AI QA CO-PILOT

Business Case, Build Plan & Cost Estimate

Prepared for review by the founding team

Document	AI QA Co-Pilot — Business Case, Build Plan & Cost Estimate
Version	2.0 — FOR REVIEW
Date	30 August 2026
Prepared by	Mr. Min Khant Kyaw — Developer / CTO
For	Ms. Khaing Myint San Oo — Senior QA Engineer / Funding Partner
Total requested	THB 400,000 over 20 weeks
Status	Internal. Not for distribution.

> *Costs are estimates at approximately THB 33 = USD 1 and should be re-checked against live vendor pricing before commitment. Figures marked (*) are the author's own estimates rather than published data.*

Contents

1. Executive Summary

This document sets out the business case, technical plan and cost estimate for AI QA Co-Pilot: a web platform that helps software teams move from requirements to test cases, manual execution, evidence capture and bug reporting, with a longer-term path into regression management and test automation. The product is English-language only.

1.1 The ask

Item	Duration	Amount (THB)
Phase 1 — Client demo (incl. 2 weeks stabilisation)	10 weeks	113,353
Phase 2 — Pilot-ready product (incl. 2 weeks stabilisation)	10 weeks	237,256
Sub-total — planned work	**20 weeks**	**350,609**
Reserve — unplanned contingency	~5 weeks runway	49,391
TOTAL COMMITMENT REQUESTED		**400,000**

Approximately USD 12,120. In exchange, the developer contributes five months of full-time engineering at THB 30,000 per month.

1.2 Key findings

- The core idea is sound, but the originally-scoped POC is a commodity. AI test-case generation from a user story shipped as a feature across most major test management platforms during 2025–2026 and is trivially reproducible with a general coding agent. The business cannot rest on that feature alone.
- The category is crowded and well-funded. Over USD 1.5 billion of venture capital has entered AI testing, with 40+ startups competing. QA Wolf alone has raised approximately USD 57 million.
- The most successful company in the category operates as a services business with a product interface, not as pure self-serve SaaS. That is a strong signal for our own model.
- Two defensible positions remain open to us, and both are structurally unattractive to the funded global players: the manual-testing evidence and bug-reporting loop that the market is abandoning in favour of autonomous automation, and PDPA compliance posture with in-country or on-premise deployment for regional enterprise buyers.
- Recommended path: launch as an AI-augmented QA services business that builds the platform as its own internal tooling, and productise only the modules customers repeatedly pay to self-serve. This produces revenue in months rather than years.

1.3 What we are asking the partner to approve

- The THB 400,000 budget, committed and available in full before week 1 rather than drawn down monthly.
- A signed founders' agreement in week 1 covering equity split, vesting, IP assignment and exit terms.
- Delivery of a 20-story golden dataset in week 1 — real user stories with the test cases a good QA engineer would write. This is the benchmark all generation quality is measured against.
- Ownership of both stabilisation windows and the test plan executed during them.

2. Market and Competition

2.1 Addressable market *

Bottom-up estimate of organisations with enough engineering headcount to run a QA function and hold a tooling budget:

Segment	Est. accounts	Realistic ACV (USD)	Segment value
Thailand — software houses & agencies (20+ staff)	400–800	2,000–6,000	USD 1.5M–4M
Thailand — enterprise IT (banking, insurance, retail, telco)	200–400	15,000–60,000	USD 4M–20M
Thailand — funded startups and scale-ups	300–600	1,500–5,000	USD 0.7M–2.5M
SEA-6 extension (VN, ID, MY, PH, SG)	4,000–10,000	2,000–15,000	USD 20M–90M

Thailand SAM approximately USD 6M–26M per year; SEA-6 approximately USD 30M–120M. A 2% share of Thailand over three years is roughly USD 150,000–500,000 ARR; 2% of SEA-6 is roughly USD 600,000–2.4M. This can be a profitable independent company, but it is not venture-scale if confined to Thailand. The enterprise segment carries four to ten times the contract value of SMB for comparable sales effort.

> *Note on the English-only decision: regional language support was identified as a potential differentiator, since global tools handle Thai, Vietnamese and Bahasa requirements poorly. Building English-only removes that wedge and means we compete on workflow, evidence quality and compliance alone. This is a defensible choice — most SEA software houses and outsourcing teams already work in English — but it should be a conscious one, and it narrows our advantage against well-funded global competitors.*

2.2 Competitive landscape

Tier 1 — test management incumbents, our direct feature competitors. All shipped AI generation during 2025–2026:

Tool	List price	AI position
TestRail	USD 37/seat/mo; Enterprise USD 74	AI test case, script and BDD generation
Qase	USD 20–36/user/mo; free to 3 users	Agentic Mode on Teams and Enterprise
Testsigma	Free-forever tier available	AI coworker, two-way Jira sync
Xray / Zephyr	Jira-native marketplace pricing	Native Jira AI features
Katalon	Tiered	MCP server — external agents drive the platform
Tricentis	Enterprise	Agentic test creation (March 2026)

Price anchor: USD 20–40 per seat per month is the SMB band, USD 70–90 enterprise. Anything above that needs hard justification.

Tier 2 — AI-native competitors. QA Wolf: approximately USD 57M raised, 130+ customers, estimated USD 15–20M ARR, selling an outcome (80% end-to-end coverage in four months, maintained, zero-flake guarantee) at roughly USD 90,000 per year, delivered by AI plus human engineers. Momentic: USD 18.7M

raised, customers including Notion, Xero and Webflow. Others include mabl, testRigor, Ranger and qa.tech.

Tier 3 — the existential risk. General coding agents already write Playwright tests, read Jira tickets and generate test plans. Katalon's MCP server lets a developer drive a commercial test platform without opening its interface. Any product whose core value is an LLM converting text into structured text has a shelf life measured in quarters. We must compete on workflow state, evidence artifacts, integrations and accountability for outcomes — never on raw generation quality.

3. Product Scope

3.1 The workflow

> *Jira ticket → AI-generated test cases → manual execution with evidence capture → structured bug report pushed back into Jira with evidence attached.*

Everything else — regression recommendation, automation script generation, ClickUp, dashboards, billing — is deferred. The generation step alone is a commodity. The bug report arriving in Jira with evidence already attached is what saves a working QA engineer twenty minutes per bug, and it is the part nobody else does well.

3.2 Phase 1 — client demo, 10 weeks

A credible end-to-end demonstration we can run live on a prospect's own Jira ticket. Everything listed works; nothing is mocked.

Module	Scope	Days
Auth + workspace	Single workspace, email login, no invites or roles	3
Jira integration	OAuth 3LO, browse project, select ticket, ADF parsing	6
Test case generation	Structured output, streaming, prompt caching	4
Evaluation harness	20 golden stories, acceptance scoring	3
Test case UI	View, inline edit, delete, group into suites	6
Test run execution	Create run, step through, pass / fail / blocked, notes	6
Evidence — screenshots	Upload and attach to a step, basic annotation	4
Bug report composer	Auto-draft from failed case, edit, push to Jira	5
Build sub-total		**37**
Stabilisation (weeks 9–10)	Bug bash, P0/P1 fixes, demo rehearsal, cross-browser	10
Float	Unallocated slack	3
TOTAL	**10 weeks / 50 working days**	**50**

Excluded from Phase 1: video recording, multi-tenant UI (though the database schema is multi-tenant from week 1 to avoid a rewrite), browser extension, billing, dashboards, SSO, regression and automation.

3.3 Phase 2 — pilot-ready, 10 weeks

What turns the demo into something a customer can run a real sprint on unsupervised.

Module	Scope	Days
Real multi-tenancy	Organisation switching, invitations, roles	5
Video evidence	Screen capture, chunked upload, storage, playback	7
Media pipeline	Transcoding worker, thumbnails, lifecycle deletion	4
Jira robustness	Custom field mapping, webhooks, error recovery, rate limits	4
PDPA and security	Retention jobs, audit log, encryption at rest, rate limiting	4
Reliability	Retries, job failure handling, monitoring, alerting	2

Module	Scope	Days
Build sub-total		**26**
Pilot support	Onboarding first customers, feedback, iteration	8
Stabilisation (weeks 19–20)	Tenant isolation, error paths, soak test, PDPA verification	10
Float	Unallocated slack	6
TOTAL	**10 weeks / 50 working days**	**50**

The dividing line: does the demo need it to be believed, or does the customer need it to be unsupervised? A prospect watching us drive the product will accept a single workspace and manually uploaded screenshots. A QA team running their own sprint will not accept losing evidence to a failed upload, or finding that bug reports cannot map to their custom Jira fields.

3.4 Testing and stabilisation

Sixty-three days of module work sit inside 100 available working days, giving roughly 24% stabilisation and float. For a single developer on a new codebase this is the right allocation; industry practice for solo delivery is 20 to 30%. Buffer only works if its contents are defined in advance, so the following work is committed and is not available for new features.

Window	Days	Activity
Phase 1, weeks 9–10	1–2	Structured bug bash. Both founders, all findings logged and triaged.
	3–5	Fix P0 and P1 defects only. P2 and below go to a backlog, not into this window.
	6–7	Demo rehearsal against three unfamiliar Jira tickets — not our prepared scenarios.
	8	Cross-browser testing, varied screen sizes, throttled network.
	9	Prompt re-tuning based on failures exposed by the bug bash.
	10	Contingency.
Phase 2, weeks 19–20	1–3	Tenant isolation. Verify organisation A cannot access organisation B's data under any circumstance. This is the single defect class that would end the company.
	4–6	Error paths: Jira API unavailable, upload interrupted mid-video, OAuth expiry, job failure mid-execution.
	7	Soak test: fifty concurrent runs, large uploads, sustained load.
	8	Verify PDPA retention deletion genuinely deletes, including object storage and backups.
	9–10	Pilot onboarding documentation and contingency.

Ownership: the QA partner owns both stabilisation windows and writes the test plan executed during them. The objective is not a defect-free product. Two narrower bars apply — at the end of Phase 1, the happy path never breaks in front of a client; at the end of Phase 2, error paths never lose customer data.

4. Technical Feasibility

Verdict: the full 20-week scope is comfortably within reach for one experienced full-stack developer. Nothing requires machine-learning research, novel infrastructure or a larger team. The risks are scope creep and prompt quality, not technical capability.

Module	Difficulty	Where the time goes
Jira integration	Medium	OAuth is straightforward. The pain is Atlassian Document Format — descriptions return as a nested JSON tree, not text. Budget a full day for a clean converter.
Test case generation	Easy	The code is one day. The quality is three weeks of iteration. This is why the evaluation harness matters more than any other three days in the plan.
Manual execution runner	Easy	CRUD plus a state machine. The easiest week in the plan.
Evidence capture	Medium-Hard	The hardest module. Browser screen capture works in Chrome and Edge but requires re-selection each session. Safari support is poor. Chunked upload with resume is fiddly.
Video pipeline	Medium	Presigned uploads, transcoding on a worker. Cap at 720p and ten minutes. Never transcode client-side.
Bug report to Jira	Medium	Creating the issue is easy. Mapping our fields to each customer's custom Jira fields generates support tickets forever.
Multi-tenancy	Medium	Get it right in week 1. Retrofitting is a rewrite.
Automation generation	Medium-Hard	Partner's domain. Deferred until after first pilot.

4.1 Stack and week-1 decisions

- Next.js 15 App Router with TypeScript; PostgreSQL with an ORM; managed auth provider (buys SSO later without rework); managed job queue; Cloudflare R2 for object storage, chosen for zero egress on video evidence.
- Build multi-tenant from day one, even for the demo.
- Abstract the AI provider. Model prices and capabilities changed several times in the past twelve months; no model name should appear outside a single file.
- Enforce evidence retention in code, not in a policy document. A scheduled deletion job is both a cost control and our PDPA answer.

5. Timeline

Week	Deliverable	Milestone
1	Founders' agreement signed. Partner delivers golden dataset. Repository, CI, auth, schema.	Legal clear
2–3	Jira OAuth, project browse, ticket detail, ADF parsing	Reads real Jira
4–5	Test case generation and evaluation harness	Acceptance rate measured
6–7	Test case management interface	Begin booking demos
8	Test run execution, screenshot evidence	
9	Bug report composer, push to Jira with attachments	Feature complete
10	**STABILISATION — bug bash, fixes, demo rehearsal**	**DEMO READY**
11–12	Video evidence capture, annotation, upload pipeline	
13	Multi-tenancy, invitations, roles	
14–15	Jira robustness — custom fields, webhooks, error recovery	
16	PDPA, security, retention jobs, rate limiting	
17	Reliability, monitoring, alerting	End-to-end loop complete
18	First pilot onboarding, support, iteration	Pilots live
19–20	**STABILISATION — tenant isolation, error paths, soak test**	**PILOT READY**

> *We do not wait until week 20 to sell. From week 6 there is enough to book meetings; by week 10 there is enough to demonstrate. Weeks 11 to 20 should be shaped by feedback from those meetings. Two or three planned modules will likely be displaced by something a prospect asks for — that is the plan working, not failing.*

6. Cost Estimate

All infrastructure is costed on paid commercial tiers rather than free plans, since free tiers pause on inactivity, prohibit commercial use, or display vendor branding on the login screen.

6.1 Phase 1 — 10 weeks (2.5 months)

Line item	Detail	THB
Developer fee	THB 30,000 × 2.5 months	75,000
QA partner	Contributing time, no salary drawn	0
Claude Code	Pro, USD 20/mo	1,650
GitHub Team	USD 4 × 2 seats	660
Secrets management	Bitwarden or Infisical, free tier	0
AI / LLM API	~1,800 generations incl. prompt iteration and demos	5,000
Server and infrastructure	THB 6,303/mo × 2.5 — see 6.1.1	15,758
Domain	One year	500
Sub-total		**98,568**
Contingency 15%		14,785
PHASE 1 TOTAL		**113,353**

> *If Claude Code Pro session limits prove restrictive during the generation-heavy weeks, upgrading to Max 5x (USD 100/month) raises the Phase 1 total to THB 120,943. Claude Code shares one usage pool with Claude's web and desktop apps, metered against a rolling five-hour session limit plus a weekly cap; third-party estimates place Max 5x at roughly five times Pro's capacity. Anthropic publishes no token figures, so actual consumption should be checked in the application's usage settings before deciding.*

6.1.1 Server and infrastructure

Service	Plan	USD/mo	THB/mo	Reason for paid tier
Vercel	Pro	20	660	Hobby prohibits commercial use
Supabase	Pro	25	825	Free tier pauses after 7 days idle
Auth provider	Pro	25	825	Removes vendor branding from login
Cloudflare R2	Pay as you go	5	165	Zero egress on evidence files
Job queue	Basic	20	660	Retries and concurrency
Redis	Fixed 250MB	10	330	Rate limiting and caching
Sentry	Team	26	858	Catch errors before a client does
Transactional email	Pro	20	660	Custom domain, deliverability
Uptime + status page	Standard	25	825	Public status page is a B2B trust signal
Worker instance	Shared CPU	15	495	Background processing
Analytics, DNS	Free tiers	0	0	Sufficient at this scale

Service	Plan	USD/mo	THB/mo	Reason for paid tier
TOTAL		**191**	**6,303**	

6.2 Phase 2 — 10 weeks (2.5 months)

Line item	THB
Developer fee (THB 30,000 × 2.5 months)	75,000
Development tooling	2,310
AI / LLM API (video, higher volume, stabilisation)	9,000
Server and infrastructure (2.5 months, pilot scale)	20,000
Contract UI designer	50,000
Brand and landing page	10,000
Legal — founders' agreement, NDA, pilot agreement template	40,000
Sub-total	**206,310**
Contingency 15%	30,946
PHASE 2 TOTAL	**237,256**

6.3 Combined

Phase	Duration	THB	Approx. USD
Phase 1 — Client demo	10 weeks	113,353	3,435
Phase 2 — Pilot-ready	10 weeks	237,256	7,190
Sub-total — planned work	**20 weeks**	**350,609**	**10,624**
Reserve — unplanned contingency	~5 weeks	49,391	1,497
TOTAL COMMITMENT		**400,000**	**12,121**

The developer fee is 66% of planned cost. Non-labour spend across both phases is THB 200,609.

6.4 Ongoing cost after week 20

Line	Monthly (THB)
Server and infrastructure at pilot scale (3–5 customers)	6,400 – 13,900
AI / LLM API (generation workload only)	2,000 – 6,000
Development tooling	924
TOTAL	**9,324 – 20,824**

7. Pricing and Unit Economics

7.1 Cost of goods sold

Workload	Tokens per operation	Cost per operation
Generate test cases from one user story	~9,000	~THB 2.50
AI agent executing one browser test	200,000 – 500,000	THB 20 – 53

A customer running fifty automated tests nightly would cost us THB 30,000–80,000 per month in inference against a THB 1,290 per-seat subscription. This is exactly how AI products discover negative gross margins. Agentic execution must be sold as metered credits, never bundled into a flat seat price. A cheaper alternative for demonstration is AI-generated Playwright scripts: the model writes the code once — roughly THB 700 for fifty scripts — and the scripts then run for free.

7.2 Proposed pricing

Plan	Global	Thailand / SEA	Includes
Free	USD 0	THB 0	1 project, 3 users, 50 generations/mo
Team	USD 29/user/mo	THB 590/user/mo	Unlimited projects, Jira sync, 90-day evidence
Business	USD 59/user/mo	THB 1,290/user/mo	Regression, SSO, 1-year retention
Enterprise	From USD 2,000/mo	From THB 60,000/mo	On-premise, DPA, custom retention, SLA
Automation credits	USD 0.30–1.00/run	Same	Metered. Never bundled unlimited.
Managed QA pod	USD 6,000–12,000/mo	THB 180,000–350,000/mo	2 QA engineers plus platform, guaranteed coverage

The managed pod is our highest-conviction revenue line. SEA QA outsourcing bills at USD 18–70 per hour. A two-engineer pod costs roughly THB 180,000 per month loaded and bills at USD 6,000–12,000 — a 45 to 65% gross margin, collected monthly, starting immediately. QA Wolf built an estimated USD 15–20M ARR business on essentially this structure. Regional pricing is essential: Thai SMBs resist USD 30 per seat per month, while THB 590 lands very differently for similar money.

8. Team and Partnership Terms

	Developer / CTO	Senior QA / Funding Partner
Cash	THB 0	THB 400,000
Labour	20 weeks nearly full-time	Part-time throughout
Expertise	Engineering, architecture	Manual and automation QA, customer perspective
Ongoing role	CTO, sole developer	Product, domain, sales, test plans, automation module

Contributions are roughly comparable in value once below-market labour is accounted for. A balanced split is defensible. What is not defensible is leaving the question undiscussed for six months and negotiating it once the product works — that is precisely when partnerships break.

8.1 To settle in writing before week 2

- Equity split, agreed explicitly.
- Whether the cash is equity, a loan, or a convertible instrument — and if equity, at what valuation.
- Whether the developer's THB 30,000 monthly draw is salary or an advance against equity.
- Vesting: four years with a one-year cliff, applying to both founders. This is the clause that protects each party if the other leaves in month five.
- IP assignment: all code, prompts and datasets assign to the company, not to either individual.
- Deadlock and exit: what happens if either party wants out, who buys whom, at what price.

Budget THB 30,000–40,000 for a Thai lawyer to draft this. It is the cheapest insurance the company will ever buy.

8.2 Division of labour

The QA partner owns, from week 1:

- The golden dataset — twenty real user stories with the test cases a good QA engineer would write. The highest-leverage contribution available to a part-time partner, and the benchmark all generation quality is measured against.
- Prompt output review and the acceptance rubric.
- Both stabilisation windows and the test plans executed during them.
- The credible domain voice in every client demonstration, and customer discovery interviews.
- The Playwright automation module after the first pilot — the partner writes automation code, so this is a natural handover rather than a hire.

The developer owns everything else.

8.3 Employment and IP clearance

Before any code is written, the developer's current employment contract must be reviewed for IP assignment, invention-assignment and moonlighting clauses. Thai employment contracts frequently assign inventions made during employment to the employer. No employer equipment, accounts, networks or working hours may be used. This is the single most likely way for the venture to end badly, and it is entirely preventable.

9. Risks and Kill Criteria

Risk	Severity	Mitigation
Core feature commoditised by incumbents or coding agents	Critical	Compete on workflow, evidence and compliance — never on raw generation quality
Negative gross margin from agentic inference	Critical	Metered credits, model routing, prompt caching, hard per-account caps from day one
English-only removes a regional differentiator	High	Compensate with depth on evidence and compliance; revisit language support after first pilots
Funded competitor enters SEA	High	Local incorporation, on-premise option, PDPA posture — all slow to copy
Enterprise sales cycle exceeds runway	High	Services revenue funds the wait; do not depend on subscription ARR in year one
Single developer — bus factor of one	High	Document as we go, conservative stack, no clever abstractions
Evidence capture overruns its estimate	Medium	Timebox it. If week 12 arrives without working capture, ship screenshot-only and add video post-pilot
Employment IP conflict	High	Resolve in writing before building
Scope creep from client demos	Medium	Every request goes on the roadmap, not into the 20-week plan
Partner's cash runs out mid-build	High	Full budget committed and available before week 1

9.1 Kill criteria

Agreed now and honoured. A project that cannot fail cannot teach us anything.

- Generated test-case acceptance rate stays below 60% after two rounds of prompt engineering.
- No design partner will commit to a paid pilot within 90 days of a working demo.
- Gross margin cannot be modelled above 60% at one hundred customers.
- Six months in with under THB 150,000 monthly revenue across services and product combined.

> *The evaluation harness in week 4 exists to surface the first of these by week 5 — at roughly THB 75,000 spent rather than THB 400,000. It is three days of work and the only module that produces a number rather than an opinion. Whatever else is cut, it should not be.*

10. Decision and Next Steps

Approve THB 400,000 over twenty weeks: THB 350,609 of planned work including four dedicated stabilisation weeks, plus THB 49,391 held as genuine reserve. Build only the Jira-to-generation-to-execution-to-evidence-to-bug-report loop, in English. Measure test-case acceptance rate by week 5. Start booking demonstrations in week 6 and demonstrating in week 10. Register a company entity when a customer wants to pay us, and build the browser extension with their money.

10.1 Immediate actions

#	Action	Owner	By
1	Review and comment on this document	Both	[DATE]
2	Agree equity split and engage a lawyer for the founders' agreement	Both	[DATE]
3	Confirm full budget is committed and available	Partner	[DATE]
4	Review employment contract for IP and moonlighting clauses	Developer	[DATE]
5	Prepare the 20-story golden dataset with expected test cases	Partner	[DATE]
6	Draw up a list of 15 target companies for discovery interviews	Partner	[DATE]
7	Agree the stabilisation scope in section 3.4 and confirm ownership	Both	[DATE]
8	Set up repository, CI, and provision paid infrastructure accounts	Developer	[DATE]
9	Week 1 build begins	Developer	[DATE]

10.2 Open questions

- Do we agree on the services-first strategy, or does the partner prefer product-only?
- What equity split do we consider fair given the contributions in section 8?
- Is the cash an investment, a loan, or founder capital?
- Which three companies do we target for the first pilots, and who owns those relationships?
- Are we comfortable with the kill criteria in section 9.1, and with shipping English-only?

End of report. Version 2.0 — DRAFT FOR REVIEW. All figures are estimates and should be verified against live vendor pricing before financial commitment.