Essential questions for evaluating LLM vendors

What should you consider? This guide serves as a checklist of questions tailored for your RFPs. Ensure your LLM vendors meet or exceed your organization’s stringent security, data handling, and compliance requirements.

Technical architecture and deployment

**LLM type:** What is the foundational technology behind your LLMs—open-source, wrapper, or proprietary?
**Deployment options:** Is deployment available in a single-tenant or private cloud?
**Infrastructure needs:** What computing resources are required for deployment? What services do you offer to aid in self-hosting?
**Data handling:** How do you manage data separation and secure processing?
**Compatibility:** Can your product integrate with multiple or client-provided LLMs?

Data lifecycle management

**Data sources:** Where do you source the data used to train your foundation model?
**Data pre-processing:** How is data integrity maintained during pre-processing?
**Customer data use:** Are security controls in place to prevent customer data from informing the broader model?
**Data customization:** How is client data used for customization or fine-tuning handled?
**Data sharing:** Is client-specific data used for fine-tuning shared with other clients?
**Data privacy:** What safeguards are in place for personal data?
**Redaction features:** Does your system support the redaction of sensitive data?
**Sensitive information:** What measures are in place for data anonymization, encryption, and access controls to protect sensitive information?

Customization and integration features

**Proprietary datasets:** Can proprietary datasets be used to fine-tune your LLM?
**Resources:** Which resources get deployed within a client’s private cloud during implementation?
**Third-party integrations:** Does your product seamlessly integrate with third-party services or applications commonly used in enterprises?

Enterprise security

**Authentication:** What authentication methods do you support?
**Compliance:** Is your product in compliance with standards like SOC2 Type II, HIPAA, GDPR, and PCI?
**Access:** Who has access to the foundation models?
**Single Sign-On (SSO):** Do you support single sign-on capabilities? Do you support SCIM?
**Prompt injection:** What measures are in place to prevent malicious actors from injecting harmful prompts or code?
**Jailbreaking:** What safeguards and monitoring mechanisms are in place to identify and mitigate jailbreak attempts?
LLM output compliance

Bias and toxicity mitigation: What mechanisms are in place to mitigate bias and inappropriate content?

Bias and toxicity detection: Do you have any industry standards, benchmarks or thresholds you adhere to for toxicity detection? How are these updated and refined over time?

Bias data and mechanisms: What are the sources of bias data, and what mechanisms and methodologies are used for training and fine-tuning the AI model to manage bias?

Output filtering: Is there any automated post-output filtering we should be aware of?

Diversity: How do you ensure the diversity of the LLM’s generated output?

Other content filters: Are there other types of content or policies that can be implemented to identify and warn/block inappropriate or harmful content?

Legal and regulatory compliance

IP considerations: How do you ensure that generated outputs do not infringe on third-party intellectual property rights?

Ownership: Who retains ownership of the data inputs and generated outputs?

Controls: Is there functionality to prevent certain kinds of data from being input?

Legal history: Are there any ongoing, past, or threatened legal actions against your product?

Independent review: Have your models undergone independent, third-party reviews such as HELM, NYC Bias Audit, or EU AI Act readiness? Provide any third-party audits or certifications related to bias detection, toxicity detection, PII handling, and security.

Third-party standards: What compliance standards, such as GDPR, are followed for PII protection?

Scalability and performance

Data scalability: How does your solution handle scalability and performance when dealing with large datasets?

Performance: How do you handle high demand scenarios?

Hallucinations: How do you address hallucinations? Does the platform have human oversight and reviews in place?

Monitoring and reporting

Transparency: What tools and features are available to provide insights into the decision-making process of the AI model? How does your software address concerns related to ‘opaque’ AI systems, including explainability and transparency?

Visibility: What level of visibility is provided for telemetry and security events, and how can this data be accessed?

Reporting capabilities: What types of reports can be generated to assess the effectiveness and accuracy of controls?

Accuracy: What reports can be generated to provide insight into the accuracy of generated outputs?

User feedback: How do you collect and respond to user feedback?

Financial and operational considerations

Associated costs: Are there extra costs for fine-tuning or customization?

Support: What types of support and training are available to personnel?

Opt-out features: Is it possible to disable certain generative AI features at an enterprise level?