



AskRegula: Evidence-Grounded Regulatory QA with Integrated Benchmarking and Compliance- Oriented Evaluation

Dr. Tuba Gokhan · Prof. Ted Briscoe

LREC 2026 Industry Day
Palma de Mallorca, Spain
May 2026



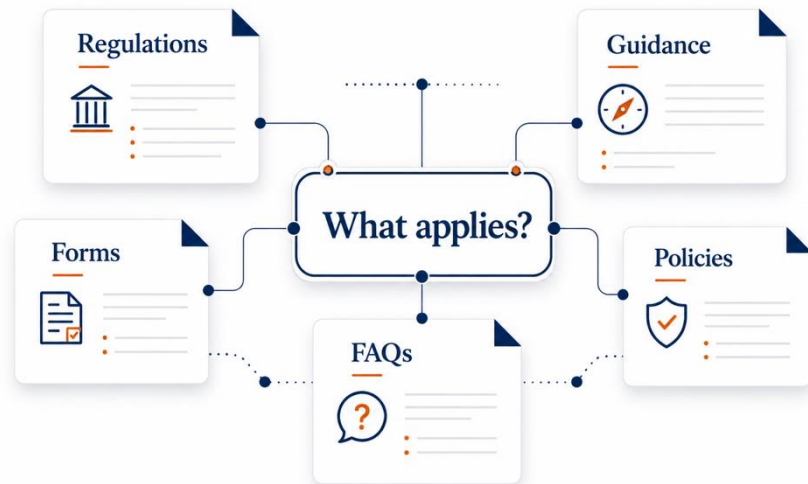
جامعة محمد بن زايد
للذكاء الاصطناعي
MOHAMED BIN ZAYED UNIVERSITY
OF ARTIFICIAL INTELLIGENCE

The Regulatory AI Challenge

Regulation is a dynamic information problem

Regulated sectors face continuously evolving compliance ecosystems:

- large and frequently updated rules
- complex, cross-referenced documents
- high cost of interpretation and operationalisation
- increasing enforcement and non-compliance risk



Why this matters

- **Regulatory change is accelerating:** AI-related regulatory developments doubled in 2025, with **157 major updates** recorded between June 2024 and May 2025.
- **Operational integration is slow:** **74% of institutions** report that integrating one new regulation takes **more than one year**.
- **Compliance is expensive:** compliance costs can reach up to **13% of operating expenses** in financial services.
- **Non-compliance is costly:** global AML/KYC/sanctions penalties reached approximately **\$3.8B in 2025**.
- **Regional enforcement is increasing:** UAE fines exceeded **380M AED** in the first eight months of 2025.

Fluent AI answers are not enough. Regulatory AI must be reliable, traceable, auditable, and evidence-grounded.

Why Regulatory QA is Hard

- Regulatory answers are rarely contained in a single passage.
- They may depend on:
 - linked information across the document structure
 - cross-references between provisions
 - multiple evidence sources combined in one answer
- related elements such as obligations, definitions, conditions, and exceptions

Regulatory QA requires more than retrieval. It requires **grounding**, **composition**, and **traceability**.

User Question (Input) How long must a firm keep records of a suspicious transaction report?

Evidence Graph

Source Passage (Retrieval Node A)

AML Rule 7.4.5 – Suspicious Transaction Reports.

A firm must submit a suspicious transaction report (STR) to the Financial Intelligence Unit and must keep a record of each STR in line with the general record-keeping requirement in Rule 3.2.1.

Explicit reference
(see Rule 3.2.1)

Target Passage (Retrieval Node B)

Rule 3.2.1 – Record-Keeping.

A firm must retain records required under these rules for at least six years from the date the record is made.

Gold Answer (Output)

A firm must keep a record of each suspicious transaction report in accordance with the general record-keeping requirement in Rule 3.2.1 [from Source], meaning the STR record must be retained for at least six years from the date the record is made [from Target].

14. SUSPICIOUS ACTIVITY/TRANSACTION REPORTS
14.1 Application and definitions
In this Chapter "money laundering" and "terrorist financing" means the criminal offences defined in Federal AML Legislation.
14.2 Internal reporting requirements
14.2.1 A Relevant Person must establish and maintain policies, procedures, systems and controls in order to monitor and detect suspicious activity or Transactions in relation to potential money laundering or terrorist financing.
14.2.2 A Relevant Person must have policies, procedures, systems and controls to ensure that whenever any Employee, acting in the ordinary course of his employment, either: (a) knows; (b) suspects; or (c) has reasonable grounds for knowing or suspecting, that a Person is engaged in or attempting money laundering or terrorist financing, that Employee promptly notifies the Relevant Person's MLRO and provides the MLRO with all relevant details.
Fraud and other Financial Crimes
3.3.38 An Authorised Person must establish and maintain effective systems and controls to: (1) deter and prevent suspected fraud and other Financial Crimes against the Authorised Person and its Customers; and (2) report suspected fraud and other Financial Crimes to the relevant authorities.
3.3.39 An Authorised Person must ensure that the systems and controls established and maintained in accordance with Rule 3.3.38: (1) enable it to identify, assess, monitor and manage Financial Crime risks; and (2) are comprehensive and proportionate to the nature, scale and complexity of its activities.
3.3.40 In identifying its exposure to risks arising from fraud and other Financial Crimes and implementing the necessary systems and controls to manage and reduce those risks, an Authorised Person must consider a range of factors, including: (1) its Customer, product and activity profiles; (2) its distribution channels; (3) the complexity and volume of its Transactions;

Question from Client: What steps must we take when a transaction is suspected of involving money laundering?

Response from Authority: When a financial institution suspects a transaction involves money laundering, you must report the suspicious activity to the Money Laundering Reporting Officer (MLRO) as required by AML 14.2.2. The MLRO is responsible for promptly investigating and documenting the circumstances, as outlined in AML 14.3.1. Based on this investigation, the MLRO must determine whether a Suspicious Activity Report (SAR) is necessary and file it with the relevant authorities. If required, as mandated by AML 14.3.1(b-c), if no SAR is filed, the reasons must be documented, as required by AML 14.3.3. The MLRO's decision to file a SAR must be made independently, without requiring approval from others, as outlined in AML 14.3.4.

You must also ensure that your institution is registered on the goAML platform to submit SARs, as required by AML 14.3.5. Additionally, you are required to maintain effective systems and controls for detecting and reporting suspicious activities, as stated in GEN 3.3.38. Failure to comply with these regulations may result in both regulatory and criminal penalties.

14.3 Suspicious Activity/Transaction Reports
14.3.1 A Relevant Person must ensure that where the Relevant Person's MLRO receives an internal notification of suspicious activity under Rule 14.2.2, the MLRO, without delay: (a) investigates and documents the circumstances in relation to which the notification made under Rule 14.2.2 was made; (b) determines whether in accordance with Federal AML Legislation a SAR/STR must be made and documents such determination; and (c) if required, make a SAR/STR as soon as practicable.
14.3.2 The MLRO must, following receipt of an internal notification of suspicious activity under Rule 14.2.2, document: (a) the steps taken to investigate the circumstances in relation to which the internal notification is made; and (b) where no external SAR/STR is made, the reasons why no such report was made.
14.3.3 Where, following a notification to the MLRO of suspicious activity under 14.2.2, no SAR/STR is made, a Relevant Person must record the reasons for not making a SAR/STR.
14.3.4 A Relevant Person must ensure that if the MLRO decides to make a SAR/STR, his decision is made independently and is not subject to the consent or approval of any other Person.
14.3.5 Relevant Persons are required to register on goAML, upon receipt of their Financial Services Permission, Recognition Order or registration licence in order to submit SAR/STRs.



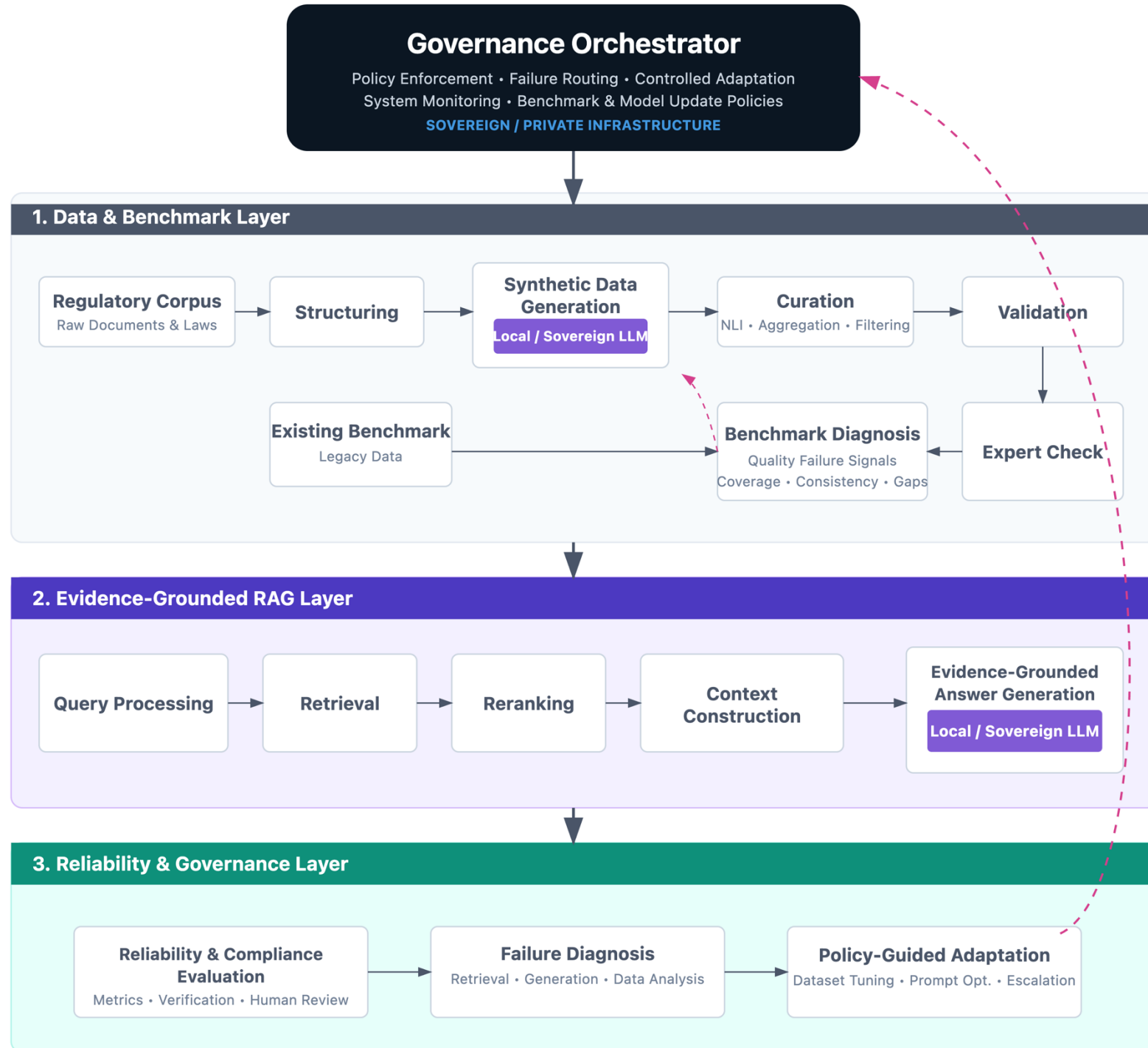
From Manual Review to Governed Regulatory AI

Feature	Manual Review	Generative AI / LLMs	AskRegula
Evidence grounding	Expert source review	May be unsupported	Retrieved regulatory evidence
Traceability	Possible but manual	Often limited	Explicit source linkage
Evaluation	Manual and hard to scale	Generic evaluation	Compliance-oriented metrics
Governance	Process-dependent	Not governance-aware	Controlled feedback loop
Deployment control	Internal workflows	External services often used	Private / sovereign-compatible
Scalability	Limited	High	High with controls



More Than a RAG System

A Governance-Oriented
Regulatory AI Framework



Research Foundations: Building Blocks for AskRegula



AskRegula Layer	Existing Foundation	Role
Dataset Layer	ObliQA, ObliQA-MP, ObliQA-XRef	Benchmark generation for obligation-oriented, multi-passage, and citation-dependent regulatory QA
RAG Layer	RIRAG Shared Task	Community evaluation of regulatory retrieval and answer generation systems
Evaluation Layer	RePASs, RePASs-L	Compliance-oriented evaluation of grounding, obligation coverage, contradiction, and linguistic quality



Potential Applications

- Evidence-grounded QA for financial regulations
- Review of relevant compliance evidence across large document collections
- AI governance and policy compliance support
- Alignment of internal policies with regulatory obligations
- Cross-reference-aware regulatory search
- Benchmarking and evaluation of regulatory AI systems

Expected Impact

- More reliable use of regulatory evidence
- Stronger traceability and auditability of answers
- Systematic evaluation and continuous refinement
- Shift from isolated prompting to governed workflows
- Better alignment with sovereign and private deployment settings
- Reusable foundation across sectors and jurisdictions

Admin Console

ACTIVE RELEASE

AGR-FIN-v1.3

12 days live

REFERENCE DOCUMENTS

Approved: 12

Pending: 2

BENCHMARK

BM-240322-07

Document Set Version: CORP-FIN-2026-03

PIPELINE VERSION

PIPE-RAG-v1.3

Evaluation Version: EV-240318-04

Reference Documents

Benchmark & Generation

Pipeline & Evaluation

Versions & Release

Current Pipeline

Current production pipeline configuration and linked versions used as the reference point.

LINKED VERSIONS

Pipeline Version	PIPE-RAG-v1.3
Document Set	CORP-FIN-2026-03
Method Version	METH-RAG-v1.3
Evaluation Version	EV-240318-04
Active Release	AGR-FIN-v1.3

TECHNICAL DETAILS

LLM Model	GPT-4.1
Embedding	BGE-large
Reranker	Cross-Encoder v2
Retrieval	Hybrid Search - Top-k 12
Answer Mode	Concise Grounded Summary
Traceability	Strict
Citation	Inline Source Citations
Evidence Thresh.	0.82
Confidence Thresh.	0.75
Retrieval Thresh.	0.68

Method Configuration

Configure the model, retrieval, response, and threshold settings for this pipeline. Save with a method version name to track your configuration.

MODEL SETTINGS

Embedding Model

LLM Model

GPT-4.1

BGE-large

Cross-Encoder v2

RETRIEVAL SETTINGS

Retrieval Method

Top-k

Context Selection

Evidence Selection

Hybrid Search

12

Balanced Context Wi

Top Evidence First

RESPONSE SETTINGS

Answer Mode

Traceability Mode

Citation Mode

Concise Grounded Summary

Strict

Inline Source Citations

THRESHOLDS

Evidence Threshold

Confidence Threshold

Retrieval Threshold

0.82

0.75

0.68

CONFIGURATION SUMMARY

LLM Model

Embedding

Reranker

Retrieval

Answer Mode

Traceability

Evidence Threshold

Confidence Threshold

Retrieval Threshold

Method Version Name

GPT-4.1

BGE-large

Cross-Encoder v2

Hybrid Search - Top-k 12

Concise Grounded Summary

Strict

0.82

0.75

0.68

e.g., METH-RAG-v1.4-exp

Run Method

Save Method Configuration

Evaluation Setup

Configure and run evaluation with selected document set and method configuration. Set acceptance criteria and select question set.

CONFIGURATION SELECTION

Document Set

Method Configuration

CORP-FIN-2026-03

Current Pipeline

ACCEPTANCE CRITERIA

Coverage (%)

Evidence Grounding (%)

Traceability (%)

Critical Issues (max)

94

95

97

0

QUESTION SET

Benchmark

Sample Size

Sample Benchmark

100

EVALUATION SUMMARY

Document Set

Method

Coverage Target

Evidence Grounding

Traceability Target

Critical Issues

Question Set

Evaluation Version Name

CORP-FIN-2026-03

Current Pipeline

≥ 94%

≥ 95%

≥ 97%

≤ 0

Sample (100 q)

e.g., EV-2026-Q1-test

Run Evaluation

Save Evaluation Configuration

Expert Check

Configure assignment and trigger expert review for this evaluation.

ASSIGNMENT SETTINGS

Assignment Mode

Sample Type

Expert Count

Double Review Rate

Stratified sample

Percentage

2 experts

20%

Conflict Resolution

Sample Size

Min Sample Size

Max Sample Size

Third expert review

e.g., 10% or 200

e.g., 50

e.g., 1000

Assignment is applied to the selected Benchmark and current method; double-review controls agreement checks.

Assign to Expert

Release Readiness Review

Current vs Proposed comparison before final release decision.

FIELD

CURRENT

PROPOSED

Document Set

CORP-FIN-2026-03

CORP-FIN-2026-03-CAND

Pipeline Version

PIPE-RAG-v1.3

PIPE-RAG-v1.4-exp

Coverage Pass Rate

93.7%

94.8%

Evidence Grounding

93.7%

96.1%

Traceability Pass Rate

97.1%

98.3%

Blocking Issues

0 critical

0 critical - 7 review-required

Recommendation

Stable current pipeline

Proceed to version comparison

New Pipeline Version Name

e.g., PIPE-RAG-v1.4

Save Pipeline Version

compliance-admin@

Prototype Interaction Model

AskRegula

Approved source version AGR-Fin-v1.3
Updated 23/04/2025

Example: Is enhanced due diligence required for onboarding a corporate client with beneficial owners in a high-risk jurisdiction under the current AML rules?

Ask

Using 4 linked sources from the approved financial regulatory scope.

Mark answer as: Needs review Ready

Summary

Based on the active regulatory scope and the documents available in this workspace, enhanced due diligence should be applied before onboarding the client. The combination of a corporate ownership structure, cross-border exposure, and beneficial ownership linked to a high-risk jurisdiction increases the AML risk profile and requires additional verification steps before the relationship can be approved.

Key points for this case

- The AML rulebook requires firms to apply enhanced due diligence where a customer or beneficial owner is connected to a high-risk jurisdiction [S1 | Core regulatory source].
- The client onboarding rules require identification and verification of beneficial owners, including escalation where ownership structures are complex or opaque [S2 | Core regulatory source].
- Workspace review notes indicate that the current file does not yet contain sufficient source-of-funds evidence for one of the controlling shareholders [S3 | Workspace document].
- Internal onboarding policy requires compliance sign-off before approval where enhanced due diligence conditions are triggered [S4 | Workspace document].

Operational guidance

- The client should not be approved for onboarding until enhanced due diligence has been completed and documented.
- Obtain additional source-of-funds and source-of-wealth evidence for the relevant beneficial owner.
- Escalate the file to Compliance for review and approval before account opening.

This answer is evidence-grounded but is not a legal opinion. Please confirm with internal legal or compliance teams before approving onboarding or client-related decisions.

* SUPPORTING SOURCES

S1. AML Rulebook - Core regulatory source

Enhanced Due Diligence for High-Risk Jurisdictions

Requires enhanced due diligence where a customer, transaction, or beneficial owner is linked to a jurisdiction classified as high risk for money laundering or terrorist financing. Supports the requirement to apply enhanced due diligence in this case.

More

S2. Client Onboarding Rules - Core regulatory source

Beneficial Ownership Verification

Requires firms to identify and verify beneficial owners and to escalate cases involving complex or non-transparent ownership arrangements. Supports the beneficial ownership verification requirement.

More

S3. AML Transaction Review Notes - Workspace document

Source of Funds Gap

Internal review notes indicate that the onboarding pack does not yet contain adequate supporting evidence for source of funds relating to one of the controlling shareholders. Supports the conclusion that the file is not yet complete.

More

S4. Internal Onboarding Policy - Compliance Requirement

Escalation Requirement

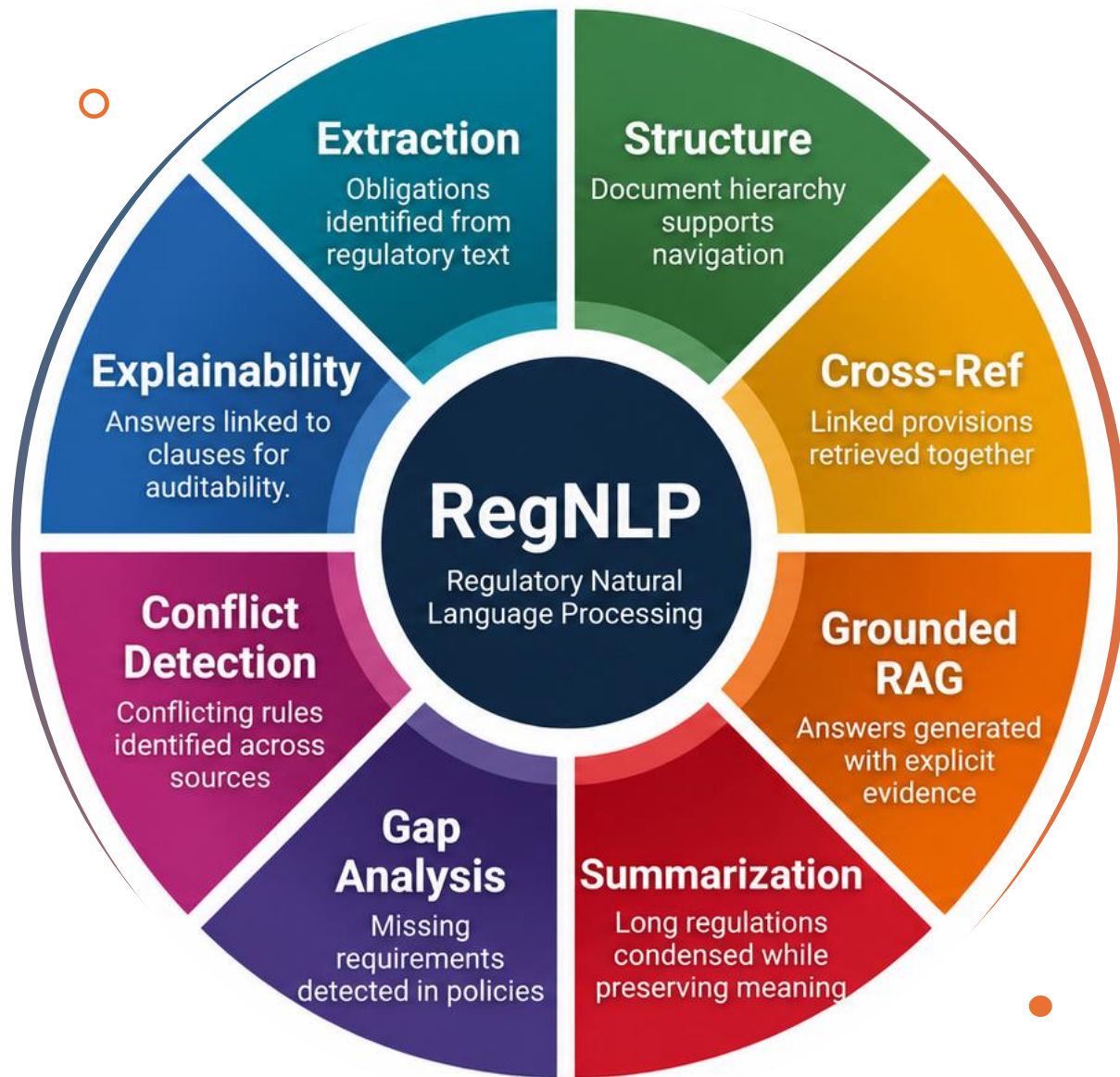
Requires Compliance approval before onboarding any client where enhanced due diligence measures have been triggered or where risk indicators remain unresolved. Supports the recommendation to escalate the file before approval.

More

AskRegula provides evidence-grounded regulatory assistance, but outputs should be reviewed by qualified professionals before being used for legal, compliance, or operational decisions.



<https://regnlp.github.io/AskREGULA/>



RegNLP is broader than QA

- Today we focused on:
 - **Evidence-grounded, evaluated, and governed regulatory QA**
- Broader directions:
 - **Extraction · Structure · Cross-References · Grounded RAG · Summarization · Gap Analysis · Conflict Detection · Explainability**
- *Open to collaborations in regulatory AI and domain-specific NLP*

Contact | tuba.gokhan@mbzuai.ac.ae