Tips for evaluating AI proposals

Who	Procurement teams
What	Best practices for evaluation

Your evaluation methods will vary depending on what and how you are buying. As you evaluate proposals, keep your end user and/or beneficiary in mind based on your organization's user research. Make sure that your evaluation criteria and selection process speak to their needs. Here are some tips to consider.

General good practices

Lifecycle and vendor management

- Plan for ongoing model maintenance, updates, and version control
- Confirm vendor commitments to transparency about changes and potential impacts
- Include rollback and incident response procedures in contracts

Risk and compliance

- Ensure adherence to applicable AI regulations and standards in your jurisdiction
- Establish explainability requirements aligned with use case needs
- Define data governance policies covering data sharing, retention, and deletion
- Assess risks related to third-party dependencies and supply chains

Documentation and governance

- Maintain detailed records of evaluation criteria, scoring, and decisions
- Require vendors to disclose model limitations and ethical considerations
- Embed responsible AI requirements into procurement contracts

Additional considerations

- Compatibility with open source AI models or platforms, where appropriate
- Support for data interoperability and integration with existing public sector systems, tools, and data repositories
- Transparency and communication plans for public-facing AI applications
- Preparedness for public records requests or transparency laws impacting AI outputs
- Strategies to mitigate risks related to hallucinations, misinformation, or misuse
- Clearly define support and escalation pathways for incidents
- If appropriate and legally feasible given your context, consider involving end-users/beneficiaries themselves in the evaluation process.

Best practices by contract method

Method	Best practices
Open solicitation	**Evaluate AI-specific technical criteria** - Suitability of the type of model for your use case (general-purpose vs. domain-specific) - Model accuracy, reliability, and robustness, including known failure modes - Scalability, latency, and availability under expected workloads - Security features aligned with your jurisdiction's regulations (e.g., GDPR, HIPAA, - FedRAMP) - Data residency, sovereignty, and cloud region options to meet compliance needs **Understand technology and vendor transparency** - Clarity on training data provenance and governance - Vendor openness about model development, fine-tuning methods, and update - cadence - Infrastructure dependencies and integration capabilities with your existing IT ecosystem - Support for open standards and interoperability to avoid lock-in (e.g., standard APIs, model interchange formats) **Review vendor and solution track record** - Demonstrated enterprise or public sector deployments - References relevant to your region and sector - Certifications and compliance attestations applicable to your regulatory environment - Support for hybrid and multi-cloud strategies or on-premises deployments **Pricing and commercial terms** - Transparent usage-based or subscription pricing models

Method	Best practices
	• Availability of cost management and optimization tools • Clear delineation of charges for pilot vs. production environments • Contract flexibility for scaling usage or changing terms as needed **Security, privacy, and compliance** • Strong data protection measures, including encryption at rest and in transit • Capabilities to comply with data protection regulations (e.g., GDPR for EU, CCPA for US) • Auditability and logging features for traceability • Support for sovereign cloud or localized data centers where required **Responsible AI and ethical considerations** • Alignment with your organization's AI ethics framework and regulatory requirements • Bias detection and mitigation strategies embedded in the solution • Explainability features for AI outputs, especially where decisions impact citizens • Human oversight mechanisms and escalation procedures

Method	Best practices
Innovation procurement	**Include all above points, plus:** - Access to sandbox or trial environments for iterative testing - Flexible contracting to support phased delivery and learning cycles - Clear frameworks for intellectual property rights and data ownership - Opportunities for co-development or joint innovation with providers - Availability of professional services and ecosystem partnerships
Framework agreement (multiple awards) or under-threshold purchase	**Focus on essential due diligence:** - Standardized compliance and security documentation - Service Level Agreements (SLAs) with clear uptime and support commitments - Transparent cost structures and usage monitoring

Evaluation criteria

As you create your evaluation criteria, align around your metrics, what the terms really mean, and the implications ↗. For example, "accuracy" often comes up during evaluation. But other metrics matter as well, such as precision (which is also sometimes confused with accuracy by non-experts) and recall. In some use cases, precision and recall are trade-offs as parts of the overall accuracy of the solution. Speak to technical experts to understand the implications of these metrics in terms of false positives/negatives, and discuss relevant operational and legal implications.



Your evaluation criteria will be case-dependent. Some evaluation criteria for open solicitations you may consider include:

- **Problem fit**, including demonstrating a strong understanding of your need and the appropriateness of the solution.
- **AI approach and explainability**, such as model algorithm and training data, model transparency, algorithmic clarity, audit readiness.
- **Environmental impact**, such as energy use and source.
- **Data governance and bias handling**, including strategies for fairness, data quality, and addressing known model limitations.
- **Integration, training and operations**, such as fit with existing infrastructure, plans for training and upskilling internal staff, and ongoing support services.
- **Risk and accountability**, like liability mechanisms, governance structures, and human-in-the-loop system design.
- **Financial value**, such as cost over the lifecycle and real ROI expectations.
- **Vendor experience**, including relevant experience with similar AI solutions.