

Metrics, Experimentation & GTM

1. NORTH STAR METRIC

North Star Metric: Weekly value of bills successfully split-and-settled through Smart Split (Rs settled within 7 days).

This single metric captures the whole value chain — a bill is scanned, split fairly by the AI, AND actually paid back. It is chosen over a count of scans because a split only delivers value when money settles. It connects user value (fast, fair splits friends actually pay) with business value (every settled split is real UPI/wallet volume, deeper ecosystem use, higher DAU/MAU). There is exactly one North Star; the metrics below support it.

2. SUPPORTING KPIS

Metric	What it measures	Why it matters
Scan-to-Confirm conversion (product)	% of scans that reach 'Confirm & Send'	Health of the core AI funnel
Split settlement rate (product)	% of split requests fully paid within 7 days	Settlement is where value is realised
Feature adoption (product)	% of split-flow users who use Scan Bill	Shows AI is replacing manual entry
OCR item-extraction accuracy (AI)	% of items + prices read correctly	Correctness and user trust
AI edit rate (AI)	% of splits where the user edits an AI item	Inverse proxy for model quality
Median scan-to-split latency (AI)	Seconds from scan to itemised split	Speed drives completion vs. abandonment

3. AARRR FUNNEL MAPPING

Funnel Stage	User Action	Metric
Acquisition	User notices the 'Scan Bill (NEW)' badge or a launch push	Badge impressions → tap-through rate
Activation	User completes their first AI scan and taps Confirm & Send	First-scan completion rate
Retention	User returns to use Scan Bill for their next group bill	W2 / W4 Smart Split repeat-usage rate
Revenue	Indirect: more settled splits = more UPI/wallet transactions, higher DAU/MAU (feature is free)	Incremental transactions per Smart Split user
Referral	Receivers see a clean itemised split and are prompted 'Split your next bill with AI'	Receiver-to-planner conversion / invites sent

4. A/B TEST PLAN

Element	Answer
Hypothesis	Showing the AI 'Scan Bill' option at the expense-entry step increases split-flow settlement rate vs. the manual-only experience.
Control version	Current manual split (type total, divide equally / by share).
Variant version	Manual split + AI Smart Split scan option.
Target users	Tier-1/Tier-2 group-bill users on OCR-supported devices, randomly assigned 50/50.
Primary success metric	Split settlement rate (requests fully paid within 7 days).
Secondary metrics	Scan-to-Confirm conversion, scan-to-split latency, CSAT (thumbs-up rate).
Sample size	~1,600 users per arm. Assumptions: baseline settlement $\approx 55\%$, minimum detectable effect +5 pp, $\alpha = 0.05$ (two-sided), power = 80%.
Success criteria	Variant beats control on the primary metric with statistical significance and no guardrail breach, sustained over a 2-week run.

Hypothesised results (stated as a bet, not a fact)

Based on the assumptions above, I expect — and would hold myself to — roughly:

- **Scan-to-Confirm completion:** 55% $\rightarrow$ ~63% (manual division replaced by tap-assign).
- **Split settlement rate:** 70% $\rightarrow$ ~78% (itemised proof makes receivers pay faster).

These are hypotheses to be proven by the test, not claimed outcomes; if the lift doesn't materialise, the assumption about scan willingness (File 2) is likely wrong.

5. DASHBOARD PLAN

Dashboard Section	Metrics Shown	User of Dashboard	Frequency of Review
Funnel health	Scan $\rightarrow$ Confirm $\rightarrow$ Settle conversion	PM / Growth	Daily
AI quality	Accuracy, hallucination, edit rate, latency	ML team / PM	Daily
Trust & support	CSAT, dispute rate, escalations	Support / PM	Weekly
Business impact	DAU/MAU lift, settled transaction volume	Leadership	Weekly

6. ROLLOUT PLAN

- **Phase 1 — Internal:** Dogfood with employees on real bills to validate accuracy and the fallback path.
- **Phase 2 — Closed beta:** Opt-in users in 1–2 metro cities; monitor accuracy, latency and disputes closely.
- **Phase 3 — Percentage rollout:** 10% $\rightarrow$ 30% $\rightarrow$ 50% $\rightarrow$ 100%, gated by region and device, advancing only when guardrails stay green.
- **Kill-switch criteria:** Roll back if hallucination rate $> 2\%$, dispute rate breaches threshold, P95 latency > 8 s, or any privacy/PII incident.

- **Red-teaming plan:** Before launch, adversarially test blurry/handwritten/regional and foreign-currency bills, prompt-injection text printed on bills, fake/manipulated totals, and PII-extraction attempts.

7. GTM STRATEGY

- **Target audience:** Urban millennial and Gen-Z group spenders in Tier-1/Tier-2 cities — students, young professionals, flatmates, travellers.
- **Positioning statement:** 'Snap the bill. PhonePe splits it fairly — taxes and all — in seconds.'
- **Launch channels:** In-app badge + coach-mark at the split step, push notification, email, Instagram reels, college ambassadors.
- **Key message:** End the calculator and the awkward math — fair, itemised splits instantly, everyone paying for exactly what they had.
- **Pre-launch checklist:** OCR-accuracy benchmark passed, fallback and human-escalation tested, privacy/redaction audit signed off, support playbook ready, kill-switch wired, analytics events instrumented.

8. PRICING OR COST NOTE

Smart Split is free — an engagement and transaction-volume driver, not a paid add-on. Cost per interaction is roughly Rs 0.10–Rs 0.40 of OCR plus small compute per scan. Against this, each successful split generates incremental UPI/wallet transactions and higher DAU/MAU; the value of a retained, more-active user far exceeds the sub-Rs 0.50 per-scan cost, so unit economics are positive even before any cross-sell. Monetisation, if ever needed, comes indirectly through ecosystem usage — never by charging users to split a bill.

9. POST-LAUNCH FEEDBACK PLAN

- **Feedback channels:** Per-split thumbs up/down, an in-app micro-survey after a split settles, periodic NPS, support tickets, app-store reviews.
- **Signals collected:** AI edit rate, dispute rate, scan abandonment points, and qualitative 'what was wrong' notes on flagged splits.
- **How feedback improves the feature:** User corrections retrain OCR/NLU on weak bill formats, prioritise the most-failed bill types, and tune confidence thresholds and escalation triggers over time.