Table 1: Comparitive analysis of Multilingual Medical Knowledge Graph System

Category	Methodology	Strengths	Limitations	LLM-KGMQA Advancements
Traditional Medical Systems	Rule-based, Database-driven	Factually reliable	Static, no reasoning	Dynamic LLM- guided KG reasoning
ML-Based Medical QA	ML models (SVM, CNN, RNN)	Better prediction	Manual features, low interpretability	Automated KG-driven reasoning
Knowledge Graph–Based Systems	Graph traversal, template-based QA	Interpretable relations	Weak NL understanding	LLM-assisted multi-hop reasoning
LLM-Based Medical QA	LLMs (Bio BERT, Clinical LLMs)	Strong language ability	Hallucinations	KG grounding for reliability
Multilingual Medical Systems	Translation + NLP pipelines	Supports multiple languages	Mapping inconsistencies	Concept-level normalization
User-Centric Medical QA	NLP chat systems	Easy user interaction	Limited accuracy	Clinically reliable, scalable QA

Table2: General Language performance metrics for LLM-KGMQA and baseline models across multilingual clinical response generation settings.

Metric	LLM-KGMQA	LLM-only Medical QA	Bio BERT-QA	Clinical BERT	Rule-Based KGQA
Perplexity (PPL	3.1	3.9	3.6	3.5	4.6
Fluency& Coherence	0.95	0.90	0.88	0.89	0.72
BLEU Score	0.86	0.82	0.82	0.81	0.68
ROUGE-L Score	0.90	0.86	0.84	0.85	0.71
METEOR Score	0.93	0.89	0.87	0.88	0.74
BERTScore	0.96	0.93	0.91	0.92	0.78
Overall Accuracy (%)	98.0	92.5	90.8	91.6	85.2

Table 3 : Medical answer evaluation of LLM-KGMQA using EM, F1, nDCG, Accuracy, and Latency.

Metric	**LLM-KGMQA**	**LLM-only Medical QA**	**Bio BERT-QA**	**Clinical BERT**	**Rule-Based KGQA**
Exact Match (EM)	91.5	85.2	86.8	85.9	78.4
F1 Score	0.94	0.90	0.91	0.90	0.82
nDCG	0.95	0.92	0.93	0.92	0.80
Accuracy	98.0	92.6	93.9	93.1	85.7
Inference Latency (s)	0.95	1.3	1.1	1.2	0.5

Table 4 :Shows how enabling key modules progressively improves accuracy and reasoning performance in LLM-KGMQA

Configuration	**EM (%)**	**F1 Score**	**nDCG**	**Accuracy (%)**
LLM-only Medical QA	85.2	0.90	0.92	92.5
LLM+ Medical KG (no fast-linking)	87.1	0.91	0.93	93.6
LLM + KG + Fast-Linking	88.6	0.92	0.95	95.4
LLM-KGMQA (Full System)	91.2	0.95	0.97	98.0

Table 5 : Comparison of Medical Question Answering Systems Across Diverse Quantitative Evaluation Performance Metrics.

Metric	**LLM-KGMQA**	**LLM-only Medical QA**	**Bio BERT-QA**	**Clinical BERT**	**Rule-Based KGQA**
Factual Consistency	97.0	93.6	95.1	93.4	90.2
Hallucination Rate	1.3	2.4	1.9	2.1	3.2
Risk Sensitivity Score	0.96	0.91	0.93	0.92	0.86
Explainability Score	0.94	0.88	0.91	0.89	0.80
Truthfulness Score	0.97	0.92	0.94	0.93	0.88

Table 6 : Benchmark Accuracy Comparison of LLM-KGMQA Across Diverse Medical QA Tasks and Clinical Reasoning Scenarios

Dataset	**Task**	**LLM-KGMQA**	**LLM-only Medical QA**	**BioBERT-QA**	**ClinicalBERT**
ICD-11	Disease classification	94.8	90.6	92.1	91.2
SNOMED-CT	Clinical terminology mapping	93.6	89.9	91.4	90.6
Medical KG	Relation-based reasoning	92.9	88.3	90.7	89.6
Multilingual Medical Text	Language-based QA	91.7	88.6	90.1	89.0
Structured Medical Corpus	Grounded medical QA	90.9	87.1	88.8	87.6

Table 7: Comparision of medical QA models across benchmark datasets, highlighting accuracy trends and reasoning effectiveness **.**

Dataset	**LLM-KGMQA (%)**	**LLM-only Medical-QA (%)**	**BioBERT-QA (%)**	**p-value vs LLM-only**	**p-value vs. BioBERT-QA**
ICD-11	98.3 ± 0.8	90.2 ± 1.2	92.0 ± 1.1	< 0.001	0.008
SNOMED-CT	97.9 ± 0.9	89.6 ± 1.3	91.1 ± 1.0	< 0.001	0.0012
Medical KG	98.1± 0.7	89.8 ± 1.1	91.3 ± 1.2	< 0.001	0.009
Multilingual Medical Text	97.8 ± 1.0	89.8 ± 1.1	91.3 ± 1.2	< 0.001	0.007
Structured Medical Corpus	98.0 ± 0.8	89.7 ± 1.3	90.6 ± 1.2	< 0.001	0.0010
Mean	98.0 ± 0.8	89.9 ± 1.2	91.4 ± 1.1	< 0.0001	0.0004

Table 8 : Ablation Study Results of LLM-KGMQA Components Across Accuracy, Reasoning, and Efficiency Metrics

Configuration	**Medical Entity Linking Accuracy (%)**	**Intent Understanding Accuracy (%)**	**Task Completion Rate (%)**	**Avg. Latency(ms)**
LLM-only Medical QA	90.2	84.1	41.2	880
LLM + Medical KG (No Fast-Linking)	93.6	88.5	79.4	720
LLM + KG + Fast-Linking	96.8	91.9	85.6	820
LLM-KGMQA (Full System)	98.0	94.3	88.0	960