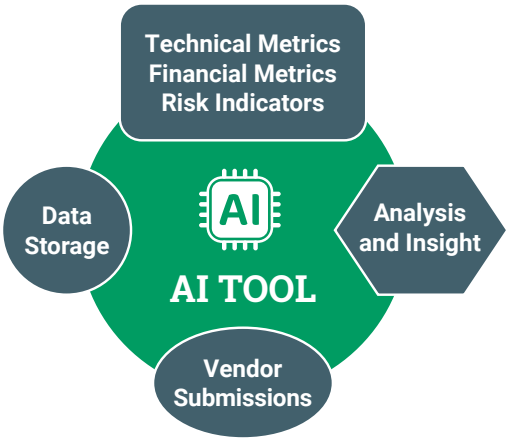
The AI tool is a custom-built instance developed to support request for proposal (RFP) evaluations. Designed with API-first architecture, it integrates seamlessly with existing procurement platforms and local data repositories. It enables secure data exchange in formats like JSON, XML, and CSV, allowing automated ingestion of evaluation spreadsheets and vendor submissions. Operating entirely within a state-owned and managed environment, the AI processes, evaluates, and returns insights or scores into the workflow ensuring that all sensitive procurement data remains fully protected and under control.

By automating these interactions, it reduces manual data entry and improves consistency and speed in vendor assessment processes.

Data Extraction and Analysis:

GTA devised a secure, AI-driven framework to transform how public sector RFP evaluations are conducted. Built on AI, this solution harnesses generative AI, intelligent document processing (IDP), and explainable machine learning (XAI) to evaluate complex vendor submissions in a fast, transparent, and defensible manner.

Figure 2
Data extraction and analysis



Key features include:

- Procurement-specific natural language processing (NLP) engine that maps vendor responses to client-defined questions with precise scoring and traceable justifications.
- Secure API-first integration with local data repositories and procurement platforms, ensuring data stays within a state-owned, auditable environment.
- Real-time evaluation capability, reducing manual workload, increasing inter-rater reliability, and drastically improving cycle time for high value bid assessments.
- Explainable confidence scoring, and health index views that empower human decision-makers with deeper insights into vendor viability.

Data Security, Privacy and Risk Assessment:

The AI tool complies with data protection regulations such as General Data Protection Regulation (GDPR), ensuring that data is handled with the utmost care and in accordance with state security features. Data anonymization and pseudonymization techniques are employed wherever appropriate to further enhance privacy.

Table 1. Security feature and AI capability

Security Feature	AI Capability	Applicable U.S. Standards / Regulations	Citation
Encryption (Data at Rest/ In Transit)	Encrypts vendor and evaluation data using AES-256 and TLS 1.2+	NIST SP 800-53, FIPS 140-2	NIST SP 800-53 Rev. 5 – Security and Privacy Controls for Information Systems and Organizations
Access Control	Role-Based Access Control (RBAC) and Multi-Factor Authentication (MFA)	CISA Zero Trust Maturity Model	CISA Zero Trust Maturity Model v2.0 (2023)
Data Anonymization / Pseudonymization	Removes vendor identifiers during evaluation to prevent bias	AI Fairness Practices, NIST AI Risk Management Framework	NIST AI RMF (2023), Section 2.2 – Data Privacy and Fairness

Smart Ingestion and Evaluation Traceability:

The platform ingests and standardizes diverse vendor documents using optical character recognition (OCR) and natural language processing (NLP), enabling structured, criteria-based evaluation. It maps extracted content to scoring rubrics and anchors insights to source text for traceable decisions. This ensures fairness and auditability in vendor assessments. To optimize alignment with technical requirements, prompt engineering tailors the AI's interpretation and response precision.

Prompt Engineering

1 Evaluation Against Criteria

Prompt:

"Review the vendor's technical/financial document and assess how each response aligns with the mandatory questions. Determine whether each response fully meets, partially meets, or fails to meet the evaluation criteria outlined in the scoring rubric."

2 Answer Mapping

Prompt:

"For each mandatory question, create a response mapping table that includes the following fields:"

- **Mandatory Question:** The official evaluation question as defined by state official.
- **Mapping Response:** The exact wording or direct quote from the vendor's document addressing the question.

3 Scoring and Justification:

Prompt:

"Assess each mapped response according to defined criteria. Assign a score and provide a justification for the rating, referencing the vendor's content. Summarize the response's strengths, weaknesses, and any gaps observed."

- **Score:** Assign a rating based on scoring rubric (No Response, Poor, Marginal, Adequate, Good, Excellent).
- **Justification for Rating:** Provide a detailed explanation, citing direct evidence from the vendor's submission.
- **Summary:** Highlight the strengths and weaknesses and identify any gaps in the vendor's response.

4 Confidence Indicators:

Prompt:

"On a scale of 1 to 5, how confident is this score based on the clarity and completeness of the vendor's response?"

- Help evaluators judge how strongly AI stands behind its assessment.

5 Supporting Quotes

Prompt:

"Provide direct quotes from the vendor's document that support the evaluation for each question. Ensure each quote is associated with the corresponding rating and mapped response."

- **Supporting Quotes:** Extract and list the key quotes that form the basis for evaluation, directly tied to the assigned rating.

6 Reference Files

Prompt:

"List all reference documents consulted during the evaluation process, including file names, document titles, and metadata (e.g., version, submission date) to ensure traceability and auditability."

- **Document Reference:** Specify relevant section titles or page numbers to ensure traceability and consistency throughout the evaluation process.

Traceability and Transparency:

To build trust in AI-assisted procurement, the team developed a **proof-of-concept (POC) traceability framework** that complements AI's capabilities. While the AI tool doesn't track scores or versions on its own, GTA embedded those safeguards through structured prompts and external tools.

Key safeguards include:

- **Confidence Indicators:** The AI tool can self-rate its confidence (e.g., 1–5 scale) when prompted. While not statistical, it gives evaluators a sense of how solid AI's response might be.
- **Version Control:** Scoring rubrics are versioned in systems like SharePoint or Excel. Each AI output is tagged to the rubric version used, ensuring consistency and audit readiness.
- **Quote Mapping & Justification:** For every evaluation, AI extracts supporting quotes and explains why a score was assigned. These are saved alongside evaluator notes for transparency and defensibility.

Responsible AI Use, Real Impact

This proof-of-concept follows responsible AI principles:

- **Consistent and fair evaluation:** Standardized prompts help ensure objective, uniform scoring across all vendor submissions.
- **Audit-Ready documentation:** All AI outputs, including justifications and supporting quotes, are securely stored to support audits, peer reviews, and appeals.
- **Balanced innovation:** The approach enables innovation in procurement while upholding strong accountability and governance standards.

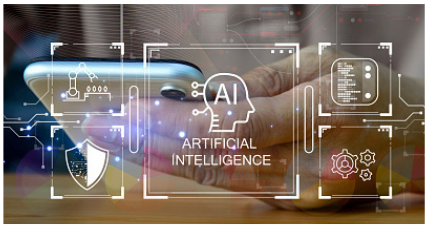
Evaluator-in-the-Loop (EITL) Model

While AI performs the initial assessments, human evaluators remain in control:

- Evaluators can review AI-generated scores, adjust them if needed, and annotate justifications.
- The system offers confidence scores and flags ambiguous responses for manual review.
- Final scores are compiled with both AI and human insights, supporting compliance and buy-in.

IMPACT

WHAT DID THE PROJECT MAKE BETTER?



The AI -powered vendor evaluation framework marked a significant advancement in public procurement by shifting from manual, subjective assessment to a transparent, automated, and data-driven process. In a real-world application supporting a multi-million-dollar state technology modernization initiative, the system analyzed thousands of pages of technical and financial documents. This reduced evaluation timelines by nearly 50% and saved evaluators considerable time and effort. The framework used large language models (LLMs) with built-in reasoning to align closely with evaluation rubrics, achieving over 90% consistency in scoring. This improved both the accuracy and efficiency of the review process while reducing the burden on evaluators.

AI acted as an enabler, supporting human judgment rather than replacing it. This approach strengthened transparency, documentation, and audit readiness.

HOW DO YOU KNOW?

Gains were recognized across a range of procurement evaluation considerations:

Evaluation Dimension	Traditional Approach	AI-Powered Approach	Impact Metrics / Benefits
Efficiency	Manual reviews, spreadsheets, long timelines	Automated scoring, real-time data processing	Approximately 50% reduction in evaluation time (from months to weeks).
Objectivity	Subjective scoring is prone to personal bias	AI ensures consistent, criteria-based scoring	Approximately 30% reduction in decision-making bias (based on scoring patterns).
Scalability	Limited to small batches of vendor proposals	Scales effortlessly across tens of vendors	Approximately 2 times more vendors processed with the same staffing level.
Data Utilization	Siloed data, manual interpretation	Leverages historical performance, risk data, and benchmarks	Improved accuracy in predicting vendor risk by up to 30 to 40%.
Transparency	Audit trails difficult to maintain	Maximum traceability with evaluation logs, justifications, and confidence scores	Audit-ready digital logs for every evaluation decision.
Continuous Improvement	Static evaluation model; no learning from past results	AI continuously learns from new data to refine evaluations	Model accuracy improves with each evaluation cycle.

WHAT NOW?

As we look ahead from the proof-of-concept to the beta phase, we anticipate:

- **Refining Prompt Library for Procurement Scenarios:** Optimize prompts tailored to procurement needs, reducing variability and improving interpretation accuracy across complex RFPs.
- **Establishing a Secure Sandbox Environment:** Build an isolated, access-controlled environment for testing AI enhancements without exposing sensitive procurement data.
- **Training Stakeholders on AI-Augmented Evaluation:** Develop targeted training for state procurement teams to interpret AI-generated insights confidently and apply informed human judgment.