DECISIONFLOW

Launch Measurement and Experiment Plan

North Star, funnel, analytics, AI evaluation, experiments, and decision cadence

Document code	LAUNCH-04
Version	2.0
Status	Final Public Portfolio Edition — Simulated Launch Scenario
Baseline date	14 July 2026
Author	Min Thu Kyaw
Role	AI Product Manager
Evidence rule	Actual artifacts are identified separately from assumed capabilities and simulated outcomes.

EVIDENCE & PUBLISHING INTEGRITY

This public portfolio edition distinguishes actual evidence, assumed continuity steps, simulated case-study outcomes, targets, and recommendations. Simulated customers, commercial results, benchmarks, incidents, quotations, and later-lifecycle outcomes are illustrative and must not be interpreted as results from a live commercial product. Evidence labels are retained throughout the package to support transparent and responsible interpretation.

Document Control

Control	Definition
Purpose	Document the Launch phase for Launch Measurement and Experiment Plan.
Primary audience	Hiring managers, product leaders, launch stakeholders, engineering, customer success, and operations.
Evidence classes	Actual artifact / assumed post-development capability / simulated launch outcome / recommendation.
Scenario assumption	All development P0 release gates are assumed complete before the simulated controlled launch begins.
Ethical use	Do not present simulated customer evidence, revenue, testimonials, or performance figures as real-world results.
Change control	Publication rule: retain evidence labels unless content is replaced by documented real-world observations.

Actual Project Artifacts

- Low-Fidelity Prototype
- High-Fidelity Prototype
- MVP Frontend Repository
- MVP Backend Repository

Contents

Measurement Philosophy

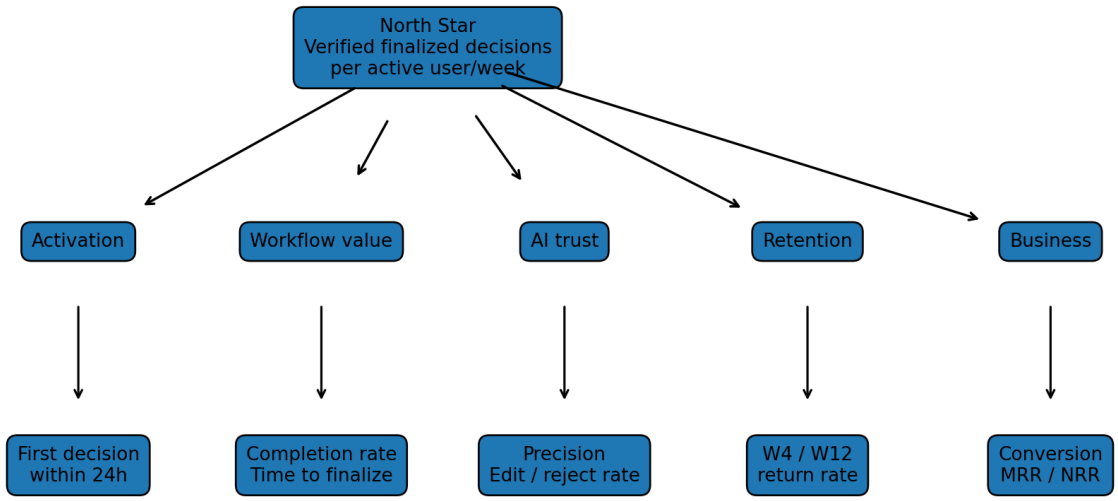
DecisionFlow should not be judged by accounts created, transcripts processed, or AI outputs generated in isolation. Launch success requires a combination of verified customer value, trusted AI performance, repeat workflow behavior, commercial signal, and operational reliability.

Evidence classification - Actual: public prototypes, repository references, uploaded MVP development baseline, and prior Product/Plan/Develop artifacts. Assumed: the production-ready v1.0 core capabilities documented in DEV-06 as completed for the launch scenario. Simulated: user behavior, commercial results, AI benchmarks, incidents, quotations, and launch outcomes.

North Star Metric

Verified Finalized Decisions per Active User per Week - A finalized decision counts only when the user has reviewed the extracted structure, explicitly selected an outcome, and stored a durable record. Normalizing by active user prevents raw signup growth from hiding weak repeat value.

Launch measurement hierarchy



Launch metric hierarchy connecting activation, workflow value, AI trust, retention, and commercial outcomes.

Launch Funnel

Stage	Definition	Primary failure question
Eligible	User belongs to an approved launch account and has access.	Was recruitment or access incomplete?
Registered	User created and verified an account.	Did identity or invitation fail?
Configured	Workspace and required settings are ready.	Was setup confusing or blocked?

Stage	Definition	Primary failure question
Started	User submitted a source for a new decision.	Did the user lack a meaningful use case?
Reviewed	User completed extraction review.	Was AI output too weak or review too costly?
Recommended	User generated and inspected a recommendation.	Did trust or context fail?
Finalized	User explicitly selected and stored an outcome.	Did comparison or finalization block value?
Retained	User returns and completes another meaningful workflow or retrieves a decision.	Did the product fail to become a habit?
Converted	Organization accepts paid commercial terms.	Was value insufficient or pricing/approval unclear?

Metric Dictionary

Metric	Formula / definition	Launch target	Owner
24-hour activation	Registered users finalizing first decision within 24h / registered users	>= 70%	Product
Start-to-finalize completion	Finalized decision workflows / started workflows	>= 70%	Product
Median time to value	Median time from first decision start to finalization	<= 12 min	Product + UX
Week-4 retention	Activated users with meaningful activity in Week 4 / activated users	>= 60%	Product
Second-decision completion	Activated users finalizing a second decision by Week 12 / activated users	>= 55%	Product
Decision precision	Correct extracted decisions / extracted decisions	>= 90%	AI
Decision recall	Correct extracted decisions / ground-truth decisions	>= 85%	AI
Hallucination rate	Unsupported material outputs / evaluated outputs	<= 5%	AI + QA
AI acceptance without material edit	Outputs accepted without meaning-changing edit / outputs reviewed	>= 70%	AI + Product
Trust rating	Average post-workflow trust rating	>= 4.0 / 5	UX Research
p95 AI latency	95th percentile processing time	<= 35 sec	AI + Platform
Failed processing rate	Failed source/recommendation jobs / jobs started	< 1%	Engineering
Paid conversion	Paid customer accounts / onboarded sponsoring customer accounts	>= 30%	GTM
Cost per finalized decision	AI and variable infrastructure cost / finalized decisions	<= $0.60	Product + Finance

Analytics Event Taxonomy

Event	Required properties	Sensitive-data rule
launch_account_invited	account_id, segment, channel, invited_at	No customer content.

Event	Required properties	Sensitive-data rule
account_registered	account_id, user_role, source_channel	No password or personal profile beyond approved analytics identifiers.
account_preferences_configured	account_id, retention_setting, ai_disclosure_acknowledged	Do not log names or emails.
decision_started	decision_id, source_type, size_bucket, project_tag_present	Do not log decision title or source text.
extraction_completed	model_version, prompt_version, latency, fallback, item_counts	No extracted strings.
review_completed	edited_field_count, rejected_item_count, review_duration	No original or edited content.
recommendation_generated	version, model, latency, token_bucket, fallback	No rationale text.
recommendation_accepted	recommendation_version, changed_selected_option_boolean	No option names.
decision_finalized	workflow_duration, action_count, recommendation_version_count	No decision statement.
decision_retrieved	age_bucket, search_used, related_record_count	No search query unless explicitly anonymized.
support_request_created	priority, workflow_stage, category	No transcript/decision text in analytics.
subscription_converted	package, seat_bucket, contract_type	Commercial amount handled under approved finance policy.

AI Quality Measurement

Evaluation layer	Method	Minimum evidence
Offline benchmark	Human-annotated representative transcripts and notes.	Precision, recall, field accuracy, hallucination, subgroup breakdown.
Adversarial tests	Prompt injection, conflicts, sparse context, sensitive data, no options, very long inputs.	No instruction override or unsupported secret disclosure.
Production sampling	Consent-aware review of anonymized/approved outputs.	Weekly sample and documented error taxonomy.
User correction signals	Meaning-changing edits, rejections, owner/date changes, regeneration.	Field-level edit rates and top failure modes.
Recommendation evaluation	Groundedness, evidence coverage, trade-off reasoning, uncertainty, actionability.	Human rubric; not agreement with a preferred answer.
Model/prompt comparison	Version-controlled A/B or shadow evaluation.	Quality, latency, cost, and safety before promotion.

Experiment Backlog

ID	Hypothesis	Design	Primary metric	Guardrail
EXP-01	Guided onboarding increases activation.	Randomize eligible accounts to guided or self-service onboarding.	24-hour activation.	Support cost and satisfaction.
EXP-02	Evidence-first recommendation layout increases trust.	Show evidence before recommendation versus current order.	Trust rating and acceptance.	Review duration and error detection.
EXP-03	A second-decision prompt improves retention.	Prompt a second decision three to five days after first finalization.	Week-4 retention.	Reminder opt-out and satisfaction.
EXP-04	Transparent scoring increases	Explain score source and allow user override.	Compare-to-finalize	Decision time and confusion.

ID	Hypothesis	Design	Primary metric	Guardrail
	completion.		conversion.	
EXP-05	Decision templates reduce time-to-value.	Offer product, architecture, hiring, and vendor templates.	Median time to finalize.	Template-induced missing context.
EXP-06	PDF/DOCX import expands qualified usage.	Enable for controlled subset after parsing validation.	Started decisions per active user.	Parsing failures and sensitive data risk.
EXP-07	Usage-based package increases paid intent.	Compare decision limits with seat-only messaging.	Qualified conversion intent.	Comprehension and cost predictability.

Review Cadence

Cadence	Review	Decision owner
Daily during W1-W2	Errors, incidents, activation failures, AI failures, latency, cost, support.	Launch owner
Twice weekly W3-W6	Funnel, edit/reject patterns, onboarding, feedback, account risks.	Product + CS
Weekly	AI benchmark sample, model/prompt changes, trust, retention, experiments.	Product + AI + Data
Monthly	Commercial pipeline, conversion, MRR, retention, support economics, roadmap.	Leadership
Week 12 gate	All customer, product, AI, operational, and business evidence.	Phase gate committee

Scale / Iterate / Stop Framework

Decision	Criteria	Action
Scale	Core gates met, no unresolved critical risk, repeat behavior and commercial signal visible.	Expand cohorts and channels while preserving monitoring and quality.
Iterate	Problem/value signal exists, but one or more correctable workflow, AI, trust, retention, or GTM gates fail.	Freeze broad expansion; run prioritized experiments and fixes.
Stop / Reframe	Weak customer value, persistent untrusted AI, unsafe operations, or no credible willingness to pay.	Stop acquisition; preserve learning; reconsider target problem or product.

Data Governance

- Never send raw transcript text, decision text, names, emails, or secrets to product analytics.
- Use stable pseudonymous IDs and role/segment categories approved by the privacy policy.
- AI evaluation datasets require a documented source, consent basis, access control, retention period, annotation guide, and deletion process.
- Simulated portfolio metrics must remain clearly labeled and separate from actual product telemetry.
- Any public benchmark or testimonial requires reproducible evidence, methodology, and permission.