Automated GRC Tool for Fintech Systems

Ansh Vora[1], Tanish Chaudhari[2], Harjas Rohra[3], Prasanna Kulkarni[4] and Shweta Shah
[12345]*Pune Institute of Computer Technology, Pune*

***Abstract*—As the volume of digital payments have increased rapidly, several financial institutions have had to deal with an increasingly intricate and dynamic regulatory environment. In the face of frequent and consistent circulars and advisories issued by authorities like the Reserve Bank of India (RBI), National Payments Corporation of India (NPCI) as well as global standards bodies such as PCI DSS and ISO/IEC 27001, the orthodox manual compliance processes are overburdened. With the expected workload of interpreting complex and lengthy legal documents, extract obligations from them and then map them to the organization's internal controls, the compliance team faces a gigantic task. Furthermore, the compliance team also must work with other teams to execute those controls and gather auditable evidence under tight deadlines. In the face of the rapidly changing modern fintech systems, this labor-intensive workflow is error-prone and ill-suited to manage the real-time risk associated with the current environment. This paper presents the design and implementation of a framework built especially for tightly regulated digital payments environment- the AI-first Governance, Risk and Compliance (GRC) framework that can automate "obligation-to-evidence" lifecycle from one end to another. The system can intake circulars and contractual documents in machine readable or non-readable formats and uses Mistral Document AI merged with a specialised language model and Retrieval-Augmented Generation (RAG) pipeline to extract obligations into schema-validated YAML.**

***Index Terms*—Agentic AI, Digital Payments, Document Processing, Governance, Risk and Compliance, Retrieval-Augmented Generation.**

I. INTRODUCTION

Domestic regulators (RBI, NPCI and SEBI) and international standards frameworks (PCI DSS and ISO/IEC 27001) are shaping a considerably fluid regulatory environment in which financial institutions have to operate nowadays. For high-volume payment systems like UPI, IMPS, AEPS, NEFT and RTGS, organisations must especially align their audit posture, internal controls and risk registers with guidelines, standards, and circulars that are published by the aforementioned parties. The process of reviewing regulatory documents and deriving obligations from them to align with the organisation's existing process or controls is done manually by compliance officers and internal auditors at present. In addition to this, they must also coordinate the new controls' implementation with operations and infrastructure teams spanning multiple verticals and also in conclusion collect the evidence for any audits that are to be done in the future. Given the expanse of the workload, it is heavily based on text analysis, spreadsheet-based tracking and unstructured communication. These characteristics make it difficult to scale and meet the needs of the ever-increasing regulatory volume and complexity, any delays or inconsistencies caused due to this might result in residual risk and regulatory penalties, and in extreme cases, loss of license as well.

II. PROBLEM SETTING AND CHALLENGES

The problem statement had the main objective of addressing gaps in regulatory change management for fintech, but it had several specific challenges with respect to explainability and scalability that was to be maintained:

- For the domain of the objective, every AI-generated obligation or control mapping had to be traceable back to robust source clauses. The high-stakes interpretation of documents had to be sufficiently rational to hold up against regulatory and audit scrutiny.
- High velocity and heterogeneity in the regulatory environment were the first difficulty encountered. In a year, several hundred circulars are published by regulators like RBI and NPCI in different formats such as multi-page PDFs or scanned notices.

- Building on the previous difficulty, a capable document AI and OCR model was required to ingest the diversity of document formats in the inputs ranging from clean PDFs and DOCX files to scanned images and partner contracts with redactions.
- Due to financial institutions' preference for on-premises deployments, some parts of the proposed pipeline had to operate under rigid data locality rules but also extract benefits from LLM capabilities against the requirement for minimal external data sharing.

Simple rule-based parsers or ungrounded LLM summarisation and other similar naïve approaches were thus inappropriate for a production-ready GRC use case the pipeline aimed to tackle.

III. LITERATURE SURVEY

Automating the GRC lifecycle for fintech institutions requires integrating advances from several subfields: document AI and OCR, legal and regulatory NLP, retrieval-augmented generation, and workflow and analytics infrastructure.

The foundational problem of recovering machine-readable text from scanned regulatory notices has a longer history than recent advances in large models might suggest. Tesseract, which combines a feature-based recogniser with adaptive classification, remains a widely deployed baseline for noisy document images in on-premise settings (Smith, 2007), owing in part to its deterministic behaviour under fixed configurations and its absence of network dependencies.

Research in document AI has since moved toward multimodal, layout-aware architectures. LayoutLMv3 jointly encodes text, layout, and visual signals within a unified transformer, achieving state-of-the-art results on form understanding, invoice parsing, and document question answering (Huang et al., 2022). Donut takes a structurally distinct approach, discarding the OCR stage entirely: its encoder–decoder maps document images directly to structured outputs, reducing the error propagation that arises when a weak OCR stage affects all downstream processing (Kim et al., 2021). For multi-page regulatory artefacts circulars and master directions that distribute definitions and obligations across dozens of pages Hi-VT5 compresses page-level visual–textual features while preserving cross-page context, supporting reasoning over obligations whose scope depends on definitions appearing considerably earlier in the document (Tito et al., 2022).

The legal NLP community has developed domain-specialised encoders motivated by the persistent finding that generic language models mishandle statutory language. LEGAL-BERT extends the BERT pre-training regime to legal corpora and consistently outperforms general-purpose baselines on named-entity recognition, classification, and similarity tasks (Chalkidis et al., 2020). Evaluating such models systematically is itself non-trivial; LexGLUE addresses this by providing a benchmark suite spanning multi-label classification, legal entailment, and related tasks, establishing a common evaluation surface where prior results had been difficult to compare across studies (Chalkidis et al., 2022).

Sequence length presents a practical constraint that clause-level models tend to sidestep by operating on extracted fragments rather than full documents. Longformer introduces a sparse attention mechanism that scales linearly with sequence length, making global context tractable for summarisation, classification, and clause-level reasoning in documents exceeding 60 pages (Beltagy et al., 2020) a property relevant when a single obligation references definitions distributed across an entire master direction.

Lewis et al. (2020) demonstrate that coupling a neural retriever with a sequence-to-sequence generator substantially improves performance on knowledge-intensive tasks while preserving the ability to trace generated responses back to retrieved passages. Dense Passage Retrieval operationalises this using dual encoders that map questions and passages into a shared embedding space, delivering higher recall than lexical retrieval on open-domain question answering benchmarks (Karpukhin et al., 2020).

In regulatory contexts, however, exact terminology abbreviations such as AEPS or RTGS, or precise statutory identifiers carries semantic weight that embedding similarity can obscure. The probabilistic relevance framework and BM25 remain strong baselines where surface form is decisive (Robertson & Zaragoza, 2009), and hybrid retrieval strategies that combine dense and lexical methods have accordingly received attention in domain-specific retrieval settings. DuckDB provides an in-process OLAP

engine with vectorised execution suited to on-premise analytical workloads, eliminating the operational overhead of a separate database server in single-tenant deployments (Raasveldt & Mühleisen, 2019). Apache Airflow has become a standard mechanism for defining and executing DAG-based workflows, providing scheduling, retry logic, and SLA monitoring for complex data and machine-learning pipelines (Kumar et al., 2024). Both are directly applicable to orchestrating document ingestion, parsing, and downstream AI components in a GRC context.

Three gaps are evident across these lines of work. Most document-AI and legal-NLP systems terminate at extraction or question answering, leaving the step unaddressed from a structured obligation to a mapped control, enforcement verdict, and auditor-traceable evidence record. RAG and agentic methods handle grounding and tool use but rarely account for the hybrid privacy constraint present in fintech GRC, where public regulatory circulars and confidential partner contracts must be processed under distinct data-handling regimes within the same pipeline. Finally, there is limited published guidance on architecture that compose agentic AI, event-driven messaging, embedded analytics, and on-premises deployment into a configuration sustainable for regulated payment institutions operating without cloud-scale infrastructure.

IV. SYSTEM OVERVIEW

The proposed system realises an end-to-end "obligation-to-evidence" pipeline by composing five tightly integrated subsystems:

1. document ingestion and parsing,
2. obligation extraction,
3. control and risk mapping,
4. enforcement and evidence collection,
5. continuous compliance analytics and audit interaction.

Together, these components implement a closed loop from raw regulatory artefacts to auditable compliance posture.

4.1 Document Ingestion and Parsing

The ingestion layer is responsible for normalising the ingestion layer normalises heterogeneous regulatory and contractual artefacts into a representation suitable for downstream NLP and agentic processing. Regulatory circulars and partner contracts may arrive as born-digital PDFs, DOCX files, or scanned images bearing stamped and signed notices; the ingestion service accommodates all three, assigning provenance metadata authority, date, and document type before scheduling parsing workflows through an orchestrator. Each ingested artefact is processed by a document AI engine that recovers both text and layout. Born-digital PDFs and DOCX files are converted to a markdown-like intermediate representation preserving headings, lists, and table structure, while scanned images are routed through OCR. Where high-capacity document AI is unavailable or incompatible with on-premise constraints, a deterministic OCR engine serves as a fallback adequate for most notices encountered during evaluation, though its handling of multi-column regulatory tables required manual post-processing in several cases. The output of this stage is a structured document object containing normalised text, coarse layout annotations, and block-level provenance offsets.

4.2 Obligation Extraction Pipeline

Built atop the normalised document representation, the obligation extraction pipeline converts free-form regulatory prose into machine-readable obligations. Clause segmentation partitions each document into semantically coherent spans sections, sub-clauses, bullet points, and tabular entries enriching each span with its position in the document hierarchy alongside any referenced regulatory identifiers. Obligation identification is performed by a specialised language model operating within a retrieval-augmented generation setup. For each candidate clause, the model classifies whether the text encodes a regulatory obligation, a definition, guidance, or purely informational content. That this classification depends so heavily on prompt design is, itself, an architectural risk the system must actively manage.

A small set of carefully tuned prompts elicits structured JSON-like outputs from the model, yielding YAML representations consumed by downstream agents a step that, while procedurally straightforward, proved sensitive to prompt phrasing in practice, particularly for obligation types that straddle definitional and prescriptive language. One prompt targets obligation extraction from regulatory text; a second, used later in the pipeline, surfaces candidate controls from existing control patterns. Both are

designed to behave deterministically under fixed model configurations, a requirement that directly supports reproducibility and auditability.

4.3 Control and Risk Mapping Agent

Materialised obligations are published onto an event bus, where each event encapsulates a single obligation, its metadata, and a cryptographic fingerprint of the originating clause. An alternative formulation considered a pull-based polling model; the event-driven approach was retained primarily for scalability under regulatory bursts, though it introduces ordering constraints on downstream agents that must be managed carefully a trade-off addressed further in Section V.

The control and risk mapping agent is implemented as an agentic workflow built on a tool-orchestration framework. For each incoming obligation event, the agent performs hybrid retrieval over an internal control library: dense vector search retrieves semantically similar controls by objective, scope, and domain, while lexical search ensures exact regulatory terminology and identifiers are preserved. Retrieved candidates are ranked by relevance and criticality before the agent proposes a mapping annotated with similarity scores and rationale from the obligation to one or more existing controls. When the library lacks an adequate match, the agent synthesises a candidate control template aligned with the organisation's canonical control schema.

Risk assessment is integrated directly into this workflow. Obligations receive criticality ratings based on regulatory authority, affected payment channel, and potential impact on confidentiality, integrity, or availability; the risk register aggregates these assessments, enabling computations of exposure across domains, authorities, and control families.

4.4 Enforcement and Evidence Collection

To close the loop from obligations to operational posture, an enforcement and evidence collection layer operates against a representative enterprise surface. In the current implementation, this surface is a mock directory service exposing users, groups, organisational units, device records, and policy objects through a RESTful API. The mock service is sufficient for proof-of-concept evaluation; it does not, however, capture the consistency anomalies and replication lag common in production IAM environments, a limitation that future work should address before deployment at scale. Controls mapped in the preceding stage are expressed as expectations over this enforcement surface. A control may require that multi-factor authentication be enabled for a class of administrative accounts, or that specific transaction-processing hosts belong to hardened security groups. The enforcement engine queries the directory service, evaluates the returned configuration against the expected posture, and records a verdict compliant, non-compliant, or not-applicable.

4.5 Continuous Compliance Analytics and Audit Interaction

The final subsystem surfaces pipeline state to human analysts through a compliance dashboard and an audit interaction interface. The dashboard is implemented as a lightweight web application layered on an embedded analytical database, computing key performance indicators: overall compliance posture score, counts of total and per-verdict controls, risk exposure segmented by regulatory authority and technical domain, and temporal trends as new documents enter the pipeline. New regulatory artefacts trigger a cascade of automated processes spanning obligation extraction, control mapping, enforcement evaluation, and analytic surfacing. Deliberate hooks for human oversight are preserved at points where judgement or exception handling is required, reflecting a design philosophy that treats automation and analyst intervention as complementary rather than competing concerns.

V. EVALUATION AND RESULTS

A development virtual machine with 16 vCPUs, 64 GB RAM and SSD-backed storage was used to evaluate the prototype replicating a realistic single-organisation on-premise deployment scenario. A small number of anonymised partner-contract samples (relating to digital payments), regulatory documents from RBI and NPCI, and excerpts from PCI DSS and ISO/IEC 27001 made up the test corpus. Both functional and non-functional requirements were assessed, and obligation extraction and control-mapping components were tested through ingestion, parsing, control mapping, and dashboarding.

5.1 Functional Testing

A suite of regression and scenario-based tests was developed to validate the functional behaviour of each subsystem, including document ingestion, obligation extraction, event publication, control mapping, enforcement and dashboard rendering. Test cases covered success paths as well as failure modes such as malformed documents, transient message-broker outages and inconsistent enforcement responses. Security-oriented tests verified that the dashboard rejected unauthenticated requests with appropriate status codes and that secrets such as API keys were not exposed in logs during full end-to-end runs. Resilience tests confirmed that restarting the message broker did not result in event loss, demonstrating the effectiveness of persistent queuing and idempotent consumers.

5.2 Performance Evaluation

Performance was measured against the non-functional targets defined in the requirements specification. Table 2 of the project report shows that all key metrics met or exceeded their thresholds on the evaluation hardware's-time constraint.

Table 1: Performance Summary

Metric	Target	Observed (Mean)
Ingest-to-extract latency	≤ 10 s	6.2 s
Extraction throughput	≥40 pages/min	53 pages/min
Queue propagation delay	≤ 3 s	0.6 s
Control test cycle duration	≤ 15 s	9.8 s
Dashboard query response time	≤ 5 s	0.9 s

5.3 AI Output Quality Assessment

To assess the quality of AI-generated obligations, a gold set of 25 obligations was constructed from the regulatory corpus and independently rated by three internal auditors. Each extracted obligation was evaluated on a five-axis rubric covering relevance, atomicity, verbatim faithfulness to the source clause, domain accuracy and actor identifiability, with scores assigned on a five-point Likert scale. Inter-rater reliability, measured using Krippendorff's alpha, reached 0.78, indicating substantial agreement among auditors. Mean scores were high across most dimensions reflecting the system's conservative behaviour of declining to hallucinate actors when the source text was ambiguous. This conservative stance is preferable in an audit context, even at the cost of slightly lower perceived completeness.

Table 2: AI Output Quality Table

Dimension	Mean score (out of 5)
Relevance	4.6
Atomicity	4.4
Verbatim faithfulness	4.8
Domain accuracy	4.5
Actor identifiability	4.1

5.4 Compliance Posture and Comparative Analysis

To characterise the system's impact on compliance operations, a 30-control deployment was used to compute a representative posture snapshot. The resulting compliance posture score was 86%, with 26 controls marked compliant and four non-compliant, and a risk exposure score of 11 on the project's internal scale; week-over-week trends showed improving posture and decreasing risk as additional obligations were onboarded. The mean time to evidence the time from triggering a control test to obtaining stored evidence was 9.8 seconds, significantly shorter than the multi-day manual cycles observed in baseline processes.

Table 3: Representative compliance-posture snapshot for 30-control deployment.

KPI	Value	Trend
Compliance posture score	86%	+3%
Total controls	30	–
Compliant controls	26	+1
Non-compliant controls	4	−1
Risk exposure score	11	−2
Critical findings	1	–
Mean time to evidence (s)	9.8	−1.4

Table 4 presents a comparative analysis positioned the system against three baseline approaches: manual audits, document-AI-only extraction pipelines, and ungrounded LLM or simple RAG setups. Manual audits achieve high accuracy but incur substantial latency and lack continuous posture metrics, while document-AI-only systems stop at extraction without automated control mapping or enforcement. LLM-only and basic RAG configurations can support question answering over regulations but typically do not integrate with enforcement surfaces or embedded analytics.

Table 4: Comparative Analysis of Methods

Capability	Manual Audit	Document-AI	Basic RAG	Our Approach
Obligation extraction	High, slow	Yes	Yes	Yes
Control mapping	Manual	No	Partial	Yes
Enforcement integration	Manual	No	No	Yes
Continuous posture KPIs	Spreadsheets	No	No	Yes
On-prem deployable	N/A	Sometimes	Rare	Yes
Hybrid privacy (SLM+LLM)	N/A	No	No	Yes

VI. OBSERVATIONS

The evaluation suggests that an AI-first, agentic architecture can substantially compress the obligation-to-evidence cycle for fintech GRC, turning what is traditionally a multi-day manual workflow into a sub-minute automated pipeline for representative regulatory inputs. This shift enables compliance teams to move from periodic, spreadsheet-driven exercises to near-continuous posture management, with KPIs such as posture score, risk exposure and mean time to evidence available as live analytical views rather than ad-hoc reports. The high auditor ratings on relevance, faithfulness and domain accuracy indicate that, under strict schemas and conservative prompting, document AI and RAG components can achieve extraction quality that is acceptable for audit contexts while avoiding hallucinated obligations or actors.

Architecturally, the results also validate a set of design choices that balance capability, privacy and operational simplicity. Routing confidential contracts to local specialized models and public regulations to a higher-capacity LLM–RAG path shows that hybrid privacy strategies are practically viable without sacrificing accuracy on the public corpus. The use of embedded analytics and event-driven orchestration, rather than heavyweight big-data stacks, keeps the platform deployable on modest on-premise infrastructure while still supporting real-time dashboards and audit chat over the same analytical store. Finally, exposing directory APIs, analytical queries and vector search as tools to an agentic layer demonstrates how unstructured audit questions can be consistently mapped to reproducible, evidence-grounded actions, improving auditor productivity over legacy dashboard-only workflows.

VII. LIMITATIONS

Domain-specific evaluation scope.: The empirical study is limited to regulations and standards in the digital-payments and information-security domain, including RBI and NPCI circulars and PCI DSS and ISO/IEC 27001 excerpts. While the architectural pattern is general, applying it to other regimes such as healthcare, telecom or data-protection law would require re-tuning prompts, adjusting obligation and control schemas, and validating performance on new corpora before similar claims could be made.

Absence of long-horizon feedback studies.: The system is designed to incorporate auditor feedback e.g., ratings or corrections on extracted obligations and control mappings but the project window did not allow a multi-quarter study of how such feedback shapes model behaviour and library evolution over time. As a result, there is no empirical evidence yet on the stability of mappings or the risk of drift when the system is exposed to evolving regulations and organisational practices.

Security and hardening gaps.: Several aspects required for a production-grade deployment are intentionally out of scope: hardware-backed key management, mutual TLS between services, zero-trust network segmentation, and formalised incident-response processes for the compliance platform itself. The current implementation should therefore be viewed as a feasibility demonstrator; adopting it in a regulated production environment would necessitate a dedicated hardening phase and an independent security review.

Limited scale and multi-tenancy considerations.: All experiments were conducted in a single-organisation setting on a single virtual machine, without stressing the architecture under multi-tenant or cross-region workloads. Scaling the framework to support a

managed compliance service for multiple fintech clients would require explicit work on tenant isolation, configuration management, and performance guarantees across heterogeneous regulatory portfolios.

VIII. CONCLUSION AND FUTURE SCOPE

The framework presented in this work takes an AI-first, agentic approach to automating the obligation-to-evidence lifecycle in fintech Governance, Risk and Compliance a domain where regulatory latency has historically been measured in days of analyst effort rather than minutes of compute. By composing document AI, a specialised obligation-extraction pipeline, a hybrid dense-sparse control and risk mapping agent, and an embedded analytical substrate, the system processes representative RBI, NPCI and PCI DSS artefacts end-to-end in under a minute, preserving full traceability from dashboard verdicts back to source clauses. Experimental results confirm that the platform meets its non-functional performance targets on modest on-premise hardware, with auditor-rated obligation quality scoring consistently high across document types though it should be noted that these evaluations were conducted against a mocked enforcement surface, and production IAM environments will introduce complexities the prototype does not yet model.

Architecturally, perhaps the most practically significant finding is that a hybrid privacy strategy routing confidential contracts to local specialised models while directing public regulatory documents through an LLM-RAG path need not come at the cost of operational complexity. Event-driven orchestration absorbs the routing logic cleanly, and the embedded analytical store eliminates the need for heavyweight big-data infrastructure that most mid-sized fintech institutions cannot justify. That this design choice also simplifies auditability is a fortunate consequence rather than an explicit goal, yet it reinforces the case for event-driven patterns in high-stakes compliance settings. The audit chat interface, operating over live compliance data rather than static reports, represents a concrete step toward explainable, tool-centric GRC tooling though its utility in adversarial audit scenarios remains to be evaluated empirically.

Several directions merit attention, though they vary considerably in both feasibility and expected payoff. The most immediate concerns are on the AI side: fine-tuning compact language models on contractual and payments-domain corpora would sharpen extraction nuance for obligation types that currently straddle definitional and prescriptive language, and incorporating long-horizon auditor feedback into prompt and embedding adaptation offers a path toward continuous model improvement without retraining from scratch. Whether such feedback loops remain stable under regulatory drift new circulars that redefine existing obligation categories is an open question the current architecture does not resolve.

From a systems perspective, native connectors to payment gateways, transaction logs, and security monitoring platforms are a prerequisite for any production deployment; the mocked directory service, while analytically convenient, defers precisely the integration complexity that determines real-world viability. Alongside this, security hardening through hardware-backed key management and zero-trust service-to-service communication warrants early attention rather than treatment as a post-prototype concern. A regulatory knowledge graph linking obligations, controls, risks, and historical incidents into a queryable structure would meaningfully extend the system's capacity for impact analysis when a single regulatory change propagates across multiple control families. Longer term, evolving the architecture toward a multi-tenant, cloud-native deployment model would make managed compliance services viable for institutions operating under diverse and overlapping regulatory portfolios, though this transition introduces data residency and isolation requirements that would substantially reshape the current design.

REFERENCES

[1] Beltagy, I., M. E. Peters, and A. Cohan. “Longformer: The Long-Document Transformer.” arXiv preprint arXiv:2004.05150, 2020.

[2] Chalkidis, I., I. Androutsopoulos, and N. Aletras. “Legal-BERT: The Muppets Straight Out of Law School.” In Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 2898–2904, 2020.

[3] Chalkidis, I., T. Fergadiotis, P. Malakasiotis, N. Aletras, and I. Androutsopoulos. “LexGLUE: A Benchmark Dataset for Legal Language Understanding in English.” In Proceedings of the

60th Annual Meeting of the Association for Computational Linguistics, pp. 4310–4330, 2022.

[4] Huang, Y., Y. Xu, C. Cui, T. Lv, J. Liu, L. Cui, Z. Tan, F. Zhang, and F. Wei. "LayoutLMv3: Pre-Training for Document AI with Unified Text and Image Masking." arXiv preprint arXiv:2204.08387, 2022.

[5] Karpukhin, V., B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-T. Yih. "Dense Passage Retrieval for Open-Domain Question Answering." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 6769–6781, 2020.

[6] Kim, G., T. Kim, S. Yun, H. Kim, S. Park, and J. Kim. "Donut: Document Understanding Transformer without OCR." arXiv preprint arXiv:2111.15664, 2021.

[7] Kumar, A., P. Sharma, and R. Gupta. "Workflow Management with Apache Airflow: A Case Study." In Proceedings of the International Conference on Advances in Computing, pp. 112–119, 2024.

[8] Lewis, P., E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." In Advances in Neural Information Processing Systems, pp. 9459–9474, 2020.

[9] Raasveldt, M., and H. Mühleisen. "DuckDB: An Embeddable Analytical Database." In Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data, pp. 1981–1984, 2019.

[10] Robertson, S. E., and H. Zaragoza. "The Probabilistic Relevance Framework: BM25 and Beyond." Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.

[11] Schick, T., S. Thakoor, T. Schuster, R. Raileanu, et al. "Toolformer: Language Models Can Teach Themselves to Use Tools." In Advances in Neural Information Processing Systems, 2023.

[12] Smith, R. "An Overview of the Tesseract OCR Engine." In Proceedings of the 9th International Conference on Document Analysis and Recognition, pp. 629–633, 2007.

[13] Tito, R., H. Nguyen, T. Bui, B. Ionescu, and J. Lladós. "Hi-VT5: Hierarchical Visual Text Transformer for Multi-Page DocVQA." arXiv preprint arXiv:2212.05935, 2022.

[14] Wang, X., J. Wei, D. Schuurmans, Q. Le, E. Chi, and D. Zhou. "Self-Consistency Improves Chain of Thought Reasoning in Language Models." In Proceedings of the International Conference on Learning Representations, 2023.

[15] Yao, S., J. Yu, J. Zhao, I. Shafran, T. Griffiths, Y. Cao, and K. Narasimhan. "ReAct: Synergizing Reasoning and Acting in Language Models." In Proceedings of the International Conference on Learning Representations, 2023.