iDharma

SAMPLE

INDEPENDENT · EXPERT-LED · STANDARDS-BASED

AI COMPLIANCE AUDIT

AI Audit Report

Client: Aarav Life Insurance (illustrative sample)

Audit tier: Compliance Audit

Prepared by: iDharma — Independent AI Audits

Overall risk rating: High

This is a sample report. It was produced for a fictional company, 'Aarav Life Insurance,' to show exactly what a real iDharma Compliance Audit report looks like — its structure, depth, and tone. Aarav Life is not a real client and the findings are illustrative. A real report follows this same shape, built entirely from your own systems and documents.

1 Executive summary

iDharma conducted a compliance-tier AI audit of Aarav Life Insurance's four production AI/ML systems: UnderwriteIQ v2.1 (underwriting risk scoring), ClaimsGuard (fraud triage), PolicyPal (customer chatbot), and RetentionRank (churn and renewal targeting). The audit assessed governance, data and privacy compliance, model fairness and validation, transparency, human oversight, and third-party risk against the EU AI Act, NIST AI RMF, ISO/IEC 42001, India's Digital Personal Data Protection (DPDP) Act, and GDPR. Evidence was drawn exclusively from the nine-section AI Audit Evidence Pack and the AI/ML Systems Overview submitted by the client.

The single most important finding: **UnderwriteIQ scores ~6,000 life-insurance applications a month** for pricing and eligibility, placing it in the **high-risk category under the EU AI Act** — with none of the high-risk obligations yet met, no independent validation, and no applicant AI-use notice. This is the area to address first.

The audit identified nine findings spanning Critical to Low severity. The most significant exposures are: (1) UnderwriteIQ's likely classification as a high-risk AI system under the EU AI Act with no conformity obligations yet met; (2) a flagged gender selection-rate disparity (0.84) that has not been re-tested through two subsequent quarterly retrains, with ethnicity and disability fairness entirely unmeasured across all models; (3) no independent model validation across any system, with all validation performed by the teams that build each model; (4) an incomplete and unsigned vendor data-processing agreement for PolicyPal, which transmits customer PII to a third-party LLM API without documented cross-border transfer mechanisms or retention controls; and (5) systemic absence of lawful-basis and consent documentation for training data across all four models under both the DPDP Act and GDPR.

Positive foundations exist: an AI Governance Committee with named owners and a formal charter is in place, dataminimisation principles are documented, an appeal path exists for underwriting decisions, and the organisation has selfidentified several of these gaps. However, governance effectiveness is undermined by a persistent pattern of action items carried forward without resolution — the fairness re-test, the transparency notice, and the DPA retrieval have each been deferred across at least two committee cycles.

The overall risk rating is High. The combination of likely high-risk AI Act classification, unresolved fairness signals in a consequential underwriting model, absent data-processing legal bases, and a live LLM chatbot handling customer PII without adequate vendor contractual controls creates a material regulatory and reputational exposure that requires structured, timebound remediation. This report is informational and does not constitute legal advice; qualified counsel should be engaged for binding compliance determinations.

2 Scope & approach

What was assessed: four AI systems — UnderwriteIQ (underwriting risk scoring), ClaimsGuard (claims fraud triage), PolicyPal (customer chatbot), and RetentionRank (renewal targeting).

What was not assessed: no live security or penetration testing, no bias testing of model outputs (that is Risk Audit scope), and no review of systems not listed above.

Standards referenced: EU AI Act, NIST AI Risk Management Framework, ISO/IEC 42001, India DPDP Act, and GDPR.

Method: the iDharma AI Audit methodology — Scope, Gather, Assess, Report, Readout — working a structured checklist across eight domains.

3 Findings

F-01 UnderwriteIQ Likely Meets EU AI Act High-Risk Classification with No Conformity Obligations Met

CRITICAL

SYSTEM: UNDERWRITEIQ · DOMAIN: REGULATORY CLASSIFICATION & GOVERNANCE

What we found. UnderwriteIQ v2.1 scores approximately 6,000 life insurance applications per month to determine pricing and eligibility. The EU AI Act Annex III, point 5(b) covers AI systems used to evaluate creditworthiness or establish credit scores, and point 5(a) covers AI systems used by insurers for risk assessment and pricing in relation to natural persons; life insurance underwriting scoring is squarely within these categories for the organisation's EU policyholders. The model card and governance charter both confirm that no independent model validation exists (validation is performed by the build team), no customer-facing AI-use notice has been deployed (recorded as 'under review' in both May and June 2026 committee minutes), decision logging is acknowledged as partial with no remediation timeline, and no conformity assessment documentation has been prepared. The AI Governance Committee has identified the transparency obligation but carried it forward unresolved across at least two consecutive monthly cycles.

Why it matters. Deployment of a likely high-risk AI system without meeting EU AI Act Article 9 (risk management), Article 10 (data governance), Article 13 (transparency), Article 14 (human oversight), and Article 17 (quality management) obligations exposes Aarav Life to regulatory enforcement action, potential fines, and reputational harm for EU-connected policyholders. The absence of independent validation and incomplete logging additionally undermines the organisation's ability to demonstrate compliance or investigate complaints.

Recommendation. Conduct a formal EU AI Act risk classification assessment with legal counsel as the immediate priority. If high-risk classification is confirmed: (1) commission independent model validation by a function separate from the build team before the next quarterly retrain; (2) implement complete decision and override logging (see Finding 4); (3) deploy a compliant applicant-facing AI-use notice covering the right to human review; (4) produce a technical documentation package per Article 11 and Annex IV; and (5) establish a quality management system per Article 17. Assign a single accountable owner with board-level escalation rights and a fixed deadline no later than the next retrain cycle.

F-02 Unresolved Gender Fairness Disparity in UnderwriteIQ; Ethnicity and Disability Not Measured Across Any Model — Model Evaluation

HIGH

SYSTEM: UNDERWRITEIQ, CLAIMSGUARD, RETENTIONRANK · DOMAIN: BIAS AND FAIRNESS

What we found. The September 2025 fairness evaluation (Section 2 of the Evidence Pack) recorded a gender selectionrate ratio of 0.84 (female/male), flagged as 'Watch — borderline' using the four-fifths rule. The same report explicitly recommended re-testing at every quarterly retrain cycle and investigating feature-level drivers (occupation and postal region identified as proxy attributes in the model card). As of June 2026, two quarterly retrains have been completed and no re-test has been conducted; the May 2026 committee minutes confirm the item remains outstanding. Ethnicity and disability were not measured in September 2025 due to data unavailability and have not been assessed since. No bias or disparate-impact testing of any kind is documented for ClaimsGuard (Section 4) or RetentionRank (Section 7). RetentionRank's features include age and location, both identified in its own documentation as carrying proxy-attribute risk.

Why it matters. A 0.84 selection-rate ratio is at the boundary of the four-fifths disparate-impact threshold and, if it has deteriorated through subsequent retrains, could constitute evidence of unlawful discrimination in insurance pricing and access decisions affecting protected groups. Absence of ethnicity and disability measurement means material disparities may exist undetected. ClaimsGuard's untested bias in fraud flagging creates additional risk of discriminatory claim-handling delays. Under the EU AI Act (high-risk obligations), GDPR Article 22, and broader anti-discrimination principles applicable in India, this is a significant legal and reputational exposure.

Recommendation. Immediately re-test UnderwriteIQ's gender selection-rate ratio using current production data; if the ratio has deteriorated below 0.80 or material feature-level drivers are confirmed (particularly occupation or postal region as ethnicity proxies), convene an emergency governance review before the next retrain. Formalise a fairness testing protocol requiring: (a) disparate-impact testing at every quarterly retrain as a release gate; (b) documented pass/fail thresholds approved by the AI Governance Committee; (c) a DPDP-compliant methodology for measuring ethnicity and disability proxy effects (e.g., geographic proxy analysis, surname-based proxy where legally permissible). Extend bias testing to ClaimsGuard and RetentionRank within 90 days, with results reviewed at the next committee meeting.

F-03 No Independent Model Validation Across Any System — Build-Team Self-Validation Only

HIGH

SYSTEM: UNDERWRITEIQ, CLAIMSGUARD, POLICYPAL, RETENTIONRANK · **DOMAIN:** GOVERNANCE & ACCOUNTABILITY

What we found. The model-owner register (Section 9) explicitly records that none of the four models has undergone independent validation: UnderwriteIQ is noted as 'validated by build team'; ClaimsGuard, PolicyPal, and RetentionRank each record 'No' independent validation. The AI/ML Systems Overview confirms this as a known organisational gap. No independent validation function, challenger model programme, or external review process is evidenced anywhere in the submitted documentation.

Why it matters. Self-validation by the team that builds a model is a fundamental conflict of interest in model risk management practice. It creates a structural inability to detect systematic errors, overfitting, or bias introduced during development. For UnderwriteIQ — which influences consequential life insurance decisions at scale — the absence of independent validation also directly undermines any future EU AI Act conformity claim and is inconsistent with mature model risk governance. For ClaimsGuard, unvalidated performance may be affecting claims handling quality and speed in ways that are currently invisible to management.

Recommendation. Establish an independent model validation function (internal second-line risk team or qualified external reviewer) with explicit authority to assess all production AI models. As an immediate action, commission an independent validation of UnderwriteIQ before its next quarterly retrain, covering performance, calibration, fairness, and feature-proxy risk. Extend independent validation to ClaimsGuard within 90 days given the absence of any testing documentation. Formalise a model governance policy requiring independent sign-off as a condition of production deployment for all high-impact models, and update the model-owner register accordingly.

F-04 Incomplete Decision and Override Logging in UnderwriteIQ — Audit Trail Deficiency

HIGH

SYSTEM: UNDERWRITEIQ · **DOMAIN:** HUMAN OVERSIGHT & MONITORING

What we found. Section 3 (Decision Logs, 3-month sample) provides direct evidence of the logging gap: applications AP2006, AP-2015, and others show underwriter overrides occurring (e.g., override_decline, override_accept) but 'override_logged = no.' The model card (Section 1) acknowledges that 'not all override events are captured' and notes there is no remediation timeline. The AI Governance Committee minutes do not record any owner or deadline assigned to this issue.

Why it matters. Incomplete override logging has three compounding consequences: (1) it prevents accurate measurement of override rates, which is essential for monitoring whether human oversight is functioning as intended; (2) it creates an incomplete audit trail that would impede any regulatory investigation or customer complaint review; (3) it means fairness analyses and model performance assessments are based on incomplete outcome data, potentially masking systematic patterns in which cases are overridden. Under the EU AI Act's high-risk human oversight requirements and NIST AI RMF Manage functions, a demonstrably incomplete logging system is a direct compliance deficiency.

Recommendation. Treat complete override logging as a time-critical remediation item with a fixed deadline of 60 days. Assign a named system owner in the model-owner register. Implement mandatory structured override capture at the point of underwriter decision (not post-hoc) with required fields: application ID, model score, model recommendation, underwriter decision, override rationale code, and timestamp. Once complete logging is in place, conduct a retrospective analysis of the partial-capture period to assess whether any systematic override patterns were missed. Report remediation completion to the AI Governance Committee.

F-05 PolicyPal Vendor DPA Incomplete — Unsigned, Missing Sub-processor Clauses, No Cross-Border Transfer Mechanism

HIGH

SYSTEM: POLICYPAL · **DOMAIN:** THIRD-PARTY & VENDOR RISK

What we found. Section 5 of the Evidence Pack documents that the DPA with the third-party LLM API provider is 'referenced in onboarding; full signed agreement not located.' Specific documented gaps include: no GDPR Article 28 sub-processor list or flow-down clauses; no documented retention or deletion SLA for prompt and response data; no specified cross-border transfer mechanism for India-to-provider and provider-to-EU data flows; and DPDP data-fiduciary obligations not mapped. The June 2026 committee minutes record that the missing DPA was flagged but no owner was assigned. PolicyPal can retrieve and display customer PII to answer account queries, meaning live personal data is being transmitted to a third-party API without an adequate contractual basis.

Why it matters. Operating a customer-facing system that transmits PII — including potentially sensitive insurance and financial data — to a third-party LLM provider under an incomplete, unsigned DPA constitutes a direct GDPR Article 28 violation for EU policyholders and a likely breach of DPDP Act data-fiduciary obligations for Indian policyholders. In the event of a data incident involving the vendor, Aarav Life would have no contractual basis for data deletion, breach notification SLAs, or sub-processor accountability. This is an acute, ongoing legal exposure.

Recommendation. As an immediate action: (1) obtain and execute the full, signed DPA with the LLM API provider; (2) ensure the DPA includes GDPR Article 28-compliant provisions (sub-processor list, flow-down obligations, audit rights, deletion SLA, breach notification timeline); (3) document a valid cross-border transfer mechanism (e.g., Standard Contractual Clauses) for all data flows crossing India-EU and India-provider boundaries; (4) map DPDP data-fiduciary obligations onto the vendor relationship and confirm contractual compliance. Until a compliant DPA is executed, consider whether PII retrieval capability in PolicyPal should be temporarily restricted to minimise exposure. Assign a named owner at the June 2026 committee level with a 30-day deadline for DPA execution.

F-06 Training-Data Lawful Basis, Consent, and Lineage Not Documented Across All Four Models

HIGH

SYSTEM: UNDERWRITEIQ, CLAIMSGUARD, POLICYPAL, RETENTIONRANK · **DOMAIN:** DATA GOVERNANCE & PRIVACY

What we found. Section 8 (Training-Data Consent & Lawful-Basis Records) confirms that: UnderwriteIQ has only partial lawful-basis documentation with derived features undocumented; ClaimsGuard, RetentionRank, and PolicyPal have no documented lawful basis or consent records under the DPDP Act; feature data lineage is incomplete for UnderwriteIQ, ClaimsGuard, and RetentionRank, and not mapped for PolicyPal. GDPR Article 6 lawful-basis records and Article 22 automated-decision considerations are explicitly noted as undocumented for UnderwriteIQ and RetentionRank. The document records that 'remediation has not yet been scoped,' indicating no active programme is underway.

Why it matters. Processing personal data for model training without a documented lawful basis is a violation of GDPR Article 6 and DPDP Act Section 4 for all data subjects covered by those regimes. For UnderwriteIQ and RetentionRank, the absence of Article 22 documentation means individuals subject to automated decision-making have no established mechanism to exercise their rights to human review, explanation, or objection. Incomplete feature lineage prevents effective bias investigation, model auditing, and data-subject rights fulfilment (e.g., right to erasure under GDPR Article 17 and DPDP Act Section 12).

Recommendation. Initiate a structured data-governance remediation programme covering all four models within 60 days. This should include: (1) mapping all training data sources and establishing a documented, valid lawful basis for each under both GDPR (Article 6) and DPDP Act (Section 4); (2) documenting Article 22 considerations and existing safeguards for UnderwriteIQ and RetentionRank, including the human review and appeal mechanism; (3) completing feature lineage documentation for all models to support bias investigation and data-subject rights fulfilment; (4) establishing a process to maintain these records through future retrains. Engage legal counsel to assess whether retrospective lawful-basis gaps require notification to data protection authorities.

F-07 No Customer-Facing AI-Use Disclosure Across Underwriting or Claims Decisions

MEDIUM

SYSTEM: UNDERWRITEIQ, CLAIMSGUARD, POLICYPAL, RETENTIONRANK · **DOMAIN:** TRANSPARENCY & DISCLOSURE

What we found. The UnderwriteIQ model card (Section 1) records that 'No AI-use notice shown to applicants; transparency obligation recorded as under review.' The AI/ML Systems Overview (Section 2 of that document) confirms 'there is currently no customer-facing notice that AI is used in underwriting or claims.' Committee minutes (May and June 2026) show this has been raised and deferred without resolution or assigned ownership. No disclosure mechanism is evidenced for ClaimsGuard, PolicyPal (beyond its functional identity as a chatbot), or RetentionRank.

Why it matters. For EU policyholders, GDPR Article 13/14 requires disclosure of automated decision-making logic, and EU AI Act Article 13 requires high-risk AI systems to provide meaningful transparency to affected persons. Under the DPDP Act, data principals must be informed of the purposes of processing. Absence of disclosure prevents individuals from exercising their rights to human review, explanation, or objection. It also increases regulatory and litigation risk if an adverse decision is challenged, as Aarav Life cannot demonstrate that affected persons were on notice of AI involvement.

Recommendation. Deploy an AI-use transparency notice for applicants affected by UnderwriteIQ as the immediate priority. The notice should: identify that an AI system is used in underwriting; describe its purpose and the role of human review; explain the right to request human reconsideration; and provide a contact mechanism. Extend equivalent disclosures to ClaimsGuard-influenced claim communications. For RetentionRank, assess whether targetingbased pricing nudges constitute automated decision-making warranting disclosure under GDPR Article 22. PolicyPal should explicitly identify itself as an AI assistant at the start of each session. All disclosures should be reviewed by legal counsel for GDPR and DPDP Act compliance before deployment.

F-08 PolicyPal Lacks Output Accuracy and Hallucination Monitoring; Prompt-Injection Controls Untested

MEDIUM

SYSTEM: POLICYPAL · **DOMAIN:** SECURITY & MONITORING

What we found. Section 6 (PolicyPal System Prompt, Controls & PII Handling) documents that there is no automated hallucination or output-accuracy monitoring, no jailbreak or prompt-injection testing has been performed, and conversation log retention is undefined. Controls consist only of a keyword blocklist and a system-prompt instruction to avoid legal or medical advice. The June 2026 committee minutes flagged absent output monitoring and assigned no owner. The system prompt grants access to customer account records and PII, amplifying the impact of any successful prompt-injection attack.

Why it matters. A generative AI chatbot handling insurance policy and billing questions with access to customer PII, operating without hallucination monitoring or prompt-injection controls, presents multiple concurrent risks: (1) misinformation about policy terms or billing that could harm customers or generate liability; (2) prompt-injection attacks that could cause the model to exfiltrate customer PII from account records; (3) undefined log retention means the organisation cannot reconstruct conversations for complaints, regulatory requests, or incident investigation; (4) without output monitoring, systematic errors or harmful outputs would go undetected. These risks are amplified by the incomplete vendor DPA (Finding 5).

Recommendation. (1) Implement automated output-quality sampling: a proportion of PolicyPal conversations should be reviewed against ground-truth policy documents, with accuracy metrics tracked and reported to the AI Governance Committee monthly. (2) Commission prompt-injection and jailbreak adversarial testing before the next quarter and remediate identified vectors. (3) Define and enforce a conversation log retention period consistent with DPDP Act and GDPR requirements (document this in the DPA). (4) Consider implementing an output guardrail layer that validates responses against a curated policy-document knowledge base before delivery. (5) Assign a named output-monitoring owner in the model-owner register.

F-09 ClaimsGuard and RetentionRank Lack Any Production Monitoring, Drift Detection, or Revalidation Schedule

MEDIUM

SYSTEM: UNDERWRITEIQ, CLAIMSGUARD, RETENTIONRANK · **DOMAIN:** MONITORING & MODEL LIFECYCLE

What we found. The ClaimsGuard model documentation (Section 4) explicitly records: no documented bias or disparateimpact testing; no drift or performance monitoring in production; no periodic revalidation schedule. RetentionRank documentation (Section 7) records no fairness review and no opt-out or transparency mechanism, with no monitoring or revalidation schedule evidenced. Neither model has a documented performance baseline against which production degradation could be measured. UnderwriteIQ, by contrast, has a PSI of 0.07 and quarterly retraining, though fairness testing is not embedded in that cycle.

Why it matters. Models deployed without drift detection or revalidation schedules can degrade silently in production — producing outputs that diverge from validated performance without triggering any alert. For ClaimsGuard, undetected degradation could result in systematic misclassification of legitimate claims as fraudulent (or vice versa), harming policyholders and increasing operational loss. For RetentionRank, model drift could cause pricing nudges to become increasingly discriminatory or commercially counterproductive. The absence of any monitoring baseline also means the organisation cannot demonstrate to regulators that deployed models remain fit for purpose.

Recommendation. Within 90 days: (1) establish a production monitoring baseline for ClaimsGuard (AUC, precision/recall on a held-out labelled set, PSI on key features) and implement automated monthly drift alerts; (2) create a formal revalidation schedule for both ClaimsGuard and RetentionRank (minimum semi-annual, triggered earlier by drift alerts or significant business changes); (3) integrate fairness metrics into ClaimsGuard's monitoring dashboard, with particular attention to claim-handling speed disparities by geographic proxy groups; (4) document all monitoring configurations and review thresholds in the model-owner register. Assign owners to both models' monitoring programmes by name.

4 Prioritised action list

#	ACTION	CLOSES	SEVERITY	SUGGESTED OWNER
1	Treat UnderwriteIQ as EU AI Act high-risk; confirm exposure with counsel; restrict use until compliant.	F-01	Critical	CTO + counsel
2	Execute a signed PolicyPal LLM Data-Processing Agreement (GDPR Art. 28, transfer mechanism, retention); confirm data isn't used for training.	F-05	High	CTO
3	Re-test UnderwriteIQ gender fairness on current data and add ethnicity/disability measurement; bind to each retrain.	F-02	High	Model-risk lead
4	Stand up an independent model-validation function, separate from the build teams.	F-03	High	CRO / Risk
5	Complete UnderwriteIQ decision and override logging.	F-04	High	Engineering lead
6	Document lawful basis, consent, and data lineage for all four models (DPDP + GDPR).	F-06	High	DPO / Legal
7	Deploy a customer-facing AI-use disclosure for underwriting and claims.	F-07	Medium	Compliance
8	Add PolicyPal output monitoring + prompt-injection controls; production monitoring & drift for ClaimsGuard / RetentionRank.	F-08, F-09	Medium	Engineering / Data

5 Framework snapshot

FRAMEWORK	OVERALL STANDING	NOTES
EU AI Act	Gaps	UnderwriteIQ likely high-risk and not yet meeting high-risk obligations; no conformity assessment or applicant AI-use notice.
NIST AI RMF	Partial	Govern committee exists, but no independent validation; Measure / Manage are weak — limited monitoring and unassigned remediation.
ISO/IEC 42001	Partial	Committee + model-owner register meet parts of Cl. 5 / 6.1; gaps in validation, monitoring, and data governance.
India DPDP Act	Gaps	Lawful basis and consent for training data undocumented; PII sent to a third-party LLM without data-fiduciary safeguards.

FRAMEWORK	OVERALL STANDING	NOTES
GDPR	**Gaps**	Processor terms incomplete (PolicyPal DPA unsigned); Art. 6 / 13 / 22 records absent for automated underwriting of EU policyholders.

6 Next steps

Immediate. The two highest-urgency items — the EU AI Act classification (Action 1) and the PolicyPal DPA (Action 2).

Getting help. If it would help to have these fixed rather than just listed, iDharma can match you with an independently verified AI / model-risk consultant for the remediation work.

Deeper assurance. Given UnderwriteIQ's likely high-risk status, a Risk Audit would add the bias / fairness and security testing beyond this compliance review.