Prototype Evaluation Report (PE-001) - AI Model

Prototype Name: Sentiment Analysis for Log Routing (SALR) v0.5
Date of Test: 2025-10-25
R&D Lead: [Data Scientist Name]
Purpose: Evaluate the initial performance of a new ML model intended to automatically route support tickets based on client log file sentiment.

1. Technical Summary

- **Model Architecture:** Bidirectional Encoder Representations from Transformers (BERT) variant, fine-tuned on 10,000 NexaCore support ticket texts (Arabic/English).
- **Training Data:** 15GB of historical, anonymized customer support interactions.
- **Compute:** AWS p3.2xlarge GPU Instance (4 hours of training time).

2. Performance Metrics

Metric	Target (Min Viable Product)	Result (v0.5)	Status	Notes
Accuracy (F1 Score)			**FAIL**	Low score in distinguishing 'Frustration' from 'Urgency' in Arabic text.
Inference Latency	per log entry		**PASS**	Meets the real-time requirements for log processing.
Training Cost	per run	$180	**PASS**	Cost is manageable for monthly retraining.
False Negative Rate			**FAIL**	Too many critical (P1) tickets are

				routed as non-urgent.

3. Findings and Recommendations

- **Finding:** The model performs well on English text (F1), but drops significantly on mixed Arabic/English logs (F1). This is a critical blocker for MENA region deployment.
- **Recommendation 1 (Model Improvement):** R&D must source an additional 5,000 labeled Arabic-only log entries to improve regional language feature extraction. **(R&D Priority: High)**
- **Recommendation 2 (Deployment:** Deploy the English-only version to the US development sandbox for integration testing, but do not promote to production until F1 score reaches across both languages.
- **Go/No-Go Decision: No-Go** for production until multi-language accuracy is improved. Proceed to an extended Research & Prototyping phase.