

SYSTEMS MANUSCRIPT

O-Lab: An Online Laboratory for Musical Intelligence

Twenty familiar classical works, precomputed interpretations and online response collection

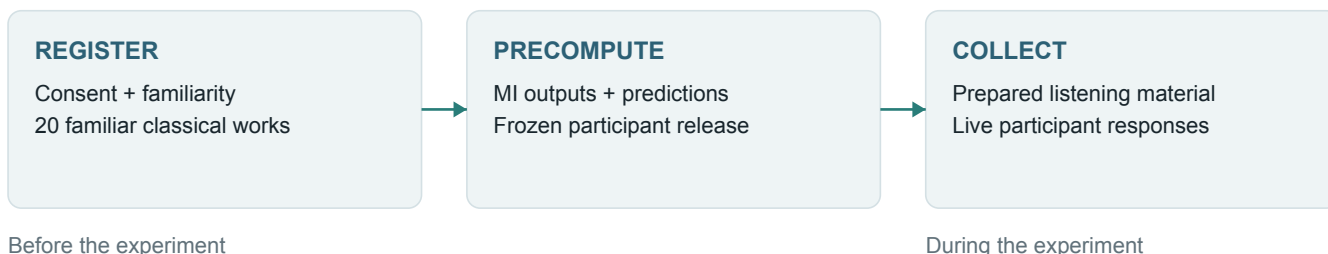
Amaç Erdem | Independent researcher

10 October 2026 | Technical systems protocol - Draft 2 / classical design

Abstract. O-Lab is a proposed online research environment for comparing interpretations of familiar classical works. The platform targets 1,000 registered people, with consent, screening, readiness and completion reported separately. Each participant confirms exactly 20 distinct familiar composition families from a curated catalogue and completes a short reference-performance calibration. No general listening-history archive or streaming-account connection is required. Shared MI audio representations are prepared before catalogue release; participant-specific predictions and trial schedules are prepared when a validated baseline snapshot arrives. During the experiment, the platform presents prepared alternatives and collects responses without MI, GPU or LLM inference. This revision specifies composition/performance identity, registration, state transitions, storage, email, computation, privacy, operational sizing and acceptance criteria. Supabase and isolated batch workers remain proposed services. No participant result or production deployment is claimed.

1. Two manuscripts, one versioned contract

The scientific companion is now the classical interpretation protocol, revision 4 [1]. Its primary outcome is the error in predicting the difference between pleasure ratings for two performances of the same passage. This systems manuscript supports that design. Earlier PDFs remain historical records; their general-profile requirement and live block-update narrative do not govern this release.



Confirmed constraints. 1,000 total registrations; exactly 20 familiar classical composition families per submitted selection; reference calibration before the session; precomputed alternatives and predictions; response collection only during the experiment.

The website currently has an editable participation email draft and public research material. The local revision aligns those explanations and adds the revised papers; it does not create participant accounts, a production catalogue, database tables, email delivery or an experiment runner. All runtime commands in this manuscript are proposed contracts.

ENROLMENT CONTRACT

Count registrations and readiness separately

2. One recruitment cap, explicit denominators

Registration means one email-verified, minimally eligible adult study account accepting the registration privacy notice. Research consent precedes background, familiarity screening and calibration. Stop after 1,000 unique registrations, regardless of later completion. Withdrawals do not silently reopen slots. A registration is not proof of screening success or statistical power.

Measure	Counted event
Registered / consented	Unique eligible account / current approved consent recorded.
Screened	At least one catalogue familiarity check; retain all attempted and exposed items.
Selection complete	Exactly 20 distinct composition families explicitly familiar before today; no duplicate-family inflation.
Calibration submitted	Required reference questions completed, including explicit cannot-rate responses where appropriate.
Precomputed / ready	Current baseline, all assets, permissions, frozen predictions and schedule pass release gates.
Started / completed	First confirmatory event / all scheduled primary trials and completion markers.
Analysis eligible	Prespecified valid-pair rules; not equated automatically with full completion.

The workflow is account pending → registered → consented → background complete → familiarity screening → selection complete → calibration submitted → preparation queued → preparing → review required or ready → active → completed. A screening-incomplete state supports saved progress or exit. Withdrawal, expired consent, asset revocation or a new baseline can block admission from relevant states.

Count allocation is transactional and idempotent. Two concurrent final-slot requests cannot both receive the same registration number. Retries return the existing result. Never discard registered people who cannot identify 20 works from the recruitment denominator.

Universal success in finding 20 familiar works is an expectation to test in the pilot, not a platform guarantee. Report the proportion reaching 20, burden and differences by musical background.

PARTICIPANT JOURNEY

The registration interface

3. A guided classical-music intake

Screen	Action and persisted evidence
Account	Email verification, adult/language eligibility and registration privacy notice; no research answers before consent.
Study information	Approved consent, recording comparisons, burden, withdrawal, storage and contact information.
Background	Training, classical-listening frequency, music-making, age band, country, language and device context.
Familiar works	Catalogue search and short reference previews; yes/unsure/no prior familiarity and 1-5 intensity. Progress counts composition families to 20.
Reference calibration	Rate reference A on 0-100 pleasure; record exact-recording familiarity and cannot-rate separately.
Review and submit	Show 20 works, selected passages and answers; freeze a versioned baseline and queue preparation.
Preparing / ready	Show honest status. Issue a session invitation only after all release checks pass.

Participants need not know catalogue numbers or formal titles. Composer/common-title search and preview-based recognition are allowed. The interface distinguishes knowing the musical passage from having heard this exact performance. Every preview and replay is recorded; listening during screening cannot establish prior familiarity by itself.

If fewer than 20 qualifying families are found, offer further candidates, save progress and permit exit. Do not require streaming-account credentials, an exported account archive, fabricated selections or a forced yes. The catalogue offers only work families that already have three reviewed, permitted performances.

A participant may edit before the confirmatory session. Each submission creates a new snapshot and invalidates older release eligibility. The page explains which part is saved and which is still preparing. Recognition answers are not scored for musical expertise, and uncertainty is an acceptable response.

MEASUREMENT AT INTAKE

Background and twenty familiar works

4. Minimal background and explicit familiarity

Background variables support description and prespecified adjustment. Collect music-training years, music-making level, classical-listening frequency, usual traditions, language, age band, country and listening setup. Make optional fields clear, keep preferred name/contact outside analysis, and avoid diagnostic free text. Catalogue choice and recruitment determine representativeness; a large sample alone does not.

Field	Meaning and validation
composition_family_id	Canonical musical work. Different movements, arrangements and recordings of that work count once toward 20.
movement_id / passage_id	The actual recognized excerpt; recognition is not generalized to the complete work.
known_before_today	yes / unsure / no; only yes counts. Must be the participant's explicit response.
work_familiarity	Required ordinal 1-5 response with frozen anchors; does not substitute for prior-familiarity confirmation.
reference_recording_id	Performance A, presented during screening and calibration; alternatives B/C remain withheld.
reference_pleasure	0-100 or explicit cannot-rate. Baseline direct feedback, not an inferred listening-history value.
recording_known_before	yes / unsure / no for exact performance A; separate from work familiarity.
exposure_receipts	Passage, playback duration, replay count, order, question version and time.

Exactly 20 distinct qualifying work families form the submitted selection. The candidate pool can be larger, but unselected or uncertain items are not counted. Selection is not an observed lifetime history, an estimate of listening frequency, or evidence that the participant knows every movement of a composition.

Missing numerical calibration ratings remain missing. A frozen shrinkage/fallback policy must still produce a valid prediction release without substituting model-generated answers. Record observed-rating coverage. Completion of a question may legitimately be a cannot-rate response.

FAMILIARITY INGESTION

Freeze a catalogue and a personal baseline

5. Catalogue-first intake

The initial release uses a curated classical catalogue. No Spotify/Apple Music OAuth connector, listening-history import or participant audio upload is required. Titles and alternate names serve search; exact waveforms, rights and passage identities come from the laboratory registry. Only families with three usable performances enter screening.

Store work, movement, arrangement, performance, recording-session, physical-master and excerpt identities separately. Duplicate encodings or remasters of one performance are not independent interpretations. Keep score edition, measure boundaries, repeat handling, instrument and performer provenance. The selection count uses composition family, while MI cache reuse uses waveform/configuration identity.

5.1 Permission and source evidence

Verify the underlying composition, arrangement and sound-recording rights for analysis, playback, excerpting, storage and any redistribution. Archive licence URL/text, rights holder, checked date, jurisdiction assumptions, attribution and expiry. Public access or an old composition is insufficient by itself. IMSLP and Musopen provide useful candidates; MAESTRO/ASAP provide additional performance resources [20-23]. A page listing is not a completed quality or rights audit.

5.2 Immutable personal snapshot

Submit exactly 20 yes-confirmed composition families with required background, reference questions and exposure receipts. Hash a canonical payload containing question versions, missingness, catalogue version, reference allocation and participant responses. Editing any included field produces a new version; it never overwrites the baseline bound to an existing session.

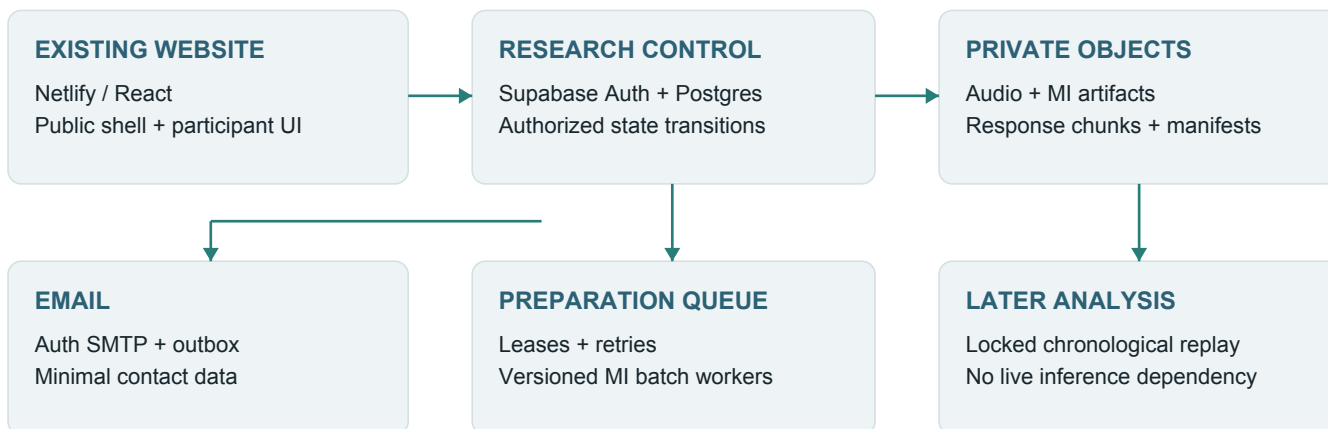
The readiness gate requires both the frozen 20-work baseline and every scheduled A/B/C audio asset, aligned passage, permission and MI artifact. Twenty familiar titles alone cannot release a session.

SYSTEM ARCHITECTURE

A reference deployment and its boundaries

6. Keep the public site; add a private research service

The proposed baseline retains the current React/Vite site on Netlify. Supabase supplies authentication, PostgreSQL, private object storage and a durable queue. Small authenticated request handlers authorize state transitions and validate batches. Separate container workers perform MI computation. These are design selections for review, not provisioned accounts. Database and worker resources are proposed in the specific Frankfurt region, subject to the eventual institution's requirements [3-9, 14-15].



Layer	Selected approach and data boundary
Public web	Netlify: site, papers, static O-Lab shell. No participant datasets, contact lists or private artifacts in deploy assets.
Identity / data	Supabase Auth and Postgres, proposed eu-central-1. Restricted identity mapping; pseudonymous research tables.
Files	Private Supabase buckets for permitted audio, MI artifacts, events and approved exports; signed access after authorization.
API / queue	Supabase request functions plus transactional database procedures and Supabase Queues. Region-pin processing where supported; verify with deployment evidence.
MI workers	Docker CPU jobs on AWS Batch/Fargate; GPU jobs on AWS Batch/EC2 when measured need justifies them. Same proposed region.
Email	Resend for Auth SMTP and transactional API mail. Recipient and provider delivery data have a separate processing footprint.
Optional LLM	Regional managed inference adapter, initially disabled; research registration and session collection work without it.

A single PostgreSQL source of truth is sufficient for this initial scale unless benchmarks show otherwise. Adding a separate vector database, Kafka cluster or Kubernetes control plane is not a requirement. This simplicity is a design judgment, not a claim that an untested smallest-tier instance will meet the session load.

IDENTITY, FAMILIARITY AND RESEARCH RECORDS

The classical-work data model

7. Separate identity, sound and evidence

Use opaque IDs; keep contact and reidentification data outside analytical exports. Familiarity and preference profiles remain personal research data. Enforce ownership, role permissions and server-authorized transitions with tested database grants and row-level policies [3].

Object group	Essential invariant
Account / consent / background	Unique study participant; versioned consent; restricted contact link; structured responses and explicit missingness.
Works / movements / arrangements	Composition-family identity is stable across excerpts, performances and arrangements; count families for the threshold.
Performances / recordings / assets	Performer, instrumentation, source/master identity, rights evidence, waveform hash, derivation and QA.
Passages / alignment	Score edition, measures, repeat path, per-recording boundaries and reviewer sign-off.
Screening / selection / calibration	Candidate exposure log; exactly 20 families; A role; familiarity and reference ratings; immutable baseline version.
MI jobs / artifacts	Waveform/configuration/engine key, status, validated output hash, retries and numerical checks.
Predictions / release	All primary predictions timestamped before session; input cutoff, model/feature versions, baseline and schedule hashes.
Trials / responses	Work, role B/C, order, playback receipt, rating/cannot-rate, pair preference, interruptions and quality flags.
Operations / analysis	Outbox, leases, withdrawal ledger, exports, locked partitions and analysis provenance.

Postgres holds structured state, responses, receipts and manifests. Private object storage holds permitted audio, MI traces and compact archives. Database backup does not cover object payloads; maintain and test independent object recovery [6]. Participants cannot read hidden predictions, another person's baseline or arbitrary catalogue assets.

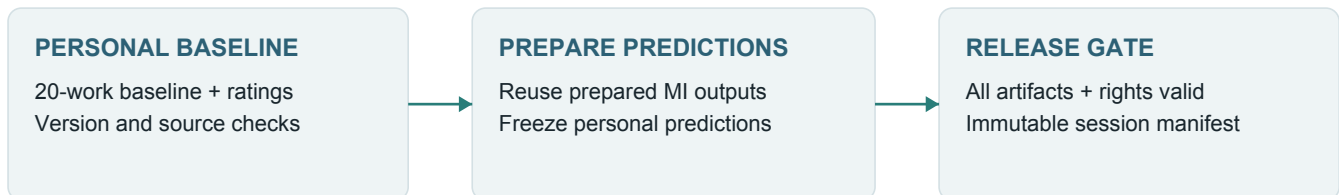
Workers receive only required source capabilities, never contact-table access. Researchers receive pseudonymous exports. Access to identity linkage requires a distinct audited role. The browser cannot write ready state, overwrite model outputs or choose its own trial sequence.

THE CENTRAL COMPUTATIONAL CONTRACT

Precompute the pool, then the personal release

8. Shared preparation before catalogue release

Compute shared MI outputs for every approved waveform and exact presentation excerpt before that family is offered for registration screening. A candidate family needs three usable performances for the primary arm. Cache keys include waveform hash, decoder/preprocessing, excerpt boundaries, sample rate/channels, gain policy, engine commit, output schema, precision and configuration.



Baseline edited? New snapshot and prediction release required.

Shared catalogue audio is prepared before participant screening.

8.1 Preparation as participants submit

A validated 20-work baseline triggers participant-specific fitting, prediction and schedule assembly. Reuse already validated audio artifacts. Fit only the frozen low-dimensional personal parameters using reference responses and allowed background fields. Population parameters and feature selection are fixed independently of confirmatory responses. Save every B/C prediction before the first test trial.

A participant-specific key includes baseline hash, catalogue, shared artifacts, model/hyperparameter hash and allocation seed. These artifacts remain private. A changed baseline, catalogue, rights status or scientific configuration requires a new release. A successful job for an obsolete baseline cannot make the participant ready.

8.2 Release and retries

Validate all permissions, hashes, finite outputs, schemas, alignment QA, exactly 20 qualifying work families, reference-question completion and all scheduled predictions. Commit a single immutable release and one invitation outbox event atomically. Jobs use leases, heartbeats, bounded retries and idempotent result commits [7]. Withdrawal cancels participant preparation and is rechecked before completion.

An LLM is not needed for deterministic catalogue selection, threshold validation or fitting. GPU work is optional and restricted to measured batch workloads. No queue, model service or generative assistant is on the live trial path.

API AND RELEASE RULES

Transactions that enforce the study contract

9. Proposed commands, not deployed endpoints

Command	Atomic behaviour
register / consent	Allocate one of 1,000 unique registrations; bind current notice/consent and eligibility versions.
record-screening	Store own candidate exposure and explicit recognition response; never infer familiarity from playback.
submit-baseline	Require exactly 20 distinct yes-confirmed work families and completed reference questions; freeze snapshot and queue preparation.
commit-job	Verify lease, current authorization, baseline/model versions and validated artifact hashes; reject obsolete completion.
release-participant	Check all A/B/C assets, alignment, permissions and primary predictions; freeze schedule; outbox one invitation.
start-session	Recheck consent/release and claim a single active lease; return stimulus access without hidden predictions.
append-response	Validate scheduled trial, role, response scale and playback receipt; persist idempotently before acknowledging.
finish / withdraw	Preserve missingness and partial progress; completion requires scheduled markers. Withdrawal revokes admission and contact.

Every mutation checks ownership, study state, schema and payload size. Idempotency-key reuse with a different payload is a conflict. Test two-tab races, last-slot races, baseline changes while a job finishes, duplicate submission, revoked audio and withdrawal during invitation delivery.

Exactly 20 means 19 fails, 21 must be reduced explicitly before submission, and duplicate performances of one work cannot inflate the count. This is a server-side invariant in the proposed implementation. It is not currently enforced by the public email-interest form.

A release points to immutable objects, not mutable latest paths. Archive compaction has a checksummed receipt and an independent recovery copy. Record revoked releases and never overwrite the evidence of what a participant was scheduled to hear.

COLLECTION-ONLY RUNTIME

During the experiment, collect responses

10. Primary paired-performance task

The runner presents 20 work pairs in four proposed blocks of five. Each pair contains the prepared alternative recordings B and C, with balanced order and a frozen passage match. After each excerpt collect 0-100 pleasure or cannot-rate; after the pair collect preference/tie/cannot-judge and exact-recording familiarity. Reference A was assessed during onboarding. All trial timing and quality thresholds are versioned.

The primary release uses discrete ratings, replacing the previous fixed one-hour continuous-slider schedule. Continuous-rating instrumentation may be developed as a separate secondary task; it is disabled in the primary contract. Derive the invitation's duration estimate from actual audio durations plus pilot-measured response/setup time.

Use a version-pinned jsPsych or equivalent runner with explicit durable server persistence [16]. Record audio-context onset/end, monotonic browser times, visibility, interruptions, replay/exposure state, trial IDs and server receipt times. Network receipt is not stimulus onset, and timestamps alone do not validate timing accuracy.

10.1 Durable acquisition

Persist each answer and playback receipt in an idempotent transaction. A bounded local IndexedDB queue retries unacknowledged events. Do not label responses saved until acknowledged. Clear local payloads after durable receipt, and expose pending data before abandonment. Stop at a safe boundary when the tested buffering limit is exceeded.

A repeated excerpt is a new exposure and keeps its interruption/replay flags. Missing answers remain missing. Do not fill gaps with MI estimates. A finished session is distinct from an analysis-eligible pair and from a complete dataset.

Primary predictions remain frozen. Any feedback-dependent relational model is run later by locked chronological replay and reported as secondary retrospective evaluation, not as online adaptation.

COMPUTE ALLOCATION

CPU, GPU and the preparation queue

11. Profile the engine before buying capacity

A CPU container is the baseline for decoding, validation, catalogue resolution, database work and any MI component whose measured throughput is adequate. The initial benchmark candidate is 2 vCPU/4 GiB, then 4 vCPU/8 GiB if needed. These are test configurations, not measured requirements. Record elapsed time per audio minute, peak RAM, storage traffic, numerical output differences and failure rate on representative durations and formats. Cache reuse must be reported separately from new computation.

Work	Execution and admission rule
Account / intake API	Short authenticated operations on the request service. Expensive parsing is queued.
Decode / MI extraction	Batch CPU containers first; GPU-capable containers only if the actual engine path supports and benefits from acceleration.
Release assembly	CPU task joins permitted shared artifacts to the participant snapshot and verifies the fixed session manifest.
Experiment collection	Browser playback and response capture; small server writes. No inference queue waits in a live block.
Post-session analysis	Separate analysis queue; locked code, partitions and chronological input cutoffs.
Optional LLM	Separate request budget and permissions; disabled without changing core onboarding or experiment availability.

Supabase request functions have constrained CPU and memory and are not the home for long MI jobs [8]. AWS Fargate does not support the GPU resource parameter; optional GPU jobs therefore use an EC2-backed Batch environment [14-15]. A 16-24 GiB GPU can be benchmarked as a candidate, but compatibility, model memory, account quotas and regional availability must be verified before selection. GPU use is not justified solely by the word AI.

11.1 Bounded workers and backpressure

Maintain separate queues for preparation, analysis and optional assistance. Preparation receives priority while recruitment is active; analysis cannot exhaust the resources needed to admit participants. Worker concurrency has a configured ceiling, daily spend cap and per-job timeout. A dispatcher leases preparation jobs, submits resource-specific Batch tasks and records the external job identifier. Queue reconciliation detects tasks that were submitted but whose submission acknowledgement was lost.

11.2 Readiness is an operational promise

Show a coarse preparing status and a measured queue estimate only when reliable. Do not promise immediate access after list submission. Rights review, missing recordings and numerical failures are distinct blockers. Invite participants only after preparation; schedule invitations in tested batches to limit simultaneous sessions. Preemptible workers may be used for restartable precomputation, with checkpoint or restart rules and a bounded on-demand fallback.

OPTIONAL ASSISTANCE, CONTROLLED INPUTS

What an LLM may do

12. Useful assistance without changing the experiment

The initial release can operate without an LLM. Where an assistant is later useful, it answers questions from approved study information, explains how to complete a familiar-work selection, or proposes a structured interpretation of participant-entered text. It must not generate a familiar-work profile, infer unreported familiarity, select consent answers, diagnose participants, or replace a human response with a prediction. Deterministic validation remains the authority for eligibility and readiness.

Task	Permitted proposal / required control
Registration help	Retrieve only approved, versioned FAQ text; display the source and route uncertain answers to staff.
Free-text normalization	Propose an instrument/tradition category; preserve original text and require confirmation where meaning changes.
Recording disambiguation	Suggest candidates from permitted metadata; deterministic checks and participant/staff confirmation establish identity.
Researcher assistance	Work on approved pseudonymous or synthetic exports; retain code and model provenance. No silent writes to study data.
Experimental instructions	Human-reviewed, frozen text. No personalized ad-lib explanations inside study blocks.
Scientific outcomes	Not generated, imputed or scored by the assistance model in this protocol.

12.1 Proposed provider boundary

A regional Amazon Bedrock adapter is a candidate because compute is already proposed on AWS. No foundation model is selected in this draft. Activation requires a named model/version, confirmed region, task evaluation, contract/retention review and a cost limit. Bedrock supports cross-region routing, so a regional-looking endpoint alone is insufficient evidence that processing stays in one region [17-18]. Keep cross-region/global profiles disabled unless the approved data plan explicitly permits their destinations. If requirements cannot be met, retain static help.

12.2 Input, output and audit controls

An application gateway strips contact information, consent details and unnecessary personal text. It supplies only the minimum permitted task fields. Participant familiar-work profiles are not automatically sent to an LLM, even if pseudonymized. Imported text is untrusted data; it cannot instruct the assistant to export records, query arbitrary URLs or modify permissions. Tool access is allowlisted and read-only by default; returned structures are schema-validated.

Log task type, prompt-template version, model identifier, configuration, input/output hashes, token counts, latency, rejection and human-correction outcomes. Retain exact sensitive prompts only under an approved necessity and retention rule. A model upgrade runs the same synthetic/approved evaluation set before release. Required evaluations include unsupported-answer rate, factual agreement with the FAQ, metadata correction accuracy and tests that attempt to extract another participant's information. No accuracy is reported until those tests have been run.

IDENTITY AND TRANSACTIONAL MESSAGING

Email is part of the research workflow

13. Two mail paths, separate purposes

Supabase Auth uses a configured Resend SMTP service for verification and login. The default Supabase mail service is unsuitable for public recruitment [2]. The application sends study messages through a server-side Resend API integration and a database outbox. No vendor key is placed in the browser. The proposed sending addresses are `login@auth.musicalintelligence.app` and `study@lab.musicalintelligence.app`; they are not configured by this document. Replies must reach a monitored mailbox with a named owner.

Trigger	Message and constraint
Verification requested	Short OTP email, expiring and rate-limited; no music profile or background information.
Consent completed	Consent receipt or authenticated link to the exact version. No implication that profile preparation has finished.
Profile submitted	Receipt and preparing status. Do not claim the experiment is ready.
Participant release ready	One invitation with a normal authenticated site link, availability instructions and support contact.
Clarification required	Generic request to review an issue on the site; avoid identifying song lists in the email body.
Incomplete onboarding	Optional reminder under the approved contact policy; stop after the bounded sequence or opt-out.
Session completed	Receipt after server-confirmed completion, with debrief/support information.
Withdrawal / security	Appropriate confirmation and operational follow-up; reminders are cancelled immediately.

13.1 Reliable sending

The transaction that creates a message-worthy state also writes an outbox row. A sender claims it, rechecks current consent/contact status, and passes a stable event key to the provider. Resend documents a 24-hour idempotency window, so O-Lab must retain its own message/event uniqueness beyond that window [12]. If the provider accepted a request but acknowledgement was lost, reconcile before resending; do not blindly retry indefinitely.

Verify webhook signatures and deduplicate provider event IDs. Delivery, delay, bounce and complaint events update the ledger; delivery does not prove that a person read the email [13]. Stop nonessential mail on complaints and hard bounces. Separate reminder preferences from authentication mail and essential study notices. Disable click/open tracking for research messages; do not treat tracking pixels as participant engagement evidence.

13.2 Launch preparation

Configure and verify SPF, DKIM and DMARC for the sending subdomains, review existing domain records before edits, and test common mailbox providers. Set application and provider rate limits for the expected recruitment bursts. A proposed reminder policy is at most two onboarding reminders, seven days apart, unless the approved protocol differs. The dates and cadence are configurable, timezone-aware and cancelled when their triggering condition no longer holds.

DATA GOVERNANCE

Where information is stored and processed

14. A specific region is a proposal, not a compliance certificate

The baseline proposes Frankfurt (eu-central-1) for primary Supabase data, MI/analysis workers and a separately controlled backup bucket. Supabase distinguishes specific regions from general region groupings [5]. The final choice depends on the responsible institution, recruitment jurisdictions and approved data plan. A region setting alone does not establish GDPR, KVKK or other legal compliance. This draft is an engineering specification and leaves legal/ethics determinations to the responsible institution.

Data class	Proposed destination / access
Email / account identity	Supabase Auth and restricted identity records. Resend also processes recipient and delivery data; its processing locations and retention require contractual review.
Consent / background / profile	Primary regional Postgres with restricted research access. No automatic disclosure through public site analytics.
Permitted audio / MI traces	Private regional object storage; only approved roles and bounded worker capabilities. Rights may further restrict replication.
Live response chunks	Recent payloads and receipts in regional Postgres; compacted archives in private object storage. Temporary retry buffer on the participant's device.
Backup copies	Encrypted independent storage in the approved region/account, with separate access and explicit expiry.
Public site / papers	Netlify CDN. The website provider still receives ordinary request metadata; global delivery is not EU-only processing.
LLM / operational logs	Minimal permitted fields only. Review actual destinations, logging and subprocessors before enabling the optional service.

14.1 Minimum security posture

Require TLS, encrypted storage, administrator MFA, least-privilege roles, protected secrets and separate development/staging/production projects. Development uses synthetic records. Enable deny-by-default row and object policies; test cross-participant reads and writes, anonymous access and staff-role boundaries [3-4]. Service-role credentials remain exclusively server-side. Support staff cannot freely browse scientific responses, and analysts cannot freely recover contact identities.

Restrict browser origins, escape participant text, apply a content security policy, rate-limit auth/intake requests and reject unexpected upload types or oversized/decompression-heavy files. Do not fetch arbitrary playlist URLs supplied by users without an allowlist and server-side request safeguards. Redact tokens and personal text from logs. Signed URLs are temporary bearer capabilities, not DRM; do not describe them as preventing copying. Their duration and cache behavior must match the study's access policy.

OPERATIONS THAT MUST BE TESTED

Retention, withdrawal and recovery

15. An explicit, reviewable retention schedule

The following periods are proposed engineering defaults, not approved research/legal obligations. The institution must adopt or replace them in the consent and data-management plan before recruitment. Changes are versioned. Research data may be retained for a different approved period than temporary imports and contact details; do not promise immediate erasure from every backup.

Record type	Draft retention rule
Unverified registration	Expire OTP normally; remove inactive unverified registration state after 7 days.
Screening / calibration	Versioned responses and exposure logs follow research retention; no account-history archives collected.
Contact linkage	Retain through participation and essential follow-up, then a proposed 90-day closeout period; optional future contact stored separately.
Consent / research records	Proposed five years after study close, subject to institution, consent, source rights and withdrawal conditions.
Operational logs	30 days for sanitized routine logs; longer security/consent audit retention requires a defined justification.
Backups / browser buffers	Proposed 35-day rolling independent backups; purge browser retry data after durable acknowledgement or controlled abandonment.

15.1 Withdrawal is a coordinated workflow

Immediately block new sessions, disable reminders and queued participant jobs, and record the withdrawal scope. Revoke new access and stop active collection at the next safe opportunity. The data steward executes deletion or restriction under the approved policy across identity mappings, rows, objects, personal derived artifacts and exports still under laboratory control. Distinguish a request for no further contact from a request to delete research data. Truly anonymous released aggregates cannot necessarily be recovered; explain the actual boundary in advance.

Keep a minimal withdrawal ledger under a defined lawful retention basis so that a restored backup does not reactivate a participant. Apply that ledger before reopening services after restoration. Existing signed URLs can remain valid until their expiry; clearing browser caches or previously downloaded material cannot be guaranteed. Avoid absolute instant-revocation claims.

15.2 Recovery objectives

The proposed live-study target is a database recovery point of at most five minutes and service recovery within four hours, demonstrated by a restore exercise. Supabase daily backups alone do not meet that target; PITR is an additional paid option, and file objects require separate backup [6]. A response acknowledgement means that the primary database transaction committed, not that catastrophic loss is impossible. Archive roughly five minutes of event batches together, verify the primary and independent object copies, then expire covered staging payloads. PITR and archive recovery must together demonstrate the five-minute recovery-point target. Manifests retain hashes and sequences so archives can be reconciled after database rollback.

Perform a restore into an isolated environment before launch and periodically during collection. Verify identities, consent, familiar-work snapshots, event receipts, artifacts and withdrawal suppression. Recovery time and data loss are measured outcomes. Backups, key rotation and incident response each need a named owner and a tested runbook.

PLANNING EXAMPLES, NOT BENCHMARK RESULTS

Size the workload from the classical catalogue

16. Families, recordings and participant links

For H people with submitted selections, store 20H selected-work links. H can be less than the 1,000 registration cap. With W released work families and three primary recordings per family, shared computation covers about 3W waveform inputs, plus separately keyed excerpts and any optional arrangement assets. A new participant reuses shared audio features and requires personal preparation only.

Planning example, not a measured requirement	Quantity
1,000 submitted profiles x 20 works	20,000 selection links; about 40 MB at a hypothetical 2 kB per link before indexes and provenance.
200 work families x 3 recordings	600 recordings. Availability, rights and familiarity coverage have not yet been verified for this example.
600 recordings x 4 min; 44.1 kHz stereo 16-bit	25.40 GB uncompressed source audio, excluding derived excerpts and replicas.
1,000 sessions x 40 excerpts x 60 sec	40,000 test-excerpt presentations; 48 GB audio delivery at 160 kbit/s, excluding onboarding, replay and retries.
20 reference ratings and 40 test ratings/person	Up to 20,000 baseline and 40,000 test numerical-or-missing responses if all 1,000 complete; pair preferences and telemetry are additional.

Audio hours in the 600-recording example equal 40. Let r be measured worker-hours per audio-hour; shared compute is approximately $40r$ worker-hours before overhead. Twenty reference ratings per participant do not justify training an unrestricted personal model. Measure batch runtime, peak memory and numerical equivalence before allocating GPU capacity.

16.1 Cost and capacity

Budget separately for hosting, Auth/database, database recovery, object storage and recovery, email, CPU/GPU preparation, egress, staff, rights and participant compensation. Recheck provider prices and quotas before provisioning [6,9,19]. No monthly total is guaranteed by this draft. Load tests should model bursty discrete answers and playback telemetry, not assume the previous continuous 10 Hz workload.

ACCEPTANCE PROTOCOL

Validate before recruitment, then measure

17. Evidence required for each release

The criteria below are proposed acceptance tests, not reported passes. Run unit/integration checks on a synthetic dataset, then authorized human usability and device pilots. Load tests must avoid sending real mail or consuming production participant slots. Retain versions, workload, environment, result logs and failure cases. A working public page or a successful deployment is not evidence that the laboratory is ready.

Area	Required evidence
Registration cap	1,001 concurrent eligible attempts produce exactly 1,000 accepted unique study registrations; retries do not increment the count.
Consent / profile	Stale consent, 19 familiar families, 21 untrimmed selections, duplicate-work variants and unsure/no recognition cannot pass release. Edits invalidate old releases.
Isolation	Participant A cannot read/write B's rows, objects or session. Anonymous and unauthorized staff requests fail. Worker scope excludes contacts.
Precompute	Golden fixtures, waveform/config cache keys, numerical tolerances, rights expiry, retries, cancellation and stale-result tests pass.
Collection	Network drop, refresh, duplicate upload, second tab and partial object commit preserve receipt integrity and expose missingness.
No runtime inference	Instrument a real session: zero MI/GPU/LLM requests during acquisition; all presented assets trace to the released manifest.
Email	OTP plus invite delivery tested; webhook signatures, outbox retries, bounce suppression and withdrawal cancellation verified.
Restore / withdrawal	Independent database/object recovery exercise passes, including deletion-ledger replay and revoked session admission.
Scientific replay	All primary predictions precede test exposure; confirmatory labels cannot update global parameters. Secondary replay never accesses future observations.

17.1 Performance targets to test

Begin load testing at 50 simultaneous sessions and stress at 100, using the actual discrete-answer and playback-event schema. Test synchronized block boundaries, slow connections and retries. Measure p95 acknowledgement latency, durable receipt integrity and pending queue growth. These are operating targets, not demonstrated support for 1,000 simultaneous users.

Report onboarding completion, valid profile coverage, preparation latency p50/p95, computation failures, invitation delivery, session interruptions, missingness, completions and analysis eligibility. Report these by prespecified background groups where sample size and privacy permit. Separate usability failures from scientific exclusion, and document corrective releases. Do not substitute completion rate for model validity.

FROM DESIGN TO OPERATED LABORATORY

Implementation sequence and unresolved decisions

18. Build in stages with concrete exit criteria

Stage	Deliverable and exit criterion
1. Institutional and scientific contract	Responsible controller/PI, recruitment scope, consent, source permissions and classical scientific revision and prospective prediction lock recorded.
2. Registration vertical slice	Staging Auth, profile, consent, 20-family familiarity and reference-calibration validation, saved progress and capped registration transaction pass synthetic tests.
3. Precompute pipeline	Pinned MI container, recording/rights registry, deduplicated jobs, private artifacts and versioned ready manifest work end to end.
4. Collection runner	Device practice, prepared stimulus playback, durable chunk receipts, retry/recovery and no-inference trace demonstrated.
5. Operational readiness	Production configuration reviewed; mail domain, logs, restore, withdrawal, budget alerts and responsible operators verified.
6. Controlled launch	Approved pilot then staged invitations. Expand only after measured capacity and participant burden support the next cohort.

18.1 Decisions that this draft does not invent

Before launch, finalize institution/controller, jurisdictions, languages, catalogue and source rights, preview and trial durations, validated wording, exact MI output subset, executable personal-model contract, independent development/pilot data, missingness thresholds, precision/power assessment, incentives, retention, cloud ownership and spending limits. The 20-familiar-work requirement and collection-only primary design are already fixed in this revision.

18.2 Roles and a stop procedure

The study lead owns consent, scientific changes and launch authorization. A data steward owns permissions, access requests and withdrawal. An operator owns deployment, mail, uptime, restore and spend controls. A researcher can hold several roles initially, but permissions and actions remain distinct and auditable. Maintain a second recovery contact and accessible backup credentials under institutional control. A collection-pause switch blocks new starts while allowing safe acknowledgement and closure of active sessions.

The incident procedure is to stop affected admissions, preserve minimal evidence, revoke exposed credentials or releases, assess affected records, restore or patch in staging, rerun relevant acceptance checks and obtain the required operational approval to resume. External notification duties depend on the responsible institution and applicable rules. A calendar date does not automatically open recruitment.

18.3 Contribution and current status

O-Lab makes an experimental plan operational by connecting identity, provenance, batch computation and timed observations through explicit release gates. The design is reusable, but its value must be established by implementation and measured reliability. This draft was prepared with AI assistance for source review, writing and document production; responsibility for scientific and operational decisions remains with the author. Only the technical manuscript and supporting design artifacts have been created in this task. No participant service has been provisioned or study mail sent.

SOURCE RECORD

References

Sources consulted for this revision or inherited from the checked 10 October technical draft. Dataset and provider conditions must be rechecked before use. Experimental choices and numerical scenarios are proposed, not observed results.

[1] Erdem, A. Musical Pleasure Across Interpretations of Familiar Classical Works. Scientific protocol revision 4, 10 October 2026. Local companion; publication pending.

<https://musicalintelligence.app/research/learning-the-listener/>

[2] Supabase. Send emails with custom SMTP. Production mail and authentication delivery constraints.

<https://supabase.com/docs/guides/auth/auth-smtp>

[3] Supabase. Row Level Security. Database grants, row policies and access tests.

<https://supabase.com/docs/guides/database/postgres/row-level-security>

[4] Supabase. Storage Buckets. Private/public access models and object policies.

<https://supabase.com/docs/guides/storage/buckets/fundamentals>

[5] Supabase. Available regions. Primary region and specific-region selection.

<https://supabase.com/docs/guides/platform/regions>

[6] Supabase. Database Backups. Database-only backup scope, PITR and recovery pricing.

<https://supabase.com/docs/guides/platform/backups>

[7] Supabase. Supabase Queues. Postgres-backed durable message queues.

<https://supabase.com/docs/guides/queues>

[8] Supabase. Edge Function limits. CPU, memory and duration constraints.

<https://supabase.com/docs/guides/functions/limits>

[9] Supabase. Pricing and fees. Plan, compute, included storage and egress.

<https://supabase.com/pricing>

[12] Resend. Idempotency Keys. Provider duplicate suppression and its time window.

<https://resend.com/docs/dashboard/emails/idempotency-keys>

[13] Resend. Webhooks. Delivery lifecycle events.

<https://resend.com/docs/webhooks/introduction>

SOURCE RECORD

References, continued

Sources consulted for this revision or inherited from the checked 10 October technical draft. Dataset and provider conditions must be rechecked before use. Experimental choices and numerical scenarios are proposed, not observed results.

[14] AWS. Amazon ECS task definition differences for Fargate. Resource constraints.
<https://docs.aws.amazon.com/AmazonECS/latest/developerguide/fargate-tasks-services.html>

[15] AWS. Run GPU jobs. AWS Batch GPU scheduling on EC2 resources.
<https://docs.aws.amazon.com/batch/latest/userguide/gpu-jobs.html>

[16] jsPsych. Running Experiments. Browser execution and permanent data storage.
<https://www.jspsych.org/v8/overview/running-experiments/>

[17] AWS. Data protection in Amazon Bedrock. Shared responsibility and processing controls.
<https://docs.aws.amazon.com/bedrock/latest/userguide/data-protection.html>

[18] AWS. Cross-Region inference in Amazon Bedrock. Geography and global routing behavior.
<https://docs.aws.amazon.com/bedrock/latest/userguide/cross-region-inference.html>

[19] Resend. Pricing. Transactional email plans and limits.
<https://resend.com/pricing>

[20] IMSLP. Beethoven Op. 27 No. 2; candidate performances and per-file rights.
[https://imslp.org/wiki/Piano_Sonata_No.14,_Op.27_No.2_\(Beethoven,_Ludwig_van\)](https://imslp.org/wiki/Piano_Sonata_No.14,_Op.27_No.2_(Beethoven,_Ludwig_van))

[21] Musopen. FAQ; composition and recording rights.
<https://musopen.org/faq/>

[22] Magenta. MAESTRO; performance audio and MIDI.
<https://magenta.withgoogle.com/datasets/maestro>

[23] Foscari et al. ASAP; score/performance alignment.
<https://github.com/fosfrancesco/asap-dataset>