Tender Analyzer: A Profile-Centric System for Automated Bid Discovery, Structured RFP Decomposition, and Submission Readiness

Utkarsh Chaudhari
Dept. of Computer Science & Eng.
MIT School of Computing, MIT ADT University
Pune, Maharashtra, India

Rohit Jaiswal
Dept. of Computer Science & Eng.
MIT School of Computing, MIT ADT University
Pune, Maharashtra, India

Tamanna Khatik
Dept. of Computer Science & Eng.
MIT School of Computing, MIT ADT University
Pune, Maharashtra, India

Disha Bendre
Dept. of Computer Science & Eng.
MIT School of Computing, MIT ADT University
Pune, Maharashtra, India

Prof. Revati Raspayle
Guide, Dept. of Computer Science & Eng.
MIT School of Computing, MIT ADT University
Pune, Maharashtra, India

Abstract*—Securing public-sector or enterprise contracts requires prospective vendors to locate relevant solicitations, verify statutory eligibility, and inspect tender packages that routinely exceed one hundred pages. Because these tasks span disconnected portals and unstructured documents, smaller firms frequently miss viable bids when they overlook regulatory exemptions or buried penalty terms. This article introduces Tender Analyzer, a locally hosted Flask and SQLite software prototype that unifies these stages into a single guided workflow. Upon registration, an enterprise provides a format-validated CIN or GSTIN alongside its business domain, product catalogue, annual revenue, operating tenure, quality certifications, geographic footprint, and MSME or DPIIT status. Solicitations ingested from simulated feeds representing 52 procurement networks are scored on a 0–100 scale via a six-attribute weighted model (domain 25, TF-IDF cosine similarity 25, turnover 15, experience 15, certifications 10, and geography 10), adjusted for small-business and startup relaxations, and assigned an Eligible, Partially Eligible, or Ineligible status. Any selected notice—or an uploaded PDF, DOCX, or TXT file—is categorized as an RFP, RFQ, EOI, or RFI and rendered as a seven-part structured briefing paired with a clause-referencing question-answering panel. The system further gates bid dispatch behind a mandatory document checklist, emulates Class-3 Digital Signature Certificate (DSC) signing, and computes an indicative win-likelihood estimate bounded between 28% and 94%. Multilingual support spans English, Hindi, and Marathi across three interface themes. Since portal polling, registry validation, and cryptographic transmission operate in simulation and empirical accuracy benchmarks have not yet been executed, this work documents architectural design and functional behavior while treating the issuing authority's original solicitation as definitive.

***Index Terms*—Bid readiness, cosine similarity, electronic procurement, MSME concessions, multilingual web application, RFP decomposition, tender recommendation, TF-IDF.**

I. INTRODUCTION

A procurement tender constitutes an official solicitation through which a purchasing organization invites vendors or contractors to submit technical and commercial proposals for a specified requirement. While every published notice represents a potential revenue opportunity, pursuing it is worthwhile only when the prospective bidder satisfies the buyer's thresholds regarding financial turnover, prior execution tenure, regulatory accreditations, regional presence, and administrative documentation. Identifying suitable opportunities is therefore an essential business-development function rather than a routine clerical task.

In practice, systematic tender discovery poses substantial operational friction. Within India, central government solicitations are hosted on the Central Public Procurement Portal (CPPP) [5] and the Government e-Marketplace (GeM) [6]; individual state administrations operate independent e-procurement portals; public-sector undertakings and large private corporations maintain proprietary vendor hubs; and third-party aggregators such as BidAssist and TenderTiger syndicate subsets of these listings. Because each platform enforces distinct taxonomies, metadata schemas, date conventions, and currency units, organizations monitoring multiple channels must repeatedly execute manual queries and reconcile heterogeneous records.

Locating a relevant solicitation is only the initial hurdle; interpreting the tender dossier itself presents a second major bottleneck. A typical bid package combines legal clauses, technical schedules, compliance annexures, and subsequent corrigenda, scattering critical parameters—such as minimum turnover, operational tenure, earnest money deposit (EMD), liquidated-damages rates, performance bank guarantees, and closing timestamps—across dozens of pages. Smaller enterprises are disproportionately disadvantaged: teams often remain unaware that Udyam-certified MSMEs and DPIIT-recognized startups qualify for EMD waivers or relaxed eligibility thresholds, while overlooking a single mandatory annexure can trigger automatic technical rejection.

Tender Analyzer consolidates these fragmented activities into a cohesive end-to-end pipeline. First, it evaluates and orders incoming solicitations against a structured enterprise profile; second, it decomposes a selected tender into a standardized seven-section analytical view; and third, it enforces pre-submission completeness checks. All platform outputs serve strictly as decision-support guidance—helping bidding teams prioritize effort while leaving final verification to the official solicitation text and qualified human review.

A. Problem Statement

Prospective suppliers currently navigate disconnected procurement websites and manually inspect lengthy tender

documents merely to determine basic qualification. Existing listing portals broadcast notices without evaluating enterprise-specific fit, whereas generic document summarizers condense text without accounting for the reader's corporate credentials or statutory privileges. This paper investigates how to engineer an integrated platform that (i) scores solicitations against a specific vendor profile while automatically applying regulatory concessions, (ii) transforms raw tender text into an instrument-aware seven-part briefing, and (iii) prevents premature submission of incomplete bid packages.

B. Research Questions

RQ1: How can domain alignment, lexical text similarity, financial capacity, operating tenure, accreditation overlap, and regional coverage be synthesized into a transparent relevance score and qualification verdict for a given enterprise–tender pair? **RQ2:** How can unstructured solicitation documents be distilled into a uniform seven-section analytical layout whose strategic recommendations adapt to whether the notice is an RFP, RFQ, EOI, or RFI? **RQ3:** How can a prototype deliver preliminary win-likelihood feedback and enforce dossier completeness without overstating predictive certainty? **RQ4:** What experimental evaluation framework can rigorously assess ranking relevance, field-extraction fidelity, and user experience without misrepresenting unmeasured benchmarks as empirical findings?

C. Objectives

- Onboard vendor accounts using structurally validated CIN or GSTIN identifiers, manage credential authentication, and support interface localization across English, Hindi, and Marathi.
- Maintain a structured corporate profile capturing primary business domain, sector specializations, product catalogue keywords, annual revenue, year of incorporation, completed project count, quality certifications, operating jurisdictions, and MSME or DPIIT registration flags.
- Populate a local tender repository from simulated multi-portal feeds and rank listings through a weighted six-factor scoring model paired with a three-tier eligibility classification.
- Ingest uploaded PDF, DOCX, TXT, or pasted solicitation text, identify the underlying procurement instrument category, and generate a structured seven-section analytical briefing.
- Provide an interactive question-answering assistant that responds to bidder queries by citing relevant clauses from the active tender text.
- Gate final bid dispatch behind a complete mandatory document checklist and present an indicative win-probability score alongside target discount guidance.

D. Scope and Contributions

The implemented prototype spans eight functional subsystems: enterprise onboarding and structural verification, multi-portal feed ingestion, concession-aware profile matching, procurement instrument classification, seven-dimension document parsing, clause-grounded Q&A assistance, gated bid dispatch with win estimation, and a configuration hub governing company profiles, visual themes, languages, notifications, documentation, support tickets, and account security. Specifically, this work offers four primary contributions: (1) a unified workflow bridging initial tender discovery to bid-readiness verification; (2) an interpretable weighted matching algorithm that incorporates MSME and startup relaxations while exposing criterion-by-criterion pass/fail explanations; (3) an instrument-sensitive document analyzer that explicitly flags whether commercial pricing is permitted; and (4) a structured validation methodology that clearly separates demonstrated prototype mechanics from future empirical benchmarking.

II. BACKGROUND

A. Tender Instruments

Issuing authorities select among distinct solicitation formats depending on procurement complexity. A Request for Proposal (RFP) seeks a customized technical solution and is typically evaluated under Quality-and-Cost Based Selection (QCBS) or a two-envelope procedure in which commercial bids remain sealed until technical qualification concludes. A Request for Quotation or Invitation for Bid (RFQ/IFB) addresses standardized goods or routine services and awards the contract to the lowest compliant bidder (L1). Conversely, an Expression of Interest or Pre-Qualification Query (EOI/PQQ) serves strictly to shortlist capable vendors; including commercial pricing at this stage can cause immediate disqualification. Finally, a Request for Information (RFI) conducts non-binding market discovery. Treating these four instruments identically misdirects preparation effort, so the platform elevates instrument classification to a core analytical dimension.

B. Statutory Identity of a Bidder

Corporate applicants are formally identified through either a Corporate Identification Number (CIN) or a Goods and Services Tax Identification Number (GSTIN). A 21-character CIN encodes listing status, a five-digit industrial classification code, state jurisdiction, incorporation year, ownership structure, and a sequential registration number. A 15-character GSTIN comprises a two-digit state code, the entity's 10-character Permanent Account Number (PAN), an entity registration counter, the fixed character 'Z', and a checksum character. Tender Analyzer validates adherence to these statutory patterns and extracts embedded metadata, but does not execute live API calls against Ministry of Corporate Affairs (MCA) or GSTN servers; consequently, its verification indicator denotes syntactic compliance rather than external registry confirmation.

C. Text Similarity with TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) encodes text as a sparse weighted vector in which a token's weight increases with its local occurrence in a document and decreases across documents where it appears frequently [8]. Similarity between two vectors u and v is measured via their cosine angle:

$$cos(u, v) = (u \cdot v) / (\|u\| \|v\|) \quad (1)$$

The matching engine constructs TF-IDF vectors from an enterprise's product-catalogue keywords and a solicitation's scope description using scikit-learn [7]. Relative to dense neural embeddings, TF-IDF is computationally lightweight, deterministic, and transparent to audit, although it cannot capture semantic equivalence across distinct synonyms and may assign elevated similarity to adjacent yet incompatible offerings (such as software consulting versus hardware procurement). Accordingly, lexical similarity serves as just one component among six evaluation signals.

D. Concessions for MSMEs and Startups

Public procurement policies frequently grant financial and qualification relief to smaller businesses by waiving bid security and lowering eligibility bars. The matching engine models three standard relaxations: an Udyam-registered MSME receives a full waiver on the Earnest Money Deposit (EMD) alongside a 50% reduction in the minimum annual turnover threshold, while a DPIIT-recognized startup is exempted from prior operational tenure requirements. Within the prototype, these adjustments operate as built-in screening heuristics; actual applicability always remains governed by the explicit terms of each solicitation.

III. RELATED WORK AND RESEARCH GAP

Government portals including CPPP [5] and GeM [6] host official tender notices and offer basic keyword search, whereas commercial platforms aggregate listings and deliver email alerts. Those systems focus primarily on notice distribution; by contrast, Tender Analyzer layers enterprise-specific eligibility scoring, concession modeling, and clause-referenced document breakdown directly atop solicitation discovery.

Within academic literature, retrieval-augmented generation (RAG), formulated by Lewis et al. [2], pairs neural text generation with evidence passages retrieved from external corpora. Zhao and Li [1] apply retrieval and large language models (LLMs) to the buyer side of procurement to draft semi-structured tender specifications from historical templates. In a bachelor's thesis supervised by Ren, da Silva [3] benchmarks multiple LLMs on contract clause extraction, observing that retrieval assists with lengthy documents but remains constrained by retriever precision. Similarly, Nascimento [4] investigates LLM-driven extraction of structured attributes from municipal procurement gazettes in Santa Catarina, evaluating performance against a manually curated gold dataset of 100 notices.

TABLE I
COMPARISON WITH EXISTING PLATFORMS AND LITERATURE

Source	Focus	Difference in Tender Analyzer
CPPP, GeM [5], [6]	Government notice hosting and basic catalog search	Layers enterprise-specific scoring and concession-aware verdicts onto notice discovery
BidAssist, TenderTiger	Private syndication of public procurement listings	Adds instrument-sensitive decomposition and a gated pre-submission workflow
RAG [2]	Text generation grounded in retrieved external passages	Omitted in current prototype; planned for future citation-grounded Q&A
Tender RAG [1]	Automated drafting of tender specifications for buyers	Vendor-side screening of published solicitations rather than buyer drafting
Contract extraction [3]	Post-award contract clause extraction via LLMs	Pre-bid tender analysis using deterministic rule-based extraction
Gazette study [4]	Structured field extraction from municipal gazettes	Couples attribute extraction with profile ranking and dossier readiness checks
Tender Analyzer	Unified discovery, decomposition, and submission readiness	Unified local prototype integrating discovery, analysis, and readiness

Research gap. Existing listing aggregators help suppliers find solicitations, while document-extraction studies examine lengthy contracts or notices in isolation. Few architectures reported in prior literature bridge both stages via an explicit vendor profile incorporating statutory MSME/startup exemptions and carry that context forward into pre-submission readiness checks. Tender Analyzer is designed as an integrated prototype to address this end-to-end workflow gap.

IV. SYSTEM OVERVIEW

Tender Analyzer is structured around eight modular subsystems backed by a shared SQLite relational store and unified session state.

A. Onboarding and Verification. During registration, applicants select English, Hindi, or Marathi and submit a CIN or GSTIN that undergoes syntactic pattern validation before account creation proceeds. Subsequent form sections capture corporate metadata (registered legal name, email address, designated contact, and operating jurisdiction), operational history (year of incorporation, three-year average turnover in INR lakhs, completed contract count, and largest executed project value), technical capabilities (primary domain across Goods, Services, or Works, eighteen sector tags, free-text product keywords, and SKU volume), and statutory flags covering MSME registration, DPIIT startup recognition, and quality certifications. Three pre-configured demonstration profiles—spanning medical devices, IT services, and highway infrastructure—enable immediate platform exploration without manual onboarding.

B. Ingestion. A background ingestion module simulates parallel polling across 52 procurement channels encompassing central government portals, state e-tender platforms, public-sector undertakings, and private enterprise sourcing hubs. Each polling pass executes four sequential operations: probing endpoints, normalizing heterogeneous currency notations into uniform INR floating-point values and dates into ISO 8601 format, gating out cancelled, expired, or malformed notices, and upserting valid records into the database. A live telemetry strip displays active portal counts, mean response latency, and the latest synchronization timestamp, updating dynamically across cycles.

C. Matching and Dashboard. Incoming tenders are evaluated against the active vendor profile using the weighted formulation in Section V-A. The main workspace presents a verified-company header card, the ingestion telemetry strip,

four summary KPI cards, and three navigation tabs (Personalized Matches, Real-Time Portal Ingest, and Audit & Submissions), supported by multi-faceted filtering across keywords, authority category, tender type, instrument class, eligibility verdict, sector, and sort order.

D. Instrument Classification and Forensic Analysis. A document parser extracts raw text from PDF (via pypdf), DOCX (via python-docx), or plain-text files, after which an analytical classifier categorizes the notice as RFP, RFQ/IFB, EOI/PQQ, or RFI and populates the seven-section report detailed in Section VI.

E. Clause-Citing Assistant. Positioned alongside the structured report, a conversational panel answers user questions regarding the active solicitation while pointing to relevant document clauses. Six one-click prompt shortcuts address instrument strategy, EMD and MSME waivers, delay penalties, required certifications, payment schedules and performance guarantees, and startup exemptions. Responses render in the active UI language and persist in the database.

F. External RFP Analyzer. Users may upload external PDF, DOCX, TXT, or Markdown files up to 32 MB—or paste raw tender text—for standalone inspection outside the ingested database. The engine produces the identical seven-part report, scores eligibility against the active company profile, and allows selecting the preferred output language.

G. Application Gateway. A mandatory-document upload dossier, commercial bid input field, emulated Class-3 DSC signing step, and win-probability estimator comprise the final bid-readiness stage (Section V-D).

H. Settings and Localisation. Seven configuration tabs manage corporate profile updates, interface themes, language selection, alert preferences, knowledge-base guides, help-desk tickets, and password security. Navigation labels, tender cards, evaluation matrices, and assistant prompts translate dynamically across English, Hindi, and Marathi, while three visual themes (Clean Light, Midnight Navy, and Cyber Slate) style the entire application.

End-to-end journey. (1) The applicant registers using a CIN or GSTIN and corporate capability details. (2) Statutory syntax is verified and the profile is persisted. (3) Tender notices arrive via simulated portal feeds. (4) The matching engine scores, orders, and labels each opportunity. (5) The user filters the dashboard and selects a solicitation. (6) The document is classified and structured into seven analytical sections. (7) The user consults the clause-citing assistant. (8) Required checklist files are uploaded; while any slot remains empty, the bid stays in a pending state. (9) The user enters a commercial quote informed by a recommended L1 discount. (10) The platform performs a simulated submission and issues an audit receipt displaying an estimated win probability. (11) The bidder cross-checks critical terms against the authoritative tender file.

V. ARCHITECTURE AND METHODOLOGY

Information flows sequentially across seven operational stages outlined in Table II. Stages 1–3 process structured profile and notice metadata, Stage 4 analyzes unstructured document text, and Stages 5–7 govern bid preparation and human verification.

TABLE II
PROCESSING STAGES, INPUTS, AND OUTPUTS

Stage	Input and processing	Output
1. Onboarding	Registration inputs; CIN/GSTIN regex validation; salted password hashing	Validated company profile
2. Ingestion	Simulated 52-portal feeds; parallel probe, normalize, gate, and upsert	Normalised tender records
3. Matching	Company profile and tender; six weighted signals; MSME/startup relief	Score (0–100) and verdict
4. Analysis	Solicitation text; instrument classifier; seven-part report; clause Q&A	Structured report and answers
5. Gateway	Checklist file uploads and commercial quote; completeness verification	PENDING or SUBMITTED; receipt
6. Estimation	Quoted bid, qualification cushion, and active competitor count	Win probability and pricing advice
7. Human review	Structured briefing compared against authoritative source tender	Final bid/no-bid decision

A. Weighted Matching

Given an enterprise c and a solicitation t, the matcher computes six normalized sub-scores within [0, 1]: domain alignment D (Goods, Services, or Works), TF-IDF cosine similarity E between product keywords and tender scope, turnover adequacy T (company revenue relative to required minimum turnover), experience adequacy X (years active relative to required tenure), certification overlap C (such as ISO, CMMI, BIS, or Class-1 contractor credentials), and geographic overlap L. The overall match score is formulated as

$$S(c, t) = 25D + 25E + 15T + 15X + 10C + 10L \qquad (2)$$

These coefficients sum to 100 and represent heuristic engineering weights rather than parameters trained on annotated historical outcomes. Statutory relaxations adjust thresholds prior to scoring: the 50% MSME turnover relief halves the denominator in T, the DPIIT startup exemption waives the tenure requirement in X, and the MSME EMD waiver is surfaced directly alongside the stated bid-security figure. Based on individual criterion outcomes, each tender receives a categorical verdict: **Eligible** indicates that all mandatory conditions are satisfied or waived; **Partially Eligible** indicates a turnover or certification shortfall where a joint venture or consortium is advised; and **Ineligible** denotes a core domain or fundamental credential mismatch. On the detail screen, every criterion carries an explicit pass, covered, or exempt status badge so bidders can audit how the verdict was derived.

B. Instrument-Aware Guidance

Instrument classification governs both strategic bidding advice and whether commercial pricing may be included in the submission (Table III).

TABLE III
TENDER INSTRUMENTS AND THE GUIDANCE ATTACHED TO EACH

Instrument	Evaluation	Price in bid	Suggested focus
RFP	Quality-and-cost scoring (QCBS) or two-envelope evaluation	Yes, in a second envelope	Technical methodology, prior case studies, and SLA compliance
RFQ / IFB	Lowest price (L1)	Yes, as an upfront schedule	Competitive unit pricing; avoid unnecessary narrative bulk
EOI / PQQ	Stage-1 prequalification	Prohibited; submitting prices risks automatic rejection	Audited financials, quality accreditations, and statutory licenses
RFI	Non-evaluative market review	No	Influencing upcoming technical specifications

C. Winning-Probability Estimate

Once a bidder initiates or completes an application, a predictive scoring module computes an indicative probability of award. Letting B represent a baseline score and d denote identified credential or liquidity deficiencies, the win estimate is calculated as

$$P = min(94, max(28, B - 12d + \Delta C + \Delta P + \Delta E)) \quad (3)$$

where ΔC adjusts for market rivalry (+15 percentage points when fewer than two competitors are active in the segment, 0 for two or three rivals, and −10 for four or more), ΔP evaluates the entered financial quote against a target L1 discount band of 6.8% to 11.0% below estimated tender value, and ΔE adds a qualification cushion when company turnover exceeds 1.5 times the minimum threshold or when completed project volume is high. Absent optional accreditations or tight financial headroom count toward d. The interface renders this figure alongside a qualitative confidence label, a target discount suggestion, and a competitive threat indicator. Because all constants are fixed heuristics in source code rather than empirically calibrated regression weights, the score serves solely as directional guidance.

D. Gated Submission

The submission module renders individual upload slots for every mandatory document required by the detected instrument; the demonstration RFP specifies eleven items. While any required slot remains unpopulated, the application persists in a PENDING state and final dispatch stays locked. For sandbox demonstrations, an auto-attach utility populates sample signed files. Once 100% of checklist items are attached, submission transitions the record to SUBMITTED and generates an audit receipt containing a UUID v4 transaction reference, target gateway endpoint (such as a GeM/CPPP direct-submission interface), ISO 8601 timestamp, final commercial quote, and simulated Class-3 DSC verification status. Both cryptographic signing and portal transmission operate locally in simulation without contacting external government servers.

E. Security Measures

User credentials are protected via salted password hashing in Werkzeug, session state relies on signed cookies, and all company profiles, bid applications, and chat histories are linked to user accounts via relational foreign keys. Because the prototype executes locally over unencrypted HTTP, production-grade TLS transport, strict multi-tenant isolation, and automated data-retention schedules remain outside the current implementation scope.

VI. ANALYSIS OUTPUT DESIGN

Each processed solicitation is organized into seven standardized panels (Table IV). Binding eligibility conditions are visually separated from general narrative text, while the pre-qualification matrix benchmarks tender thresholds directly against stored enterprise attributes. Strategic recommendations dynamically adapt to the detected instrument class, and the source tender file remains accessible for download and manual verification.

TABLE IV
SEVEN-SECTION REPORT LAYOUT AND SCREENING PURPOSE

Section	Content	Screening purpose
1. Summary	Core procurement objective and concise executive synopsis	Quick relevance check
2. Pre-qualification	Turnover, tenure, EMD, accreditations, and geography vs. vendor data	Eligibility triage
3. Scope	Technical deliverables, standards, delivery sites, and acceptance tests	Capability fit
4. Schedule	Publication, pre-bid meeting, closing deadline, opening, and delivery	Bid planning
5. Commercial terms	Performance bank guarantee, milestone billing, and payment terms	Liquidity check
6. Penalties and risk	Weekly liquidated damages, maximum penalty cap, and risk clauses	Risk assessment
7. Checklist	Instrument-specific mandatory documents and upload dossier	Submission compliance

VII. IMPLEMENTATION

The system is implemented as a monolithic Python web server exposing RESTful JSON endpoints for bid submission, pending upload persistence, external document parsing, clause-citing Q&A, and feed synchronization, paired with HTML page routes for onboarding, authentication, dashboard views, tender details, standalone analysis, and settings. Table V summarizes the software stack and Table VI details the core SQLite schema.

TABLE V
COMPONENTS AND THEIR REALISATION

Component	Responsibility	Realisation
Frontend	Pages, filters, themes, localisation	Jinja2 views; CSS custom properties for three themes; vanilla ES6 JS; embedded EN/HI/MR dictionaries
Backend	Routing, sessions, engines	Python 3.9–3.12, Flask 3+, Werkzeug; modular engines for ingestion, verification, matching, parsing, Q&A, prediction, and dispatch
Database	Persistent state	SQLite3 relational schema with seven tables, foreign keys, and JSON fields
NLP	Text similarity	scikit-learn TfidfVectorizer and cosine similarity

Component	Responsibility	Realisation
Documents	Text extraction	pypdf and python-docx extractors; native reading for TXT and Markdown
Not present	—	Zero external LLM APIs, vector databases, OCR engines, or cloud dependencies

TABLE VI
PRINCIPAL DATABASE TABLES

Table	Principal fields
users	Primary key, unique login email, salted password hash, enterprise name, locale preference, creation timestamp
company_profiles	Foreign key to users, CIN/GSTIN strings, syntactic verification flag, entity type, annual revenue, incorporation year, primary sector, MSME/DPIIT flags, JSON lists for accreditations and jurisdictions
tenders	Unique tender ID, solicitation title, buyer name and sector, procurement mode, instrument class, estimated budget, bid security amount, closing date, category, region, raw document text
tender_applications	Foreign keys to user and tender, workflow state, quoted price, UUID receipt token, target endpoint, JSON dossier manifest, dispatch timestamp
chat_messages	Foreign keys to user and tender, author role, dialogue text, message timestamp
support_tickets	Help-desk support inquiries submitted via the configuration hub

VIII. TESTING AND EVALUATION METHODOLOGY

System verification combines unit tests covering text parsing, scoring math, and identifier regex validation, integration tests checking cross-module state transitions, and structured user acceptance scenarios. Table VII outlines eleven functional test cases specifying expected prototype behavior; empirical pass rates and user-study metrics are reserved for future evaluation rather than reported here.

TABLE VII
FUNCTIONAL VALIDATION PLAN (TC01–TC11)

ID	Scenario	Expected result
TC01	Register using a syntactically valid CIN or GSTIN and complete metadata	Account persisted; structural verification badge displayed
TC02	Submit an invalid statutory identifier or omit required fields	Error prompt shown; profile creation blocked
TC03	Navigate to dashboard and toggle domain, type, or status filters	Displayed solicitation cards dynamically update to match criteria
TC04	Evaluate tenders against an MSME and DPIIT startup profile	EMD exemption and threshold relaxations applied; verdict updated
TC05	Submit a valid PDF, DOCX, or TXT file to the standalone analyzer	Seven-section structured report generated and scored
TC06	Attempt uploading an unsupported format or file exceeding 32 MB	Upload rejected with descriptive validation feedback
TC07	Query the Q&A sidebar regarding bid security or delay penalties	Response returned citing the corresponding tender clause
TC08	Process sample RFP, RFQ, EOI, and RFI solicitation documents	Accurate instrument badge and pricing eligibility rule shown
TC09	Toggle active interface language and visual theme in settings	All pages, tables, and prompts update to selected locale and palette
TC10	Attach partial checklist files first, followed by remaining documents	Dispatch disabled while pending; audit receipt issued upon completion
TC11	Compute win probability across varied bid discounts and rival counts	Estimated score remains strictly bounded within 28%–94%

Quantitative protocol. Subsequent empirical evaluation should: (1) benchmark tender ranking using Precision@k, Recall@k, Mean Reciprocal Rank (MRR), and NDCG against keyword-only and domain-only baselines using relevance judgments from at least two independent domain annotators; (2) assess document extraction fidelity via field-level precision, recall, and F1 scores against manually annotated RFP ground truth; (3) measure end-to-end review speed against timed manual screening while gathering System Usability Scale (SUS) ratings; and (4) calibrate win-probability estimates against historical award outcomes using proper scoring rules such as the Brier score, ensuring strict deduplication between development and test splits.

IX. DISCUSSION, LIMITATIONS, AND FUTURE WORK

Observations. Stepping through the seeded demonstration data illustrates the intended workflow (Appendix C): an enterprise reporting ₹1.40 crore in annual turnover is classified as eligible for a ₹3.80 crore medical-equipment RFP, its ₹7,60,000 EMD is marked as waived under MSME rules, and bid dispatch remains gated until all eleven checklist documents are uploaded. These walkthroughs verify functional logic on synthetic records rather than constituting empirical accuracy or time-savings claims.

TABLE VIII
FAILURE MODES AND MITIGATIONS

Risk	Effect	Mitigation or status
Incomplete profile	Low or skewed scores	Mandatory registration inputs; per-criterion pass/fail explanations
Scanned PDF	Little or no text extracted	Unsupported in current build; OCR integration planned
Unusual clause wording	Field missed or misplaced	Mandatory manual verification against authoritative source PDF
Parser artefacts	Empty bullets or repeated headings in a section	Regex and post-processing cleanup scheduled
Stale or simulated data	Outdated deadline or scope	Cross-verify deadlines and corrigenda on official portals
Over-trusting the estimate	Misplaced confidence in a bid	Explicitly labeled as heuristic; statistical calibration planned

Limitations. Portal ingestion currently relies on simulated feeds and seeded tender records; CIN and GSTIN verification validates syntactic format without querying live MCA or GSTN databases; Class-3 DSC signing and gateway transmission are emulated; scanned image-only PDFs cannot be parsed due to the absence of an OCR pipeline; document extraction uses

deterministic rule-based heuristics rather than LLM reasoning; matching weights, concession thresholds, and win-probability constants are uncalibrated engineering heuristics; and deployment targets local execution without HTTPS, container orchestration, or cloud hosting.

Future work. Key next steps include: (1) integrating live API and scraper connectors for CPPP, GeM, and state procurement portals; (2) implementing automated corrigendum diffing alongside proactive deadline alerts; (3) adding OCR support for scanned tender PDFs and expanding regional language coverage; (4) introducing a retrieval-grounded LLM assistant inspired by [2] with page-level source citations; (5) learning matching weights from labeled bid outcomes and statistically calibrating win-probability estimates; (6) hardening security and containerizing the application for cloud deployment; and (7) releasing a public benchmark dataset for procurement tender matching.

X. CONCLUSION

This article introduced Tender Analyzer, an integrated software prototype uniting profile-driven tender recommendation, instrument-sensitive RFP analysis, and gated bid-readiness verification. By embedding MSME and startup concessions directly into the scoring pipeline, surfacing criterion-level eligibility rationales, and preserving immediate access to the source solicitation, the platform reduces early-stage screening friction for smaller enterprises. Its current contribution lies in its cohesive system design and structured validation plan, setting a clear foundation for live portal integration, empirical accuracy benchmarking, and statistical calibration.

ACKNOWLEDGMENT

The authors express their sincere gratitude to Prof. Revati Raspayle for her continuous mentorship and valuable insights throughout the development of Tender Analyzer, as well as to the Department of Computer Science and Engineering, MIT School of Computing, MIT Art, Design and Technology University, Pune, for providing the academic infrastructure supporting this work.

REFERENCES

[1] Y. Zhao and D. Li, "A large language model-based framework for semi-structured tender document retrieval-augmented generation," arXiv:2410.09077, Oct. 2024.

[2] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, pp. 9459–9474.

[3] W. A. da Silva, "Evaluation of large language models in contract information extraction," B.Sc. thesis (advisor: T. I. Ren), Centro de Informática, Univ. Federal de Pernambuco, Recife, Brazil, Oct. 2024.

[4] F. V. Nascimento, "Extração de dados a partir de publicações do Diário Oficial dos Municípios de Santa Catarina usando large language models," B.Sc. thesis, Computer Science, Univ. Federal de Santa Catarina, Florianópolis, Brazil, Jul. 2025.

[5] National Informatics Centre, Government of India, "Central Public Procurement Portal (CPPP)." [Online]. Available: https://eprocure.gov.in/

[6] Government of India, "Government e-Marketplace (GeM)." [Online]. Available: https://gem.gov.in/

[7] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[8] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Inf. Process. Manage., vol. 24, no. 5, pp. 513–523, 1988.

APPENDIX A: LOGICAL DATA MODEL

Each company profile is linked to a single user account and records a CIN or GSTIN with a structural-verification flag, legal entity category, annual turnover, incorporation year, primary domain, MSME and DPIIT startup indicators, JSON arrays of certifications and operating states, and registration metadata covering completed project count, peak project value, sector tags, catalogue keywords, and SKU count. Each tender record stores a unique reference identifier, title, issuing authority name and sector type (government or private), tender category, instrument type, estimated contract value, EMD requirement, closing deadline, domain category, location, and full extracted solicitation text. Application records associate a user with a tender and track submission state (PENDING or SUBMITTED), quoted commercial amount, UUID transaction reference, gateway endpoint label, JSON manifest of uploaded dossier files, and submission timestamp. Chat records persist user-tender dialogue turns with sender role, message body, and timestamp to support session continuity.

APPENDIX B: PSEUDOCODE

```
PROCEDURE RankTenders(profile, tenders):
  validate profile; read MSME / startup flags
  FOR each tender t:
    thr_turnover = t.min_turnover * (0.5 if MSME else 1)
    thr_exp = 0 if startup else t.min_experience
    D = domain_match(profile, t)
    E = cosine(tfidf(profile.catalogue), tfidf(t.scope))
    T = ratio(profile.turnover, thr_turnover)
    X = ratio(profile.years, thr_exp)
    C = cert_overlap(profile, t)
    L = location_overlap(profile, t)
    S = 25D + 25E + 15T + 15X + 10C + 10L
    verdict = Eligible | Partial | Ineligible
              (from criterion outcomes)
  RETURN tenders sorted by S descending

PROCEDURE AnalyseDocument(file_or_text):
  check type (PDF/DOCX/TXT/MD) and size <= 32 MB
  text = extract(file_or_text)
  kind = classify(text)  # RFP|RFQ|EOI|RFI
  report = parse_seven_sections(text)
  attach strategy_advice(kind); price_allowed(kind)
  RETURN report

PROCEDURE Submit(user, tender, bid, uploads):
  IF any mandatory slot is empty:
    save status = PENDING; RETURN locked
  p = clamp(B - 12*d + dC + dP + dE, 28, 94)
  save SUBMITTED, uuid4, timestamp, bid
  RETURN receipt + p + pricing_advice
```

APPENDIX C: WORKED EXAMPLE

The medical-devices demonstration enterprise is a DPIIT-recognized startup established in 2021 with ₹1.40 crore in annual revenue and active Udyam MSME registration. Inspecting the matched solicitation for advanced medical diagnostic equipment and ICU patient monitors (estimated contract value ₹3.80 crore, floated by a central healthcare authority) produces the criterion breakdown in Table IX.

TABLE IX
PRE-QUALIFICATION VIEW FOR THE SAMPLE RFP

Criterion	Tender requires	Company has	Result
Annual turnover	Minimum ₹120 lakh (3-year average)	₹140 lakh verified	Satisfied

Criterion	Tender requires	Company has	Result
Operational age	Minimum 3 years in domain	5 years (incorporated 2021)	Satisfied
EMD	₹7,60,000 bid security	Registered Udyam MSME	Full waiver applied
Location	New Delhi / Nationwide	Pan-India, Maharashtra, NCR	Jurisdiction matched
Overall	—	Composite score 89%	Qualified

The structured briefing also highlights a 5% performance bank guarantee and a liquidated-damages rate of 0.5% per week capped at 10% of total contract value. Eleven mandatory submission documents are enumerated in the dossier, starting at 0 of 11 uploaded so dispatch remains locked. The commercial quote input defaults to ₹353.4 lakh, reflecting the estimated value discounted by the suggested 7% L1 margin.

APPENDIX D: RESPONSIBLE USE

Because automated screening tools can shape commercial bid/no-bid commitments, Tender Analyzer enforces seven responsible-use guardrails: (1) all match scores and verdicts are presented strictly as advisory indicators accompanied by per-criterion explanations; (2) statutory verification badges explicitly disclose that checks are syntactic rather than live registry lookups; (3) simulated subsystems (portal feeds, DSC signing, and gateway dispatch) are transparently labeled; (4) win-probability outputs are clearly framed as uncalibrated heuristics; (5) users are instructed to monitor official portals for corrigenda and verify every binding clause in the original solicitation; (6) account credentials are salted and hashed with strict foreign-key record ownership; and (7) unmeasured performance metrics are never presented as empirical results.