

DECISIONFLOW

Technical Scalability and Operational Qualification Report

Reliability, performance, security, recovery, capacity, and operational limits

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-04
Version / status	2.0 — Final Public Portfolio Edition — Simulated Qualification Scenario
Baseline / author	14 July 2026 — Min Thu Kyaw
Role / purpose	AI Product Manager — Technical Scalability and Operational Qualification Report
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

01	Qualification scope
02	Production reliability
03	Load and capacity
04	Security and privacy
05	Recovery and continuity
06	Observability and support
07	Scale limits
08	Technical decision

Qualification Scope

Technical qualification tests the assumed production-ready v1.0 core defined in DEV-06. It does not qualify collaboration, enterprise tenancy, PDF/DOCX ingestion, SSO, native integrations, or regional processing. The scenario evaluates normal pilot load, controlled growth load, recovery, security controls, AI-provider dependency, and support operations.

Production Reliability

Measure	Target	Simulated result	Status
Availability	≥99.9%	99.95%	Pass
p95 non-AI API latency	<500 ms	390 ms	Pass
p95 AI latency at qualified load	≤35 sec	32 sec	Pass
Failed processing rate at qualified load	<1%	0.8%	Pass
SEV-1 incidents	0	0	Pass
SEV-2 incidents	Declining / contained	1 during Weeks 13–24	Pass with action
Duplicate finalization	0	0 after idempotency correction	Pass

Load and Capacity



Profile	Concurrent sessions	Concurrent AI jobs	p95 API	p95 AI	Failure rate	Result
Qualified operating range	100	20	460 ms	32 sec	0.8%	Pass
Controlled growth test	250	40	690 ms	56 sec	1.9%	Fail
Stress test	500	80	1.4 sec	112 sec	4.8%	Fail

Capacity decision — The current architecture is qualified for controlled growth only. Before increasing cohort size materially, AI generation should move to an asynchronous job model with idempotency, queue depth metrics, back-pressure, cancellation, and provider-health routing.

Security and Privacy

Control	Simulated evidence	Qualification
Authentication/session	HttpOnly secure session assumption from DEV-06; expiry and logout tested.	Pass within scenario.
Authorization	Owner isolation tests across decision, transcript, and recommendation records.	Pass; zero cross-user access in test set.
Secrets	Production default-secret checks and managed environment configuration.	Pass.
Rate limits	Auth and AI generation thresholds tested.	Pass at qualified range.
Data retention	Published policy and user deletion path.	Pass for individual accounts.
Provider disclosure	AI processing and fallback provenance visible.	Pass.
Security review	Two medium findings closed; no critical/high open findings.	Pass with annual review.
Enterprise controls	SSO, organization policy, regional controls not implemented.	Out of scope; do not market.

Recovery and Continuity

Test	Target	Simulated result	Status
Database restore RTO	≤60 min	42 min	Pass
Recovery point objective	≤5 min	4 min	Pass
Rollback drill	≤30 min	24 min	Pass
Provider outage recovery	Actionable status + retry/fallback	Pass in scenario	Pass
Migration rehearsal	Clean and upgrade path succeed	Pass	Pass
Finalized-record integrity after restore	100%	100% sampled records consistent	Pass

Observability and Support

Area	Coverage	Residual gap
Application telemetry	Request errors, latency, job failures, auth, workflow funnel.	Add queue metrics before async processing.
AI telemetry	Provider, model, prompt, usage, fallback, latency, cost.	Improve calibration monitoring and subgroup quality.
Business telemetry	Activation, completion, retention, conversion, MRR.	Formalize metric definitions and data tests.
Incident response	Severity model, commander, communication, RCA, follow-up.	Run cross-functional game day every quarter.
Support operations	Priority routing, knowledge base, privacy-safe reproduction.	Reduce guided onboarding and high-touch support.

Scale Limits

Boundary	Qualified now	Not qualified
User model	Individual account and private decision records.	Shared workspaces, role permissions, enterprise tenancy.
Input	Pasted text and UTF-8 TXT within defined limits.	PDF/DOCX and meeting-platform ingestion.
Language	English AI-content quality.	Commercial multilingual AI-quality claims.

Boundary	Qualified now	Not qualified
Capacity	100 concurrent sessions / 20 concurrent AI jobs.	High-volume public traffic and enterprise concurrency.
Operations	Controlled production, monitoring, backups, support.	Multi-region active-active and formal enterprise SLAs.

Technical Qualification Decision

Technical decision — QUALIFIED FOR CONTROLLED GROWTH within the documented individual v1.0 scope and tested operating range. NOT QUALIFIED for broad scale or enterprise commitments until asynchronous AI processing, higher-load performance, provider resilience, and organization-level controls are demonstrated.

Next technical gate	Threshold
40 concurrent AI jobs	p95 ≤45 sec; failures <1%
Queue recovery	No duplicate job; safe retry and cancellation
Provider failover drill	No silent provenance change; recovery within runbook target
90-day security review	No open critical/high findings
Restore drill	RTO ≤45 min; RPO ≤5 min
Observability	Metric-definition tests and alert coverage for all release gates

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