

Table 1: Comparative Analysis of Existing Systems and the Proposed Knowledge Centric Framework

Category	Existing Systems	Core Limitation	Evaluation Parameters	Proposed Framework
Knowledge Graph + LLM Framework	Li et al. [28] (TKDE 2026)	Global planning limited to static KG snapshots	Accuracy, reasoning depth, KG coverage	Dynamic institutional KG updates
Semantic Query Answering	Atzeni et al. [4] (2026)	Weak on OCR derived text	Semantic precision, F1	Retrieval over heterogeneous corpora
Information Retrieval via KG Embeddings	Rawat et al. [18] (2026)	No multi turn context	Retrieval accuracy, MRR	Context aware dialogue retrieval
RAG based Knowledge Discovery	Zhao et al. [20] (ACM TKDD 2026)	Survey oriented; lacks system level integration for real time conversational settings	Retrieval accuracy, coverage, faithfulness	Unified RAG framework with integrated validation and institutional grounding
Generative Graph Analytics	Shang & Huang [26] (TKDE 2025)	No document ingestion pipeline	Query accuracy, latency	OCR integrated graph pipeline
Dual Process Reasoning	Liu et al. [27] (TKDE 2025)	Domain specific (math)	Reasoning accuracy	Policy aware reasoning
Graph RAG for Dataset Discovery	Diamantini et al. [16] (2025)	Limited to datasets	Explainability, precision	Dialogue level explainability
KG Extended RAG for QA	Linders & Tomczak [11] (2025)	Requires structured input	Precision, Recall	OCR enabled KG construction
Multi hop Medical QA	Wang et al. [23] (2025)	Domain restricted	Multi hop F1	Institutional QA adaptation

Table 2: Composition and Characteristics of the Heterogeneous Institutional Document Corpus Used for Experimental Evaluation

Document Type	Description	Data Format	Processing Method	Number of Documents
Structured Documents	Academic calendars, course regulations, timetables, tabular records	Digital (Text / Tables)	Direct text parsing	300
Unstructured Documents	Policy documents, guidelines, notices, announcements	Text / PDF	Text extraction and normalization	600
Semi Structured Documents	Forms, circulars, procedural documents	PDF / Mixed layout	Layout aware parsing	200
Scanned / Image Based Documents	Printed notices and legacy records	Images / Scanned PDFs	OCR (Tesseract + preprocessing)	100

Table 3: Comparative Analysis of Baseline and Proposed Knowledge Centric Conversational Frameworks Across Retrieval, Grounding, and Reasoning Dimensions

System / Method	Source & Venue	Retrieval Strategy	Grounding	Reasoning
LLM Only (GPT-3.5)	OpenAI, 2023	Parametric inference	None	None
Naive RAG	NeurIPS 2020	Sparse dense retrieval	Document level	None
Self RAG	ICLR 2024	On-demand adaptive retrieval	Chunk level	Self-critique tokens
RAG Meeting LLMs	Fan et al. [32]; KDD 2024	Modular retrieval pipeline	Chunk level	None
FoRAG	Cai et al. [33]; KDD 2024	Factuality-constrained retrieval	Chunk level	Factuality scoring
KG + LLM Roadmap	Pan et al. [21]; IEEE TKDE 2024	KG-augmented dense retrieval	Triple level	Symbolic KG inference
Zero-Shot Link Prediction	Li et al. [25]; ICDM 2024	Zero-shot KG traversal	Triple level	Inductive KG reasoning
Generative Graph Analytics	Shang & Huang [26]; IEEE TKDE 2025	Graph-driven query retrieval	Entity level	Graph traversal
Dual-Process Reasoning	Liu et al. [27]; IEEE TKDE 2025	Embedding-based retrieval	Chunk level	Dual-process inference
Link Prediction across KGs	Cui et al. [29]; IEEE TKDE 2025	Cross-graph embedding retrieval	Triple level	Transfer-fusion reasoning
KG-Extended RAG	Linders & Tomczak [11]; Appl. Intell. 2025	KG-extended semantic retrieval	KG + chunk level	KG-guided inference
Retrieval & Structuring RAG	Jiang et al. [30]; KDD 2025	Structure-augmented retrieval	Structured chunk level	Structuring-guided reasoning

Table 4: Comprehensive OCR and Text Extraction Performance Evaluation Across Document Types

Document Type	Avg. Pages	OCR Accuracy (%)	Character Error Rate	Word Error Rate	Text Completeness	Noise Sensitivity	Processing Time (ms/page)
Scanned PDFs	12	93.4	4.8	6.6	High	Moderate	410
Image Based Notices	4	91.2	6.9	8.8	Moderate to High	High	390
Printed Circulars	8	94.1	4.2	5.9	High	Low	370
Handwritten Forms	2	86.7	11.5	14.2	Moderate	Very High	460
Low Resolution Scans	10	88.9	9.3	11.1	Moderate	Very High	445

Table 5: Effect of Knowledge Representation Strategies on Retrieval Readiness

Representation Strategy	Chunk Size	Metadata Used	Embedding Dim	Index Size (MB)	Recall	Precision	Context Relevance
Document Level Storage	1200	No	768	180	0.58	0.61	0.55
Chunk Level Storage	300	No	768	2452	0.69	0.69	0.69
Chunk + Structural Metadata	300	Yes	768	270	0.72	0.72	0.71
Chunk + Temporal Metadata	300	Yes	768	278	0.73	0.73	0.72
Chunk + Full Metadata	300	Yes	768	290	0.75	0.75	0.73

Table 6: Retrieval Strategy Comparison and Effectiveness Evaluation

Retrieval Strategy	Top k	Precision@k	Recall@k	MRR	NDCG	F1 score	Avg. Time (ms)
Keyword Based	5	0.60	0.55	0.49	0.52	0.57	35
Dense Retrieval	5	0.71	0.69	0.63	0.66	0.70	85
Sparse + Dense (Union)	5	0.78	0.76	0.69	0.71	0.77	110
Hybrid (Rank Fusion)	5	0.83	0.81	0.75	0.78	0.82	125

Table 7: Evaluation of Response Generation Quality under Different Configurations

Generation Strategy	Grounding Source	Accuracy (%)	Context Relevance	Faithfulness	Hallucination Rate (%)	Avg. Tokens Generated
LLM Only	None	68.2	0.58	0.55	22.5	180
Dense Retrieval + LLM	Retrieved Chunks	78.4	0.71	0.68	10.9	195
RAG without Validation	Retrieved Chunks	82.9	0.75	0.73	7.6	205
Proposed RAG Pipeline	Validated Context	89.6	0.82	0.81	4.1	210

Table 8: Effect of Reasoning and Validation on System Performance

Validation Strategy	Policy Rules	Logical Checks	Neuro Symbolic Reasoning	Accuracy (%)	Hallucination Rate (%)	Response Rejection (%)
No Validation	No	No	No	82.9	7.6	0
Rule Based Only	Yes	Yes	No	86.7	5.3	6.2
Neuro Symbolic Only	No	Yes	Yes	87.9	4.8	7.1
Rule + Neuro Symbolic	Yes	Yes	Yes	89.6	4.1	8.5

Table 9: System Efficiency and End to End Performance Evaluation

System Configuration	Avg. Latency (ms)	Retrieval Time (ms)	Generation Time (ms)	Validation Time (ms)	Memory Usage (MB)
Keyword Based System	180	35	120	0	620
Dense Retrieval Only	260	85	150	0	740
RAG without Validation	290	125	150	0	810
Proposed System	310	1251	150	35	860

Table 10: Overall Performance Comparison of Baseline Systems and Proposed Framework

System	Accuracy (%)	Precision	Recall	F1	Hallucination Rate (%)	Latency (ms)
LLM Only	68.2	0.65	0.62	0.63	22.5	420
Naive RAG [6]	72.1*	0.69*	0.67*	0.68*	11.3*	250*
Self RAG [32]	79.4*	0.76*	0.74*	0.75*	8.2*	340*
CRAG [33]	82.1*	0.79*	0.77*	0.78*	6.7*	310*
GraphRAG [16]	80.6*	0.77*	0.75*	0.76*	7.9*	485*
KG+LLM [28]	76.3*	0.73*	0.71*	0.72*	9.4*	390*
Hybrid RAG (no reasoning) [30]	85.6	0.83	0.81	0.82	6.2	300
Proposed OCR Knowledge Centric System	89.6	0.86	0.84	0.85	4.1	310

Table 11: Detailed class level performance analysis across document categories, query types, and pipeline modules based on Accuracy, Precision, Recall, and F1 Score

Category / Class	Documents	Queries	Accuracy (%)	Precision	Recall	F1 Score	Hallucination Rate (%)
Module A — Knowledge Acquisition (OCR & Document Processing)							
Scanned PDFs	120	240	91.3	0.89	0.87	0.88	5.1
Image based notices	80	160	88.7	0.86	0.84	0.85	6.4
Printed circulars	150	300	93.1	0.91	0.90	0.90	4.2
Handwritten forms	40	80	83.4	0.81	0.79	0.80	9.7
Low resolution scans	60	120	85.9	0.83	0.81	0.82	8.3
Module B — Query Type Performance (Online Pipeline)							
Factual queries	—	80	92.5	0.91	0.89	0.90	3.2
Procedural queries	—	75	88.0	0.86	0.84	0.85	5.3
Policy related queries	—	45	87.1	0.85	0.83	0.84	5.8
Multi hop queries	—	—	81.3	0.79	0.77	0.78	8.9
Deadline / schedule queries	—	—	90.4	0.88	0.87	0.87	4.6
Module C — Document Domain / Administrative Category							
Academic regulations	200	—	91.2	0.89	0.88	0.88	4.0
Examination schedules	150	—	93.7	0.92	0.91	0.91	3.5
Administrative notices	180	—	88.4	0.86	0.84	0.85	5.9
Fee & financial policies	120	—	87.6	0.85	0.83	0.84	6.1
Hostel & campus circulars	90	—	86.9	0.84	0.82	0.83	6.8
Course & syllabus documents	160	—	90.1	0.88	0.86	0.87	4.7
Grievance & support records	100	—	85.3	0.83	0.81	0.82	7.4

Table 12: Statistical Significance Test Results for the Proposed System vs. Baseline Systems

Comparison (Proposed vs.)	Metric	t statistic	p value	Effect Size (Cohen's d) / Significance
Self RAG [32]	Accuracy (%)	4.82	<0.001	1.31 (large) — Significant
Self RAG [32]	F1 Score	5.17	<0.001	1.42 (large) — Significant
CRAG [33]	Accuracy (%)	3.64	0.003	0.98 (large) — Significant
CRAG [33]	F1 Score	3.91	0.002	1.05 (large) — Significant
GraphRAG [16]	Accuracy (%)	4.11	0.001	1.12 (large) — Significant
Naive RAG [6]	Hallucination Rate (%)	6.23	<0.001	1.68 (large) — Significant
McNemar's Test (vs. CRAG [33])	Response Correctness	chi2 = 8.94	0.003	Significant

Table 13: Impact of System Components on Accuracy and Hallucination Rate

System Configuration	Active Components	Accuracy (%)	Hallucination Rate (%)
Proposed System	OCR, Hybrid Retrieval, RAG, Reasoning	89.6	4.1
Without Reasoning	OCR, Hybrid Retrieval, RAG	85.6	6.2
Without RAG	OCR, Hybrid Retrieval, LLM	78.4	10.9
Dense Retrieval Only	OCR, Dense Retrieval, LLM	82.1	7.8
Keyword Retrieval Only	OCR, Keyword Retrieval, LLM	74.3	12.6
Without OCR	Text Corpus, Hybrid Retrieval, RAG, Reasoning	80.7	8.9
LLM Baseline	LLM Only	68.2	22.5