

Marqib — Model governance note

Version 1.0 · September 2026

1. Where models are and are not used

- **Detection does not use a language model.** Rules are SQL and TypeScript with thresholds held in one configuration file. The same data and thresholds always produce the same alerts, and every alert lists the records it rests on. Each rule's red flags, parameters and required data are shown on the Coverage page.
- **Decisions are made by people.** An analyst proposes; an MLRO approves or returns. Closing a case or deciding not to report requires a reason code. The system never files an STR.
- **A language model is optional and only writes text.** It rewrites the reasons and the STR narrative from the evidence table the rule already produced. Without it, Marqib drafts from templates filled with the same records.

2. Controls on generated text

1. **Tokenisation before every call.** Client names, client ids, IBANs, card numbers and wallet addresses are replaced by tokens. Originals are restored inside the firm's environment after the response.
2. **Evidence binding.** The model must cite evidence rows as [E<n>]. Any sentence without a valid citation is removed, and the number removed is logged.
3. **Side-by-side review.** The reviewer sees the rule text and the model text together and chooses which to file. Either can be edited.
4. **Audit.** Each generation records the provider, model, prompt hash, time and number of tokenised values.
5. **No training.** Marqib does not train or fine-tune models on client data. With the endpoints below, the provider's no-training and no-retention terms apply.

3. Endpoint options (chosen by the firm)

- Azure OpenAI or AWS Bedrock in the firm's own region and account.
- An open-weights model on the firm's own hardware, via an OpenAI-compatible server (for example vLLM or Ollama).
- None: template drafting only.

4. Change management

- Thresholds live in one versioned configuration file, so a change is visible in the code history.
- **(planned)** Threshold changes approved in the application, with above- and below-the-line backtesting.

- **(planned)** QA sampling of closed alerts.

5. Limitations the firm should know

- Rules find only the patterns they are written for. The Coverage page lists typologies not yet covered.
- Output quality depends on the completeness of the exported data. Rules whose inputs are missing produce no alerts, and Coverage shows this.
- goAML indicator codes must be mapped to the FIU's list before live submission. The XML structure should be validated against the FIU's XSD during the pilot.