أكاديمية كاوست
KAUST ACADEMY

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Tadqeeq: An Explainable Multi-Agent System for ZATCA E-Invoicing Compliance

Sara Alqahtani, Ghady Almaghrabi, Ahmed Radwan

sara.alqahtani@kaust.edu.sa, ghady.almaghrabi@kaust.edu.sa, ahmedyradwan02@gmail.com

Abstract

Electronic invoice compliance with ZATCA requirements is a difficult challenge for small-to-medium enterprises, given the difficulty of manually checking invoices against the applicable rules and regulations. This project, **Tadqeeq**, is a multi-agent system that verifies an existing e-invoice – image, scan, or PDF, from any source – against ZATCA's e-invoicing regulations. It combines OCR/vision-language field extraction, a deterministic rule engine for the compliance decision, and Retrieval-Augmented Generation (RAG) grounded in official ZATCA documentation for every explanation. The key architectural feature: the compliance decision (Pass/Fail) is issued exclusively by the deterministic rule engine, while the language model explains that decision using citations verified, before delivery, against the retrieved regulatory evidence itself. Across evaluation, Tadqeeq achieved 100% retrieval recall, 98.3% rule-application accuracy, and a Groundedness score of 4.29/5 under independent LLM judging – against a score of approximately 1.0/5 for baseline systems and for an ablated variant of Tadqeeq with retrieval disabled, isolating grounding as the direct, measured contribution of the RAG component.

Date: August 24, 2026

Code: https://github.com/consar21/TADQEEQ

Project Page: https://github.com/Gha0-0/tadqeeq-invoice-compliance

Contents

1 Introduction **2**

2 Related Work **2**

3 Methodology **3**

3.1 Design Principle . . . 3

3.2 Pipeline Architecture . . . 3

4 Results **5**

4.1 Retrieval and Citation Layer . . . 5

4.2 Extraction Layer . . . 6

4.3 Compliance Decision Layer . . . 6

4.4 Baseline Comparison . . . 6

5 Discussion **7**

5.1 Decoupling Extraction Confidence from Compliance Verdict . . . 7

5.2 Precision-Recall Trade-offs in Rule Design . . . 7

5.3 Invoice-Type-Conditional Rule Application . . . 8

5.4 Dual-Path QR Code Detection . . . 8

5.5 Query Wording as a First-Class Retrieval Variable . . . 9

5.6 Grounded Conciseness versus Verbose Completeness . . . 9

6 Limitations **9**

6.1 OCR Model Reliability . . . 9

6.2 Rule Coverage . . . 9

6.3 External Dependency on a Third-Party LLM API . . . 10

6.4 Scope Boundaries . . . 10

7 Future Work **10**

8 Conclusion **10**

References **11**

1 Introduction

In the modern financial landscape, regulatory compliance is paramount, particularly in Saudi Arabia where the Zakat, Tax and Customs Authority (ZATCA) has pioneered strict electronic invoicing mandates. Non-compliance carries severe real-world impacts, including heavy financial penalties and reputational damage; yet, traditional manual auditing remains slow, costly, and highly prone to human error. To bridge the gap between unstructured financial documents and rigid legal frameworks, there is a critical need for an automated, end-to-end mechanism that can rapidly ingest, verify, and query invoice data. To address this challenge, this work proposes **Tadqeeq**, an autonomous, AI-driven audit assistant that fuses generative AI with deterministic logic to verify *existing* invoices, from any source. The system utilizes computer vision and OCR to structure raw invoice files into a machine-readable representation, processes them through a deterministic compliance validation engine to verify that all mandatory fields are present, correctly formatted, and arithmetically consistent, and grounds every resulting explanation in a Retrieval-Augmented Generation (RAG) pipeline indexing official ZATCA documentation – so that a compliance narrative is never presented to the user without a real, checkable citation behind it. A dedicated verification stage confirms every generated citation against the retrieved evidence itself before a report is finalized. Ultimately, Tadqeeq minimizes manual overhead while ensuring audit accuracy, strict compliance alignment, and explanations that are auditable rather than merely confident-sounding.

2 Related Work

Several categories of existing tools intersect with ZATCA e-invoicing, but each addresses a different problem than the one Tadqeeq solves.

Invoice generation libraries. `php-zatca-xml` [1] and `mrsool/zatca` [2] build, sign, and submit ZATCA-compliant XML invoices, assuming the underlying invoice data is already correct and structured. They extract nothing from an existing document, verify nothing against an already-issued invoice, and provide no citation-grounded explanation of a compliance outcome.

Accounting platforms with built-in compliance. Commercial platforms such as Qoyod perform arithmetic and format validation at the moment an invoice is created within the platform itself [3], and `lavaloon-eg/ksa_compliance` [4] integrates a similar generation-and-submission flow into an ERP system. These checks operate only on invoices generated inside the platform's own structured interface; they are not designed to audit an arbitrary, externally-sourced invoice image or PDF after the fact, and their validation feedback is not paired with a citation to the specific regulatory source underlying each check.

ZATCA's official clearance platform. ZATCA's own Fatoora platform clears and validates invoices submitted to it in real time as part of mandatory Phase Two integration, and separately offers a public invoice-validity lookup service. This is the regulatory authority itself, not a business-facing compliance-auditing tool; it validates invoices actively being submitted through its integration, not an arbitrary existing invoice a business wishes to check independently, and its output is a clearance decision rather than a cited explanation.

Across all three categories, no existing tool reads an arbitrary, already-issued invoice from any source and produces a compliance verdict paired with a citation to the specific regulatory text supporting that verdict. Tadqeeq addresses this gap directly: verification of existing invoices, source-agnostic, with every explanatory claim grounded in a real, checkable citation rather than an unverified assertion of confidence.

3 Methodology

3.1 Design Principle

Tadqeeq is built around a single architectural constraint: **rules decide Pass/Fail; the language model only explains**. Field extraction and compliance rule-checking are deterministic and fully auditable. Every explanatory claim the system produces is grounded via Retrieval-Augmented Generation (RAG) in official ZATCA regulation documents and vetted regulatory sources, and a dedicated verification stage checks every generated citation against the actually-retrieved evidence before it reaches the final report.

3.2 Pipeline Architecture

The operational pipeline is implemented as a graph-orchestrated multi-agent workflow (LangGraph) with seven stages and two conditional feedback loops, rather than a single feed-forward pass.

1. Extraction.. Given an input invoice $\mathcal{I}$, the system first attempts native text extraction directly from the PDF, without invoking any vision-language model. If native extraction is unavailable or yields insufficient text (a scanned or image-based document), the system falls back to OCR, and subsequently to direct vision-language model (VLM) extraction on the raw image, producing a structured field set $\mathcal{D}_S$:

$$f_{\text{Extract}}(\mathcal{I}) = \mathcal{D}_S \tag{1}$$

2. Rule Check.. $\mathcal{D}_S$ is evaluated against six deterministic ZATCA compliance rules – arithmetic total reconciliation, seller VAT number format, buyer VAT number format, seller address presence, standard-invoice mandatory fields, and mandatory QR code presence – each independently verified against the primary ZATCA regulation text [5, 6] rather than assumed. This produces a rule-outcome vector $\mathbf{v}_{\text{checks}}$:

$$g_{\text{Rules}}(\mathcal{D}_S) = \mathbf{v}_{\text{checks}} \tag{2}$$

3. Tier Check.. Each individual extracted field is assigned a confidence tier (high, medium, or low) based on extraction source and rule outcome. Fields extracted from native document text, or extracted via OCR but passing their associated rule, are tiered high. Fields extracted via OCR and failing their associated rule are tiered medium, since the failure may originate from misreading rather than a genuine invoice defect. Missing fields are tiered low.

4. Field Recheck (conditional).. If any field is tiered medium, a targeted VLM recheck is issued for those specific fields only – re-grounding in the original image rather than iterating on an already-extracted, potentially lossy text representation – and control returns to Rule Check with the corrected field(s).

5. Retrieval (RAG).. For every rule that fails, a fixed, independently-tested query retrieves supporting evidence from a Chroma vector index built over official ZATCA regulation PDFs (table-aware and section-aware chunking) and vetted regulatory web content [3, 7], using Maximal Marginal Relevance (MMR) retrieval to balance relevance against redundancy:

$$R(\text{query}) = \mathcal{E} \tag{3}$$

6. Solver.. The Solver drafts the audit report – summary, explainability narrative, citations, and actionable suggestions – conditioned on the extracted fields, the rule-outcome vector, and the retrieved evidence:

$$\mathcal{R}_{\text{draft}} = \mathcal{M}_{\text{Solver}}(\mathcal{D}_S, \mathbf{v}_{\text{checks}}, \mathcal{E}) \tag{4}$$

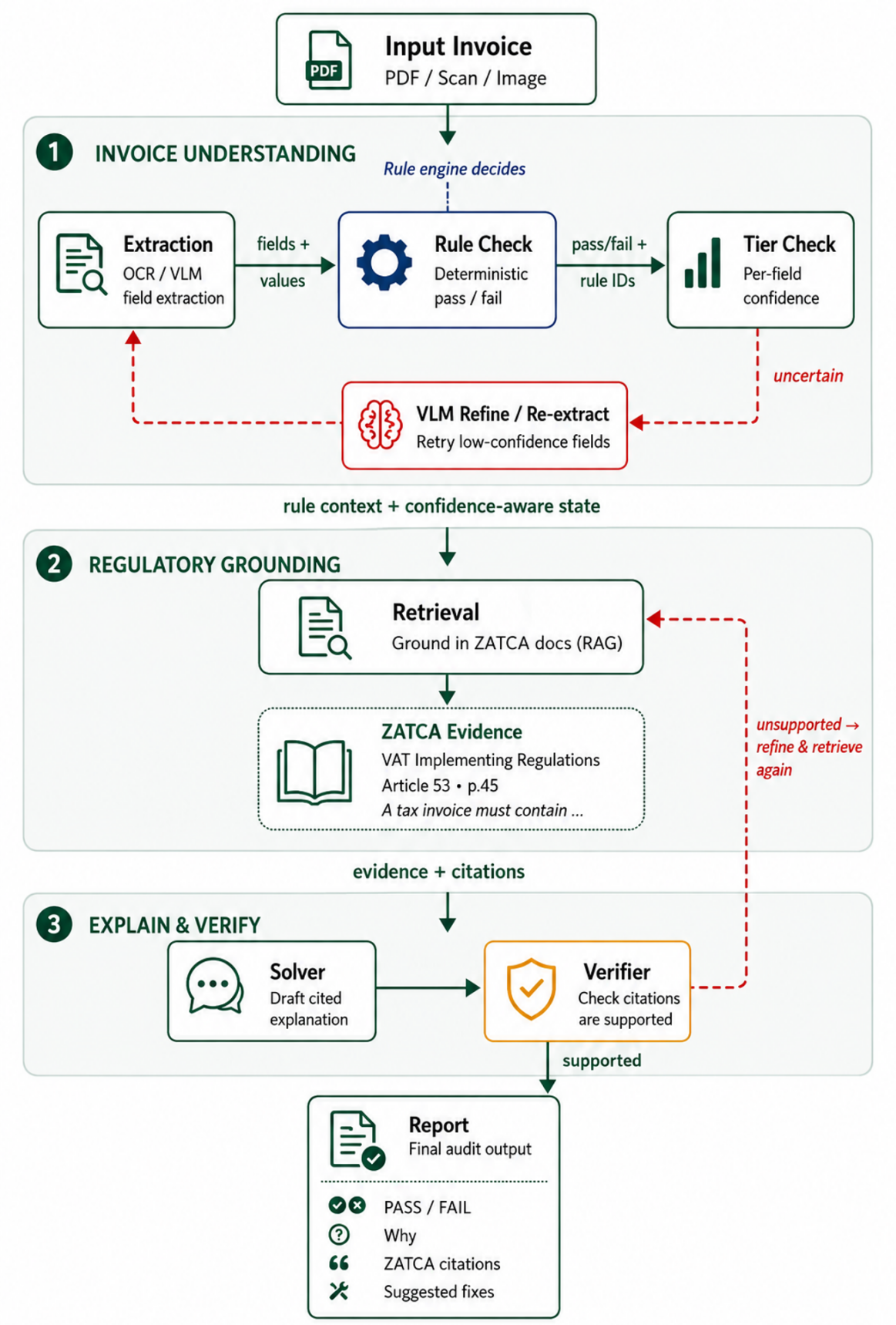


Figure 1 The Tadqeeq pipeline: *Three phases: rules decide Pass/Fail — the LLM only explains, verified against real regulatory evidence before it reaches the report.(Section 3.2).*

The Solver never re-derives or overrides $\mathbf{v}_{\text{checks}}$; its role is restricted to natural-language synthesis over already-decided outcomes.

7. Verifier.. Every citation in $\mathcal{R}_{\text{draft}}$ is checked against $\mathcal{E}$ to confirm it is genuinely supported by the retrieved evidence, not merely plausible-sounding. If any citation is unsupported, the draft is rejected and control returns to the Solver with the specific issue identified, up to a bounded number of retries, before falling through to the final report regardless.

Output.. The final report is emitted only once verification succeeds , and includes the overall Pass/Fail status, a natural-language explainability narrative, every citation used with its exact source, and concrete suggestions where applicable.

subsectionGround-Truth

To ensure evaluation integrity, field values were independently annotated through human review. Expected rule outcomes were subsequently recomputed by passing these reviewed values through Tadqeeq's `RuleEngine`, rather than relying on external rule checklists. This design guarantees that evaluation metrics reflect Tadqeeq's true rule-application accuracy against an independent human-reviewed baseline.

4 Results

Tadqeeq was evaluated across three independent layers – retrieval, extraction, and end-to-end compliance decisioning – against ground truth that was independently verified .

4.1 Retrieval and Citation Layer

The retrieval layer was evaluated against a hand-curated set of rule-linked queries, each paired with a source independently confirmed against the raw ZATCA regulation text or the underlying website content.

Table 1 Retrieval and citation evaluation

Metric	Result
Retrieval Recall@8	100%
Retrieval Mean Reciprocal Rank (MRR)	0.92
Citation Grounding Accuracy	83%
Hallucinated citations (across all real invoices tested)	0

These figures reflect a deliberate, evidence-driven improvement cycle rather than a single evaluation pass. An initial audit of the retrieval corpus identified concrete defects: PDF tables split mid-structure by character-count chunking, a website scraper capturing full-page navigation and footer content alongside genuine article text, and ground-truth entries whose reference source was, on direct re-reading of the primary regulation text, topically mismatched to the rule it was meant to support. Correcting these – table-aware and section-aware PDF chunking, content-density-based web extraction [3], and re-verified ground truth against the primary source [5] – raised Retrieval Recall@8 from 67% to 100% and MRR from 0.58 to 0.92.

Zero hallucinated citations were observed across all real invoices evaluated, attributable to the Verifier Agent, which checks every citation the Solver generates against the actually-retrieved evidence before it reaches the final report, and requests a retry when a citation cannot be substantiated.

4.2 Extraction Layer

Field-level extraction accuracy was measured against invoices with known ground-truth field values .

Table 2 Per-field extraction accuracy (100 synthetic invoices)

Field	Accuracy
Seller VAT Number	100.0%
Total (Ex-VAT)	100.0%
Issue Date/Time	96.9%
Buyer VAT Number	95.9%
Total VAT	95.9%
Invoice Type	99.0%
Invoice Serial Number	99.0%
Total (Inc-VAT)	90.7%
Seller Name	77.3%
Buyer Name	73.5%
Seller Address	32.7%
Buyer Address	34.0%
Overall	**88.5%**

A consistent pattern emerges: structured, numeric, or short-form fields (VAT numbers, totals, dates, invoice type) extract near-perfectly, while long free-text Arabic fields – particularly addresses – are extracted with markedly lower fidelity. Manual inspection of address mismatches confirmed these are predominantly single-character OCR misreadings between visually similar Arabic letters and occasional structural reordering of the address string, rather than random noise.

4.3 Compliance Decision Layer

End-to-end report-layer evaluation ran the full pipeline against a verified sample of real and synthetic invoices, comparing predicted field values, predicted rule outcomes, and predicted overall compliance status against ground truth in which expected rule outcomes were recomputed through Tadqeeq's own `RuleEngine` from trusted field values, never taken directly from any external labeling.

Table 3 Report-layer evaluation

Metric	Result
Field Extraction Accuracy	88.5%
Rule-Application Accuracy	98.3%
Compliance Status Accuracy	100%

4.4 Baseline Comparison

Tadqeeq was compared against three alternative systems on the same real invoices: a Gemini 2.5-flash baseline generating a report directly with no rule engine or retrieval, a Qwen (qwen3-vl-flash) baseline under identical conditions, and Tadqeeq with its RAG retrieval stage disabled (an ablation

holding every other component constant). Each generated report was scored by an independent LLM judge (GPT-4o-mini) on five dimensions, using a judging rubric that explicitly floors the Groundedness score to its minimum value when a report contains zero real, checkable citations to an external regulatory source – regardless of how confident or well-written the surrounding text is.

Table 4 Baseline comparison, average scores (1–5, higher is better)

System	Quality	Halluc.	Groundedness	Completeness	Actionability
Baseline Gemini 2.5-flash	4.03	4.65	1.00	3.61	4.48
Baseline Qwen	3.93	2.48	1.15	4.71	4.84
Tadqeeq without RAG	4.05	4.90	1.00	4.17	4.51
Tadqeeq (full)	**4.28**	**5.00**	**4.29**	4.28	4.78

The central finding is the isolated effect of retrieval: Tadqeeq without RAG scores 1.00 on Groundedness – statistically indistinguishable from both baseline systems – while Tadqeeq (full), differing from the no-RAG variant only in the presence of the retrieval stage, scores 4.29. Since every other component of the pipeline is held constant across this comparison, this is direct evidence that grounding, not any other architectural feature, is the specific cause of the improvement.

Baseline Qwen scores marginally higher than Tadqeeq (full) on Completeness and Actionability. Manual inspection of the underlying report text attributes this to Qwen's tendency to produce longer, more numerically detailed analysis – including recalculated arithmetic breakdowns with specific figures – which reads as thorough even where, per the judge's own reasoning on individual cases, some of that specific-sounding detail is not independently verifiable. Tadqeeq's reports are comparatively more concise, restricting stated claims to what is directly supported by a retrieved citation; this discipline is the direct cause of its Groundedness and Hallucination scores, at a small, quantified cost to perceived verbosity-driven completeness.

5 Discussion

5.1 Decoupling Extraction Confidence from Compliance Verdict

A central design decision was treating *extraction confidence* and *compliance verdict* as independent outputs, not one conflated signal. Early routing logic sent any failed rule to human review regardless of how clearly the field was read – so a seller VAT number read with high confidence, but genuinely malformed, was routed for human review instead of being reported as a definitive failure. "Do we trust what was extracted?" and "does it comply?" are distinct questions, and conflating them undermines explainability.

The final design routes missing or severely degraded fields to human review, routes ambiguous OCR-sourced failures to a secondary verification agent, and reports rule failures on confidently-extracted fields directly as compliance defects. This keeps the output auditable: a reviewer can distinguish "the system is unsure what this says" from "the system is sure, and it violates a rule."

5.2 Precision-Recall Trade-offs in Rule Design

Compliance rules initially included exact arithmetic reconciliation: verifying that the declared subtotal matched the sum of line items, and that subtotal, discount, and tax summed to the declared total. Iterative testing against real invoices showed that these checks produced a high

rate of false failures, driven by OCR and LLM extraction noise on genuinely compliant documents rather than by real invoice defects.

Table 5 summarizes this trade-off for the two rule categories most affected.

Table 5 Evolution of arithmetic compliance rules under empirical testing

Rule	Initial design	Final design
Subtotal–total reconciliation	Verify subtotal − discount + tax = total within tolerance	Presence check only (`total_present`)
Line-item reconciliation	Verify $\sum$(qty × price) matches declared subtotal	Removed for simplified invoices; retained selectively for standard invoices where line-item VAT rates are disclosed

This shift favors precision over recall: the system is less likely to flag a compliant invoice incorrectly, at the cost of being less able to catch genuine arithmetic inconsistencies introduced by the invoice issuer. Given that Tadqeeq's target use case is automated first-pass screening rather than final audit determination, this trade-off was judged acceptable – a high false-positive rate would undermine user trust in the tool more than a reduced sensitivity to arithmetic fraud.

5.3 Invoice-Type-Conditional Rule Application

ZATCA distinguishes simplified (business-to-consumer) invoices from standard (business-to-business) invoices, with materially different mandatory-field requirements [8]. Tadqeeq's rule engine applies this distinction directly: buyer identification, subtotal disclosure, and seller/buyer address are enforced only for standard invoices, while both invoice types share common requirements (seller identification, VAT amount, QR code presence). Uniformly applying standard-invoice requirements to simplified invoices during early development produced systematic false failures on legitimate retail receipts, which by design omit buyer information. This confirmed that regulatory-aware conditional logic, rather than a single flat rule set, is necessary for a compliance checker operating across heterogeneous invoice types.

5.4 Dual-Path QR Code Detection

QR code presence verification required two distinct detection strategies depending on the extraction path. For native PDFs, embedded image objects are extracted directly via PyMuPDF and evaluated with a standard QR decoder, which performs reliably on the high-resolution, artifact-free images typical of digitally generated documents.

For OCR-derived paths, direct QR decoding proved unreliable on photographed or scanned invoices, where QR codes are frequently small, skewed, or affected by lighting artifacts. Tadqeeq instead uses the vision-language OCR model's own grounding output – which localizes embedded image regions during text extraction – combined with a lightweight heuristic filter based on aspect ratio (QR codes are approximately square) and edge-density texture (QR modules exhibit high-frequency black/white alternation, distinguishing them from smoother logo graphics). This design accepts a weaker guarantee – confirming that a QR-*shaped*, QR-*textured* region exists, rather than decoding and validating its content – in exchange for substantially higher detection reliability on low-quality real-world images.

5.5 Query Wording as a First-Class Retrieval Variable

Retrieval accuracy improved substantially – Recall@8 from 67% to 100%, MRR from 0.58 to 0.92 – through a combination of corpus-quality fixes and, notably, the wording of the retrieval queries themselves. Two queries that failed to retrieve their correct, independently-verified source under their original phrasing succeeded at rank 1 once reworded to more closely match the vocabulary of the source material, with no change to the underlying corpus or retrieval mechanism. This motivated a design decision to route production retrieval through a small set of fixed, pre-tested queries per rule, rather than a dynamically-constructed query built from the rule name and that invoice's specific error message: a fixed query can be evaluated and improved systematically in the way the retrieval layer above was, while a query that differs on every invoice cannot be validated in advance at all.

5.6 Grounded Conciseness versus Verbose Completeness

In the baseline comparison, Tadqeeq (full) leads decisively on Groundedness (4.29 vs. 1.00–1.15) and ties for the highest Hallucination score, but scores marginally below the Qwen baseline on Completeness and Actionability. Direct inspection of the underlying report text and the judge's own per-invoice reasoning attributes this to a genuine, quantifiable trade-off rather than a weakness in either direction: Qwen's reports are longer and include specific-sounding recalculated figures that read as thorough, some of which the judge itself flags as not independently verifiable; Tadqeeq's reports are shorter and restrict every stated claim to what is directly supported by a retrieved citation. The same discipline responsible for Tadqeeq's Groundedness and Hallucination performance is the direct, identifiable cause of its narrower margin on metrics that reward descriptive volume.

6 Limitations

6.1 OCR Model Reliability

- **Non-deterministic failures.** The vision-language OCR model occasionally entered a non-terminating, repetitive generation state on certain inputs. This behavior was not consistently reproducible: the same input image succeeded in one run and failed in another under otherwise identical conditions. A thread-safe timeout guard mitigates the operational impact by skipping affected files rather than blocking the pipeline, but the underlying cause was not diagnosed within the project timeframe.
- **Localization gaps in QR detection.** The grounding-based QR detection heuristic depends on the OCR model successfully localizing an embedded image region during generation. In a non-trivial fraction of tested invoices, no image region was localized at all, even when a QR code may have been visually present – a limitation of the underlying model's grounding behavior rather than of the detection heuristic itself.
- **Threshold sensitivity.** Aspect-ratio and edge-density thresholds for QR detection were calibrated against a limited set of manually inspected samples. Low-contrast or partially degraded QR codes near the threshold boundary may be misclassified in either direction.

6.2 Rule Coverage

The removal of arithmetic reconciliation rules, discussed in Section 5.2, reduces Tadqeeq's ability to detect genuine numerical inconsistencies introduced by an invoice issuer, in exchange for eliminating extraction-noise-driven false positives. This is a deliberate trade-off rather than an

oversight, but it means the current rule set validates *presence and format* more rigorously than it validates *arithmetic correctness*.

6.3 External Dependency on a Third-Party LLM API

Both structured field extraction and secondary verification depend on a commercial generative AI API. During development, this introduced operational instability independent of Tadqeeq's own logic: free-tier rate limits, periodic model deprecations requiring manual reconfiguration, and intermittent server-side errors. Large language model inference at scale is also a known source of infrastructure cost and latency pressure, particularly where model size and memory bandwidth constraints govern throughput; this constrains Tadqeeq's practical throughput under free-tier conditions and would require a paid or self-hosted inference tier for production-scale deployment.

6.4 Scope Boundaries

- **Single-page processing.** The pipeline currently processes only the first page of multi-page documents. Multi-page invoices are not yet supported.
- **No cryptographic or portal-level validation.** Tadqeeq validates document-level field presence and format but does not verify ZATCA-specific cryptographic elements (invoice UUID, invoice counter value, previous invoice hash, digital signature), which require live integration with ZATCA's clearance and reporting infrastructure rather than document-level analysis alone. This was an explicit scope boundary, not an implementation gap.

7 Future Work

Several extensions are identified as natural next steps for continued development of Tadqeeq.

- **OCR reliability and diagnosis.** The non-deterministic OCR failures observed (Section 6.1) were mitigated, not resolved.
- **Per-line VAT cross-validation.** The schema captures a VAT rate per line item but not a per-line VAT amount. Adding this and validating it against unit price × quantity × VAT rate would restore finer-grained arithmetic verification without reintroducing the invoice-level noise issues from Section 5.2.
- **QR code content validation.** Detection currently confirms structural presence only – not the QR's embedded content against ZATCA's required fields (seller name, VAT number, timestamp, total, VAT amount). Targeted image enhancement or a dedicated QR-specialized model are candidate approaches for reliable decoding.
- **Production-scale infrastructure.** The system currently relies on commercial generative AI APIs. Enterprise-scale deployment would need a dedicated inference tier – paid or self-hosted – sized for consistent throughput and latency under sustained load.

8 Conclusion

Tadqeeq implements an end-to-end, multi-agent pipeline for verifying existing ZATCA e-invoices – combining native PDF text extraction, vision-language OCR, a deterministic compliance rule engine, and Retrieval-Augmented Generation grounded in official ZATCA documentation. Its principal contribution is the integration logic connecting these components: extraction confidence and compliance verdict are treated as independent signals rather than conflated into a single output, retrieval grounds every explanatory claim in a real, checkable regulatory source, and

a dedicated verification stage confirms every generated citation against the retrieved evidence before a report reaches its final form.

Evaluation across retrieval, extraction, and end-to-end compliance decisioning demonstrated strong, independently-verified performance: 100% retrieval recall, 98.3% rule-application accuracy, 100% compliance status accuracy, and zero hallucinated citations across all real invoices tested. A controlled comparison against baseline systems – including an ablated variant of Tadqeeq with retrieval disabled, differing from the full system only in that one component – isolated retrieval as the direct, measured cause of a Groundedness improvement from 1.00 to 4.29 on a 5-point scale, providing concrete evidence that grounding compliance explanations in real regulatory text, rather than relying on a language model's unsupported confidence, is what distinguishes Tadqeeq from systems that generate plausible-sounding but uncitable compliance narratives.

Development proceeded empirically throughout, with each reported result grounded in a specific, reproducible finding: query wording measurably affects retrieval accuracy independent of corpus quality; rule-outcome labels must be recomputed through a system's own deterministic logic rather than trusted from external sources, to avoid circular evaluation; and grounded conciseness and verbose-but-unverifiable completeness represent a genuine, quantifiable trade-off rather than a simple ranking of system quality. Tadqeeq is positioned as a first-pass automated compliance screening tool that reports its reasoning at the level of individual, citable rules rather than as an opaque verdict – reducing manual audit overhead while keeping every compliance claim it makes open to direct verification against the regulation it cites.

References

[1] Saleh7. *php-zatca-xml*. https://github.com/Saleh7/php-zatca-xml. Accessed: 2026-07-09.

[2] mrsool. *zatca*. https://github.com/mrsool/zatca. Accessed: 2026-07-09.

[3] Qoyod. *Why E-Invoices Get Rejected in Saudi Arabia*. https://www.qoyod.com/en/learn/rejected-invoices/. Accessed: 2026-08-22.

[4] lavaloon-eg. *ksa_compliance*. https://github.com/lavaloon-eg/ksa_compliance. Accessed: 2026-07-09.

[5] Zakat, Tax and Customs Authority (ZATCA). *E-invoicing Detailed Technical Guideline*. Official ZATCA publication.

[6] Zakat, Tax and Customs Authority (ZATCA). *E-Invoicing Resolution*. Official ZATCA publication, 19 May 2023.

[7] Qoyod. *The QR Code in the E-Invoice*. https://www.qoyod.com/en/learn/qr-code-in-e-invoice/. Accessed: 2026-08-22.

[8] Zakat, Tax and Customs Authority (ZATCA). *What is E-Invoicing?* https://zatca.gov.sa/en/E-Invoicing/Introduction/Pages/What-is-e-invoicing.aspx. Accessed: 2026-07-09.