NAME: ERNEST BAFFOE

CORRECTED DATA FLOW DIAGRAM

KEY FLAWS IDENTIFIED IN ORIGINAL DATA FLOW

1. ***Excessive Collection*** at Step 1 (User Mobile App)
2. ***No Consent or Compliance*** between Step 2 and Step 3 (API Gateway to Raw Data DB)
3. ***No Classification or Retention*** at Step 3 (Raw Data DB)
4. ***No Denied Handling*** at Step 5 (Preprocessing Service)
5. ***No Transparency or Logging*** at Step 7 (Decision Service)
6. ***No Masking or Anonymization*** at Step 9 and 10 (Analytics DB and Third-Party Partners)

DETAILED CORRECTIONS AND ANNOTATIONS

1. CORRECTION 1: **Implement Data Minimization at Step 1 (User Mobile App)**

   What Changed:
   Reduced data collection to ONLY essential fields: Name, Phone Number, Email, Loan Amount, and Income Verification. Removed collection of entire contact lists, continuous GPS tracking, and device logs.

   Why it's Necessary:
   Ghana's Data Protection Act (DPA) 843 Principle 3 (Minimality) requires collecting only data necessary for the stated purpose. Collecting users' entire contact lists and continuous GPS data is excessive for assessing creditworthiness. This violates user privacy, creates security risks and exposes QuickLoan to Data Protection Commission penalties.

2. CORRECTION 2: **Add Explicit Consent Capture Between Steps 2 to 3 (API Gateway to Raw Data DB)**

   What Changed:
   Implement a Consent Management Module between the API Gateway (Step 2) and Raw Data DB (Step 3). This module presents users with clear, plain-language consent forms explaining what data is collected, why it's needed, how it will be

used, and retention periods. Users must actively opt-in, and consent is logged with timestamp and policy version.

Why it's Necessary:
The original flow showed "no consent capture", violating Ghana's Data Protection Act Principle 4 (Consent), which mandates freely given, specific, informed, and unambiguous consent before processing personal data. Without consent, QuickLoan faces regulatory fines, potential shutdown orders from the Data Protection Commission, and violations of user autonomy rights.

3. CORRECTION 3: **Implement Data Classification and Retention Policies at Step 3 (Raw Data DB)**

   What Changed:
   Tag all data fields with classification levels: SENSITIVE (Name, Phone, Email, Income – encrypted at rest), CONFIDENTIAL (Loan amounts, decisions), INTERNAL (Session IDS), PUBLIC (Aggregated stats). Implement automated retention rules: PII deleted 7 years after account closure, consent records retained per legal requirements, analytics data anonymized after 90 days, device logs deleted after 30 days.

   Why it's Necessary:
   Without classification, all data receive uniform protection regardless of sensitivity, violating the Data Protection Act Principle 8 (Security). Without retention policies, data is kept indefinitely, violating Principle 7 (Retention). Classification enables differential security controls (encryption for sensitive data), while retention rules prevent indefinite data accumulation, reducing breach of risks, and demonstrating DPA compliance.

4. CORRECTION 4: **Add Transparency and Explainability Logging at Step 7 (Decision Service)**

   What Changed:
   Enhance the Decision Service to log comprehensive explainability data for every loan decision: the decision itself (Approve / Reject), model confidence score (0-100%), top 5 contributing features with weights, demographic flags for fairness monitoring, timestamp, and model version used. This creates a complete audit trail.

Why it's Necessary:
The original "black box" decision-making with "no transparency/loggin" makes it impossible to detect algorithmic bias, respond to customer inquiries about rejections, or demonstrate "fair and lawful processing" (DPA Principle 2). Transparency logging enables ongoing fairness audits, provides evidence-based explanations to customers, and operationalizes ethical AI governance by making bias measurable and correctable.

5. CORRECTION 5: **Implement PII Masking and Anonymization at Steps 9-10 (Analytics DB and Third-Party Partners)**

   What Changed:
   Add data masking/anonymization process before data flows to Analytics DB (step 9). Techniques include Tokenization (names/IDS => random tokens), Generalization (exact age => age ranges), Aggregation (individual GPS => regional statistics), and Suppression (removal of direct identifiers). Step 10 (Third-Party Partners) receives only aggregated insights, never raw customer PII.

   Why It's Necessary:
   The original flow showed full PII flowing to analytics and external partners without masking, violating DPA Principle 3 (Minimality) and 8 (Security). Analytics don't need customer names to identify trends. Every system storing raw PII is a breach point. Masking preserves analytical value while eliminating privacy risks. If breached, attackers get anonymized tokens, not real identities. For third parties, only aggregated insights prevent data mishandling and reduce liability.

Summary: Flaws Corrected and Compliance Achieved

Flaw #	Original Problem	Correction Applied	Data Protection Act Principle Addressed
1	Excessive Collection at Step 1	Data Minimization	Principle 3: Minimality
2	No Consent between Steps 2 to 3	Consent Management Module	Principle 4: Consent
3	No Classification or Retention at Step 3	Classification Tags and Retention Policies	Principle 7 & 8: Retention and Security

4	No Transparency or Logging at Step 3	Explainability Logging	Principle 2: Fair and lawful Processing
5	No Masking or Anonymization at Steps 9-10	PII Masking and Anonymization	Principles 3 & 8: Minimality and Security