DATA GOVERNANCE & INTEGRITY

MODULE LAB

Emmanuel Kwabena Ansu

3rd February, 2026

DELIVERABLE 1: GOVERNANCE REVIEW CARD (PRIMARY TASK)

SECTION	ISSUE/DEFINITION	IMPACT	FIX/MITIGATION
Data Quality Risk	Incomplete and inconsistent customer input data (e.g., missing income fields, inconsistent phone number formats from mobile app inputs)	Poor-quality inputs directly degrade ML loan-scoring accuracy, leading to incorrect approvals or rejections and increased default risk	Enforce mandatory field validation at app level, standardize formats (phone, income, ID), and implement preprocessing validation checks before model scoring
Legal & Compliance Risk	Collection of excessive PII without explicit consent (e.g., entire contact list)	Violates Ghana Data Protection Act (Act 843) principles of data minimization and lawful processing, exposing company to regulatory penalties and reputational damage	Implement explicit, purpose-specific consent capture and remove non-essential data collection
Data Classification	It is the process of categorizing data based on its confidentiality to determine the level of access that should be granted to it and the level of protection it requires against unauthorized access or disclosure	Sensitive	Financial data, contact data, behavioural data require highest protection under Act 843
Bias & Fairness Risk	Source of Bias: ML model trained on historical loan outcomes that may reflect socio-economic or geographic bias	Automated decisions may systematically disadvantage certain demographic groups, creating unfair denial patterns	Introduce fairness testing by demographic segments and periodic bias audits before model redeployment
Storytelling/ Reporting Recommendation	Metric to Monitor: Approval Rate Disparity Ratio	Ratio of loan approval rates between highest-approved and lowest-approved demographic groups	Enables transparent, ethical oversight of ML decisions
	Visualization Type	Grouped Bar Chart	Supports clear comparison across groups
	Why It Matters	makes algorithmic bias visible and measurable, enabling ethical oversight of automated decisions	Prevents unethical or unlawful automated decisions

DELIVERABLE 2: CORRECTED DATA FLOW DIAGRAM (ANNOTATED FIXES)

i. Below are the required corrections, mapped directly to the flawed pipeline steps:

ii. Excessive Collection at **Step 1** (User Mobile App)

Correction: Remove access to entire contact list; collect only data strictly required for loan eligibility.

Why: Enforces data minimization under Act 843 and reduces privacy exposure.

iii. No Consent Between **Step 2 and Step 3** (API Gateway → Raw Data DB)

Correction: Insert explicit consent capture and logging before data persistence.

Why: Ensures lawful processing and auditability of consent.

iv. No Classification or Retention at **Step 3** (Raw Data DB)

Correction: Apply data classification tags (Sensitive) and define retention periods per data category.

Why: Prevents indefinite storage of high-risk personal data.

v. No Defined Handling at **Step 4** (Preprocessing Service)

Correction: Add validation, normalization, and rejection rules for incomplete or malformed records.

Why: Protects ML model from garbage-in / garbage-out risk.

vi. No Transparency at **Step 7** (Decision Service)

Correction: Log decision inputs, outputs, and model version for audit purposes.

Why: Enables explainability and regulatory review of automated decisions.

vii. No Masking at **Step 9 & 10** (Analytics DB & Third-Party Partner)

Correction: Apply anonymization or masking before analytics and external sharing.

Why: Limits exposure of sensitive financial and personal data.

DELIVERABLE 3: SUMMARY OF REVIEW PROCESS (≈230 WORDS)

This governance review evaluated Quick Loan Mobile's data pipeline using data lifecycle and data classification principles to identify governance, compliance, and ethical risks. Each stage from data collection to third-party sharing was assessed to determine whether the data processed was necessary, lawful, accurate, and protected. Applying data lifecycle analysis, risks were identified at collection (excessive data capture), storage (lack of classification and retention rules), processing (poor validation), and decisioning (absence of transparency). Data classification principles were used to establish that customer financial and behavioural data qualifies as Sensitive under Ghana's Data Protection Act (Act 843), requiring strict safeguards, consent, and limited retention. The proposed Approval Rate Disparity Ratio metric ensures ethical governance by making automated decision outcomes transparent and measurable across demographic groups. Instead of relying on abstract claims of fairness, this metric provides concrete evidence of whether the ML system disproportionately approves or rejects certain populations. Visualizing this metric over time allows management to detect emerging bias early and intervene before harm occurs.

Together, these recommendations shift QuickLoan from a purely growth-driven, opaque automation model to a governed, compliant, and ethically defensible system. The approach balances operational efficiency with regulatory accountability and responsible AI practices, ensuring sustainable fintech innovation.