COMPTROLLER AND AUDITOR GENERAL OF INDIA
Office of the Comptroller & Auditor General

EXPRESSION OF INTEREST

EOI No.: CAG/AI-PLATFORM/EOI/05052026

Establishment of a Sovereign AI & Data Platform
with Agentic AI Applications

Infrastructure as a Service | Data & AI Platform | Intelligent Agents

Issuing Authority	Office of the Comptroller & Auditor General of India
EOI Reference	CAG/IS/AI-PLATFORM/EOI/05052026
Date of Issue	05 May 2026
Last Date for Submission of Queries / Clarifications	15 May 2026
CAG Response to Queries (Published)	20 May 2026
Last Date for EOI Response	26 May 2026
Tender Type	Open
EMD	NIL
Tender Fee	NIL
Contact	Sushant
Email	aao4is@cag.gov.in, diris@cag.gov.in

DISCLAIMER

This Expression of Interest (EOI) is not an agreement, nor an offer or commitment of any kind by the Office of the Comptroller & Auditor General of India (CAG). It is issued solely as a market engagement exercise to gauge the interest, capability, and readiness of qualified organisations in responding to the envisaged procurement. No contractual obligation whatsoever shall arise from this EOI process unless and until a formal contract is signed and executed by a duly authorised officer of CAG with the selected vendor.

The information contained herein is provided in good faith and to the best of the CAG's current knowledge. CAG reserves the right to amend, supplement, or withdraw this EOI at its absolute discretion and without incurring any liability to any party. Responses to this EOI will be used only to inform the design of a subsequent Request for Proposal (RFP). Submission of a response does not confer any preferential right or guarantee of participation in the RFP process.

No Earnest Money Deposit (EMD) is required for submission of this EOI. EMD requirements, if any, will be specified in the subsequent RFP. No Tender Fee is applicable for this EOI.

Information submitted in response to this EOI may be subject to disclosure under the Right to Information Act, 2005, subject to applicable exemptions under Section 8 thereof. Respondents should clearly mark any genuinely confidential commercial information.

The successful bidder(s) shall be required to sign an Integrity Pact with CAG as per the guidelines of the Central Vigilance Commission (CVC), applicable for procurements above the prescribed threshold.

TABLE OF CONTENTS

1. BACKGROUND AND CONTEXT

1.1 About the Comptroller and Auditor General of India

The Comptroller and Auditor General of India (CAG) is a constitutional authority established under Article 148 of the Constitution of India. CAG is the supreme audit institution of the country, entrusted with the audit of accounts of the Union, the States, and Union Territories with Legislature. Its mandate encompasses financial, compliance, and performance audits across the entire public expenditure ecosystem of India.

CAG operates through a network of field offices across all States and Union Territories, processing vast quantities of financial records, pension documents, tax returns, and compliance data. The scale, complexity, and analytical depth demanded by these functions create a compelling case for deploying advanced Data, AI, and Machine Learning capabilities in a governed, sovereign manner.

1.2 The Digital Transformation Imperative

CAG has undertaken a structured digital transformation journey with the objective of enhancing audit quality, reducing processing timelines, and enabling data-driven decision-making. Several targeted AI applications have already been piloted, based on which a larger objective of designing a self-sustained platform with some preliminary use cases is now being sought. Some of the immediate objectives include:

- Pension Pre-Scrutiny (AI-PARAS): An AI-enabled system to automate first-level scrutiny of pension proposals across Accountant General (A&E) offices, reducing processing time and improving detection of document deficiencies.
- AI-Based ITR & Assessment Order Audit: A system for automated analysis and audit of Income Tax Return documents.
- AI-Based Supplementary Financial Audit: An extension of the ITR audit capability to include supplementary financial statements and related compliance documents.
- Commercial Financial Audit Agent: AI-assisted audit of financial statements of Public Sector Undertakings (PSUs), Autonomous Bodies (ABs), and Urban Local Bodies (ULBs), supporting Supplementary Audit, preparation of Audit Comments and Separate Audit Reports (SARs), and automated preparation of chapters of the General Purpose Financial Report (GPFR).

1.3 The Need for a Unified Sovereign Platform

CAG now seeks to establish a Sovereign AI & Data Platform that will serve as the central digital backbone for all present and future AI and analytics workloads. This platform is intended to consolidate infrastructure, data, and AI services under a single governed environment — eliminating fragmentation, enforcing institutional ownership of data and models, and enabling rapid deployment of new intelligent applications without repeated ground-up investment.

As part of this initiative, CAG intends to commission four priority agentic AI applications — covering Pension Pre-Scrutiny, ITR Audit, Supplementary Financial Audit, and Commercial Financial Audit — as the first wave of intelligent workloads to be designed, built, and operated on this unified platform.

2. VISION AND OBJECTIVES

2.1 Vision

To establish a sovereign, scalable, and governed AI & Data Platform that empowers the Comptroller and Auditor General of India to deliver higher-quality, more timely, and data-driven audit outcomes — while retaining full institutional control over its data, models, and intelligent systems.

2.2 Strategic Objectives

- Establish a unified, cloud-native data foundation (Lakehouse architecture) that consolidates structured and unstructured institutional data across audit domains.
- Build shared, reusable AI and Machine Learning services — including document processing, OCR, NLP, embedding generation, and model management — accessible to all current and future applications.
- Enable agentic AI orchestration, allowing intelligent, multi-step autonomous workflows to be designed, deployed, and governed within a controlled institutional environment.
- Identify and integrate four priority agentic application use-cases (Pension Pre-Scrutiny, ITR Audit, Supplementary Financial Audit, and Commercial Financial Audit) as the first wave of use-cases on the platform.
- Enforce data sovereignty, ensuring all data, models, and AI processing remain within institutionally controlled infrastructure, with no dependency on uncontrolled external environments.
- Achieve vendor-agnosticism and open-standards alignment, preventing technology lock-in and enabling long-term evolution.
- Establish end-to-end governance, audit trails, and explainability mechanisms compliant with Government of India data protection and security frameworks.

2.3 Operational Objectives

- Reduce manual processing time for pension, audit, tax document scrutiny, and financial statement analysis of commercial entities through intelligent automation.
- Enable analytics and reporting workloads across CAG's national operations from a single data platform.
- Support up to 20,000 concurrent users at peak institutional load with 99.9% platform uptime during critical audit periods.
- Provide a scalable infrastructure capable of ingesting over 10 million documents annually and growing to at least 2 PB of managed data.
- Ensure a 5-year data retention capability with full audit reproducibility.

3. SCOPE OVERVIEW

The envisaged scope of this initiative is structured across three integrated pillars. Respondents may offer capabilities covering all three pillars as a prime contractor, or respond to demonstrate specialised capabilities in one or more pillars, subject to the consortium or subcontracting guidelines to be defined in the subsequent RFP.

Pillar	Component	Summary Description
Pillar 1	**Infrastructure as a Service (IaaS)**	Sovereign, scalable cloud or on-premises infrastructure including compute, GPU, storage, networking, and security. May include NIC/MeitY empanelled cloud or private cloud deployment.
Pillar 2	**AI & Data Platform**	Lakehouse data foundation, data ingestion, governance, metadata management, AI/ML services, model registry, vector database, RAG framework, agentic orchestration, low-code workflow designer, and observability layer.
Pillar 3	**Agentic AI Applications**	Development, deployment, and integration of four priority agentic use-cases: (i) Pension Pre-Scrutiny Agent (AI-PARAS), (ii) ITR & Assessment Order Audit Agent, (iii) Supplementary Financial Audit Agent, and (iv) Commercial Financial Audit Agent — each operating as a native consumer of the common platform.

The three pillars are interdependent by design. The IaaS layer provides the sovereign compute and storage substrate. The AI & Data Platform runs on top of this infrastructure and exposes shared services. The Agentic Applications are consumers of the platform capabilities, not independent technology stacks. Respondents are expected to demonstrate awareness of this layered architecture and how their proposed solutions align with it.

4. PILLAR 1 – INFRASTRUCTURE AS A SERVICE

4.1 Deployment Model

CAG seeks a sovereign infrastructure deployment that can operate in one or more of the following configurations, subject to detailed technical evaluation during the RFP stage:

- Government Cloud (GovCloud): Deployment on NIC Cloud (National Informatics Centre) or MeitY-empanelled cloud service providers, ensuring data residency within India.
- Private Cloud / Hybrid Cloud: Deployment on dedicated private cloud infrastructure hosted at government-approved data centres, with potential hybrid connectivity to public cloud for burst workloads. This should include on-premises services too.
- On-Premises at CAG-designated facilities: Subject to infrastructure readiness and security requirements.

All infrastructure must ensure that data sovereignty is maintained at all times — no data should reside on or transit through infrastructure outside India's geographic boundaries. The data should also be anonymised as per standard and should incorporate retention policies.

4.2 Compute Requirements

- Kubernetes-based containerised deployment environment.
- High-availability multi-zone (Multi-AZ) compute clusters with automated failover.
- GPU compute capability for AI/ML training and inference workloads — either provisioned directly or through secure, metered access to a GPU cluster via private VPC connectivity.
- Auto-scaling capability to handle peak loads of up to 20,000 concurrent users.
- Secure, isolated tenancy with namespace separation for different functional domains and applications.

4.3 Storage Requirements

- High-throughput object storage suitable for data lake and lakehouse architectures (minimum 2 PB scalable).
- Tiered storage: hot, warm, and cold tiers with automated lifecycle policies.
- Minimum 5-year retention with legal hold and secure deletion capabilities.
- Encrypted at rest (AES-256) and in transit (TLS 1.2+).

4.4 Networking and Security

- Private VPC with strict ingress/egress controls.
- Web Application Firewall (WAF), DDoS protection, and intrusion detection/prevention.
- Secrets management and certificate lifecycle management.
- Compliance with CERT-In guidelines, IT Act 2000 (as amended), and applicable Government of India cybersecurity frameworks.

4.5 SLA Expectations

SLA Parameter	Requirement
Platform Uptime (Baseline)	99.5%
Platform Uptime (Peak Audit Periods)	99.9%
Recovery Point Objective (RPO)	15 minutes or less
Recovery Time Objective (RTO)	2 hours or less
Backup	Automated with validated restore procedures

5. PILLAR 2 – SOVEREIGN AI & DATA PLATFORM

The AI & Data Platform is the core intellectual and operational layer of the initiative. It provides shared services across all institutional applications, ensuring that AI capabilities are built on a governed, reusable data foundation rather than isolated, application-specific silos.

5.1 Data Foundation – Lakehouse Architecture

- Open-standards Lakehouse (e.g., Apache Iceberg, Delta Lake, or equivalent) supporting ACID transactions, schema evolution, and time-travel queries.
- Unified storage for structured, semi-structured, and unstructured data (financial records, scanned documents, legal texts, pension registers, ITR data).
- Data ingestion framework supporting: batch ingestion, real-time streaming, API connectors, database connectors, file and document ingestion (PDF, scanned images, Excel), email ingestion, and web crawler integration.
- OCR and document intelligence services for processing scanned physical records at scale (minimum 10 million documents annually).
- Data partitioning, compaction, and lifecycle management for cost-efficient storage at petabyte scale.

5.2 Metadata & Governance Layer

- Centralised metadata catalog providing data discovery, lineage tracking, and impact analysis across all datasets and pipelines.
- Automated data quality profiling, monitoring, and alerting with configurable rules.
- Policy-driven Role-Based Access Control (RBAC) and Attribute-Based Access Control (ABAC) with fine-grained dataset permissions.
- Data masking, tokenisation, and anonymisation capabilities for sensitive fields (PII, financial data).
- Reproducible datasets — all datasets must be versioned and replayable for audit purposes.
- Compliance with data retention policies and legal hold requirements.

5.3 AI & Machine Learning Services

- Model-agnostic AI inference framework supporting self-hosted open-source models (e.g., Llama, Mistral, Qwen) and integration with third-party APIs where permissible.
- Model registry with version control, model locking, drift detection, and evaluation pipelines.
- Embedding generation and vector database for semantic search and similarity matching across large document corpora.
- Hybrid search combining dense vector search with traditional keyword-based retrieval.
- Retrieval-Augmented Generation (RAG) framework with grounding enforcement, re-ranking, and hallucination mitigation controls.
- Fine-tuning infrastructure for domain adaptation of foundation models on institutional datasets.
- Feature engineering pipelines and feature store for ML workloads.

5.4 Agentic Orchestration & Workflow

- Multi-step agentic workflow orchestration supporting autonomous, human-in-the-loop, and hybrid execution modes.
- Tool integration framework allowing agents to call external APIs, internal data services, and third-party systems under controlled governance.
- Low-code workflow designer enabling domain experts (non-engineers) to design, configure, and deploy agents with institutional guardrails.
- Event-driven workflow triggers (schedule, data event, API call, document arrival).
- Execution replay and audit trail for all agentic operations — every agent action must be logged, reproducible, and explainable.

5.5 Security & AI Guardrails

- Prompt injection defence and adversarial input detection.
- Toxicity and content moderation filters for all AI-generated outputs.
- Grounding enforcement — AI responses must be attributable to specific source documents or data.
- Token rate limiting and API gateway controls.
- Prompt logging and model output logging with tamper-proof audit trails.

5.6 Observability & Monitoring

- Centralised logging, telemetry, and cost monitoring across all platform components.
- Pipeline execution tracking, failure diagnostics, data freshness monitoring, and SLA dashboards.
- Data drift monitoring and model performance degradation alerts.
- Usage metering for compute, storage, and AI API calls to enable internal cost attribution.

6. PILLAR 3 – AGENTIC AI APPLICATIONS

Four priority agentic use-cases have been identified for Phase 1 deployment on the platform. These use-cases span CAG's core operational domains — pension processing, income tax return audit, supplementary financial audit, and commercial financial audit — and represent the highest-priority areas for intelligent automation within CAG's institutional mandate. Respondents are expected to demonstrate their capability to design, build, and deploy these agents as native consumers of the unified Sovereign AI & Data Platform.

6.1 Agent 1 – Pension Pre-Scrutiny Agent (AI-PARAS)

Use-Case: Automated first-level scrutiny of pension proposals submitted to Accountant General (A&E) offices across States. CAG has conducted internal assessments and exploratory work on this use-case. This EOI invites organisations to demonstrate their capability to deliver a production-grade implementation on the unified platform.

Core Agent Capabilities Required:

- Ingestion of pension case documents (service books, salary certificates, PPO forms, pension calculation sheets) — both digitally submitted and scanned physical documents.
- OCR and semantic extraction of structured fields from multi-format pension documents.
- Automated validation against pension rules — CCS Pension Rules, state-specific adaptations, and variants for Family Pension, DCRG, Commutation, Disability, and Special Pension.
- Detection of document gaps, calculation errors, rule mismatches, and inconsistencies with explainable findings presented to pension processing officers.
- Multi-state, multi-office deployment with logical data isolation per state.
- Integration with existing case management workflows and officer hierarchy (Dealing Hand, AAO, SAO, DAG).
- Human-in-the-loop design — the agent assists but does not sanction; all final decisions remain with designated officers.
- Full audit trail of agent findings, officer reviews, and case outcomes.

6.2 Agent 2 – ITR & Assessment Order Audit Agent

Use-Case: Automated audit and analysis of Income Tax Return (ITR) documents as part of CAG's statutory audit mandate. CAG has conducted internal assessments and scoping exercises for this use-case. The intent is to deliver a production-grade ITR & Assessment Order Audit Agent as an integral part of the unified Sovereign AI Platform.

Core Agent Capabilities Required:

- Ingestion and parsing of Income Tax Return documents (ITR-1 through ITR-7 forms and all associated schedules) in digital and scanned formats.
- Ingestion and parsing of Assessment Orders (including summary assessment orders under Section 143(1), scrutiny assessment orders under Section 143(3), and reassessment orders under Section 147/148) — extraction of assessed income, additions, disallowances, penalties, and demand raised.
- Cross-referencing of ITR-declared figures against Assessment Order-determined figures: identification of under-assessments, omitted income, incorrectly allowed deductions, and cases where AO's additions were reduced or dropped at appellate stages.
- Analysis of compliance with Income Tax Act provisions — detection of anomalies against applicable provisions (Sections 10, 14A, 37, 40, 43B, 80-series deductions, etc.) and identification of high-risk patterns.
- Generation of structured audit observations with dual-source attribution — each observation linked to the specific ITR schedule line and the corresponding Assessment Order paragraph, with explainable rationale and applicable statutory provision cited.
- Bulk processing capability with configurable sampling strategies (full population, risk-based stratified sample, or random sample).
- Risk-scoring and prioritisation of cases based on multiple signals: quantum of assessed income, magnitude of additions/disallowances, appellate history, sector-specific risk indicators, and year-on-year variance.

6.3 Agent 3 – Supplementary Financial Audit Agent

Use-Case: AI-assisted audit of supplementary financial statements and related compliance documents submitted by audited entities. CAG has identified this as a priority use-case and conducted preliminary scoping.

Core Agent Capabilities Required:

- Ingestion and analysis of supplementary financial statements, appropriation accounts, and related audit-relevant documents.
- Automated reconciliation and consistency checks across financial data sets.
- Detection of material misstatements, classification errors, and non-compliance with applicable accounting standards and government financial regulations.
- Structured output of audit observations linked to source document sections for auditor review.
- Support for multi-year trend analysis and comparative analytics across audit entities.
- Tight integration with the ITR Audit Agent for unified, cross-domain financial audit workflows.

6.4 Agent 4 – Commercial Financial Audit Agent

Use-Case: AI-assisted audit of financial statements of Public Sector Undertakings (PSUs), Autonomous Bodies (ABs), and Urban Local Bodies (ULBs), supporting Supplementary Audit, preparation of Audit Comments and Separate Audit Reports (SARs), and automated preparation of chapters of the General Purpose Financial Report (GPFR).

Core Agent Capabilities Required:

- Audit Universe Management: Automated creation and maintenance of a comprehensive audit universe covering PSUs, Autonomous Bodies, and Urban Local Bodies, populated through API-based ingestion from MCA-21, Department of Public Enterprises, SAMARTH e-Gov Suite, OIOS, and CAG Connect.
- Financial Data Capture and Repository: Structured ingestion and storage of financial statements at the line-item level via OCR-based extraction, API-based pull from MCA-21, and structured manual entry.
- Risk Assessment and Audit Planning: AI-computed risk scores based on materiality thresholds, year-on-year variances, ratio analysis, and trend deviations. Early warning indicators for significant financial movements, persistent losses, and abnormal patterns.
- Audit Intelligence and Analytics: Cross-entity comparison, sector-wise benchmarking, multi-year trend analysis, and identification of outliers and anomalies. Integration with CEDAR for GSTN data and CoEFA for data analytics outputs and risk models.
- GPFR Preparation: Automated, auditable preparation of General Purpose Financial Report chapters for Union (CPSEs) and States (SPSEs), with validation rules, maker-checker controls, and comprehensive audit trails.
- Role-Based Dashboards and MIS: Configurable, role-specific dashboards with drill-down from aggregate to entity-wise, sector-wise, and period-wise views.
- Inter-Audit Linkages: Automatic tagging of financial audit findings relevant to compliance or performance audits, with integration with OIOS, SFR portal, and CAG Connect through secure APIs.
- Human-in-the-Loop Design: All AI-generated risk scores, audit observations, and GPFR draft content are subject to maker-checker review by designated officers before finalisation.

6.5 Common Agent Requirements

- All agents must execute on the Sovereign AI & Data Platform and consume shared platform services (OCR, embedding, vector search, RAG, governance).
- Each agent must implement human-in-the-loop controls — no automated output is to be treated as a final determination without officer review.
- All agent outputs must be fully explainable, traceable to source data, and logged in the platform audit trail.
- Agents must support configurable rule sets to accommodate jurisdictional and procedural variations.
- Performance and accuracy benchmarks for each agent will be defined and agreed upon during the RFP stage.

6.6 AI Transformation Office

CAG aims to transform its business operations leveraging Agentic AI and achieve continuous service excellence. Vendors are required to propose an AI Transformation Office to cover the following scope:

- Central body to oversee, align, and guide all AI initiatives required to deliver the contracted services.
- Work with CAG stakeholders to define a 5-year AI-driven transformation roadmap.
- Identification and prioritisation of use cases to drive intelligent automation, enhance operational efficiency, and reduce dependency on manual intervention through Agentic AI capabilities.

7. GUIDING PRINCIPLES

The following principles shall govern the design, implementation, and long-term operation of the platform. Respondents are expected to demonstrate alignment with these principles in their capability statements.

Principle	Interpretation
Data-First Architecture	Data ingestion, governance, and lineage precede application and AI deployment. All AI workloads operate on governed datasets within the Lakehouse. Applications must not maintain independent data silos.
Platform-First Enablement	The platform is a shared institutional foundation. All applications are consumers of platform capabilities, not independent stacks. This eliminates duplication and ensures consistent operational standards.
Modular Design and Pluggable Tools	The platform shall be designed as a collection of loosely coupled, independently replaceable components. Each functional layer must expose well-defined APIs and be substitutable without requiring changes to adjacent layers. Open-source and standards-based tooling is preferred.
Governance by Design	Governance controls are embedded in the architecture — not implemented through manual oversight. This includes automated lineage, audit trails, RBAC/ABAC, and reproducibility of AI outputs.
Metadata-Driven Operations	Datasets, pipelines, models, and workflows are discoverable, traceable, and governed through centralised metadata services. No process should operate outside the governed metadata layer.
Sovereignty & Residency	All data, models, and AI processing must reside within institutionally controlled, India-based infrastructure. Dependence on uncontrolled external cloud or AI services is not permitted.
Open Standards & No Lock-In	Technology choices must avoid proprietary lock-in. Preference for open-source standards (Apache Iceberg, Kubernetes, OpenTelemetry, etc.) that can be operated, migrated, and evolved by CAG or its future partners.
Human-in-the-Loop	AI and agentic systems are decision-support tools. Statutory and administrative decisions remain with designated human officers. AI findings must be presented as structured inputs to human judgment, not as autonomous determinations.
Explainability & Auditability	Every AI output — whether from an analytical model or an agentic workflow — must be explainable, attributable to source data, and reproducible for audit scrutiny.

8. INDICATIVE PHASING

The implementation is envisaged across four phases. Detailed timelines, milestones, and deliverables will be defined in the RFP. The following represents the indicative sequencing:

Phase	Title	Key Deliverables
Phase 1	**Core Platform Foundation**	IaaS deployment; Lakehouse foundation; data ingestion framework; OCR & document intelligence; baseline AI services; governance and metadata layer; security hardening.
Phase 2	**Application Enablement**	Onboarding of all four agentic use-cases onto the platform; data pipelines for pension, ITR, and financial audit datasets; system integration and user acceptance testing.
Phase 3	**Advanced Intelligence**	Agentic workflow maturation; cross-domain analytics; advanced RAG and multi-step reasoning; model fine-tuning on institutional data; low-code agent designer enablement.
Phase 4	**Institutional Scale**	National rollout across all AG offices; performance optimisation; capacity expansion; transition to steady-state operations; knowledge transfer to CAG team.

9. INDICATIVE SCALE AND CAPACITY REQUIREMENTS

The following table provides indicative scale parameters to assist respondents in assessing their capability and preparing meaningful responses. These are current estimates and will be refined through the EOI engagement and subsequent RFP process.

Parameter	Indicative Requirement
Baseline Concurrent Users	1,000
Peak Concurrent Users	Up to 20,000
Annual Document Ingestion Volume	10 Million+ documents per year
Annual Data Growth	150 TB per annum (minimum)
Platform Lifecycle Storage Capacity	Scalable to 2 PB Raw and beyond
Daily AI Inference Calls (Peak)	5 Million+ per day
Data Retention Period	5 Years (minimum)
Platform Uptime (Baseline / Peak Audit)	99.5% / 99.9%
Recovery Point Objective (RPO)	15 minutes
Recovery Time Objective (RTO)	2 hours
Number of Priority Agentic Applications	4 (Phase 1); extensible
Number of States / AG Offices in Scope	All States and UTs (28 + 8)

10. ELIGIBILITY, PRE-QUALIFICATION AND TECHNICAL QUALIFICATION CRITERIA

The criteria set out in this section are indicative and will be formalised in the subsequent RFP. They are published at the EOI stage to allow prospective respondents to self-assess their eligibility and to inform CAG's understanding of the market. Respondents are encouraged to respond even where they meet criteria through consortium arrangements, as described in Section 10.1.

10.1 Consortium Arrangements

- A single Lead Entity (Prime) must be designated and will hold primary contractual responsibility for the full scope of the engagement throughout its duration.
- All Pre-Qualification (PQ) criteria must be met individually by the Lead Entity. PQ criteria cannot be aggregated across consortium members or subcontractors.
- The Lead Entity may engage subcontractors for defined portions of the scope. The Lead Entity remains solely accountable for all subcontracted work and must disclose the identity, role, and scope of all proposed subcontractors.
- Technical Qualification (TQ) criteria may be demonstrated by the Lead Entity, any named consortium member, or any named subcontractor — provided that entity is committed to the engagement for the scope area to which the credential relates.
- Where the Lead Entity, a consortium member, or a subcontractor has itself delivered work as a subcontractor to a prime SI in a prior engagement, such credentials are fully eligible provided the relevant subcontract document names the end client and describes the scope of work performed.
- Acceptable evidence for TQ credentials includes: (a) a work order or purchase order from the end client; (b) a work order or purchase order from a prime SI clearly naming the end client and the scope of work; or (c) a completion certificate issued by the prime SI or the end client. Respondents may redact commercial and financial values from submitted documents.

10.2 Pre-Qualification Criteria (Lead Entity)

#	Criterion	Specification	Evidence Required	Type
PQ1	**Legal registration in India**	Company incorporated under Companies Act 1956/2013, or LLP registered in India	Certificate of Incorporation	Pass/Fail
PQ2	**Minimum years of operation in India**	At least 5 years in active operation as on the date of EOI submission	Certificate of Incorporation / GST registration	Pass/Fail
PQ3	**Information security certification**	Valid ISO 27001:2022 ISMS certification as on date of submission	ISO certificate (number and validity)	Pass/Fail
PQ4	**Average annual IT/digital services turnover**	₹1,500 Crore or above, averaged over the last three audited financial years	Audited financial statements or CA-certified turnover certificate	Pass/Fail
PQ5	**Financial health**	Positive net worth in the latest audited financial year; profitable in at least two of the last three financial years	Audited financial statements	Pass/Fail
PQ6	**Not blacklisted or debarred**	Not blacklisted/debarred by any Central/State Govt, PSU, or autonomous body as on date of submission	Self-declaration on letterhead	Pass/Fail

PQ7	**No disqualifying litigation**	No pending litigation that may materially affect the ability to execute the contract	Self-declaration on letterhead	Pass/Fail
PQ8	**Minimum IT workforce**	At least 500 full-time IT professionals as on the bid submission date	Self-declaration on letterhead with HR certificate	Pass/Fail
PQ9	**Minimum AI/ML workforce**	At least 50 full-time AI/ML professionals as on the bid submission date	Self-declaration on letterhead with HR certificate	Pass/Fail

11. EOI RESPONSE REQUIREMENTS

Interested organisations are invited to submit a structured capability statement addressing the following sections. Responses should be concise, factual, and focused on demonstrating relevant capability and experience. The total response (excluding annexures) should not exceed 30 pages.

Section A: Organisational Profile

- Legal name, registered address, nature of entity, year of incorporation.
- Overview of the organisation's core business and areas of practice relevant to this EOI.
- List of key personnel (name, designation, brief bio) who would be involved in this engagement.
- ISO 27001:2022 certification status (provide certificate number and validity).

Section B: Technical Capability

- Approach to Sovereign AI & Data Platform implementation architecture philosophy, preferred technology stack, and alignment with the guiding principles described in this EOI.
- Experience with Lakehouse architectures (specify open-source components used Iceberg, Delta, Hudi, etc.).
- AI/ML services capability model-agnostic frameworks, self-hosted LLMs, fine-tuning, RAG, vector databases.
- Agentic AI capability frameworks used (LangGraph, CrewAI, Autogen, custom orchestration), governance and human-in-the-loop implementation approach.
- IaaS capability NIC Cloud empanelment status, private cloud offerings, GPU provisioning.

Section C: Relevant Project Experience

- Up to five (5) reference projects most relevant to this scope. For each project, provide: Client name and type, scope summary, technologies used, scale (users /data volumes), your role (prime/sub/consortium), and current status.
- Specifically highlight any experience with Government of India audit institutions, constitutional bodies, or large-scale public sector data platforms.

Section D: Proposed Approach to the Four Agentic Use-Cases

- High-level approach to designing and deploying the Pension Pre-Scrutiny Agent (AI-PARAS) on the platform including document processing architecture, pension rule engine design, and multi-state data model.
- Approach to ITR & Assessment Order Audit Agent dual-document ingestion architecture, cross-referencing logic, IT Act provision rule engine, and structured audit observation generation.
- Approach to Supplementary Financial Audit Agent cross-document reconciliation, anomaly detection, and integration with ITR Agent.
- Proposed development approach for each agent architecture, tooling, and rationale.

Section E: Consortium and Subcontracting Details (if applicable)

- Names and roles of all consortium members, if any.
- Names, roles, and scope of all proposed subcontractors, if any.
- Allocation of responsibilities across the three pillars (Infrastructure, Platform, Agents) for each entity in the proposed delivery structure.
- Confirmation of Lead Entity and its programme management and accountability framework for the consolidated scope.

Section F: Questions and Clarifications

Respondents are encouraged to raise questions or seek clarifications about the requirements, scope, or procurement approach. CAG will use these questions to improve the RFP and may share anonymised clarifications with all respondents.

12. PROOF OF CONCEPT (POC) REQUIREMENTS

As part of the RFP evaluation process, CAG intends to require shortlisted bidders to conduct a time-bound Proof of Concept (POC) prior to commercial bid opening. The detailed POC framework will be specified in the RFP; the following describes the indicative structure to allow respondents to plan and prepare.

12.1 Purpose and Scope of the POC

- Validate that the bidder can deploy a functional, integrated Sovereign AI & Data Platform within a defined timeframe.
- Demonstrate end-to-end operation of at least one agentic use-case on the platform, from data ingestion through AI processing to a structured, explainable output with human-in-the-loop controls.
- Provide CAG with a scored, comparative basis for differentiating shortlisted bidders beyond their written proposals.

12.2 Who Conducts the POC

- Only bidders shortlisted after technical evaluation of written proposals will be invited to participate. CAG anticipates shortlisting two to three bidders for the POC stage.
- The POC must be conducted by the Lead Entity, with active participation from named subcontractors or consortium members contributing to the relevant scope areas.
- All costs of conducting the POC shall be borne entirely by the bidder. CAG will not compensate any bidder for POC participation.

12.3 Platform Layer - Mandatory Features to Demonstrate

#	Platform Feature	Minimum Demonstration Standard	Scoring Weight
P1	**Data Ingestion – multi-format**	Ingest at least two document formats (e.g. PDF, scanned image, structured data file) into the Lakehouse within a single pipeline run. Demonstrate schema detection and metadata tagging on ingestion.	8 marks
P2	**Lakehouse – ACID transactions and time travel**	Demonstrate a data update operation followed by a time-travel query that retrieves the prior state of the dataset. Must use an open-source Lakehouse format (Iceberg, Delta Lake, or Hudi).	8 marks
P3	**Metadata catalog and data lineage**	Show automated lineage tracking from raw ingestion through to a processed dataset. Demonstrate dataset discovery via the metadata catalog with at least two searchable attributes.	6 marks
P4	**Data governance – access control and audit trail**	Demonstrate RBAC-based access restriction (at least two roles with different dataset visibility). Show a tamper-evident audit log of data access and pipeline execution events.	6 marks
P5	**OCR and document intelligence**	Process a batch of at least 50 scanned or image-based documents through the OCR pipeline. Demonstrate structured field extraction with confidence scoring.	8 marks
P6	**Vector database and hybrid search**	Load a document corpus of at least 10,000 documents into the vector store. Execute a hybrid search query (semantic + keyword) and demonstrate ranked retrieval with source attribution.	8 marks
P7	**Self-hosted LLM inference**	Demonstrate inference on a self-hosted open-source LLM (not a third-party API). Show a prompt-response cycle with latency metrics and prompt logging.	6 marks

P8	**Observability dashboard**	Show a live monitoring view with at least: pipeline execution status, data freshness indicator, AI inference latency, and an active alert triggered by a simulated anomaly.	5 marks

Platform layer total: 55 marks. A bidder must score a minimum of 38 marks (70%) on the platform layer to proceed to the agent demonstration scoring. Failure to meet this threshold will be treated as a non-qualifying POC.

12.4 Agentic Use-Case Demonstration Bidder's Choice

Each shortlisted bidder must select and demonstrate one of the four priority agentic use-cases. The bidder declares their chosen use-case at the time of submitting their written technical proposal. Changing the selected use-case after proposal submission is not permitted.

12.5 Agentic Use-Case Scoring Criteria

#	Scoring Dimension	What the Evaluator Looks For	Marks
A1	**End-to-end pipeline integration**	Data flows from ingestion through the Lakehouse, through the AI services layer, into the agent orchestration layer, and produces a structured output — without manual intervention at any stage.	10 marks
A2	**Accuracy and relevance of AI output**	AI-generated findings, observations, or anomaly detections are accurate against the synthetic test dataset. Output is relevant, specific, and attributable to source data fields.	15 marks
A3	**Explainability and source attribution**	Every AI finding is traceable to a specific source document, field, or rule. The system can present a human-readable rationale for each finding. No black-box outputs.	10 marks
A4	**Human-in-the-loop controls**	A human reviewer can view, accept, modify, flag, or override any AI finding through a defined workflow. Override actions are logged. The system does not auto-finalise any output without officer action.	8 marks
A5	**Platform-native operation**	The agent uses the platform's shared services (ingestion pipeline, vector search, LLM inference, audit log) rather than independent or external components. No data leaves the platform boundary during operation.	2 marks

Agent demonstration total: 45 marks. Combined POC total: 100 marks (55 platform + 45 agent). The POC score will constitute 30% of the overall technical evaluation score in the RFP. The remaining 70% will be based on the written technical proposal.

12.6 Data, Environment, and Rules of Engagement

- CAG will provide synthetic or suitably anonymised sample datasets for each use-case prior to the POC. Bidders must not use real CAG data or real taxpayer or pensioner data at any stage of the POC.
- Bidders may conduct the POC on their own cloud or on-premises infrastructure.
- The POC demonstration will be conducted in-person at a venue designated by CAG in New Delhi, or via a live remote session at CAG's discretion. Pre-recorded demonstrations will not be accepted.
- Each bidder will be allocated a fixed time slot: a half-day for the platform layer demonstration and a half-day for the agent demonstration including a Q&A session.

- Bidders may not materially change their infrastructure, software components, or demonstration dataset between written proposal submission and the POC. Any material divergence may result in disqualification.
- The evaluation panel's scores on each POC dimension are final. No re-demonstration will be permitted after the POC session concludes.

13. PROCUREMENT PROCESS

The envisaged procurement process following this EOI is described below. CAG reserves the right to modify, extend, or abbreviate any stage of this process without prior notice.

Step	Stage	Description
1	**Expression of Interest (This Document)**	Market engagement to gauge capability and interest. Responses used to shape the RFP.
2	**EOI Evaluation & Market Dialogue**	CAG will review all EOI responses with support from its advisors. Selected respondents may be invited for structured dialogue to clarify and refine requirements.
3	**RFP Issuance**	A detailed RFP will be issued incorporating learnings from the EOI process. The RFP will define technical specifications, evaluation criteria, and commercial framework.
4	**Pre-Bid Meeting**	Formal pre-bid interaction with all eligible bidders. Queries and clarifications will be addressed and published.
5	**Technical & Commercial Bid Submission**	Bidders submit detailed proposals. Evaluation will follow QCBS (Quality cum Cost Based Selection) methodology.
6	**Evaluation & Shortlisting**	Technical evaluation of bids. Shortlisted bidders (typically top 2–3) are invited to proceed to the POC stage.
7	**Proof of Concept (POC)**	Shortlisted bidders demonstrate the platform and at least one agentic use-case on CAG-provided synthetic data. POC performance is scored and forms part of the final technical score.
8	**Commercial Bid Opening**	Commercial bids of technically qualified bidders opened. Final QCBS score computed.
9	**Contract Award**	Contract execution with the selected System Integrator. Implementation commences as per agreed Phase 1 milestones.

Submission of a response to this EOI does not confer any entitlement to participate in the subsequent RFP, nor does it constitute pre-qualification.

14. KEY DATES

S.No	Activity	Date
1	Publication of EOI	**05 May 2026**
2	Last Date for Submission of Queries / Clarifications	**15 May 2026**
3	CAG Response to Queries (Published)	**20 May 2026**
4	Last Date for Submission of EOI Responses	**26 May 2026**

Note: All dates above are confirmed for this EOI publication. Respondents are advised to monitor the Central Public Procurement Portal (CPPP) at https://eprocure.gov.in/eprocure/app for any updates, corrigenda, or extensions.

15. INSTRUCTIONS FOR SUBMISSION

15.1 General Conditions

- Tender Type: Open Tender | Category: Services | EMD: NIL | Tender Fee: NIL
- Responses must be submitted in English only. Responses in any other language will not be evaluated.
- Responses should be submitted as a single PDF document, clearly structured as per the sections described in Section 11 of this EOI.
- Supporting documents (ISO certificates, client references, financial statements) should be compiled as annexures with clear labelling.
- The total response (excluding annexures) should not exceed 30 pages.
- CAG is not obligated to acknowledge or respond to any EOI submission.
- CAG reserves the right to accept or reject any response without assigning any reason.
- Submission of a response does not incur any obligation on the part of CAG to proceed with a procurement or to include the respondent in any future process.
- All costs associated with preparation and submission of a response are to be borne entirely by the respondent.
- All shortlisted respondents shall be required to execute a Non-Disclosure Agreement (NDA) with CAG before any sensitive institutional data, process details, or datasets are shared during subsequent stages of the procurement process. The NDA template will be provided by CAG at the appropriate stage.
- Information submitted in response to this EOI may be shared internally within CAG and with its empanelled advisors for the purpose of procurement planning.
- By submitting a response, the respondent confirms that the information provided is true, accurate, and complete to the best of their knowledge.