

Mina: A Multilingual LLM-Powered Legal Assistant Agent for Empowering Access to Justice in Bangladesh

Azmine Toushik Wasi, Wahid Faisal, Mst Rafia Islam, Md Rizwan Parvez



Motivation

A JUSTICE SYSTEM AT BREAKING POINT

Bangladesh has a backlog of cases its courts cannot absorb, and a legal market its citizens cannot afford.

3.7M⁺

Pending cases across Bangladeshi courts as of 2024, up to 4.4 million by some estimates.

1.90k

Judges per citizen. ~2,100 judges serve the entire country.

10–60_{ys}

Typical case-resolution timeline. Lawyer fees are unregulated; legal aid is underfunded.

Motivation

WHY EXISTING LEGAL NLP FALLS SHORT

Bangladesh is a low-resource, common-law, bilingual jurisdiction. Generic legal assistants miss on all three.

LANGUAGE

Bengali has scarce annotated data and limited digital content. English-centric models struggle with its morphology and script.

JURISDICTION

Common-law statutes inherited from British rule, with Farsi-influenced legal terminology that generic models cannot interpret.

ACCESS

The population most in need has limited digital literacy, poor connectivity, and no affordable formal legal channel.

Method

SYSTEM ARCHITECTURE

An orchestrator routes every query; specialists handle retrieval and tools.

01 · DECIDE

Orchestrator Agent

Inspects the query, chat history, and uploaded documents.
Decides whether internal context suffices or RAG is required.

- direct answer
- chat-context lookup
- delegate to RAG

02 · RETRIEVE

RAG Agent

Owens end-to-end retrieval: keyword generation, two-stage vector lookup, and quality assessment against the query.

- act-level recall
- section-level grounding
- relevance re-check

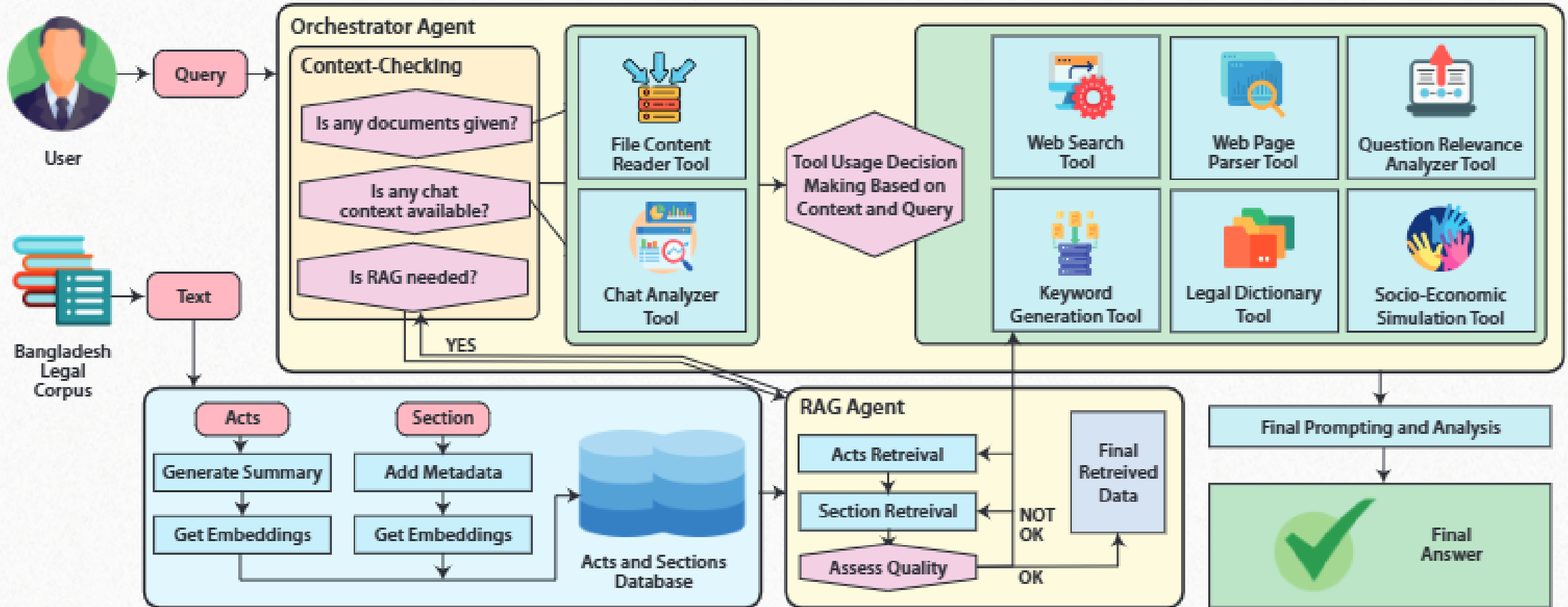
03 · AUGMENT

Tool Suite

Specialized utilities invoked only when needed, kept narrow to preserve latency and legal integrity.

- legal dictionary
- web search · parser
- document reader

Method



Method

TWO-STAGE RAG

Retrieve acts first, then sections, to keep unrelated statutes from being conflated.

STAGE 01 / RECALL

Act-level retrieval

Cohere multilingual keywords search summarized representations of every Bangladeshi act. Top n=5 acts pass through.



STAGE 02 / GROUND

Section-level grounding

Section chunks are filtered to the retained acts, then re-ranked. Top n=10 sections feed final reasoning.



STAGE 03 / REASON

Orchestrator answer

Grounded sections, dictionary lookups, and optional tool outputs are composed into a context-aware legal draft.

Method

AGENTS AND TOOLS

A small, specialized toolkit invoked only when the orchestrator demands it.

RETRIEVAL

Keyword Generator

LLM produces 5–10 semantic search terms; regex fallback for low-context queries.

RETRIEVAL

Relevance Analyzer

Embedding-based detection of semantic continuity across multi-turn legal consultations.

EXTERNAL

Web Search · Page Parser

DuckDuckGo retrieval; BeautifulSoup extracts up to 5,000 chars of clean page content.

LINGUISTIC

Bengali Legal Dictionary

Custom resource for colonial-era and Farsi-influenced terms, enables plain-language explanation.

USER INPUT

File Content Reader

Parses .pdf, .docx, .pptx uploads into structured text for downstream reasoning.

SIMULATION

Socio-Economic Simulator

Explores how demographic variables interact with legal access, for diagnostics and policy planning.

Evaluation

EVALUATION SETUP

We benchmark MINA against the actual Bangladesh Bar Council Exam, all three stages, both years.

STAGE 01

Preliminary MCQ

Optical-mark replicated. Each model $\times$ setup $\times$ year repeated 5 runs, averaged.

PASS THRESHOLD

$\geq 50 / 100$

STAGE 02

Written Exam

13 questions across 6 legal domains.
 ≥ 2 law-faculty evaluators per answer; composite holistic scoring.

INTER-ANNOTATOR K

0.827 · Cohen

STAGE 03

Viva Voce

Conducted as live chat dialogue. 5 evaluators per session; coherence, accuracy, legal reasoning.

EVALUATORS

law faculty, anonymized

Results

MCQ, PRELIMINARY STAGE

Two-stage RAG plus tools is what clears the human pass bar.

Model	Year: 2022				Year: 2023			
	W/o RAG	Naive RAG	2-Step RAG	Tools	W/o RAG	Naive RAG	2-Step RAG	Tools
<i>Random Choice</i>	25.00	25.20	25.40	25.20	25.60	25.40	25.20	25.00
<i>Proprietary Models</i>								
GPT-4o	18.60	62.40	69.20	73.60	19.20	58.80	67.80	72.20
Gemini-2.0-Flash	12.40	61.20	68.60	69.20	12.20	59.40	69.00	70.20
Gemini-2.5-Flash	<u>30.20</u>	<u>68.80</u>	<u>75.60</u>	<u>77.00</u>	<u>32.40</u>	<u>69.20</u>	<u>76.40</u>	<u>77.00</u>
<i>Open-Source Models: Small</i>								
Llama3.2-1B-Instruct	6.20	6.00	6.40	8.20	7.00	7.20	7.40	9.20
Llama3.2-3B-Instruct	9.20	9.00	11.20	13.40	11.20	11.40	14.00	15.20
Gemma-3-4B-it	12.40	15.20	20.60	23.20	14.40	16.60	22.20	26.40
Qwen3-4B-Instruct-2507	<u>14.20</u>	<u>28.40</u>	44.60	<u>49.80</u>	<u>16.20</u>	<u>29.40</u>	45.40	<u>52.40</u>
Command-A-8B	8.20	25.20	<u>47.00</u>	47.40	11.20	23.40	<u>49.20</u>	50.20
Llama3.1-8B-Instruct	15.20	18.40	30.20	32.40	16.20	18.60	33.40	35.20
<i>Open-Source Models: Large</i>								
Gemma-3-12B-it	20.40	35.20	48.40	50.20	22.60	36.20	52.40	54.20
Gemma-3-27B-it	28.20	45.40	60.20	64.40	30.20	46.60	62.40	66.20
Qwen3-30B-A3B-Instruct-2507	<u>34.20</u>	<u>50.40</u>	<u>65.60</u>	<u>70.80</u>	<u>36.20</u>	<u>52.40</u>	<u>67.20</u>	<u>72.40</u>
Llama3.1-70B-Instruct	24.20	30.40	40.20	42.40	26.20	32.20	44.40	46.20

2022 COHORT · 5-RUN
AVERAGES HUMAN PASS
RATE 2023: 17.96%

Results

WRITTEN EXAM, SECOND STAGE

Model	Year: 2022				Year: 2023			
	W/o RAG	Naive RAG	2-Step RAG	Tools	W/o RAG	Naive RAG	2-Step RAG	Tools
<i>Proprietary Models</i>								
GPT-4o	19.20	55.40	69.80	71.20	21.20	60.80	72.60	75.60
Gemini-2.0-Flash	18.40	58.20	68.60	70.00	18.20	65.40	74.80	75.20
Gemini-2.5-Flash	<u>35.20</u>	<u>70.40</u>	<u>78.60</u>	<u>81.00</u>	<u>36.20</u>	<u>71.20</u>	<u>79.40</u>	<u>81.80</u>
<i>Open-Source Models: Small</i>								
Llama3.2-1B-Instruct	6.20	6.00	6.20	7.00	7.00	7.00	7.20	8.00
Llama3.2-3B-Instruct	10.20	11.00	14.20	16.40	12.00	14.20	17.40	20.00
Gemma-3-4B-it	15.20	20.20	24.40	28.00	18.20	22.40	26.00	30.20
Qwen3-4B-Instruct-2507	22.20	56.40	<u>71.60</u>	68.00	26.20	60.40	64.20	70.40
Command-A-8B	25.00	<u>60.20</u>	71.00	<u>74.40</u>	25.20	<u>73.00</u>	<u>74.20</u>	<u>76.00</u>
Llama3.1-8B-Instruct	<u>28.20</u>	34.40	46.20	50.00	<u>30.20</u>	66.00	68.20	71.20
<i>Open-Source Models: Large</i>								
Gemma-3-12B-it	30.00	42.20	55.40	58.00	32.20	44.40	58.00	60.20
Gemma-3-27B-it	38.20	55.00	68.20	72.40	40.00	56.20	70.00	74.20
Qwen3-30B-A3B-Instruct-2507	42.20	60.40	74.00	78.20	44.00	62.20	76.40	79.40
Llama3.1-70B-Instruct	<u>45.00</u>	<u>62.20</u>	<u>75.00</u>	<u>79.80</u>	<u>47.20</u>	<u>65.00</u>	<u>77.20</u>	<u>80.20</u>

2023 COHORT · W/ FULL
TOOLS HUMAN AVG: 40–60%

Results

VIVA VOCE, CHAT-MEDIATED ORAL EXAM

Strong reasoning under conversational pressure, the gap to human ceiling is here.

Model	W/o RAG	N. RAG	2-S RAG	Tools
Proprietary Models				
Gemini-2.0 Flash	32.80	60.80	74.40	76.60
Gemini-2.5 Flash	<u>36.20</u>	<u>70.40</u>	<u>79.20</u>	<u>81.00</u>
Open-Source Models: Small				
Llama3.2-1B-Instruct	6.20	6.40	6.20	7.00
Llama3.2-3B-Instruct	11.20	12.40	15.20	17.40
Gemma-3-4B-it	16.20	50.40	65.20	69.20
Qwen3-4B-Instruct-2507	22.20	<u>56.40</u>	67.20	70.20
Command A	<u>27.80</u>	55.60	<u>70.40</u>	<u>71.20</u>
Llama3.1-8B-Instruct	24.20	52.20	66.40	67.20
Open-Source Models: Large				
Gemma-3-12B-it	31.20	43.20	56.40	59.20
Gemma-3-27B-it	39.20	56.00	69.20	72.40
Qwen3-30B-A3B-Instruct	42.20	61.20	75.20	79.40
Llama3.1-70B-Instruct	<u>46.00</u>	<u>63.20</u>	<u>77.20</u>	<u>80.20</u>

5-EVALUATOR AVERAGES
HUMAN PASS: 96–97%

DEPLOYMENT

COST RELATIVE TO HUMAN CONSULTATION

99.4–99.9% cheaper.

MCQ-STYLE QUERY

~ 0.2 – 0.6 \$

WRITTEN-LENGTH QUERY

~ 0.8 – 2.0 \$

FULL MULTI-TURN SESSION

≤ 10 \$ · ~12 BDT

Reference: Basic Legal Consultation in Bangladesh Runs 2,000 – 10,000 BDT.

Analysis

ERROR PATTERNS

Failure modes cluster around procedure, not language, and around silence, not error.

MCQ	Bengali conjunction conflation	Models flatten "O" (and) against "ba" (or) , losing the disjunctive semantics that govern statutory conditions.
MCQ	Procedural hierarchy mis-ranking	Order 1 Rule 8 versus Rule 13; Res Judicata mis-mapped onto Section 151. Surface lexical cues over structural reasoning.
WRITTEN	Hallucinated citations / missing conditions	Fluent answers that omit procedural prerequisites or cite statutes that do not exist. Arithmetic errors in limitation-period calculations.
VIVA	No clarifying questions	When facts are ambiguous, the model commits to an answer rather than probing. Section misattributions count as major oral faults.



Computational Intelligence
and Operations Laboratory

Thank You!

If you have any questions, please ask!



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute

ACL 2026 Findings