DECISIONFLOW

Qualification Risk, Evidence and Assumption Register

Evidence strength, unresolved uncertainty, next tests, ownership, and decision triggers

QUALIFY PHASE | Version 1.0 | 14 July 2026

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
Document code	QUAL-05
Version / status	2.0 — Final Public Portfolio Edition — Simulated Qualification Scenario
Baseline / author	14 July 2026 — Min Thu Kyaw
Role / purpose	AI Product Manager — Qualification Risk, Evidence and Assumption Register
Lifecycle continuity	Actual MVP → DEV-06 illustrative productionization → simulated v1.0 launch → qualification decision.
Evidence rule	Actual artifacts, assumed capabilities, simulated outcomes, and recommendations are labelled separately.

Actual Project Artifacts

Artifact	Reference
Low-Fidelity Prototype	View low-fidelity prototype
High-Fidelity Prototype	View high-fidelity prototype
MVP Frontend Repository	Open frontend repository
MVP Backend Repository	Open backend repository

Contents

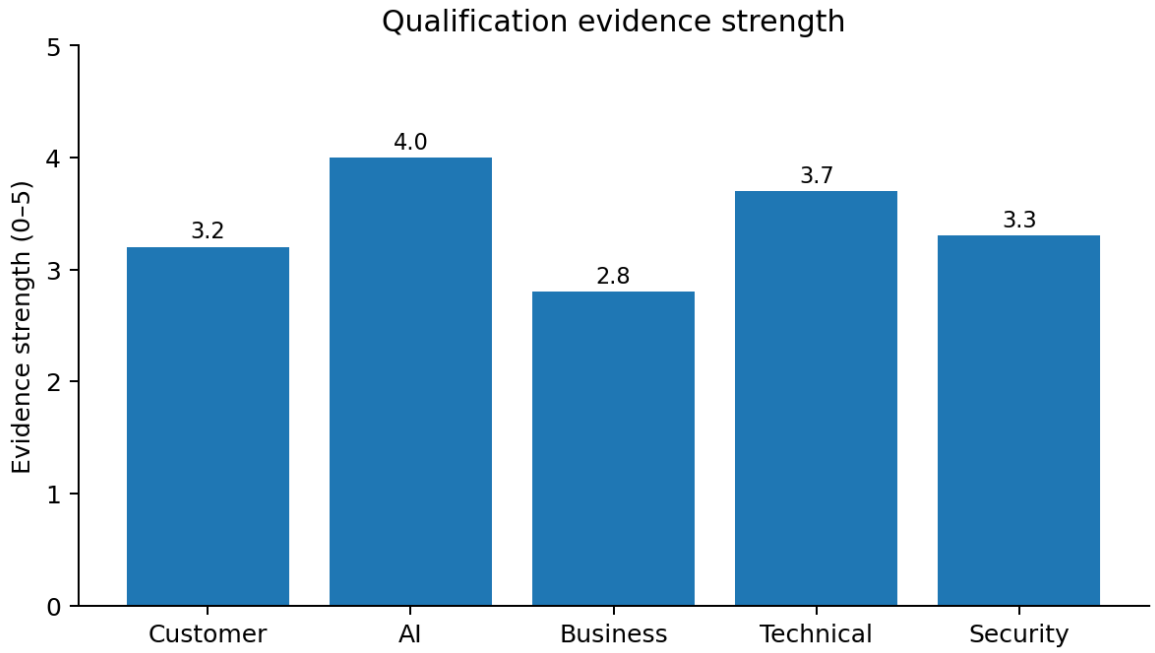
Register Purpose

This register prevents simulated launch success from becoming an unsupported scale claim. Each uncertainty is linked to available evidence, evidence strength, impact if wrong, the next test, an accountable role, and a decision trigger. The register should be updated whenever a metric definition, model version, package, segment, or operating boundary changes.

Evidence Quality Model

Rating	Definition	Decision use
Strong	Reproducible, representative, longitudinal, and independently reviewable.	Can support scale decisions within the tested boundary.
Moderate	Consistent but limited by sample size, duration, simulation, or segment concentration.	Supports controlled investment and further testing.
Weak	Directional, modeled, self-reported, or based on few observations.	Cannot justify broad scale.
Missing	No valid evidence or an invalid metric for the product scope.	Must be tested or removed from the decision.

Evidence Strength



Evidence strength

Domain	Rating / 5	Reason
Customer	3.2	Behavior and interviews align, but sample is controlled and role-concentrated.
AI	4.0	Structured benchmark, adversarial tests, production sample, and human rubric.
Business	2.8	Only eight original paying organizations; CAC is modeled.
Technical	3.7	Normal-load, stress, restore, and incident evidence; high-load gate fails.
Security	3.3	Core controls reviewed, but no enterprise scope or long operating history.

Qualification Risk Register

ID	Risk	Evidence	Strength	Impact	Next test / owner	Trigger
QR-01	Repeat use may decline after	58.9% W12; 49.3% W24 retention.	Moderate	High	Templates + recurring-review cohort /	W24 <45%

ID	Risk	Evidence	Strength	Impact	Next test / owner	Trigger
	novelty.				Product	
QR-02	Due-date suggestions may mislead accountability.	87.8% accuracy; 34% edited.	Strong	High	Relative-date benchmark / AI	Accuracy <90%
QR-03	Confidence may be over-trusted.	ECE 0.12; comprehension 83%.	Moderate	High	Calibration and copy test / AI+UX	High-confidence error >2%
QR-04	Commercial conversion may not repeat in open channels.	8 of 15 controlled accounts paid.	Weak	High	Repeatable-channel cohort / GTM	CAC payback >12m
QR-05	High-touch onboarding reduces margin.	47% support-adjusted contribution.	Moderate	High	Self-service onboarding test / Product+CS	Contribution <55%
QR-06	AI concurrency creates latency and failures.	40-job test failed.	Strong	High	Async queue load test / Platform	p95 >45s or fail ≥1%
QR-07	Provider dependency interrupts core value.	One primary provider.	Moderate	High	Failover game day / AI+Platform	Outage >30m
QR-08	Non-English AI quality may be overstated.	All three tested languages below gate.	Strong	Medium	Language-specific benchmark / AI	Any marketing claim before pass
QR-09	Integration demand may distract from core retention.	Six interview requests; no causal proof.	Weak	Medium	Export/webhook discovery / Product	Build starts without experiment
QR-10	Simulated data may be mistaken for real traction.	Portfolio case-study format.	Strong	High	Disclosure audit / Author	Any unlabeled public reuse

Assumption Register

ID	Assumption	Current evidence	Status	Next evidence
QA-01	High-frequency decision owners are the best beachhead.	70% PM retention; 57% founder/lead retention.	Supported	Repeat in second cohort.
QA-02	Durable retrieval drives retention.	44% weekly retrieval; strongest interview theme.	Supported	Causal onboarding experiment.
QA-03	Professional pricing reflects value.	Five of eight describe price as acceptable.	Partially supported	15+ paying accounts and price test.
QA-04	Templates increase habit formation.	Seven interview requests; no experiment result.	Unproven	Run recurring-template experiment.
QA-05	Integrations increase retention.	Conversion blocker/request only.	Unproven	Test export/webhook before native build.
QA-06	Collaboration will expand account value.	1.6 paid users/account; collaboration not built.	Unproven	Discovery only; no launch claim.
QA-07	English quality generalizes across domains.	240-record benchmark with domain mix.	Partially supported	Add legal/finance/high-stakes exclusions and subgroup gates.
QA-08	Async jobs unlock controlled scale.	Current synchronous load fails at 40 jobs.	Hypothesis	Implement and rerun capacity gate.

Decision Triggers

Trigger	Required action
Week-24 retention below 45% for two qualified cohorts	Pause acquisition and prioritize recurring-value redesign.
Material hallucination above 5% or high-severity unsupported output	Freeze model/prompt promotion; review affected workflows.
Confirmed unauthorized data access	Stop affected processing and activate security incident response.
Support-adjusted contribution margin below 45%	Freeze broad onboarding; automate or reprice.
40-job load gate fails after async architecture	Reduce growth cap and reassess provider/architecture.
Paid logo retention below 75% at 90 days	Reassess ICP, package, recurring value, and pricing.
Simulated evidence reused without disclosure	Correct publication immediately and add disclosure control.

Governance Cadence

Cadence	Review
Weekly	Customer activation, retention, AI failures, edits, latency, incidents, support, and experiments.
Biweekly	Risk score, assumption changes, model/prompt versions, and roadmap evidence.
Monthly	Commercial cohort, margin, CAC evidence, renewal, capacity, and security actions.
Quarterly gate	Scale / Iterate / Pivot / Stop decision with evidence-quality review.

Evidence discipline: replace all simulated values with observed, reproducible evidence before using this material as a real product-performance claim.