Implementation Roadmap

Last Updated: 10 July 2026, 13:11

Prepared for SkincareLab

Project SkinSync App

Contents

Introduction

The strategy and brand work produced across this engagement is specific, grounded, and unusually honest about where the risk sits. This document exists to translate that work into a delivery sequence that protects what matters most. It is not a task list. It is an argument about order, describing which decisions need to be made before which others can be made well, and why getting that sequence wrong is more costly than getting any individual feature wrong.

Three principles drive the sequencing throughout. Emotional dependencies come first. SkinSync's value rests entirely on a user trusting the product with something personal, and trust is not a feature that can be added later. The interactions and structural decisions that build or break trust in the first sixty seconds, and in the difficult weeks that follow, must be designed and tested before anything that depends on trust being present. Surface work built on top of unresolved emotional architecture tends to look fine and feel wrong, and that gap is expensive to diagnose once it is baked into a shipped product. The second principle is structural before surface. Data models, insight logic, streak mechanics, and the pacing of progress information are all structural decisions. Notification copy and milestone animations are surface decisions. Surface decisions made before the structural layer is stable often have to be redone. The third principle is that the highest-risk assumptions are tested earliest. The two-week drop-off window, the emotional handling of flat progress, the balance between honest data and calm framing: these are the places the product is most likely to fail quietly. They should be the first things to face real users, not the last.

This document does not replace sprint planning, ticket management, or team-level delivery tooling. Those live elsewhere and operate at a different level of granularity. What this document provides is the reasoning that should sit above those tools, the rationale for why phases are structured as they are, what each phase must produce before the next can begin, and what success looks like at each boundary. Anyone making sequencing decisions during build, whether that is prioritising which features enter a sprint or deciding whether a phase is genuinely complete, should be working from this document alongside whatever delivery tooling the team uses.

Phase Overview

The four phases below do not follow a conventional agile sprint structure. They follow the emotional logic of the product itself, which means the sequence is determined by what needs to be true before something else can work, not by what is quickest to build.

Phase	Name	Duration	Focus
1	Emotional Foundation	Weeks 1 to 4	Onboarding architecture, first-session experience, trust baseline, structural data model

Phase	Name	Duration	Focus
2	Journey Architecture	Weeks 5 to 9	Middle-weeks experience, insight delivery, streak and pacing mechanics, anxiety handling
3	Identity Layer	Weeks 10 to 13	Progress visibility, milestone design, pattern recognition, the before and after moment
4	Polish and Pre-Launch	Weeks 14 to 16	Notification refinement, copy audit, motion and interaction consistency, launch readiness

The reasoning behind this structure is the same reasoning that shaped every other document in this engagement. SkinSync's value proposition depends on a sequence of emotional events happening in the right order, and the build sequence must mirror that. A user who does not make it through the first session cannot be retained by a well-designed week-three insight. A week-three insight that arrives without the structural pacing logic in place will land as noise or, worse, as anxiety. A before and after comparison that appears before the honest framing around timelines is established will produce exactly the deflation the Aspiration Gap work named as the single biggest retention risk. Each phase creates the conditions that make the next phase possible. Skipping forward, or treating phases as parallel workstreams that can resolve simultaneously, undermines that dependency chain. Phase 1 must be genuinely complete before Phase 2 begins in earnest, and the completion criteria at each boundary are defined in the sections that follow.

Phase 1: Emotional Foundation

The sequencing choice here is not a preference. It is a conclusion drawn from the research. Across every workshop session, one finding recurred with enough consistency to treat it as structural fact: SkinSync's ability to deliver anything useful depends entirely on whether a user makes it through the first session feeling understood rather than interrogated. Everything else in this product, the insights, the streak tracking, the before and after comparison, the pattern recognition that turns logging into learning, is only available to someone who stayed. And whether they stay is decided, largely, in the first sixty seconds.

That is why this phase exists before any other. Not because onboarding is the most exciting design problem, but because it is the one that must work before anything else can. A well-designed week-three experience is meaningless if the week-one experience sends people away. And the data model that powers honest pacing and contextualised insights needs to be established now, not retrofitted around features that assumed a different structure.

The Aspiration Gap work named the emotional state at first open with precision. Users arrive carrying curiosity, self-consciousness, mild scepticism, and the particular kind of vulnerability that comes

them. It needs to meet them where they are, get out of the way, and give them one clear thing to do. Relief, not excitement, is the target feeling.

The Heartbreak Scale placed trust at two out of five at the start of the experience. That score is not a criticism. It is simply what arrives with any new user who has tried things before. The trust deficit is earned by other products, not by SkinSync. But the product still has to close it, and it closes it through behaviour, not through copy. An onboarding that asks for account creation before demonstrating value confirms the scepticism the user already carried in. One that asks a single useful question, pre-fills a starter routine, and shows a clear first action communicates something different: that the product already has something to offer before it asks for anything in return. That sequence, give first, ask later, is the structural expression of the trust principle that runs through every other document in this engagement.

The First 60 Seconds exercise was specific about the decisions required to create that feeling. Ask skin type before anything else. Pre-fill a starter routine based on the answer. Show a single "today" screen rather than a dashboard. Delay account creation until after first use. Use calm visuals with space to breathe. Each of these is a design decision, but they are also emotional decisions. They are the product communicating, through its structure rather than its copy, that it is here to help rather than to process.

Phase 1 also establishes the data model that everything in later phases depends on. Honest insight delivery, proportionate streak mechanics, contextualised progress framing, none of that is possible without a structural foundation that can support it. Building the onboarding experience on top of an improvised or incomplete data model means either redoing the model when the insight layer arrives, or accepting constraints that compromise the product's core promise. The structural work happens now, not because it is glamorous, but because later phases cannot begin without it.

Deliverables

Deliverable	Priority	Notes
Onboarding flow (skin type question through to first routine view)	Critical	Must reach first completed action without account creation. Delay sign-up until after first routine is logged.
Pre-filled starter routine logic	Critical	Routine should be populated from skin type answer with no blank-slate starting point. User can edit; they should not have to build from scratch on day one.
"Today" home screen (single action, uncluttered)	Critical	Leads with the routine checklist only. No dashboard, no competing metrics. Trend data and insights are accessible but not visible on first open.
Core data model (users, routines, products, skin condition logs, streak records)	Critical	Structural foundation for all subsequent phases. Must be designed to support honest pacing logic, pattern queries, and contextualised insight delivery before those features are built.
Routine check-off interaction	Critical	Checkbox mechanic with calm, proportionate feedback. Missed days shown as neutral gaps, not broken chain visuals. This must be correct from the start; changing streak mechanics after users have formed expectations is costly.
Calm visual language applied to first screens	High	Spacing, type weight, and colour consistent with the visual direction. Soft, uncluttered, no aspirational imagery. This is the baseline the rest of the product will hold to.
Screen transition behaviour	High	Forward and back navigation using horizontal slide with partial exit, as specified in the animation guide. Tab switches as cross-fade only. Establishes the motion baseline before more components are added.
Honest pacing copy for onboarding	High	A single plain-language acknowledgement that visible skin change takes weeks, not days. Sets the expectation before the user can form a wrong one. This copy must be in place before the first real user sees the product.
Account creation flow (deferred)	Standard	Follows first completed routine. Must not appear before. Build the flow now, but position it correctly in the sequence.
Empty and loading states for core screens	Standard	Skeleton placeholders, not shimmers. Copy that names the emptiness plainly ("Nothing here yet, which makes sense.") rather than decorative filler.

Deliverable	Priority	Notes
Basic reminder setup (local notifications only)	Standard	AM and PM reminder configuration, scheduled on-device. No server-side push at this stage. The notification copy must meet the voice guidelines before this ships.

Success criteria

- The majority of users who open the app reach a completed first routine step without encountering an account creation prompt
- Time from first open to first logged routine action stays consistently low, with no significant delays introduced by onboarding screens
- Onboarding drop-off, measured as the proportion of users who open the app and do not complete the first routine view, falls below an agreed threshold confirmed with the product team before this phase closes
- Users who complete the first session return within 48 hours at a rate that holds above the baseline established in the first week of live testing
- No user-reported confusion or support query mentions feeling pressured to sign up before trying the product
- The data model successfully supports the creation of user records, routine structures, product entries, skin condition logs, and streak records without schema changes required when Phase 2 begins
- The honest pacing copy is present in the onboarding flow before any external user sees the product, confirmed in QA sign-off
- All screen transitions on the core flow (onboarding through to home screen) match the motion specifications without deviation, confirmed by animation review before phase close
- Routine step completion triggers the correct checkbox interaction and dim-row behaviour, with no broken-chain visual appearing on any missed-day state
- Empty and loading states are present on all screens that can exist in an unpopulated state, confirmed in QA before phase sign-off

What Phase 1 does not include

This phase is specifically scoped to what must be true before anything else can work. The following are out of scope and should not be designed, built, or tested during this phase, regardless of how closely they relate to the features above.

Skin pattern insights and trend analysis belong to Phase 2. The data model built in Phase 1 is designed to support them, but surfacing insights before there is enough data to support honest ones would contradict the principle that insights should only appear when the data genuinely warrants them. Building the insight logic now, even in a dormant state, risks premature surfacing under delivery pressure.

moment, named across multiple sessions as the product's most significant trust event, deserves its own phase of consideration rather than being treated as an extension of the onboarding build.

Streak milestone celebrations and any designed acknowledgement of longer-term consistency are Phase 3 work. The streak mechanic is established here, including the neutral missed-day state, but milestone moments require the product to have earned them with enough data to support honest celebration. Designing the celebration before the tracking is mature would be sequencing surface before structure.

Ingredient conflict warnings are not in scope for Phase 1. The conflict logic requires a validated ingredient dataset and a contextualisation layer that turns a raw flag into a calm, specific, plain-language explanation. Building the flag without the context layer would produce exactly the anxiety-inducing experience the Design Principles Builder identified as a critical failure mode.

Push notification infrastructure is out of scope. Local, device-side reminder scheduling is included. Server-initiated push, event-triggered notifications, and the logic that determines when there is something meaningful to say, all of that belongs to Phase 2 once the data layer is in place to power it.

Community features, retailer integration, and dermatologist-reviewed templates are Phase 2 and beyond. They are not referenced in this phase and should not influence any architectural decision made during it.

Phase 2: Journey Architecture

Phase 2 begins at the point where Phase 1's hardest work becomes visible. The onboarding is done. The first routine has been logged. The data model is in place. And now the product faces the stretch that the research named, repeatedly and from different directions, as the place where everything is most likely to go quietly wrong. Weeks two and three. The novelty has gone. The habit is fragile. Skin has not yet changed in any way a user can point to. And the daily log, which felt purposeful in the first session, is starting to feel like administration for an outcome that may or may not be coming.

Phase 2 exists to design for that stretch deliberately, not to paper over it, but to hold the user through it honestly. The Aspiration Gap work was precise about the mechanism of failure: if a user does not see or feel any change in the first few weeks, they will assume the app is not working and quietly stop opening it. The operative word is assume. The product cannot accelerate the biology. What it can do is make the slow middle feel purposeful, surface small honest signals before the large ones arrive, and handle setbacks in a way that treats them as information rather than evidence of failure. That is what this phase builds.

Four distinct problems need solving here, and they are related. The first is insight delivery: the structural and emotional logic that governs when the product surfaces an observation, how it is framed, and what obligation accompanies it. The second is streak and consistency mechanics: the visual and copy language around missed days, gradual progress, and the difference between

rewarding consistency and punishing interruption. The third is anxiety handling: specifically the ingredient conflict warning system, which the Heartbreak Scale named as the single most confusing element in the product, and the broader question of how negative or ambiguous signals land without leaving the user alone with them. The fourth is the early positive signal layer: the small, honest moments of recognition that bridge the gap between starting and seeing.

None of these are surface problems. They are structural ones, which is why they sit in their own phase rather than being folded into a later polish pass. Insight delivery logic built on shaky pacing assumptions will produce exactly the premature insight problem the Design Principles Builder identified as a critical failure. Streak mechanics that look calm but behave punitively will quietly teach users to dread the app. Anxiety handling written as an afterthought, rather than designed as a core behaviour, will produce the kind of cold, context-free warnings that the research named as a trust-breaking failure mode. These decisions are architecture. They need to be right before anything is built on top of them.

Deliverables

Deliverable	Priority	Notes
Insight delivery framework (data threshold logic, surfacing rules, plain-language framing system)	Critical	Insights must only surface when the data genuinely supports them. The framework defines what constitutes sufficient data, how the observation is constructed, and the obligation to accompany every signal with context and a next step. No insight ships without a plain-language explanation and a clear direction alongside it.
Ingredient conflict warning system (flag plus contextualisation layer)	Critical	The flag alone is not acceptable. Every conflict alert must specify the nature of the concern (sensitivity risk, efficacy question, or stop-and-check), distinguish clearly between "worth being aware of" and "worth pausing on", and arrive in the same calm, unhurried register as every other part of the product. This is the highest-anxiety interaction in the product and must be designed to reduce anxiety, not amplify it.
Missed-day and streak gap handling	Critical	Neutral gap visual, not a broken chain. "Welcome back" copy on return, not a streak-recovery prompt. The interaction and copy must be confirmed against the design principles before build begins. Changing streak mechanics after users have formed expectations is disproportionately costly.
Early positive signal layer (week one to two observations)	Critical	Small, honest observations surfaced before statistically significant trend data exists. Consistency noted, skin condition patterns flagged where even early data supports them, first-week behaviour reflected back in plain language. These signals are what bridge the gap between starting and seeing, and they must be genuinely grounded rather than manufactured.
Skin condition log and trigger correlation (linking observations to habits and products)	High	The turning point named in the Aspiration Gap work is the first time a user sees that their skin is not random but responding to something they can identify. This deliverable builds the connection between a logged condition and a recent product or habit, surfaced in plain language at the right moment.
Week two and week four photo check-in prompts	High	Scheduled prompts that invite the user to capture a progress photo at the points where comparison becomes meaningful. The prompt copy must set honest expectations about what is likely to be visible at each stage. These are not triggered by the app having something impressive to show; they are triggered because two and four weeks are the moments when a comparison, however subtle, is more useful than no comparison.

Deliverable	Priority	Notes
Notification logic and event-triggered push framework	High	Server-initiated push now in scope. The framework defines which events warrant a notification, what that notification is permitted to say, and what it must never say. Rate limiting at the user level must be in place from the start. No notification sends unless there is something specific and meaningful to communicate.
Contextualised progress framing for flat weeks	High	When a week produces minimal visible change, the product does not go quiet. A short, plain acknowledgement that this is normal at this stage, paired with what the user has actually built (consistency, data, growing pattern), must be designed and in place before flat weeks are possible to encounter.
Gentle reminder refinement (copy and timing)	Standard	The basic reminder structure was established in Phase 1. This phase refines the copy against the voice guidelines, sets appropriate defaults for timing and frequency, and builds the opt-down configuration so users can reduce notification volume without turning off local reminders entirely.
Skin condition history view (log review, not trend analysis)	Standard	A simple review of what has been logged, without interpretation. Interpretation arrives later when the data genuinely supports it. This view is about giving the user access to their own record, not about the product drawing conclusions.
Basic push notification infrastructure (server-side)	Standard	The backend logic, token management, and delivery layer that powers event-triggered notifications. Built to support the notification logic framework above, not ahead of it.

Success criteria

The measurable targets below reflect the design principles established during the discovery work. Retention through weeks two and three is the primary signal, because that is the window the research identified as the highest-risk period. Secondary signals track whether the insight and conflict systems are working as intended, keeping users engaged rather than pushing them away at the moments that matter most.

- Retention rate at day 14 to day 21 holds within an agreed threshold of day 0 to day 7 retention, with no sharp drop at the two-week mark. This is the single most important signal for this phase and should be reviewed weekly from the point live users enter the product.
- App open rate in the 48 hours following a flagged insight or ingredient conflict warning holds within a defined range of open rate following a positive observation. A large gap between those two numbers indicates the conflict or insight system is creating avoidance rather than understanding.

- Users who encounter a missed-day state return to the app within three days at a rate that stays above an agreed threshold. A sharp drop in return rate after a missed day indicates the streak handling is behaving punitively, even if the visual design looks calm.
- No insight surfaces in the product before the data threshold defined in the insight delivery framework is met. This should be a hard rule enforced in QA, not a guideline.
- Every ingredient conflict warning that reaches a user includes a plain-language explanation of the specific concern, confirmed in QA review of the contextualisation layer.
- Users who receive a week-two photo check-in prompt open the app again within 48 hours at a rate consistent with their pre-prompt return behaviour, indicating the prompt is welcome rather than disruptive.
- No notification sends to a user more than once in a defined rolling window without a new triggering event, confirmed via rate-limiting audit before the push framework goes live.
- Qualitative review and in-app survey language from users in weeks two and three reflects understanding and reassurance rather than confusion or discouragement. This is a softer signal but worth tracking as a directional indicator alongside the quantitative measures.

Key risk: the insight delivery system surfaces observations before the data honestly supports them

This is the most significant implementation risk in Phase 2, and it is worth naming plainly. The pressure to surface insights early is real. An app that has been open for ten days with no insight delivered can feel empty, and the instinct during build and QA is to lower the threshold so something appears. Every product principle established in this engagement argues against that instinct. The trust that SkinSync is built on depends on the product being honest about what it does and does not yet know. An insight that arrives before the data supports it is not a kindness to the user. It is the beginning of the pattern that erodes credibility slowly, the same pattern the Rejection Letter exercise identified as the defining failure mode of every product SkinSync is trying not to be.

The mitigation is structural rather than cultural. The data threshold logic must be defined before any insight feature is built, reviewed as a design document rather than a development concern, and enforced in QA as a hard acceptance criterion, not a guideline. The rule is simple: if the product cannot point to the specific data that supports an observation, the observation does not surface. Empty states during this window must be handled by the honest framing established in Phase 1, acknowledging that nothing meaningful has emerged yet, rather than by generating a low-confidence observation to fill the space. The design checklist established in the Design Principles Builder exercise should be applied to every insight before it is approved to ship: would this still feel honest if results were disappointing this week? If the answer is anything other than yes, the insight needs to be reworked or withheld.

Phase 3: Identity Layer

Phase 3 is the phase where the product either delivers on its central promise or reveals that the promise was decorative. Everything built before this point, the onboarding, the pacing mechanics, the insight delivery framework, was preparation. Phase 3 is where that preparation is tested in the most demanding way: by asking whether a user who has been in the product for ten or eleven weeks feels genuinely different about their own skin than they did on day one. Not just more organised. Not just more consistent. Actually different, in the specific way the Identity Shift exercise named: someone who used to feel reactive and uncertain, now capable of explaining why they use what they use, trusting what they observe, and recognising their own patterns.

That is a harder target than any feature can hit on its own. It requires the product to have held its honesty through the slow middle weeks, surfaced observations at the right moments, and accumulated enough meaningful data to reflect something real back to the user. None of that is possible without Phases 1 and 2 being genuinely complete. A progress photo comparison delivered without the honest pacing framing already in place will produce the deflation the Aspiration Gap work named as the single biggest retention risk. A milestone that arrives before the data supports it is not a kindness, it is the beginning of the credibility erosion that makes everything else worthless. Phase 3 has the hardest emotional design problem in the product precisely because it can only be solved on top of everything that came before it.

What makes this phase difficult

The difficulty here is not technical. It is the problem of designing for a felt experience that cannot be manufactured. The Identity Shift exercise described the emotional arc the product needs to complete: from confused and reactive, through cautious and uncertain, through reassured, to genuinely confident and informed. The first three stages of that arc are managed by Phases 1 and 2. Phase 3 is responsible for the final movement, and the final movement depends on something the product cannot control: whether enough time has passed, and enough honest data has accumulated, for the before and after comparison to carry real weight.

The temptation at this phase is to design around that dependency rather than through it. To soften the comparison, to surface a milestone slightly earlier than the data supports, to write copy that implies more has changed than the product can honestly claim. Every one of those moves feels like a kindness in the moment and constitutes a quiet betrayal of what the product is. The Anti-Principles exercise was direct about what this looks like in practice: the team is tempted to soften or delay the photo comparison so it looks more flattering, or to slip in a "boost your results" product suggestion right at the moment of doubt. Either move trades long-term trust for a short-term feel-good fix, which is exactly the pattern this entire engagement was designed to prevent.

The other difficulty is specificity. The identity shift the product is trying to create is not generic. It is personal to each user, built from their own data, their own patterns, their own observations over their

own time. Generic milestone copy, blanket celebration states, and one-size insights all undermine it by implying the product is handing the user something it prepared for everyone. The Heartbreak Scale was precise about what would make the experience feel more empowering: language like "you now know your skin reacts well to X", not "congratulations on reaching 30 days". The first attributes the knowledge to the user's own observation. The second attributes it to showing up. Both matter, but only the first creates the identity shift.

This phase requires the product to hold two things simultaneously: genuine honesty about what the data does and does not show, and genuine warmth about what the user has built. Those are not in tension if the design is right. They become in tension only when the product tries to manufacture one at the expense of the other.

Deliverables

Deliverable	Priority	Notes
Progress photo comparison view (side by side, honest framing)	Critical	The comparison must show what is actually there, without enhancement or selective framing. Where change is minimal, the surrounding copy names the realistic timeline rather than leaving the user alone with an underwhelming result. The design of this screen should be reviewed against the Anti-Principles exercise before it ships.
Week four photo check-in prompt and framing copy	Critical	Distinct from the week two prompt established in Phase 2. By week four, the framing shifts: the copy acknowledges that this is the window where first visible change becomes possible, without promising it. Sets honest expectation without either raising or dampening it.
Pattern recognition surfacing (first clear cause-and-effect observation)	Critical	The moment named in the Aspiration Gap work as the turning point: the first time a user sees that their skin is not random but responding to something specific. This observation must be tied to real data, expressed in plain language, and delivered in a way that attributes the insight to the user's own tracking rather than to the product's cleverness.
Milestone system (genuine data-threshold triggers, personal copy)	Critical	Milestones are triggered by what has actually happened, not by calendar time. First month of logging, first visible change identified, first ingredient concern resolved, first informed product decision made. Each milestone copy references something specific to that user's data. Generic celebration copy is a failure state, not a fallback.
"Look back" summary view (what has been learned, not just logged)	High	A simple view that shows what the user now knows about their own skin: which products have shown a positive pattern, which conditions have improved, which triggers have been identified. Framed as accumulated understanding, not as a data report. This is the output the Heartbreak Scale noted could be worth showing to a dermatologist, and it should be designed with that usefulness in mind.
Informed product decision moment (first product swap based on tracked data)	High	When the data supports a recommendation that a user could act on with products they already own, the product surfaces it. Not as a sales prompt, not as a generic tip, but as a specific observation tied to what the user has logged. The design principles are explicit: the same insight should read identically whether or not the app sells anything.

Deliverable	Priority	Notes
Dermatologist-ready summary export (structured log output)	Standard	A structured summary of tracked conditions, product history, and observed patterns that would be genuinely useful in a clinical conversation. Noted in the Heartbreak Scale as a current missed opportunity. This is not a vanity feature; it is a tool that repositions the user as someone with knowledge worth sharing.
Streak milestone acknowledgement (restrained, proportionate)	Standard	Longer-run consistency milestones, thirty days, sixty days, receive acknowledgement that is warm without being effusive. The number is not the achievement. The understanding built during that time is. The copy should say so.
Voice and copy audit of all Phase 3 screens against Read It Aloud criteria	Standard	Every screen introduced in this phase is read aloud against the voice guidelines before it ships. Any copy that would tip into outcome promise, hype language, or manufactured urgency when spoken is rewritten.

Success criteria

The measures below are deliberately weighted toward qualitative signals alongside quantitative ones. The identity shift this phase is trying to create is not fully captured by a retention number. It requires evidence that something real has changed in how users understand their own skin.

- Users who reach the thirty-day mark at a rate that holds above an agreed threshold, with the proportion who then continue for a further fourteen days indicating that milestone moments are building momentum rather than creating a satisfaction peak followed by drop-off.
- The before and after photo comparison is opened by the majority of users who have reached week four with photos logged, and return rate in the 48 hours following the comparison holds above the baseline established in Phase 2, confirming the comparison is landing as evidence rather than disappointment.
- Qualitative review language from users at thirty days and beyond includes specific references to understanding their own skin, recognising their own patterns, or being able to explain why they use what they use. This language is distinct from generic positive sentiment and should be tracked as a separate signal.
- Pattern recognition observations are opened and engaged with at a rate that indicates users are finding them genuinely informative rather than dismissing them as irrelevant or generic. Click-through or read rate on pattern observations should be measured and compared against the equivalent rate for earlier, less specific insight cards.
- Milestone moments that are triggered do not prompt a support query or a negative review response about the timing feeling wrong or premature. Any milestone that triggers early due to a data threshold miscalibration must be caught in QA before it reaches a live user.

- The informed product decision moment, when it surfaces, does not appear alongside or immediately before a product recommendation. Confirmed by design review before the feature ships.
- Users who receive the dermatologist-ready summary feature engage with it at a rate that suggests genuine usefulness rather than casual curiosity. This is a directional signal rather than a defined threshold, but it should be reviewed at thirty days post-launch.
- App open rate remains consistent or improves between weeks ten and thirteen relative to weeks six through nine, indicating that Phase 3's identity layer is sustaining rather than concluding engagement.
- The copy audit of all Phase 3 screens produces a written record of every piece of text reviewed and its pass or fail outcome. No screen enters production with a failed item unresolved.

Phase 4: Polish & Pre-Launch

By the time Phase 4 begins, the product is functionally complete. The architecture is in place, the emotional logic has been built into the experience, and real users have been moving through the journey. What this phase is not is a grace period for finishing features that were not ready. It is the sustained, deliberate work of making sure that everything already built is as honest, as calm, and as trustworthy as the strategy promised it would be. The distinction matters. Feature additions at this stage are a signal that the earlier phases were not properly closed. Quality work at this stage is the sign that they were.

The word "polish" can be misleading here. It implies cosmetic finishing, a pass of shine applied to something already good. That is not what this phase asks for. What it asks for is rigour, a careful, systematic review of the product as a whole to confirm that nothing has drifted between the design intent and the built reality, and nothing has been quietly deprioritised under delivery pressure in a way that erodes the trust the product is supposed to build. Skin is personal. The users who will open SkinSync carry real vulnerability with them. They deserve a product that has been reviewed from that perspective, not just from a functional one.

Deliverables

Deliverable	Priority	Notes
Performance audit across target devices	Critical	Cold launch time, screen transition smoothness, routine check-off response time, and photo comparison load on a mid-range Android device. The product opens at moments of low attention. Nothing should feel slow. Specific targets: cold launch to first interactive state under three seconds on mid-range Android; routine step tap response under 100ms; no dropped frames during any screen transition specified in the animation guide.
Copy audit across all screens and states	Critical	Every piece of user-facing text reviewed against the Read It Aloud criteria and the Voice Sort outputs. Includes error states, empty states, loading copy, notification strings, and milestone copy, not just the primary onboarding and insight screens. Any line that would tip into outcome promise, hype language, or guilt framing when spoken aloud is rewritten before this phase closes. A written record of every item reviewed and its pass or fail status must exist at sign-off.
Emotional review of the full user journey	Critical	A structured walkthrough of the product from first open to the week four milestone, conducted with fresh eyes against the Aspiration Gap and Heartbreak Scale findings. Specifically reviewing the two-week progress moment, the missed-day response, the ingredient conflict warning experience, and the pattern recognition delivery. The question at each moment is not whether it works functionally but whether it lands emotionally as the design intended.
Accessibility review	Critical	WCAG 2.1 AA compliance confirmed across all screens currently in scope. Minimum tap target sizes, colour contrast ratios, screen reader label accuracy, and prefers-reduced-motion behaviour all verified. The prefers-reduced-motion fallback specified in the animation guide must be tested on device, not assumed from code review.
Full regression test across iOS and Android	Critical	Every flow that was working at Phase 3 sign-off is confirmed still working. Particular attention to the streak mechanic, the photo comparison trigger and display, the insight surfacing conditions, and the notification delivery behaviour. Any regression that touches a user's data, streak record, or logged history is treated as a launch blocker regardless of apparent severity.

Deliverable	Priority	Notes
App Store and Play Store preparation	High	Screenshots, metadata, app description, privacy policy link, data handling declarations. The app description must pass the same voice review as in-product copy. It represents the first piece of copy a potential user encounters and it should sound like the same product they are about to open. Health data declarations and permission explanations must be accurate and complete.
Notification audit (timing, frequency, and copy)	High	Every notification string reviewed against the notification logic framework established in Phase 2. Confirmed that no notification sends without a meaningful triggering event, that rate limiting is functioning correctly, and that the copy for the missed-day return notification specifically uses "Welcome back" framing rather than any streak-recovery language.
Motion and interaction consistency review	High	A screen-by-screen check that the animation guide has been applied correctly throughout, with no one-off durations, no spring physics, no transitions that deviate from the specified token values. Any component that was built in a different phase and not yet reviewed against the guide is confirmed or corrected here.
Load testing of backend insight and photo endpoints	Standard	Simulated load at expected launch volume, with particular attention to the photo upload and processing path and the insight delivery endpoint under concurrent requests. Not a stress test for scale, but a confirmation that the experience stays fast at realistic launch numbers.
GDPR and data handling documentation review	Standard	Final confirmation with the legal team that consent flows, data deletion paths, photo storage handling, and privacy policy content are accurate and complete. This is not new work, but it is a required sign-off condition before store submission.

Success Criteria

These targets define what "launch ready" means in measurable terms. They are not aspirational. A product that has not met them has not completed Phase 4.

- Cold launch to first interactive state is under three seconds on a mid-range Android device, tested on physical hardware rather than a simulator.
- Routine step tap response is under 100ms in normal use conditions, with no perceptible lag during rapid multi-step check-off.
- All screen transitions match the specified durations and easing values from the animation guide, confirmed by frame-rate review rather than visual estimation.

- The prefers-reduced-motion fallback has been tested on a physical device with the setting enabled, and all transitions reduce correctly to opacity fades at 150ms with no movement.
- Every piece of user-facing copy has been through the Read It Aloud review and any failed items have been resolved. The written record of this review is complete and signed off.
- No notification string in the product uses guilt language, streak-recovery framing, or outcome promises. Confirmed by the notification audit.
- The missed-day return notification reads "Welcome back" or equivalent, with no reference to a broken streak. Confirmed on both platforms.
- The two-week progress photo comparison screen includes honest timeline framing copy and no product recommendation is present anywhere on the screen or within one tap of it. Confirmed by design and copy review.
- Full regression testing has been completed and all launch blockers resolved. Any remaining known issues are documented with severity classification and a clear decision on whether each is a launch blocker or a post-launch fix.
- App Store and Play Store metadata, privacy declarations, and health data handling statements have been reviewed and approved by the legal team.
- Backend insight and photo endpoints have been load tested at expected launch volume with no performance degradation observed.

Pre-Launch Emotional Review

The pre-launch emotional review is a structured walkthrough of the product conducted as a team, with at least one person present who was not involved in building the feature being reviewed. Its purpose is not to find bugs. QA handles those. Its purpose is to ask a harder question at each stage: does this feel like what we said it would feel like?

The review is conducted in a single session, moving through the product in sequence from first open to the week four milestone state. The session should use a device and account that has been populated with realistic data, ideally several weeks of logged routines, at least one missed day, at least one ingredient conflict flag, and a set of progress photos that show genuine but modest change. Reviewing the product on a clean account misses everything that matters most about the experience, because everything that matters most happens after the first session.

The session works through five specific moments, each drawn from the research. These are the points the workshops identified as the ones where trust is most fragile and the emotional handling is most consequential.

The first is the onboarding end-state, specifically whether a new user reaches their first completed routine step feeling relief rather than completion of a task. The question is not whether the flow works. It is whether a person arriving with the emotional state described in the Aspiration Gap exercise, curious, slightly self-conscious, quietly sceptical, would feel understood by what they encountered.

The second is the missed-day state. The team logs a routine, then skips the following day, then returns. The question is whether the product's response makes coming back feel easy or whether it carries any trace of guilt, even subtly, even in the visual weight of the gap display rather than the copy.

The third is the ingredient conflict warning. A product conflict is triggered and the team reviews the full experience: the arrival of the flag, the contextualisation copy, the distinction between "worth being aware of" and "worth pausing on". The question is whether a person seeing this for the first time would feel informed or alarmed.

The fourth is the two-week progress photo comparison. The comparison is opened with deliberately modest change visible. The team reads the surrounding copy aloud, not silently. Anything that sounds like a reassurance that would not have been credible the week before, or any copy that implies the product has something to hide, is flagged for rewriting.

The fifth is the first pattern recognition observation. The product surfaces a cause-and-effect observation tied to the test account's data. The team asks whether it sounds like something a trusted, knowledgeable friend would say, whether it attributes the insight to the user's own tracking, and whether there is any commercial intent detectable anywhere on the screen.

Any finding from the emotional review that changes something on a screen should be treated with the same process rigour as a QA bug: documented, assigned, and resolved before the launch sign-off meeting. Emotional review findings are not soft feedback. In a product built on the premise that how it feels is inseparable from what it does, they are functional requirements.

The launch sign-off is not a meeting where the team decides whether to launch. It is a meeting where the evidence for launch readiness is reviewed. All success criteria must be confirmed met, all launch blockers resolved, and the emotional review sign-off document in hand. If those three things are true, the product is ready. If they are not, the phase is not complete.

Post-Launch

The things that happen after a product launches are not less important than the things that happened before. They are just harder to plan for, because they depend on evidence that does not exist yet. The purpose of a structured post-launch review cadence is not to create reporting overhead. It is to ensure that the real-world behaviour of real users becomes the primary input into every subsequent decision, and that the distance between what the product was designed to do and what it is actually doing gets measured, named, and acted on at regular intervals rather than accumulating quietly until something breaks.

SkinSync's design is built on a specific set of assumptions about how users feel, what they need at each stage of the journey, and where trust is won or lost. Some of those assumptions will prove correct. Some will not. The product that launches is not the product that will serve users best in month six, and the gap between the two is closed not by intuition but by honest, structured

engagement with what the evidence shows. The review cadence below defines when that engagement happens, what it covers, and how it feeds forward.

Review cadence

Review	Timing	Output
30-Day Review	Four to five weeks post-launch	Retention and drop-off analysis across the first three weeks; qualitative review of user-reported experience; audit of the two-week progress moment against the emotional review criteria; notification performance and opt-out rates; identification of any launch blockers that were accepted as post-launch fixes and their resolution status; a written list of confirmed issues and a prioritised fix list for the first update cycle
90-Day Review	Twelve to thirteen weeks post-launch	Full retention cohort analysis through the journey arc; qualitative signal tracking for identity shift language in reviews and in-app survey responses; insight delivery performance against the data threshold framework; streak mechanic behaviour and missed-day return rates; assessment of whether the Phase 2 emotional targets are being met in practice; first structured decision on whether any design principle is being tested in a way that requires a product response
6-Month Review	Twenty-four to twenty-six weeks post-launch	End-to-end journey review against all original success criteria; assessment of whether Phase 3 identity shift outcomes are visible in the data; commercial readiness assessment for Phase 2 features (dermatologist templates, community sharing, retailer integration) against trust and retention baseline; structured decision on Phase 2 scoping and sequencing; voice and copy drift review to confirm the product has not accumulated language that contradicts the tone guidelines under delivery pressure
Annual Review	Fifty-two weeks post-launch	Full strategic review of the product against the original aspiration gap framing; assessment of whether the brand position (patient, evidence-led, honest about pace) has held or drifted; user research to capture identity shift outcomes at scale; commercial performance review including any Phase 2 features delivered; updated product roadmap for the following twelve months grounded in evidence rather than assumption

How each review feeds forward

The 30-day review is primarily diagnostic. It answers the question the team cannot answer before launch: is the product behaving as designed when real people use it in real conditions? The retention and drop-off analysis at this stage is weighted toward the first three weeks, because that is the window the research identified as the highest-risk period, and because problems surfaced at thirty days can be addressed in a first update cycle before they compound. The most important output of this review is not the data itself but the prioritised fix list it produces. That list drives the first post-launch development sprint, and it should be treated with the same rigour as the pre-launch phase

deliverables. Any finding that touches the two-week progress moment, the missed-day experience, or the ingredient conflict handling should be escalated immediately regardless of when the review formally takes place. Those are not issues that can wait for a scheduled cycle.

The 90-day review shifts from diagnostic to evaluative. By this point there is enough cohort data to understand whether users are progressing through the emotional arc the product was designed to support, or whether they are dropping off at points the design did not anticipate. The qualitative signal tracking matters at this stage because the identity shift the product is trying to create is not fully visible in a retention number. A user who reaches day 90 is retained, but the question the review asks is whether they feel informed and in control of their own skin, or whether they feel like someone who has been following instructions. The distinction shows up in how they talk about the product, in review language, survey responses, and the specific words they use to describe what they have learned. This review also produces the first structured decision about whether any of the five design principles are being actively tested by user behaviour in a way that requires a product response. If the honest pacing principle is under pressure because users are interpreting flat progress as product failure rather than biological timeline, that is a design problem that requires a design response, not a copy change.

The 6-month review is the moment when the decision about Phase 2 features moves from theoretical to grounded. Dermatologist-reviewed templates, community sharing, and retailer integration were all designed with a clear dependency: they require a trust baseline established by a product that has held its principles under real-world pressure. The 6-month review confirms whether that baseline exists. If retention through the identity layer is strong, if the brand position has held, and if the commercial case for the Phase 2 features is supported by the data, the review produces a scoping and sequencing decision for what comes next. If the baseline is not yet established, the review produces a different decision: address what is not yet working before adding complexity. The voice and copy drift review at this stage is a quiet but important check. Products that move fast under delivery pressure tend to accumulate language that drifts from the original tone without anyone making a conscious decision to change it. Six months is the right point to catch that drift before it defines the next phase.

The annual review is the most consequential because it asks the hardest question. Not whether the product is working, but whether it is still the product it set out to be. The original aspiration gap framing describes a user who arrives confused and reactive and leaves feeling informed and genuinely in control of their own skin. The annual review asks whether that transformation is happening at scale, and whether the brand position that makes it possible, patient, evidence-led, and honest about pace, has been maintained through every decision made in the twelve months since launch. The output is not a retrospective. It is a forward-looking document: an updated product roadmap grounded in a year of evidence, a refreshed understanding of where users start and where the product is capable of taking them, and a clear-eyed account of what the next twelve months should be designed to achieve. That document is, in effect, the beginning of a new version of this one.

Dependencies & Risks

The risk register below covers the most significant delivery threats across the four-phase sequence. These are not hypothetical concerns. Each one reflects a genuine point of vulnerability identified either in the product's technical architecture, the emotional complexity of what is being built, or the dependency chain that makes the phased approach necessary. Mitigations are specific rather than generic. A mitigation that could apply to any project is not a mitigation; it is a placeholder.

Risk	Likelihood	Impact	Mitigation
Ingredient conflict data source proves insufficient in coverage or quality, requiring a mid-build pivot away from the initially selected third-party API	Medium	High	Evaluate at least two candidate sources (Open Food Facts Cosmetics and Skincarisma) against a sample of twenty to thirty real product combinations before committing to one. Define the minimum acceptable coverage threshold in writing before Phase 1 closes. If neither source meets it, the decision to build a proprietary dataset or defer the conflict feature must be made before Phase 2 begins, not during it.
WatermelonDB sync behaviour produces data conflicts when a user logs routines offline and reconnects, resulting in duplicate or missing check-off records	Medium	High	Build and test the offline sync path against a deliberately adversarial scenario: log routines across three days without connectivity, reconnect, and confirm that the backend receives all records with correct timestamps and no duplicates. This test must pass before Phase 1 closes. The backend must treat client-side writes as additive and never overwrite a local record with an older server-side version. Document this constraint in the API contract before the sync layer is built.
Screen transition and element choreography performance drops below 60fps on mid-range Android devices, making the experience feel janky rather than calm	Medium	High	All animation work must be tested on a physical mid-range Android device throughout build, not only in a simulator or on a flagship handset. Only `transform` and `opacity` properties should be animated, with `useNativeDriver: true` enforced wherever Reanimated is used. A frame-rate review of all transitions is a Phase 4 sign-off requirement. Any transition that fails the 60fps test must be simplified, not tuned, because complexity under pressure produces inconsistency.

Risk	Likelihood	Impact	Mitigation
Navigation pattern behaviour diverges between iOS and Android, particularly around bottom sheet dismiss gestures and back navigation, producing an experience that feels inconsistent across platforms	Medium	Medium	Specify the exact gesture behaviour for each platform in the interaction guide before build begins. Bottom sheet swipe-to-dismiss follows platform conventions on each OS but the animation duration and easing must match the specified tokens regardless. Assign one person the explicit responsibility of cross-platform consistency review at the end of each phase. Do not leave platform divergence to be caught at Phase 4.
Expo managed workflow encounters a limitation during build that requires a native module, introducing scope and timeline impact that was not anticipated	Low	High	Identify any feature that carries a non-trivial risk of hitting an Expo managed limitation before Phase 2 begins. The most likely candidates are advanced camera controls for progress photos and specific background processing requirements. If a limitation is identified, make the bare workflow decision immediately rather than attempting workarounds that compound debt. Swift and Kotlin are available as the fallback path; confirm team capability against both before Phase 1 closes.
Copy review creates a bottleneck late in each phase when multiple screens require voice-guideline sign-off simultaneously, delaying QA and phase close	High	Medium	Build copy review into the phase schedule as a parallel workstream rather than a gate at the end. Assign a named reviewer against the voice guidelines from Phase 1 onwards. Any screen that reaches the build-complete state without its copy having been reviewed against the Read It Aloud criteria and the Voice Sort outputs is treated as incomplete, not as ready for QA. The copy audit table required for Phase 4 sign-off should be a living document maintained throughout the build, not created retrospectively.

Risk	Likelihood	Impact	Mitigation
The streak mechanic introduces visual or copy elements during build that carry guilt framing, contradicting the neutral missed-day design principle even when the intent was neutral	Medium	High	The streak mechanic interaction and all associated copy states must be reviewed against the design principles before the feature enters QA. The review must specifically test the missed-day return state using a named person who was not involved in building the feature. A broken-chain visual or any copy that implies recovery rather than simple return is a failure state, not a preference. This review is a Phase 1 close requirement; the mechanic is too foundational to correct in a later phase without user-facing disruption.
Progress photo storage and retrieval fails under concurrent load at launch, causing users to encounter errors or delays at the product's most emotionally significant moment	Medium	High	Load test the photo upload, processing, and pre-signed URL retrieval path at expected launch volume before Phase 4 closes. The test scenario must include concurrent uploads from multiple users and pre-signed URL expiry handling. The mobile client must handle a 403 response gracefully by requesting a fresh URL rather than surfacing an error. Any performance degradation on this path is a launch blocker. The photo comparison moment was named across multiple workshop sessions as the product's most consequential trust event; it must not feel slow or broken.
Quality gate failures at phase boundaries, where deliverables are marked complete without all success criteria genuinely met, allow structural problems to carry forward into subsequent phases	Medium	Critical	Phase close requires a written record of each success criterion and its confirmed status, signed off by both the delivery lead and the product lead. Partial completion is not completion. Any criterion that cannot be confirmed met at the phase close meeting must be either resolved before the next phase begins or reclassified as a launch blocker with a documented escalation path. The emotional review conducted in Phase 4 and the pre-launch review are both designed to catch drift that accumulated across phases, but they are a last line of defence, not a planned catch mechanism.

Risk	Likelihood	Impact	Mitigation
Supabase Auth session management or JWT revocation fails to handle edge cases (expired tokens not refreshed gracefully, revoked sessions not cleared on the client), causing authentication errors that interrupt a user's routine at an inconvenient moment	Low	Medium	The JWT revocation Redis blocklist must be implemented and tested before launch. Token expiry and refresh behaviour must be tested against a scenario where a user's session expires mid-routine: the mobile client must handle a 401 response by refreshing silently rather than surfacing a login prompt. This scenario must be included in Phase 4 regression testing as a named test case, not covered by general auth testing.
PostHog self-hosted infrastructure experiences availability issues that interrupt event capture during the critical first-week retention window, leaving the team without the data needed to identify early drop-off patterns	Low	Medium	PostHog should be deployed and load tested before launch, with CloudWatch alarms on the ClickHouse instance covering both availability and disk usage. Define the minimum acceptable data capture rate and confirm it is being met in the two weeks before go-live. If self-hosted stability cannot be confirmed, the fallback decision should be made before launch: either PostHog Cloud with a documented data handling justification, or a delayed analytics launch with manual review as the interim approach.