International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI-Powered Intelligent Invoice Automation: A Survey of Document Understanding, Validation, and Accounts Payable Workflows

Kshitij Jagtap[1], Aryan Bankar[2], Rohan Gawade[3], Shreyash Atole[4], Supriya Gavhane[5]
[1,2,3,4,5] Department of Computer Engineering, SVPM's College of Engineering, Malegaon Bk., Pune, India

Abstract **-** Organizations receive invoices as structured electronic messages, digitally generated PDFs, scans, and sometimes photographs. Manual handling across these formats can delay payment, introduce keying errors, and make policy checks difficult to scale. Recent document AI combines optical character recognition (OCR), layout analysis, language models, and workflow rules to extract invoice fields and line items. Yet reliable automation requires more than reading text: it must reconcile totals, match purchase orders and receipts, route exceptions, preserve evidence, and post approved records to financial systems. This paper surveys AI-powered invoice automation across document intake, field and table extraction, validation, exception handling, and enterprise integration. It compares rule-based, machine-learning, multimodal-transformer, and structured e-invoice approaches, then presents a modular framework with human review and audit controls. We outline evaluation measures for extraction, workflow quality, and operational risk, and identify open challenges in generalization, privacy, fraud detection, and compliance. This paper proposes a framework and evaluation plan; it reports no original system performance results.

Key Words: invoice automation, accounts payable, document AI, OCR, information extraction, e-invoicing, workflow automation, human review.

1. INTRODUCTION

Invoices are routine business records, but processing them requires several linked decisions: identify the supplier and invoice, capture header fields and line items, verify arithmetic and tax treatment, match the invoice to purchasing evidence, obtain approvals, and record the liability. In high-volume settings, staff may receive these documents through email, portals, and electronic data interchange. Manual re-entry and fragmented checks create delays and make consistent control difficult [1].

Document automation has progressed from fixed templates and OCR to statistical extraction and multimodal document understanding. OCR converts page images into text, while layout-aware models use word position and visual structure to distinguish fields such as invoice number, date, supplier, tax, and total [2]–[6]. Structured e-invoices can carry machine-readable data directly, reducing dependence on image recognition when the format and business network are supported [10]–[12].

A reliable accounts-payable system must connect document interpretation to business rules and downstream actions. A correctly read amount can still be assigned to the wrong supplier, duplicated, inconsistent with a purchase order, or routed under the wrong approval policy. Generative models may help interpret unfamiliar layouts, but outputs must be constrained, checked against source evidence, and reviewed when confidence or policy conditions require it.

This paper makes three contributions:

- A structured survey of invoice automation methods, from OCR and templates to machine-learning and multimodal document models, with attention to header fields and line-item tables.
- An analysis of validation, matching, approvals, auditability, and human review as essential stages beyond extraction.
- A modular AI-powered intelligent invoice automation framework and an evaluation plan spanning extraction quality, workflow outcomes, and operational controls.

The paper is structured as follows. Sec. 2 reviews the invoice workflow and user requirements, while Sec. 3 surveys invoice automation approaches, and Sec. 4 examines document AI and structured e-invoicing. Sec. 5 discusses validation and decision support, while Sec. 6 compares representative approaches. The proposed framework is presented in Sec. 7, its evaluation, open challenges, and future directions in Secs. 8-10, and finally concluding remarks in Sec. 11.

2. BACKGROUND AND USER NEEDS

2.1 Accounts Payable Workflow

Invoice processing spans receipt, classification, data capture, verification, matching, approval, posting, payment, and archival. The exact sequence varies by organization, but the central task is to transform a supplier document into a controlled financial transaction. A purchase-order invoice may be checked against order and receipt records; a non-PO invoice may need coding and approval by a budget owner. Credit notes, recurring bills, multi-currency invoices, and tax variations introduce additional cases [1].

2.2 Requirements from Users

Accounts-payable teams, approvers, auditors, suppliers, and finance-system owners need automation that meets several requirements:

- Accurate capture: supplier identity, invoice identifiers, dates, currency, tax, totals, and line items must be linked to evidence in the source.
- Control before posting: duplicate detection, arithmetic checks, policy validation, and purchase-order matching must be explicit and reviewable.
- Efficient exception handling: uncertain or conflicting cases should be routed to the right person with a clear reason and supporting evidence.
- Auditability: the original document, extracted values, corrections, decisions, and downstream actions should remain traceable.
- Interoperability and accessibility: the system should connect with email, procurement, ERP, and e-invoice channels while supporting staff with appropriate access.

3. INVOICE AUTOMATION APPROACHES

3.1 Data Sources and Document Types

Invoice inputs range from native PDF files with selectable text to scanned images, photographs, email attachments, and structured XML messages. Image-based inputs require OCR and layout interpretation; digitally generated PDFs may expose text but still require field identification. Structured formats such as UBL and UN/CEFACT Cross Industry

Invoice encode business data for automated exchange [10], [11]. Table -1 summarizes common input sources.

Table -1: Invoice input sources

Input source	Strengths	Limitations
Structured e-invoice	Machine-readable fields; avoids OCR	Format and network support required; business checks still needed
Native PDF	Often contains selectable text; widely received	Layout varies; text reading order may be unreliable
Scanned image or photo	Captures legacy paper workflows	OCR affected by blur, skew, stamps, handwriting
Email and attachments	Fits existing supplier submission channels	Duplicate attachments, mixed documents, and security risks

3.2 Template and Rule-Based Capture

Template systems locate fields using fixed coordinates, labels, or supplier-specific patterns. They can be fast and predictable when layouts are stable and are straightforward to test. Their maintenance burden grows as suppliers change formats, documents arrive at different resolutions, or line-item tables vary. Rules remain useful for deterministic checks such as required-field presence, arithmetic, date ranges, and approved supplier lists.

3.3 OCR and Machine Learning

OCR engines recognize text from images, after which classifiers, sequence models, or key-value extraction methods assign semantic roles. CloudScan demonstrated recurrent-neural extraction for key invoice fields across a large invoice corpus [3]. Layout-aware models combine text with two-dimensional position and visual information, supporting extraction across variable formats [4], [5]. These methods can reduce per-supplier template work, but performance depends on training coverage, annotation quality, language, image quality, and the distinction between unseen suppliers and unseen document layouts.

3.4 Line-Item Table Recognition

Line-item extraction is harder than reading a small set of header fields because the system must find table boundaries, associate values with columns, preserve row relationships, and handle wrapped descriptions or merged cells. Invoice-focused surveys identify table structure as a central challenge [1]. Two-stage methods separate table detection from cell and field interpretation [13]. Evaluation should therefore score row and cell structure as well as text, since a correct token assigned to the wrong quantity or price column can create an incorrect posting.

3.5 Data Quality and Leakage

Invoice datasets often contain repeated suppliers, templates, or near-duplicate documents. Randomly splitting pages can place the same layout in training and testing and overstate generalization. Evaluation should separate by supplier, template, or time period where the intended deployment requires adaptation to new layouts. Duplicate and near-duplicate invoices should also be kept in one partition. Training annotations must distinguish evidence visible on the page from values inferred from purchase orders or ERP records, or the model may appear to extract information it did not actually read.

3.6 Workflow and Enterprise Integration

Extraction models typically operate as one service within a broader workflow. A production pipeline must accept documents, retain originals, apply versioned models and rules, expose review queues, and post approved transactions to an accounting or ERP system. Idempotent APIs, retry handling, access control, and reconciliation are as important as model selection. Standards-based e-invoicing can simplify data exchange, while legacy and mixed-format environments still require document capture [10]–[12].

4. DOCUMENT AI AND STRUCTURED INVOICES

We define intelligent invoice automation as a system that converts invoice documents or structured messages into validated, traceable financial records and advances them through an approval workflow. Four capabilities are relevant.

4.1 OCR-Based and OCR-Free Understanding

OCR-based pipelines explicitly recognize text and then classify or relate it to fields. Their intermediate text and coordinates can support inspection, correction, and reuse, but OCR errors can propagate into extraction. OCR-free methods such as Donut map document images to structured outputs without a separate OCR stage [6]. They may reduce pipeline coupling, but generated fields still need schema checks and links to source regions for human verification.

4.2 Layout-Aware Models

Models such as LayoutLM and LayoutLMv3 combine textual, spatial, and visual signals to represent document structure [4], [5]. This is useful when a field is identified by both its wording and its position relative to headings, totals, or supplier details. Invoice layouts remain highly variable, and models trained on receipts or forms do not automatically establish financial correctness. Fine-tuning and evaluation should reflect the target invoice population and exact field definitions.

4.3 Structured E-Invoicing

Structured e-invoices encode invoice fields in a machine-readable syntax and can avoid OCR for the data they contain. OASIS UBL and UN/CEFACT CII provide business-document models for electronic exchange [10], [11]. The European Union's 2025 VAT in the Digital Age directive establishes a changing regulatory context for e-invoicing and digital reporting, with staged application dates [12]. Structured data improves machine readability but does not by itself verify supplier legitimacy, purchase receipt, tax treatment, or authorization.

4.4 Toward Intelligent Automation

An intelligent system combines document understanding with deterministic controls and workflow orchestration. AI is suited to variable text, layout, and ambiguous descriptions; explicit rules are suited to arithmetic, policy thresholds, and required approvals. The system should keep these responsibilities visible, maintain provenance for each field, and support correction without silently changing the original evidence.

5. VALIDATION, CONTROLS, AND DECISION SUPPORT

Extraction is only one stage of invoice processing. Before posting, the system should assess whether the captured values form a valid and authorized transaction. Useful controls include:

- Field confidence and evidence: show the extracted value with its source page or region and separate model confidence from business-rule results.
- Arithmetic and schema checks: recompute line extensions, tax, and totals; validate currency, mandatory fields, date formats, and structured-message syntax.

- Duplicate and fraud screening: compare invoice identifiers, supplier, amount, date, bank details, and document similarity against prior records; treat alerts as candidates for review rather than proof of fraud.
- Three-way matching: compare invoice lines with purchase orders and goods-received records using configurable quantity, price, and tolerance rules.
- Approval and exception routing: direct unresolved mismatches, low-confidence fields, non-PO invoices, and policy exceptions to authorized reviewers, preserving a reason and decision history.

6. COMPARATIVE ANALYSIS OF EXISTING APPROACHES

6.1 Comparison of Representative Approaches

Table -2 compares common approaches along the dimensions most relevant to invoice processing. The comparison is qualitative and describes typical capabilities and constraints.

Table -2: Representative invoice automation approaches

Approach	Input	Typical capability	Main constraint
Template and rules	PDF or image	Predictable capture for known layouts	Supplier-specific upkeep
OCR and layout-aware AI [3]–[6]	Scans and PDFs	Key fields across variable layouts	Needs labeled data, validation, and evidence
Structured e-invoice [10], [11]	UBL or CII message	Direct mapping to invoice schema	Coverage and interoperability vary
Proposed framework	Mixed inputs	Capture, checks, review, and ERP posting	Requires integration and governed controls

Three observations emerge. First, OCR and template systems can automate stable cases but require maintenance as supplier layouts change. Second, machine-learning and multimodal models can generalize across visual variation, but their outputs need field-level evidence and business validation. Third, structured e-invoices reduce capture effort where both parties and networks support compatible formats, while workflow controls remain necessary for matching and approval.

6.2 Summary of Related Literature

Table -3 summarizes twelve representative studies, datasets, methods, and standards relevant to invoice document understanding and automation. Together they show a progression from OCR and key-value extraction toward multimodal models and structured exchange. Four patterns emerge.

First, invoice-specific surveys emphasize that extraction includes layout and table structure, not just OCR [1]. Second, neural and transformer approaches reduce dependence on fixed coordinates, but data coverage and supplier-level generalization remain important [3]–[6]. Third, benchmark datasets support reproducible research but do not capture every enterprise workflow or private invoice population [7]–[9]. Fourth, structured standards make data exchange more interoperable but do not replace accounting validation and authorization [10]–[12]. These observations motivate a framework that connects capture, verification, and posting.

Table -3: Summary of related literature

Sr. No	Paper title	Author(s)	Year	Existing problem statement	Existing problem solution	Future scope / gap
1	An Overview of Data Extraction From Invoices [1]	T. Saout et al.	2024	Invoice processing combines recognition, extraction, and structure recovery.	Surveys the workflow and methods, including table processing.	Organization-specific workflow validation remains necessary.
2	Document Information Extraction: Invoice Anatomy [9]	M. Hamri et al.	2024	Invoice components appear in varied forms and layouts.	Studies component variants using generated templates.	Broader real-world supplier coverage is needed.
3	LayoutLMv3 [5]	Y. Huang et al.	2022	Document models need combined text and visual representations.	Uses unified text and image masking for document AI.	Invoice-specific controls and posting are outside model scope.
4	OCR-free Document Understanding Transformer [6]	G. Kim et al.	2022	OCR pipelines can propagate recognition errors.	Maps document images to structured sequences without external OCR.	Outputs still need evidence links and business validation.
5	Information Extraction from Invoices [17]	T. Saout et al.	2021	Invoice key-field labels can be costly to produce.	Studies neural extraction with limited training data.	New supplier and layout generalization must be tested.
6	LayoutLM [4]	Y. Xu et al.	2020	Text-only methods omit document geometry.	Pre-trains text and layout representations.	Business-level correctness remains a separate task.
7	DocBank [14]	M. Li et al.	2020	Document layout research lacks broad annotations.	Introduces a large-scale layout analysis dataset.	Layout labels do not validate invoice transactions.
8	FUNSD [7]	G. Jaume et al.	2019	Noisy forms challenge entity and relation extraction.	Provides annotated scanned forms.	Forms differ from diverse commercial invoices.
9	CORD [8]	S. Park et al.	2019	Post-OCR receipt parsing needs structured data.	Provides receipt images and annotations.	Receipts do not represent full AP workflows.
10	CloudScan [3]	R. B. Palm et al.	2017	Invoice layouts make fixed templates brittle.	Uses recurrent neural networks for key-field analysis.	Deployment still requires validation and integration.
11	OASIS UBL 2.4 [10]	OASIS	2024	Business documents need interoperable data syntax.	Defines standardized electronic business documents.	A schema does not establish authorization or receipt.
12	VAT in the Digital Age [12]	Council of the EU	2025	Digital reporting and e-invoicing requirements evolve.	Amends EU VAT rules with staged application.	Organizations must monitor local implementation and dates.

7. AI-POWERED INTELLIGENT INVOICE AUTOMATION: A FRAMEWORK

The proposed framework illustrates how document AI can support accounts-payable processing from receipt through approved posting. It is a design proposal for organizing capabilities described in the literature, not an evaluated

commercial deployment. It does not claim autonomous payment authorization or measured performance.

7.1 Architecture

Fig -1 depicts the processing pipeline and its review path.

- **Intake layer:** accepts email attachments, portal uploads, scanned files, and structured e-invoices; records source and receipt metadata.
- **Document understanding layer:** classifies document type, performs OCR when needed, detects fields and line-item tables, and returns typed values with page and region evidence.
- **Validation layer:** applies schema, arithmetic, supplier, duplicate, tax, and purchase-order / receipt checks using versioned rules and reference data.
- **Decision layer:** assigns the invoice to a straight-through, review, exception, or rejection queue according to confidence and policy results; it does not infer approval authority.
- **Human review layer:** presents the source beside extracted values, highlights mismatches, captures corrections and decisions, and preserves the audit trail.
- **Integration layer:** sends approved records to the ERP or accounting system, stores processing status, and reconciles acknowledgements and retries.

A workflow service connects these layers. Every field should retain provenance, model and rule versions, and the status of related checks so the organization can reconstruct why a record advanced or stopped.

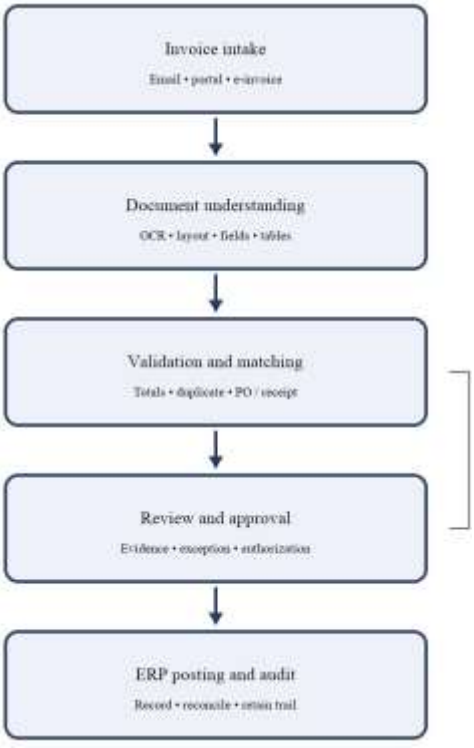


Fig -1: Layered architecture of the AI-powered intelligent invoice automation framework.

7.2 Design Principles

- **Evidence-linked extraction:** every important value can be traced to a document region or structured source field.
- **Separation of prediction and control:** probabilistic extraction is distinct from deterministic accounting rules and approval authority.
- **Review by exception:** confidence thresholds and policy outcomes determine which cases require human attention.
- **Supplier-aware adaptation:** corrections may improve future handling, with validation before a model or template change is deployed.
- **Measured performance:** report field and table accuracy on held-out suppliers or layouts and report workflow outcomes separately.
- **Safe posting:** only authorized, validated records reach the financial system; automation does not bypass segregation of duties.

The framework handles native electronic invoices and image-based documents through separate intake paths that converge on a common invoice schema. Structured data can be validated directly; scans require recognition and field linking. In both cases, purchase-order matching, approval routing, and audit logging remain explicit workflow steps.

7.3 Positioning

The framework's contribution is the integration of AI-based capture with validation, evidence, human review, and ERP posting in one workflow. It does not propose a new document model. Its research questions concern generalization to unseen suppliers, line-item structure, calibration of review thresholds, and the effect of automation on exception workload and control quality, discussed in Sec. 9.

8. EVALUATION METHODOLOGIES AND DATASETS

Invoice automation should be evaluated at several levels because high field accuracy does not guarantee a correct or efficient financial workflow. A rigorous evaluation should combine:

- **Extraction-level:** exact match and normalized field accuracy for supplier, invoice number, dates, currency, tax, and totals; cell and row metrics for line items; and confidence calibration on held-out documents.
- **Workflow-level:** duplicate detection, match precision and recall, exception routing, straight-through processing rate, correction effort, posting success, and end-to-end processing time, with results segmented by supplier and document type.
- **Control-level:** audit completeness, policy-rule coverage, access and approval enforcement, recovery from retries, and the proportion of records that post without required evidence or authorization.

Public resources support parts of this evaluation. FUNSD provides annotated forms for document understanding [7]; CORD provides receipt images and structured annotations [8]; FATURA focuses on varied invoice layouts [9]. These datasets are useful for model comparison, but they do not establish performance on an organization's supplier base, ERP data, tax policies, or approval workflow. Private evaluations should use documented sampling, supplier-aware splits, and privacy safeguards.

9. CHALLENGES AND OPEN ISSUES

- **Layout diversity:** supplier formats, languages, and document quality vary widely.
- **Line-item complexity:** table boundaries, wrapped text, discounts, and tax columns are difficult to interpret reliably.
- **Limited representative data:** invoices contain sensitive commercial information, which restricts sharing and annotation.
- **Semantic ambiguity:** similar labels can denote different values, and totals may include freight, discounts, or multiple tax rates.

- **Confidence and calibration:** model scores may not correspond to actual error likelihood and should be calibrated on representative data.
- **Fraud and duplicate risk:** document similarity can flag suspicious records but cannot establish intent or legitimacy by itself.
- **Integration and change management:** supplier master data, purchase orders, receipt records, and ERP interfaces change over time.
- **Regulatory variation:** invoice content, retention, tax rules, and e-invoicing requirements differ by jurisdiction and may change.
- **Privacy and security:** invoices can expose personal, financial, and commercially sensitive information that requires controlled access and retention.
- **Human factors:** excessive or poorly explained alerts create workload and can lead reviewers to accept incorrect suggestions.

10. FUTURE DIRECTIONS

- Supplier- and time-aware benchmarks that better measure generalization beyond random document splits.
- Multimodal extraction that links fields and table cells to visual evidence and supports reviewer correction.
- Privacy-preserving collaboration and synthetic invoice generation for expanding training coverage.
- Calibrated confidence and risk-based review policies that account for error costs and approval controls.
- Robust cross-language and cross-jurisdiction support for tax fields and structured e-invoice standards.
- Fraud-screening research that combines document signals with transaction history while keeping alerts explainable.
- Reliable ERP integration with idempotent posting, reconciliation, and auditable recovery from failures.
- User studies that measure correction effort, trust, exception workload, and control outcomes in real accounts-payable teams.

11. CONCLUSION

AI-powered invoice automation combines OCR, layout-aware extraction, multimodal document models, structured e-invoicing, and workflow rules. The main challenge is not simply reading a document; it is converting evidence into a valid, matched, approved, and traceable financial record. This survey organized the field around input formats, extraction approaches, validation, and enterprise workflow, and proposed a modular framework that preserves human review and accounting controls. Reliable deployment depends on representative evaluation, line-item accuracy, supplier-aware testing, secure integration, and clear separation between model suggestions and authorized business decisions.

REFERENCES

1. Saout, T., Lardeux, F., Saubion, F.: An Overview of Data Extraction From Invoices. IEEE Access 12, 19872–19886 (2024). https://doi.org/10.1109/ACCESS.2024.3360528
2. Ha, H.T., Horák, A.: Information Extraction from Scanned Invoice Images Using Text Analysis and Layout Features. Signal Process. Image Commun. 102, 116601 (2022). https://doi.org/10.1016/j.image.2021.116601
3. Palm, R.B., Winther, O., Laws, F.: CloudScan: A Configuration-Free Invoice Analysis System Using Recurrent Neural Networks. In: Proc. 14th IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), pp. 406–413. IEEE (2017). https://doi.org/10.1109/ICDAR.2017.74
4. Xu, Y., et al.: LayoutLM: Pre-training of Text and Layout for Document Image Understanding. In: Proc. ACM SIGKDD, pp. 1192–1200. ACM (2020)
5. Huang, Y., et al.: LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. In: Proc. ACM Multimedia. ACM (2022)
6. Kim, G., et al.: OCR-free Document Understanding Transformer. In: Proc. ECCV (2022)
7. Jaume, G., Ekenel, H.K., Thiran, J.-P.: FUNSD: A Dataset for Form Understanding in Noisy Scanned Documents. In: Proc. ICDAR Workshops (2019)
8. Park, S., et al.: CORD: A Consolidated Receipt Dataset for Post-OCR Parsing. In: Proc. Document Intelligence Workshop, NeurIPS (2019)
9. Hamri, M., Devanne, M., Weber, J., Hassenforder, M.: Document Information Extraction: An Analysis of Invoice Anatomy. Appl. Comput. Intell. Soft Comput. (2024), Art. 7599415
10. OASIS: Universal Business Language Version 2.4. OASIS Standard (2024)
11. UN/CEFACT: Cross Industry Invoice (CII). United Nations Economic Commission for Europe (2024)
12. Council of the European Union: Council Directive (EU) 2025/516 as regards VAT rules for the digital age. Official Journal of the European Union, 25 March 2025
13. Saout, T., Lardeux, F., Saubion, F.: A Two-stage Approach for Tables Extraction in Invoices. In: Proc. 35th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI), pp. 10–15. IEEE (2023). https://doi.org/10.1109/ICTAI59109.2023.00010
14. Li, M., et al.: DocBank: A Benchmark Dataset for Document Layout Analysis. In: Proc. COLING (2020)
15. Li, D., et al.: DiT: Self-supervised Pre-training for Document Image Transformer. In: Proc. 30th ACM Int. Conf. on Multimedia, pp. 3530–3539. ACM (2022)
16. Mathew, A., Karatzas, D., Jawahar, C.: DocVQA: A Dataset for VQA on Document Images. In: Proc. WACV (2021)
17. Saout, T., Lardeux, F., Saubion, F.: Information Extraction from Invoices. In: Proc. ICDAR (2021)
18. NIST: Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 (2023)