

Real-Time AI Optimization Worksheet

Tech Champions 5 – Jeddah Cohort

1. Decision Point

*What is one real-time decision your startup needs to make?
(e.g., BNPL approval, fraud flag, donation routing, phishing detection)*

2. Current Metrics

- p95 Latency: _____
- Accuracy / Precision: _____
- Cost per 1k decisions (approx): _____

3. Bottleneck

Where does it slow down or break today?

- ☐ Model too large / slow
- ☐ Infra inefficiency (no cache, no batching)
- ☐ Decision unclear (no thresholds, too many false positives)
- ☐ Other: _____

4. Lever to Apply

Which lever can help most right now?

- ☐ Model Choice → smaller, task-specific model
- ☐ Infra Tricks → caching, batching, ANN
- ☐ Decision Design → thresholds, fallback, human-in-loop

5. Next One-Week Step

What small action will you test this week?

6. Owner

Who is responsible for trying this?
