ARTIFICIAL INTELLIGENCE

# AI-Enabled Enterprise Document and Records Management Systems in Saudi Arabia

Saleem Al Mohtasib

Published November 2026


## Abstract

Enterprise records are central to accountable public services, regulatory compliance, institutional memory and evidence-based management. Saudi Arabia's digital transformation creates opportunities to improve document capture, classification, correspondence handling, search, retention and access through artificial intelligence (AI). Yet technical demonstrations of optical character recognition, natural language processing and large language models often neglect authenticity, provenance, Arabic-language performance, privacy, operational integration and the long-term obligations of records management. This narrative review examines how AI-enabled document and records management systems can support Saudi government and enterprise organisations while maintaining defensible human oversight. It synthesises research on intelligent document processing, information retrieval, enterprise architecture, model evaluation and digital governance alongside relevant Saudi data policies. A six-layer socio-technical framework is proposed, linking records governance, ingestion, trusted metadata, bounded intelligence services, workflow decisions and continuous assurance. The review argues that successful implementation should be judged by retrieval reliability, record integrity, processing quality, service outcomes and compliance, rather than automation volume alone. A staged adoption and evaluation roadmap is presented for municipalities, universities and large enterprises. The paper is conceptual, not a report of original field testing, and specifies priorities for empirical validation in Arabic-English institutional environments.


**Keywords:** artificial intelligence; enterprise document management; electronic records management; Saudi Arabia; digital transformation; optical character recognition; records governance

## 1 Introduction

Saudi organisations manage correspondence, decisions, contracts, requests, approvals and evidence across increasingly interconnected services. In municipalities, document records link applications to permissions and decisions; in universities, they support admissions, research administration and personnel processes; and in enterprises, they preserve contractual and operational accountability. Digital repositories reduce dependence on paper, but the existence of a repository is not equivalent to a reliable records system. Without consistent metadata, retention rules and authorised access, a larger digital collection may make information harder to interpret and govern. Saudi Vision 2030 makes improved public services and digital capability strategically significant, while national data governance requirements add obligations to protect the information handled through those services.

Conventional enterprise content management combines capture, indexing, storage, version control, search, routing and disposition. Its weaknesses often emerge at boundaries: scanned files lack searchable text, different departments use incompatible terminology, correspondence is repeatedly re-entered, and the same decision appears in multiple systems. Employees may resort to personal folders or email attachments when enterprise search is unreliable. Such workarounds create duplica-

tion, obscure provenance and delay disclosure or service responses. Artificial intelligence can address bounded components of this problem by extracting information, assigning candidate classifications, detecting duplicates, matching entities and assisting retrieval.

This review focuses on AI-enabled enterprise document and records management in Saudi settings, drawing explicit connections to government delivery, bilingual content, workflow automation and institutional oversight. It develops a practical architecture and evaluation approach instead of assuming that performance measured on general public datasets transfers automatically to real Saudi correspondence or sensitive case files. The framework positions human accountability, documented controls and measurable service value as conditions for trustworthy adoption.

## 2 Aim, objectives and research questions

The aim is to develop an evidence-informed framework for deploying and governing AI-enabled enterprise document and records management systems in Saudi Arabia, with particular attention to service efficiency, information integrity and institutional accountability.

Five objectives guide the review: identify recurring limitations in traditional document and records workflows; evaluate suitable AI functions for ingestion, classification, retrieval and quality assurance; define the metadata, interoperability and governance foundations needed for reliable deployment; propose a risk-based operational architecture for Saudi institutions; and formulate meaningful technical, organisational and policy evaluation measures.

The guiding questions are: Which document tasks offer credible value from AI without transferring inappropriate authority to automated outputs? How can records authenticity, confidentiality, retention and provenance be preserved from capture to authorised disposition? How should Arabic and English document quality be validated across heterogeneous templates and handwriting? What governance and change-management arrangements allow organisations to scale intelligence while monitoring errors, user behaviour and legal obligations?

## 3 Review methodology

This narrative review develops a conceptual synthesis from the supplied manuscript bibliography and a targeted editorial source check on 24 November 2026. The original database search dates, exported results, duplicate counts and screening decisions were not supplied; consequently, exhaustive database coverage, systematic selection and independent screening cannot be claimed. No pooled effect estimate is attempted. The original bibliography is retained as a provisional source inventory pending full-text verification. Sources are distinguished as empirical studies, conceptual works, standards and official guidance. The editorial checks used public web discovery and publisher or government pages; the accompanying register records the checks and unresolved items. The proposed framework and measurement protocols are design propositions, not observations from a Saudi deployment.

## 4 Enterprise records and the Saudi digital context

Enterprise documents are not merely files; they are evidence of actions and obligations. A record's value depends on content, context, structure, authenticity and traceability. A purchase approval, for instance, is meaningful only when connected to the appropriate request, authorised signatory,

version and decision time. Records programmes therefore require documented classification schemes, access rules, retention schedules and disposition controls. ISO 15489 provides an international conceptual basis for records management, while ISO/IEC 27001 describes information-security management requirements. AI must operate inside these foundations rather than substitute for them.

The Saudi setting introduces specific operational demands. Public-sector organisations frequently connect service portals, correspondence systems, archival repositories, digital signature tools and back-office applications. Citizens and institutional users may submit documents in Arabic or English, with mixed numerals, stamps, tables and varied scan quality. Historic records can include poorly aligned scans or incomplete metadata. A model optimised for clean English invoices may fail on a handwritten Arabic annotation or an agency-specific approval form. Transfer learning helps, but institution-specific testing remains indispensable.

Saudi data governance is not an optional add-on. The National Data Management Office promotes consistent rules and practices for data as a national asset, and the Personal Data Protection Law imposes obligations on relevant controllers and processors. Records systems may contain identities, educational histories, financial information, correspondence and third-party details. Accordingly, AI services require controls over purpose, access, transfer, logging, retention and human review. Whether a particular processing arrangement is permissible must be assessed under applicable rules and current official guidance, rather than inferred from a technology vendor's claims.

An electronic document management system (EDMS) manages working documents and versions; enterprise content management (ECM) coordinates content across business applications; an electronic records management system (ERMS) controls declared evidence across its retention lifecycle. In the proposed design, the records officer authorises declaration, retention classification, legal-hold release and disposition; working copies cannot overwrite authoritative evidence. Capture hashes, version identifiers, custody events, role-based permissions and approval logs form a proposed audit trail.

## 5 Document lifecycle and AI opportunities

Classification follows ingestion. Supervised models may suggest a business class or retention category based on text and metadata; unsupervised methods can support discovery of recurrent themes. The distinction between document type and record function matters. A document labelled "letter" could belong to procurement, legal proceedings or a citizen complaint, each with a different retention rule. Consequently, records classification should incorporate business process, originating organisation, transaction context and authoritative schedules. High-risk classification decisions need explicit human verification.

Search is a further opportunity. Keyword matching fails when query terms differ from document language or use agency-specific synonyms. Semantic retrieval can connect meaning across paraphrases and languages, while hybrid search combines metadata filtering, lexical retrieval and vector similarity. Retrieval-augmented generation may summarise retrieved records, but it should cite controlled source identifiers and refuse to infer unsupported facts. Users must be able to open the underlying authoritative record, see version history and distinguish a generated answer from an official decision.

Correspondence management illustrates the larger opportunity. An incoming request can be linked to a case identifier, routed to the correct department, checked for missing attachments and monitored

against service targets. AI may suggest a routing destination or response draft, but escalation, signature and final approval remain governed by existing delegations. Intelligent automation is most useful where it makes a traceable human workflow faster and more consistent without hiding the rationale for action.

Retention and disposition require particular caution. Models can detect duplicate files or propose records for review, but deletion must never depend solely on a probabilistic classification. Legal holds, litigation, investigations and records of enduring value may override routine retention settings. The system must preserve evidence of who approved a disposition, which rule applied, what was deleted and whether the approved retention period had elapsed.

**Table 1.** Document lifecycle controls and appropriate AI roles.

| Stage | AI-enabled task | Essential control |
|---|---|---|
| Receipt | Language detection and extraction | Preserve original and receipt event |
| Classification | Suggested category and metadata | Approve against records schedule |
| Retrieval | Semantic and hybrid search | Enforce permissions and show source |
| Action | Routing and drafting assistance | Authorised officer approves outcome |
| Disposition | Duplicate detection and review | Legal hold and authorised deletion |

**Table 2.** Proposed record-lifecycle responsibility mapping.

| Stage | Control | Accountable role |
|---|---|---|
| Capture | Original, hash, receipt time and source | Records officer |
| Declare | Version and retention schedule | Records manager |
| Use | Role-based access and custody log | Information owner |
| Hold | Suspend deletion; document release | Legal officer |
| Dispose | Two-person approval and disposition certificate | Records manager and approver |

## 6 Technical approaches to intelligent document processing

Optical character recognition is a basic building block, yet its quality depends on script, scan resolution, fonts, table structures, stamps and page geometry. Classical OCR uses image processing and character recognition; contemporary approaches increasingly integrate deep learning, document layout and contextual language modelling. OCR word-level accuracy cannot establish that a table, legal clause or signature location was understood correctly. Evaluation should include character error rates, field extraction performance and the preservation of reading order.

Document layout models distinguish headings, tables, key-value pairs and text regions. LayoutLM and related architectures demonstrate how textual and spatial information can be combined for document understanding. Transformer-based document models can improve the handling of varied page formats, though computational requirements and annotation costs can be substantial. For Saudi institutions, the challenge is less about adopting the newest architecture than maintaining validated performance across Arabic text, bilingual forms and local seals, while retaining fallbacks for out-of-distribution material.

Natural language processing adds entity recognition, topic classification, cross-document matching and extraction of dates, names and references. Entity resolution requires safeguards against similar names and repeated identifiers. Where two records are linked incorrectly, a decision may be attributed to the wrong case. A reliable system therefore exposes confidence, supporting evidence and the original text span. Business rules can constrain extraction: impossible dates, invalid reference formats and conflicting identifiers should trigger review rather than silent acceptance.

Generative AI can assist document summarisation, question answering and preparation of routing notes. Its ability to produce fluent language also creates a distinct risk: a plausible answer may omit qualifications or assert facts absent from the underlying record. Retrieval grounding, limited output templates, source citations, redaction and approval checkpoints reduce but do not eliminate this risk. Generative components should initially be advisory, especially for legal correspondence, citizen decisions or protected records. Sensitive information must not be sent to unauthorised external services.

Model selection should follow use-case consequence and available evidence. A rules engine may outperform a complex model for highly standardised forms. A modest classifier may be easier to explain and maintain than a large multilingual model. Conversely, irregular archives may justify stronger layout and language methods. The proper comparison is among validated alternatives using realistic documents, measurable costs and defined review capacity, not between vendor feature lists.

Xu et al. (2021) report improved FUNSD form-understanding performance for LayoutLMv2 relative to LayoutLM (84.20 versus 78.95 on the reported benchmark). The result concerns the benchmark, not Saudi records or Arabic handwriting. Evaluate OCR separately using character error rate (substitutions + insertions + deletions)/reference characters and word error rate using the corresponding word denominator. Report field F1 = 2PR/(P+R), retrieval precision at k = relevant retrieved/k, and grounding = supported answer claims/all answer claims. Stratify by Arabic/English, handwritten/printed, mixed numerals, stamps, tables and scan quality; protected files require permission tests. Unsupported answers must abstain rather than create a false record.

**Table 3.** Evidence and proposed technical comparison.

| Approach | Evidence / scope | Limitation and proposed test |
|---|---|---|
| LayoutLM / LayoutLMv2 | Xu et al. (2021), FUNSD benchmark comparison | No local Arabic validation; field F1 by language |
| OCR | Proposed local transcription benchmark | Stamps and handwriting; CER and WER |
| Hybrid retrieval / RAG | Proposed controlled evidence retrieval | False grounding; precision@k and supported claims |

## 7 Integrated architecture and interoperability

A dependable platform needs a separation between authoritative records, derived machine outputs and operational decisions. The proposed six-layer architecture begins with governance and information ownership; moves through controlled ingestion and source preservation; applies metadata and quality services; offers bounded AI processing; integrates results with business workflows; and finishes with assurance and feedback. The order is conceptual rather than a requirement to buy six separate products. Existing enterprise content management systems may fulfil several layers if their controls can be demonstrated.

Trust boundaries deserve explicit design. A search index and vector database might contain excerpts of confidential documents even when the source repository is protected. An access filter must operate before retrieval and again before information reaches a summarisation model. Training datasets, prompts, logs and cached outputs can also expose sensitive content. Least-privilege access, encryption, secure key management, retention limits and audit trails should cover every representation of a record, not merely the original file.

The architecture should remain resilient when AI functions become unavailable. Employees must be able to receive, register, route and retrieve critical records using approved fallback procedures. Models should not silently change classification policies after an update. Versioned models, controlled configuration and reversible deployment allow errors to be investigated. A service-level design should distinguish business-critical actions from convenience features so that an indexing delay does not block essential approvals.

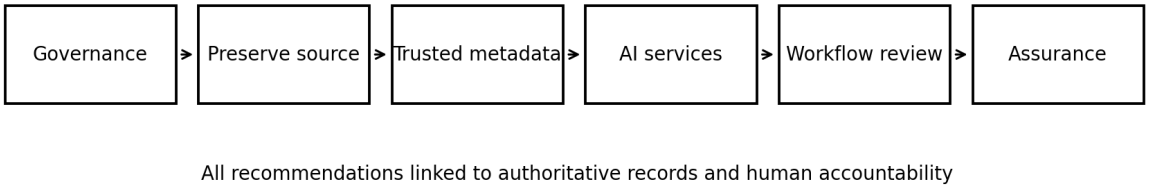


**Figure 1.** Governance-centred architecture for AI-enabled records management.

The six-layer proposal extends conventional records lifecycle controls by placing an authorisation gateway before the lexical/vector index and an approval gateway before official action. Unlike a document-understanding model, it links every suggestion to a source version and decision log. Unreadable or unauthorised content goes to a restricted exception queue; failed indexing leaves the trusted repository intact. Rollback disables the affected model or index version and restores the last approved configuration, without deleting custody evidence.

## 8 Data quality, security and responsible AI

The most immediate AI risk is often poor data rather than an exotic algorithm failure. Duplicate files, missing page sequences, inconsistent record identifiers and inaccurate scans create errors before a model runs. Data-quality checks should measure completeness, uniqueness, validity, consistency and freshness. Quality reports need accountable owners and correction workflows. If a missing field has legal significance, the appropriate outcome is an exception rather than an invented value.

Responsible deployment requires an explicit answer to who is accountable for every automated recommendation. Confidence scores are imperfect; they should be calibrated and accompanied by clear exception routes. An operator should know whether an item was extracted automatically, checked against a rule or approved by an authorised officer. Periodic sampling of apparently successful cases is important because a system that measures only flagged errors may conceal widespread low-level inaccuracies.

Privacy and fairness also require attention. Names, locations and service categories can reveal sensitive information or create differential error patterns. Arabic dialects, older document templates

and scans from particular departments may experience higher error rates than common formats. Evaluation should therefore report relevant language and document subgroups, without constructing intrusive demographic profiles. Data minimisation, purpose limitation and controlled retention should be reviewed with legal and governance specialists using current Saudi requirements.

## 9 Implementation roadmap and evaluation

Implementation should begin with a bounded problem such as automatic indexing of incoming correspondence or extraction of reference numbers from standard forms. Project teams should establish an operational baseline before a pilot: manual processing time, error rates, rework, retrieval delay, service completion and staff experience. A multidisciplinary group defines the valid task, records impact, authority levels, acceptable errors and exception ownership. A shadow trial then runs the AI without controlling the live workflow, allowing side-by-side review of machine suggestions and existing practice.

After initial validation, a limited production trial can be introduced with human verification. Metrics should separate technical accuracy from operational value. OCR character error rate, classification precision and recall, field-level accuracy, retrieval relevance and response latency describe model behaviour. Meanwhile case throughput, rework, time to locate a record, escalation frequency and user trust describe organisational outcomes. High extraction accuracy is not proof of value if verification consumes more time than the original manual step.

The team should evaluate minority document types and changed templates rather than reporting only an average. A classifier that performs well on common correspondence may still fail on confidential exceptions or attachments. Confidence thresholds should be linked to the impact of errors and the cost of review. For example, proposed tags for internal search may tolerate more uncertainty than a change to a statutory retention class. Review sampling should continue after deployment because data distributions change as forms and departments evolve.

Scaling calls for an operating model, not only a project closeout. A records owner maintains taxonomies, a data steward manages metadata quality, technical teams monitor services, security specialists assess threats, and business owners decide on exceptions. Contracts should clarify rights over training data, model improvements, audit evidence and exit arrangements. Clear ownership is particularly important when multiple vendors provide OCR, content management, workflow and generative AI functions.

Operational readiness and baseline design. A pilot should establish baseline processing time, error rates, search success, backlog, and exception handling. Sampling must include varied document types, scanned files, and low-frequency but consequential cases. Project teams should document the source of labels, acceptable error thresholds, human review rules, and rollback arrangements before live automation. Performance should be checked by document category and organisational unit. A reduction in processing time is insufficient if the system introduces missing records or inaccurate retention decisions.

Arabic and bilingual testing. Saudi workflows frequently combine Arabic and English, including mixed-script references, handwritten annotations, low-quality scans, and formal administrative terminology. Evaluation should report extraction and classification errors separately for both languages, using realistic institutional samples. Human reviewers should verify names, dates, legal

references, and high-impact decisions. Domain-specific templates and controlled vocabularies may improve reliability, but updates require renewed testing and documented approval.

**Table 4.** Evaluation dimensions for controlled pilots.

| Dimension | Measures | Interpretation |
|---|---|---|
| Document quality | Character error; field accuracy | Stratify by language and form type |
| Classification | Precision; recall; review rate | Report consequences of mistakes |
| Retrieval | Relevant results; source traceability | Measure permission compliance |
| Workflow | Resolution time; rework; backlog | Compare with pre-pilot baseline |
| Governance | Audit completeness; incidents | Verify corrective-action closure |

Hypothetical municipal correspondence pilot: admit only authorised, de-identified Arabic and English correspondence with documented retention classes and representative legacy scans. Compare manual handling with shadow-mode extraction on the same cases, with independent reference labels and recorded reviewer minutes. Require zero unauthorised disclosure in the test set, no uncontrolled deletion and compliance-owner acceptance of risk-specific error tolerances before deployment. Failed gates trigger quarantine, root-cause correction and a fresh evaluation. These are proposed acceptance conditions, not measured results.

## 10 Discussion

The evidence suggests that AI can reduce repetitive processing and improve access to institutional knowledge, but its value depends on well-defined information practices. In reality, capture, records classification, correspondence workflow and archive disposition make different decisions with different failure consequences. A capable extraction model cannot correct ambiguous authority or the absence of a retention schedule. Governance maturity is thus a prerequisite for responsible intelligence, rather than an automatic consequence of deploying it.

The proposed framework places the authoritative record at the centre. The system may generate text representations, candidate tags and summaries, but all should remain linked to source evidence. This design is especially valuable for public accountability: an administrator should be able to reconstruct not only the final decision, but also the information that informed it and any machine suggestion that was accepted or rejected. Such traceability also makes model improvement possible because reviewers can identify systematic errors.

An important managerial implication is that dashboards should not reward automation for its own sake. Counting processed pages or accepted suggestions encourages volume even when quality deteriorates. Balanced reporting should include verified accuracy, user time saved, exceptions, records retrieved correctly, overdue requests and governance incidents. Context is essential: an increase in flagged errors can indicate more effective checking rather than poorer performance. Managers need interpretable measures and the ability to investigate exceptions instead of relying on a single performance score.

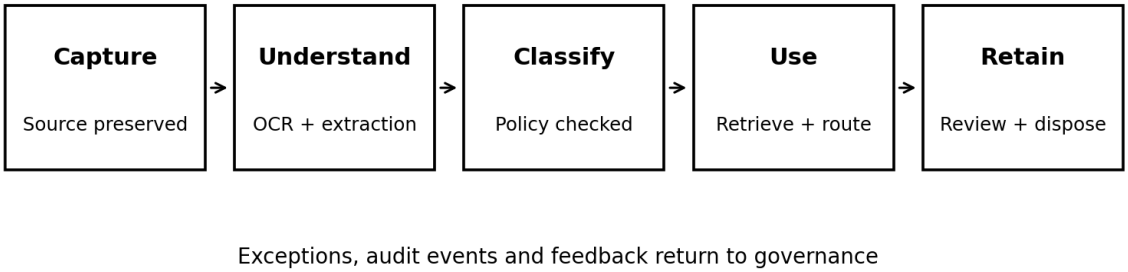


**Figure 2.** AI-assisted records lifecycle with control checkpoints.

## 11 Research gaps and future research

Published evaluations frequently use benchmark datasets that do not capture the diversity of Saudi administrative archives. More studies are needed on Arabic-English mixed documents, right-to-left layout, dated scans, stamps, signatures, tables and inconsistent filing conventions. Publicly shareable benchmark corpora could help, but must be carefully de-identified and governed. In many cases, synthetic examples can supplement rather than replace genuine operational validation.

A second gap concerns the relationship between technical improvements and public service outcomes. Research should measure whether better indexing reduces request delays or whether AI-assisted retrieval improves decision accuracy. Quasi-experimental comparisons, stepped deployment and interrupted time-series designs may test value without disrupting essential services. Researchers should disclose baseline processes, review burden, selection criteria and false-positive costs, rather than attributing every improvement to AI.

Third, long-term evidence is limited for model drift, vendor change and preservation. A records system may operate for decades while AI platforms evolve rapidly. Studies should address durable identifiers, exportability, reproducible classifications, archival formats and the treatment of generated annotations when an underlying model is retired. Another important area is explainable retrieval: investigators and officers need to understand why a document was matched to a case and whether relevant records were omitted.

Finally, governance research should explore employee competence, trust and bilingual workflow design. Training that covers exception handling, data sensitivity and evidence verification may improve outcomes more than additional model complexity. Future work should compare human-only, AI-assisted and AI-led workflows under explicit authority controls, with meaningful measurement of workload, consistency, security and service fairness.

## 12 Practical managerial and policy implications

Executives should treat intelligent records management as an information-governance programme linked to service reform, not as an isolated artificial intelligence procurement. Begin with an inventory of record categories, applications, access permissions and retention obligations. Appoint accountable business owners and establish a cross-functional governance forum. Require a business case that specifies the decision being improved, source documents, target users, exception procedures and evidence of success.

Project managers should coordinate phased delivery. The sequence should include discovery and baseline measurement, data preparation, model and rule comparison, shadow validation, controlled production, and independent review before wider expansion. Procurement specifications should request demonstrable Arabic support, records provenance, secure integration, auditability, model-change controls and data exit provisions. A vendor demonstration on attractive sample documents is insufficient evidence for production readiness.

Records professionals should define metadata standards and classification policies; data teams should quantify and correct quality problems; cybersecurity and privacy teams should assess data flows; and operational leaders should ensure that staff retain authority to correct AI suggestions. Users should have simple ways to report misclassifications, missing pages, inappropriate recommendations and access failures. That feedback must be analysed without penalising good-faith reporting.

Policy makers and sector bodies can encourage shared reference architectures, minimum evaluation requirements and carefully governed test datasets. Interoperability standards and practical guidance on records provenance may provide durable value across institutions. Evaluation should emphasise accountable decisions, lawful information use, service quality and institutional resilience. Artificial intelligence can support these objectives only when its outputs remain challengeable and tied to trustworthy records.

## 13 Limitations

This paper is a narrative review rather than a primary empirical study. Its conclusions are based on published research, technical standards and institutional guidance; they have not been validated using confidential Saudi enterprise records or controlled deployments. Consequently, proposed benefits for processing time, retrieval accuracy and service quality are plausible hypotheses, not measured outcomes. The evidence base combines technical benchmarks with organisational studies that use different document types, languages, evaluation measures and deployment environments, limiting direct comparison. English-language datasets are overrepresented in document AI research, and reported performance may not transfer to Arabic documents, mixed scripts, historical scans or municipal case files. Rapid changes in generative AI models, vendor platforms and regulatory implementation may also affect applicability. The proposed architecture assumes participating organisations can establish trustworthy metadata, retention rules, access controls and human review capacity; these preconditions may differ substantially across institutions. Future work should conduct prospective, multi-site Saudi evaluations, report failures and exception-handling costs, and independently assess long-term preservation, fairness and accountability before claiming operational impact.

## 14 Conclusion

AI-assisted document processing may support Saudi enterprise records management when extraction, classification and retrieval remain subordinate to records authority and access controls. This narrative review proposes a six-layer architecture that preserves original records, separates advisory outputs from authoritative decisions and makes disposition subject to approved retention rules and legal holds. The proposed bilingual evaluation protocol and correspondence pilot provide a route to testing technical quality, staff workload and record integrity. No Saudi field data were analysed;

implementation benefits remain hypotheses. Adoption should depend on independently reviewed pilot results and documented compliance rather than automation volume.

## 15 Declaration of AI use

Generative AI assisted editorial rewriting, conceptual control design and the preparation of the comparison tables. No original participant data or empirical intervention is reported.

## References


[1] Kingdom of Saudi Arabia. (2016). Saudi Vision 2030. https://www.vision2030.gov.sa/

[2] Saudi Data and Artificial Intelligence Authority (SDAIA). *National Data Management Office: Mandate and objectives*. https://sdaia.gov.sa/en/Sectors/Ndmo/Pages/default.aspx

[3] Saudi Data and Artificial Intelligence Authority (SDAIA). *Guide to the Saudi Personal Data Protection Law for controllers and processors*. https://dgp.sdaia.gov.sa/

[4] AI-based Integrated Approach for the Development of Intelligent Document Management System (IDMS). (2023). Procedia Computer Science, 230, 725–736. https://doi.org/10.1016/j.procs.2023.12.127

[5] Camargo Niño, C., & Zambrano Rivera, S. (2024). Use of artificial intelligence in documentary information systems: A systematic literature review 2013–2023. https://doi.org/10.5281/zenodo.13179067

[6] Alotaibi, N. N., Saberi, M., Bandara, M., & Porntaveetus, T. (2026). Intelligent Business Document Processing Using AI- and NLP-Based Techniques: A Systematic Literature Review. *Analytics*, 5(3), 31. https://doi.org/10.3390/analytics5030031

[7] International Organization for Standardization. (2016). ISO 15489-1:2016, Information and documentation: Records management—Concepts and principles.

[8] International Organization for Standardization / International Electrotechnical Commission. (2022). ISO/IEC 27001:2022, Information security management systems—Requirements.

[9] Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020). LayoutLM: Pre-training of text and layout for document image understanding. *Proceedings of KDD* 2020. https://doi.org/10.1145/3394486.3403172

[10] Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., et al. (2021). LayoutLMv2: Multi-modal pre-training for visually-rich document understanding. *Proceedings of ACL* 2021. https://doi.org/10.18653/v1/2021.acl-long.201

[11] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. https://doi.org/10.6028/NIST.AI.100-1

[12] National Data Management Office. *National Data Management and Personal Data Protection Standards*. https://sdaia.gov.sa/en/SDAIA/about/Documents/PoliciesEN.pdf

[13] International Organization for Standardization. (2017). ISO 23081-1:2017, Information and documentation—Managing metadata for records—Part 1: Principles.

[14] International Organization for Standardization. (2016). ISO 16175-1:2016, Information and documentation—Processes and functional requirements for software for managing records.

[15] National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative AI Profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

[16] Appalaraju, S., Jasani, B., Kota, B. U., Xie, Y., & Manmatha, R. (2021). DocFormer: End-to-End Transformer for Document Understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 993–1003.

[17] Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022). LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking. *Proceedings of the* 30th ACM International Conference on Multimedia, 4083–4091. https://doi.org/10.1145/3503161.3548112

[18] Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., et al. (2022). OCR-Free Document Understanding Transformer. *Proceedings of the European Conference on Computer Vision*.

[19] Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., et al. (2023). TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(11), 13094–13102. https://doi.org/10.1609/aaai.v37i11.26538

[20] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.

[21] National Institute of Standards and Technology. (2020). Security and Privacy Controls for Information Systems and Organizations. *NIST Special Publication* 800-53 Revision 5. https://doi.org/10.6028/NIST.SP.800-53r5

[22] International Organization for Standardization. (2019). ISO 30301:2019 Information and documentation — Management systems for records — Requirements.

[23] International Organization for Standardization. (2023). ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management.