EMLYON BUSINESS SCHOOL · EMBA DISSERTATION · SUPERVISION #2

Progress update · supervision #2

Innovating from the core: a design-thinking approach to an AI-adaptive music product for anxiety relief, validated from inside TDMusic — a profitable AI music-distribution company. Working product name: **Lilt**.

AUTHOR
Qiao Lu
EMBA candidate, emlyon; Founder & CEO, TDMusic

DATE
6 September 2026
Previous briefing: 10 July 2026

PHASE
Design thinking · Test
Decision memo due early October

VERDICT TODAY
NOT YET
3 of 5 posterior gates met; pilot not run

Hub http://38.180.150.12/ · **Dissertation** /pages/dissertation.html · **Reasoning chain** /pages/journey.html · **Decision engine** /pages/decision.html
App (installable) https://38.180.150.12.sslip.io/app/

Update 7 Sep 2026. The two pre-registered survey rows flagged by the audit were applied on the owner's decision: A3 0.72 → 0.65, A1b 0.80 → 0.77; concept joints C3 0.65 → 0.63, C1-neutral 0.46 → 0.41, C1-real 0.26 → 0.23. No gate flips; verdict unchanged (NOT YET). Figures elsewhere on this page are as of 6 Sep unless stated.

02 Where the project stands, in one breath

Empathize, Define and Ideate are closed; Prototype went further than planned — the product is a shipped app in App Store review, not a mock-up; Test is half in. The survey and the two focus groups are analysed and their likelihood ratios applied; the pilot, the rights check and the landing A/B have not reported. The pre-registered verdict is therefore **NOT YET**: engagement (A3 = 0.72 ≥ 0.55), willingness to pay (A4 = 0.82 ≥ 0.60) and rights (A5 = 0.65 ≥ 0.60, check open) clear their gates, while the two efficacy beliefs sit below gates deliberately set above what the meta-analytic literature alone can reach (A2a 0.85 against 0.90; A2b 0.56 against 0.70), so PROCEED cannot be cleared without running the pilot. One pre-registered rule has fired: the content pivot — real music in the acute state fell to **8 per cent**, so the acute product is engineered neutral sound and the catalogue becomes the wind-down mode. Built: a two-mode app (web v0.5.1, App Store 1.0 build 4 in review), a live Bayesian decision engine, a pre-registered survey pipeline, a pilot analysis pipeline with 28 tests, and the dissertation drafted in full. Pending: E0 and E1, the rights check, the landing test, the Bayesian update, the memo.

03 Since supervision #1 (10 July) — promised versus delivered

Left: what the 10 July briefing said would happen next. Right: what the repository log records on 6 September; the last row was not promised in July.

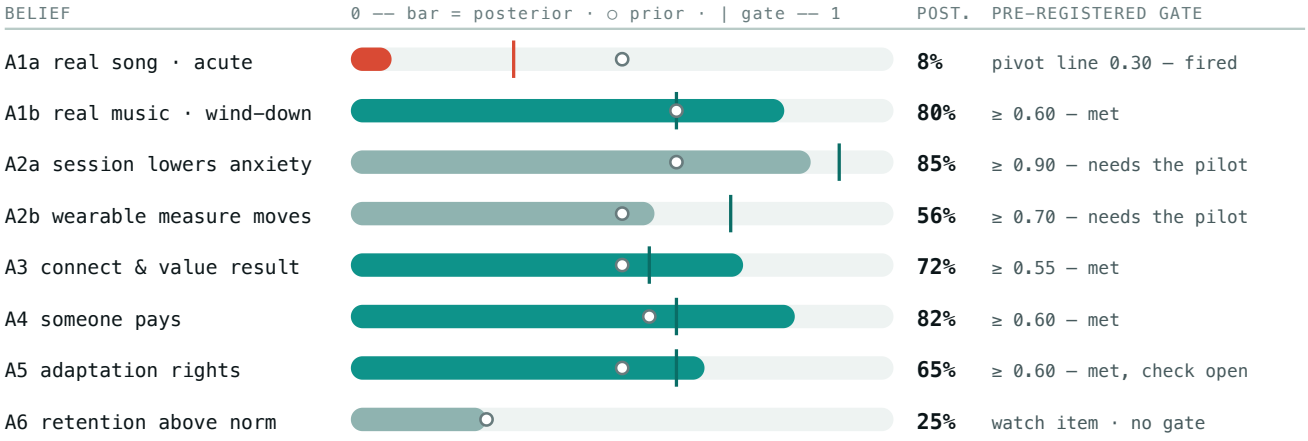
PROMISED ON 10 JULY	DELIVERED BY 6 SEPTEMBER	STATUS
Scale interviews to n = 8–12; add a psychiatrist and a music therapist	Not done as stated: still 1 expert + 5 users and caregivers, and neither extra expert interview was run. The qualitative weight was carried instead by **two focus groups, 13 participants**, with a blind audio stimulus test and a coded synthesis — a design the July plan did not contain.	SUBSTITUTED
Netnography and a **mini-survey n ≈ 100** with Van Westendorp pricing	Exceeded. Netnography at 200 coded rows; the survey became **v2, seven pre-registered blocks** with every item mapped to a construct, a threshold and a decision node — fielded to 150, analysed at **n = 134** (GAD-2, PSS-4, ODI, TAM, Kano, Van Westendorp, purchase intent).	DELIVERED
Lock persona, POV, HMW; finalise the scoring arithmetic	All three on the hub. The point score (4.55 / 2.65 / 1.80) became a **Monte-Carlo re-score** — triangular cells, weights perturbed ±40 per cent, 5,000 draws: anxiety scores 4.30 and ranks first in over 99 per cent of them, with asset fit and wearable synergy revised *down*.	DELIVERED
Methodology fit for a methods-heavy dissertation	Methodology v2 (17 Aug): convergent mixed methods, a triangulation matrix fixing what each stream may say, identification, measurement error, power in advance, insider-bias controls, stage gates — plus a biometric and ML research design (E0–E4).	DELIVERED
Efficacy micro-pilot, n ≥ 20 — two sessions, adaptive vs control, STAI-S + HRV, paired t-test	Redesigned upward and **not yet run**: a three-condition Latin-square crossover — T1 adaptive neutral, T2 adaptive real song, C the participant's own playlist as an *active* control — 20–30 people × 3 sessions, ANCOVA-form mixed model plus Bayesian re-analysis, preceded by E0 calibration (n ≈ 8). Scheduled 1–25 September.	DESIGNED · PENDING

PROMISED ON 10 JULY	DELIVERED BY 6 SEPTEMBER	STATUS
The decisive A/B: calm real song vs adaptive neutral sound	Answered ahead of the pilot, and it reversed the founding assumption: the real song is chosen by **24 per cent** in the acute scenario against **47 per cent** for wind-down (McNemar $\chi^2 = 14.29$, $p < .001$); the blind stimulus test agrees (10/13 acute → neutral, 9/13 wind-down → real). The contrast stays in the pilot as β1 − β2.	DELIVERED
Landing-page smoke test, ≥ 5 per cent visitor→waitlist	Built, not reported. The hub lists the A/B running to 20 September; the prototype README records the build as verified locally and deliberately not deployed pending sign-off, and the copy still promises "real music", which the pivot contradicts. **No conversion number exists.**	OPEN
One or two B2B conversations for C2	No result recorded; still a pre-registered pending row on A4 (LR 1.8 / 0.70).	OPEN
Pre-registered proceed / pivot / kill memo	The decision layer is built and live rather than promised: eight beliefs with priors and rationales, every evidence item a sourced, quality-tagged LR shrunk in log space, pending rows pre-registered with the LR they will contribute, posterior gates and concept joint probabilities. The memo follows the pilot.	DELIVERED
ADDED **Beyond the July plan** — a shipped product, a pilot pipeline and two written documents	**Product:** blueprint v1 (3 Sep) → app v0.1 the same night → v0.5.1 (4 Sep) — two modes, adaptive engine, tag-based recommendation over 69 Unwind tracks, research mode, safety screen, offline shell; TestFlight via GitHub Actions; App Store 1.0 build 4 submitted 3 September, **resubmitted 4 September after a Guideline 2.1 information request, waiting for review**. **Pipeline:** export → CSVs → model, Bayesian re-analysis, LR table and gate check, printed mechanically; 28 tests, exercised on labelled synthetic data before recruitment. **Documents:** the reasoning chain (16 bilingual sections, 13 diagrams, 214 facts checked) and the dissertation page (ten chapters, APA 7, appendices, the full framework set; 615 facts checked, 33 fixed).	SHIPPED

Promises: presentation/supervision-briefing.html and tutor-briefing.html (10 Jul 2026). Deliveries: TODO_2026-08-19.md §1, §5–§19; site/index.html roadmap (Done 15 · In progress 3 · Next 6).

04 Evidence and beliefs — where each of the eight now stands

Each belief carries a prior with a written rationale; each evidence item a likelihood ratio shrunk by its quality; the posterior is the odds product. Gates were fixed before the data and have not been moved.



Verdict NOT YET · 3 of 5 posterior conditions met · the efficacy pilot is a structural requirement, not a formality
As of 6 Sep 2026: survey and focus-group rows applied; E1 pilot and legal check pending. Computed by site/assets/decision.

Figure 1 · The belief register. A1a is 8 per cent by the page's rounding and "≈ 0.07" in the model's prose. The A2 gates were set above what the meta-analytic prior alone reaches (0.85 and 0.56) precisely so that PROCEED cannot be cleared by literature.

BELIEF	PRIOR	POST.	THE ONE THING THAT MOVED IT
A1a	0.50	0.08	Three acute interviews rejecting melody and lyrics took it only to ≈ 0.20 after shrinkage; the **pre-registered survey row** (24 % acute preference → LR 0.33) carried it across the pivot line. By design, qualitative evidence moves a belief but cannot cross a gate alone.
A1b	0.60	0.80	Endel's pivot to artist-attached functional music (1.40), a commercial survey, netnography complaints about generative sameness. The survey's wind-down row landed at 47 % — middle band, LR 1.0, no movement.
A2a	0.60	0.85	Five meta-analyses up (Cochrane −5.72 STAI-S; de Witte d = .545); a transfer discount for a self-administered phone session (0.75) and "calming playlists made it worse" (0.85) down. Only the pilot reaches 0.90.
A2b	0.50	0.56	Slow tempo raising vagal tone up; a non-significant psychophysiological meta-result (0.70) and the measurement facts — Apple Watch HRV MAPE ≈ 29 %, HealthKit lag, Oura sleep-only (0.60) — held it near the prior.
A3	0.50	0.72	The survey did the work: combined LR ×2.92 from willingness to connect (57 %), value of a measured result (54 %), relative advantage (60 %) and Kano (K1 attractive, K2 one-dimensional), plus ×1.3 from the focus groups. The action gap — 38 % do nothing with their stress reading — pulls back.

BELIEF	PRIOR	POST.	THE ONE THING THAT MOVED IT
A4	0.55	0.82	Survey combined LR ×2.4: the acceptable price range [$5.12, $8.55] contains the $6.99 test price and purchase intent is 35 % (18 % after the top-2 discount), against 49 % who do not want another subscription (0.8).
A5	0.50	0.65	Nothing new since July — label appetite for functional and AI-adaptive licensing (1.50) against unclear per-track adaptation rights (0.80). The legal check is the only decisive pending row: LR 8.0 or 0.15, a negative taking A5 to ≈ 0.22.
A6	0.25	0.25	Unchanged and untestable in this phase: a 4.7 % 30-day category retention base rate against hardware-anchored daily habits. A watch item, not a gate — and the most load-bearing commercial variable the phase cannot reach.

THE VERDICT AND THE CONTENT PIVOT

The evidence split the product by arousal state rather than settling the original bet: the acute moment gets an engineered neutral bed — melody-free, lyric-free, beat-free, adapting silently and showing the number only afterwards — which needs no catalogue rights, while the evening wind-down keeps real, artist-linked music, which does. The company's principal asset is the wrong content for the moment of highest need; the right response was to segment, not to abandon the asset or ignore the finding.

TWO ROWS I INTEND TO CONCEDE, NOT DEFEND

An adversarial pass found two pre-registered rows with a negative branch left unapplied in the engine. **E1 behavioural intention → A3**: ≥ 40 % overall or ≥ 55 % of owners → 1.8, else 0.70; observed 25 % / 32 %, so 0.70 applies. **C5 → A1b**: ≥ 40 % → 1.3, else 0.85; observed 35 %, so 0.85 applies.

If applied: A3 0.72 → 0.65 (gate 0.55, still cleared), A1b 0.80 → 0.77 (gate 0.60, still cleared); concept viability reads C3 63 · C1-neutral 41 · C2 45 · C1-real 23 %. **No gate flips, no verdict change** — sensitivity ±0.07 on A3, ±0.03 on A1b. Question 3 asks whether to apply it before the viva.

Provenance, stated once and repeated wherever the figures appear: the survey and focus-group numbers are the pre-registered analysis pipeline's output as published on the site, not fieldwork — the results file carries a simulated flag, the focus-group synthesis is a pre-field template, and fielded data replace both in a single pass. The E1 pilot has not been run. Sources: research/DECISION_MODEL.md §3–§5; site/assets/decision.js; site/pages/dissertation.html §5.3–§5.5; dissertation.audit.md residual risks R-1 and R-5.

05 The product — what Lilt is now

Two modes, one gesture

Settle — the acute mode the evidence chose: Web Audio synthesises a neutral bed with no melody, lyrics, beat or sharp frequencies and steps its low-pass cutoff, gain, density and pulse every 60 seconds from a published v0 rule table; it needs no catalogue rights. **Unwind** — the wind-down mode the catalogue serves: one transport over the Spotify Web Playback SDK (PKCE, Premium) with an embed fallback, owned files through Web Audio with a 1.5-second crossfade, or SoundCloud; Like, Slower, Next, Shuffle always available.

The adaptive loop

Heart rate is the live control signal, because HRV is not available live on these platforms: it arrives every 1–5 seconds at about ±6 per cent, is motion-masked from a 50 Hz accelerometer, held on jumps above 12 bpm and smoothed so lag stays under five seconds; sessions below a 0.70 quality index are flagged, not silently dropped. The step is taken on the **residual from the listener's own baseline** — "calmer than your usual 10 pm", not "your heart rate is 84". Relaxation raises HRV and lowers heart rate; HRV is a before-and-after measure only.

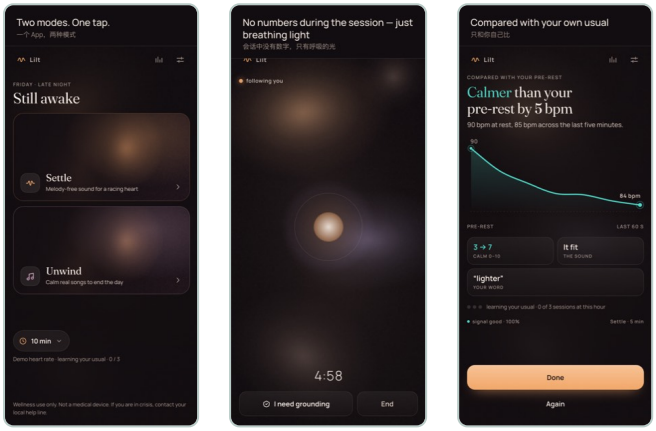
Recommendation and research mode

The Unwind catalogue holds 69 tracks, 53 of them slow piano; selection scores tags against the moment, the heart-rate band, stated liking, freshness and audio quality, shuffles with weights, constrains transitions and prints a "why this track" line of ≤ 60 characters. **Research mode** is built into the shipped app: a participant code fixes the Latin-square condition, the pre-rest extends to five minutes, STAI-6 runs before and after with two expectancy items, and the condition travels in the export — pilot instrument and product are one binary.

REAL VERSUS SIMULATED, SAID PLAINLY

The heart rate in the store screenshots and in the 72-second demo recording is the app's **labelled demo source**, not a measurement of anyone; a real signal today means a Bluetooth chest strap (Polar H10) on the Android and desktop web builds. **One recorded inconsistency:** the iOS build ships without the Bluetooth plugin while the store description mentions a chest strap — both corrections are written down (reword the description for 1.1, or add the plugin first), and until one is done the description overstates the iOS build. No accounts, no analytics, nothing leaves the phone; safety screen at HR > 130 bpm for ten seconds or on request; wellness framing only.

Status by channel. Web (PWA) — v0.5.1, installable, offline shell; tests green at 70 unit, 11 end-to-end, 6 audio and 14 scene-engine; nine owned tracks ingested at –16 LUFS AAC (32 MB). **TestFlight** — internal group live, builds from a GitHub Actions pipeline on macOS 26 with Xcode 26.2 for the iOS 26 SDK requirement, certificates valid to September 2027. **App Store — 1.0, build 4, resubmitted 4 September after a Guideline 2.1 information request, waiting for review**; six answers filed with a recorded walkthrough, and the decision not to add accounts in this version documented in the reply. The installable web build is the contingency and is gated by no store.



Home · session · result. Two modes and one duration control; no numbers during the acute session — the design response to the focus-group finding that a live figure can itself raise anxiety; then a result compared with the listener's own pre-rest and own usual. The status line reads "demo heart rate · learning your usual · 0/3", and every value shown comes from that labelled demo source.

`app/PLAN.md; app/store/REVIEW_REPLY.md; TODO_2026-08-19.md §6–§16; dissertation §7.1–§7.5; screenshots app/store/screenshots/69-1, 69-3, 69-5.`

06 Method notes — what I would like checked

The identification strategy for E1

A randomised within-subject crossover with an **active** control. Three conditions — T1 adaptive neutral sound, T2 an adaptive calm arrangement of a real song the participant chose, C a generic relaxing playlist the participant already knows — in one of six Latin-square orders fixed by the participant code; three sessions per person, ≥ 24 hours apart, same time of day, seated, phone only. A session runs five minutes of pre-rest (baseline HR, RMSSD where RR is available, STAI-6, two expectancy items) → fifteen minutes of audio → immediate STAI-6 and a check-in → three minutes of post-rest. The control is deliberately not silence, because silence changes expectancy and confounds "any audio" with "our audio". The model is fitted identically to all three outcomes, in ANCOVA form:

$$Post_{ij} = \beta0 + \beta1 \cdot T1 + \beta2 \cdot T2 + \beta3 \cdot Pre + \beta4 \cdot Order + \beta5 \cdot Expect + \beta6 \cdot Session\# + u_i + \varepsilon_{ij}$$

β1 is the pre-registered primary coefficient; the contrast **β1 – β2 answers A1a directly**, which is why July's "decisive A/B" survives inside the pilot rather than beside it. Primary outcome STAI-S (STAI-6 scaled 20–80); secondaries residual heart rate and log-RMSSD, Holm-corrected; sessions below the 0.70 quality index leave the physiological models. A Bayesian re-analysis against a meta-analytic prior then produces the posterior that drives the gate. **E0 comes first** because A2b's measurement premise must hold before its effect can be read: about eight participants wear an Apple Watch and a Polar H10 concurrently over three sessions, giving ICC and MAPE for heart rate and RMSSD and $\lambda = \mathrm{Cov}(watch, polar) / \mathrm{Var}(watch)$ to de-attenuate any model in which watch HRV is a *regressor*. Where HRV is an *outcome*, non-differential error costs power rather than causing bias.

POWER, STATED BEFORE THE DATA RATHER THAN AFTER

Paired design, α = .05 two-sided, 80 per cent power: $d_z = 0.5$ needs n = 34; 0.6 needs 24; 0.65 needs 21; 0.8 needs 15. Cochrane (–5.7 STAI-S, SD ≈ 10 → d ≈ 0.55) implies n ≈ 28 for self-report;

physiology at d ≈ 0.4 needs n ≈ 50. **So n = 20 is a feasibility signal for A2a and under-powered for A2b**, said in advance. Remedies: three sessions each, and a posterior rather than a p-value driving the gate.

SYNTHETIC DEMONSTRATION — FABRICATED, REPORTED DELIBERATELY

The pipeline was run end to end on **24 fabricated participants × 3 sessions** (seed 20260904) before any recruitment, so the code, thresholds, figures and gate arithmetic all existed before the data. On that draw A2a fired at LR 4.0 and reached 0.958 while A2b returned 0.30 and fell to 0.276 — NOT YET on A2b alone. Across 40 replications A2a fired in 78 per cent of runs, matching the 80 per cent power target, and A2b in 45 per cent, reproducing the declared under-powering rather than flattering it. **These numbers are evidence about the pipeline and never about TDMusic.**

The pipeline, and the pre-registered likelihood-ratio mapping

Session exports flow through `app/tools/export_to_csv.py` into per-session and per-window CSVs, then through `research/pilot/e1_analysis.py`, which fits the pre-registered model, runs the Bayesian re-analysis and prints the LR table, the gate check and the figures. Twenty-eight tests cover the schema, the exclusion rule, order inference, effect direction and gate arithmetic — one asserting that the Python Latin square is identical to the app's, so a change to randomisation fails a test rather than mis-assigning conditions.

THE SEVEN PENDING ROWS, WITH THE LR EACH WILL CONTRIBUTE

A5 · legal check on ≥ 50 tracks — 8.0 / 0.15, 1–2 weeks, the only decisive row and the cheapest, so it goes first. **A2a · pilot STAI-S** — 4.0 / 0.30, two weeks, n = 20–30 × 3 sessions. **A2b · pilot HRV and residual HR** — 4.0 / 0.30, same sessions. **A1a · pilot A/B content preference** — 3.0 / 0.33, same sessions, the β1 − β2 contrast. **A3 and A4 · landing conversion ≥ 5 %** — 2.5 / 0.40 each, ¥500–1,000 of ads, the only behavioural signal in the design. **A3 · think-aloud, n = 5–8** — 1.5 / 0.60, one week. **A4 (B2B) · partner conversations** — 1.8 / 0.70, ongoing.

Positive for the pilot means p < .05 and d ≥ .3 in the correct direction. The strength rubric is published (decisive 8–10, strong 3–5, moderate 1.8–3, weak 1.2–1.8, reciprocals against) and every LR is shrunk in log space by a quality exponent: 1.0 meta-analysis, 0.75 GRADE-low or a large survey, 0.5 industry-funded or single-source, 0.35–0.5 qualitative with n ≤ 5, ×0.6 again where correlated with an item already counted. An LR changes only with a written reason and a date; a threshold is never moved after the data.

research/METHODOLOGY_DESIGN.md §3–§4; research/pilot/README.md §1–§4; research/DECISION_MODEL.md §2, §4, §7. Precedent for subjective LRs: Fairfield & Charman (2017); Humphreys & Jacobs (2015); grading after GRADE.

07 Timeline to submission, and the critical path

THE CRITICAL PATH

E0 calibration → **E1 pilot** → Bayesian update and gate check → decision memo → results and discussion chapters → submission. All of it depends on recruitment, the one input I do not control: the hub listed pilot recruitment as running to 31 August and E1 starts 1 September, so a slip moves the pilot, the update and the results chapter together.

OFF THE CRITICAL PATH BUT DECISIVE

The **legal check** does not block the write-up, but it selects which concept survives: a negative at LR 0.15 takes A5 from 0.65 to ≈ 0.22, closes the real-music route and makes C1-neutral the only live app concept — at which point no more engineering time should go into parametric adaptation of catalogue tracks. It runs first because it is the cheapest decisive test.

The contingency is already written down. If the pilot returns null, the pre-registered rule is KILL/PARK on A2a < 0.50 and the phase is written up as a negative result — a valid outcome of this dissertation, and one the chapter structure absorbs by changing the verdict, not the argument.

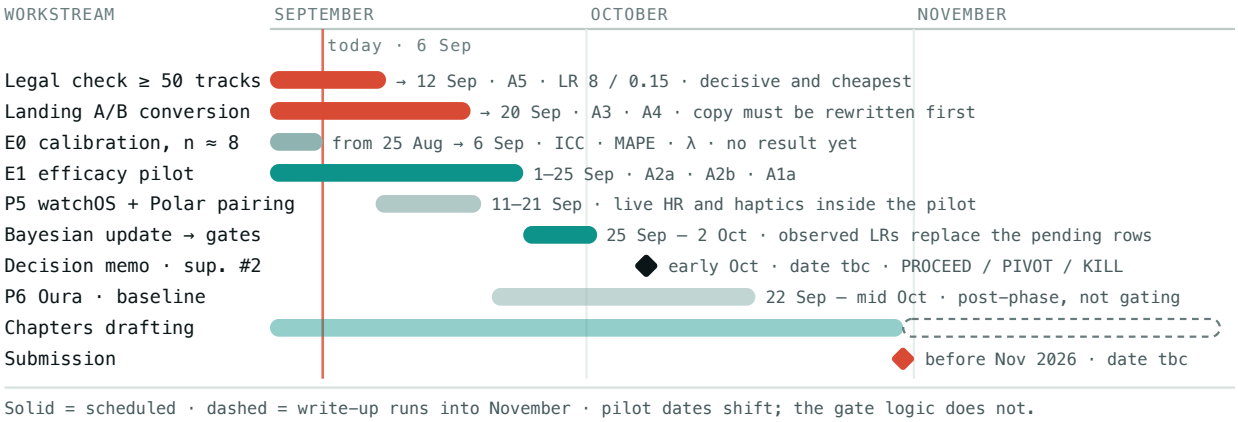


Figure 2 · The remaining phase. Dates from the hub roadmap and app/PLAN.md (P5 11–21 Sep; P6 22 Sep – mid Oct); the chapter bar carries methods, findings and discussion. Two dates are genuinely open and are marked as such — supervision #2 and the decision memo ("early October"), and the submission date; fixing both is item 3 on the repository's own to-do list.

08 Questions for you

1. **Subjective likelihood ratios.** Is a Bayesian layer with quality-shrunk *subjective* LRs and pre-registered posterior gates an acceptable Evaluate step for insider action research at EMBA level? How much of the calibration rubric belongs in the methods chapter rather than an appendix?
   METHODOLOGY · DECISION ANALYTICS

2. **Presenting pipeline-output figures before fieldwork.** The survey and focus-group numbers are the pipeline's pre-registered output as published on the site, not fielded data. Keep them in chapter 5 with the provenance statement at every use, move them to an appendix as a worked pre-registration, or hold the chapter — which reads as rigour rather than a gap?
   EVIDENCE · PROVENANCE · VIVA RISK

3. **Whether to concede the two survey rows.** Applying the two unapplied negative rows gives A3 0.72 → 0.65 and A1b 0.80 → 0.77, changing no gate and no verdict. Apply them in the engine before the viva and report the dated correction, or leave the published state and concede it verbally?
   PRE-REGISTRATION DISCIPLINE

4. **The wearable-claim wording.** Consumer heart-rate data is honest about arousal but not about anxiety, so self-report stays the construct, heart rate is the control signal and co-primary physiological outcome, and HRV is a before-and-after measure only — relaxation raises HRV and lowers heart rate, but consumer HRV carries ≈ 29 % error and arrives late. Right claim for a product promising "show me it worked"? And is "calmer than your usual 10 pm" safely inside the wellness line?
   MEASUREMENT · CLAIMS · REGULATION

5. **Word budget, and the unit of analysis.** What split between the corporate-innovation framing (VRIO, PESTEL, five forces, Ansoff, Three Horizons, ambidexterity) and the design-thinking evidence chain (six streams, the belief register, the pilot, the gates)? The frameworks currently occupy one chapter, the evidence chain four. Relatedly: should the arousal-state split be framed as one product with two modes or as two concepts — and does the acute mode weaken the asset-leverage story or sharpen it by moving the asset from the catalogue to the engine and the measurement?
   STRUCTURE · WEIGHTING · CORPORATE FRAMING

6. **If the pilot returns null.** A documented KILL/PARK written up as a negative result is a pre-registered outcome of this design. Is that a full-credit dissertation, and how much of the discussion should be pre-committed to that branch now?
   OUTCOME RISK · PRE-COMMITMENT

09 Appendix — where each artefact lives

ARTEFACT	REPOSITORY PATH · LIVE URL
Hub — the whole phase on one page	`site/index.html` · 38.180.150.12/
The dissertation (ten chapters, APA 7) · the reasoning chain	`site/pages/dissertation.html` · `journey.html` · `dissertation.audit.md` · `/pages/dissertation.html` · `/pages/journey.html`
Bayesian decision engine (live, contestable)	`site/pages/decision.html` · `assets/decision.js` · `/pages/decision.html`
Methodology v2 · ML design · decision model	`research/METHODOLOGY_DESIGN.md` · `ML_RESEARCH_DESIGN.md` · `DECISION_MODEL.md` · `/pages/methodology.html` · `/pages/ml-design.html`
Survey v2 · focus groups · pilot pipeline	`surveys/` · `research/FOCUS_GROUP_KIT.md` · `research/pilot/` · `/pages/survey-results.html` · `/pages/focus-group.html`
Product blueprint · the app · App Review reply	`site/pages/product-design.html` · `app/` · `app-ios/` · `app/store/REVIEW_REPLY.md` · 38.180.150.12.sslip.io/app/
This update	`site/notes/progress-update-2026-09-06.html` · `.pdf` · `/notes/progress-update-2026-09-06.html`