TECHNOLOGY DOC · ARCHITECTURE + ROADMAP

Kashi 可視 — technology

What's actually running behind Kashi today, and where we see room to improve. Split audience: investor-readable on top, engineer-level detail toward the bottom. Every claim traces to a file path in the live codebase.

2026-04-21 · Companion to business.html and governance · Live product: kashi-lilac.vercel.app

---

IN THIS DOCUMENT

TIER 1 · INVESTOR-READABLE (§1–§3)

Conceptual flow, honest capability list, what's running, what matters. Skip straight to §7 if you want the gap analysis first.

01 Executive summary

Kashi is a deterministic meeting-governance pipeline. It takes meeting transcripts (already produced by Zoom/Teams/Meet), runs them through a stack of structural detectors, and surfaces repeated interaction asymmetries for human review — never as harassment labels, always as contestable signals with explicit uncertainty.

DETECTORS LIVE
7
3 structural + 4 hybrid text-informed

CROSS-MEETING WRAPPERS
2
dyadic continuity + baseline drift

DETERMINISTIC CORE
100%
same input → same output · no LLM in production path

TYPESCRIPT, STRICT MODE
~3,400 LOC
covers pipeline + data model + routes

SEED EVAL PASS RATE
3/3
zero false-positives on healthy control team

RESEARCH DOCUMENTS BEHIND DESIGN
42
24 business + 18 technical-dev consideration

The one-sentence architecture

A Next.js 16 + TypeScript + Supabase application with a pipeline of deterministic detectors organized into 3 lanes (structural-only · hybrid text-informed · refused), protected by row-level security, with commitments declared in typed enums and a machine-readable detector registry — and several honest gaps we're naming on the way to pilot-ready.

02 The stack

LAYER	TECHNOLOGY	WHY
Frontend framework	Next.js 16 (App Router) · React 19 · TypeScript strict	Server-rendered pages for SEO + auth-aware demo routes + client-side interactivity where needed
Styling	Tailwind v4 + shadcn/ui components	Utility-first; matches the serious/governance aesthetic not playful-SaaS
Charts	Recharts 3.8	BarChart with LabelList, LineChart with ReferenceLine · covers the Manager Mirror + Executive Brief visuals

LAYER	TECHNOLOGY	WHY
Auth	Supabase magic-link OTP via `@supabase/ssr`	No passwords. Works through corporate email infrastructure. JP-friendly.
Database	Supabase Postgres (Tokyo region)	Row-level security native. Tenant-scoped. **Pinned to ap-northeast-1.**
Storage	Supabase Storage	Transcript files · per-tenant buckets with policy
Hosting	Vercel production	Auto-deploys from main · edge network · `honest gap` primary compute is US-based (see §6)
Detector pipeline	Pure TypeScript (no LLM in prod path)	Deterministic · auditable · fast · no per-API-call cost at runtime
LLM (seed + reasoning only)	Claude Sonnet 4.6 (seed authoring) + Claude Opus 4.7 (reasoning-heavy detector validation)	Used only for seed-data authoring and offline detector tuning · **not** called on production traffic
Transcript parsers	VTT · TXT · SRT · CSV · JSONL	Normalizes platform-specific formats into Turn[] · `src/lib/pipeline/transcript-parser.ts`
JP linguistic parsing	Regex-based surface grammar (no MeCab dependency)	Sufficient for keigo classification on meeting-length transcripts · tokenizer upgrade is a Sprint-2 candidate

03 The detector pipeline

Every transcript flows through the same 6-layer pipeline. Every layer is deterministic. Every layer's output is typed. No hidden model calls, no black-box scoring.

```
[Zoom / Teams / Meet transcript] (uploaded via /app/admin/upload) | ▼ Layer 1: Ingest & normalize VTT/TXT/SRT/CSV/JSONL parsers → canonical Turn[] array src/lib/pipeline/transcript-parser.ts | ▼ Layer 2: Structural detectors (deterministic, lane 1) · Intrusive interruption (overlap + truncation timing) · Chilling delta (post-trigger participation drop vs own baseline) · Floor-time Gini (speaking-share inequality) src/lib/pipeline/layer1-deterministic.ts | ▼ Layer 2b:
```

Hybrid text-informed (lane 2, tenant opt-in) · Unanswered-question rate (lexical) · Topic-credit ignored turns (embedding similarity) · Agreement asymmetry 同調圧力 (lexical + directional) · Keigo (敬語) peer-asymmetry (surface-grammar classification) *src/lib/pipeline/{topic-credit, agreement-asymmetry, keigo}.ts* | ▼ **Layer 3: Meeting-level metrics** Gini · asymmetry matrix · directive density · takeover count · reciprocity | ▼ **Layer 4: Longitudinal aggregation** Rolling 90-day window · per-speaker OWN baseline · dyadic-continuity test *src/lib/pipeline/{aggregate, continuity}.ts* | ▼ **Layer 5: Review-worthy event construction** Composite scored from directionality × baseline-drop × persistence | ▼ **Layer 6: Role-based presentation** RBAC · k-anonymity (declared) · Manager Mirror · Executive Brief · Victim view (planned)

What it means in plain English

- **Layer 1–2**: the raw signal extraction. Counts, durations, overlaps, who-addressed-whom.
- **Layer 3**: per-meeting rollup. "In this meeting Gini was 0.34; Kenji's interruptions landed on Mira 82% of the time."
- **Layer 4**: "is this one bad meeting or a 63-day pattern?" — the longitudinal wrapper.
- **Layer 5**: the composite that decides whether a pattern is review-worthy, using thresholds (directionality ≥ 3×, baseline drop ≥ 0.4).
- **Layer 6**: three different views rendered from the same event — private to the employee, private-self-mirror to the manager, k≥5-anonymized to the executive.

TIER 2 · ENGINEER-LEVEL (§4 ONWARD)

File paths, algorithms, type signatures, thresholds, RLS. Everything below can be audited against the actual codebase. Investors who want to skip to the gap analysis: jump to §7.

04 The 7 detectors, one lane at a time

Each detector declares its lane at compile time in `src/lib/pipeline/detector-registry.ts`. Employer-facing surfaces default to lane 1 (structural-only). Lane 2 requires tenant opt-in via `semantic_lane_enabled` feature flag.

Auto-ingest architecture (production path).

Kashi does not record meetings or upload transcripts manually in production. The detectors below are fed by webhooks from Zoom / Microsoft Teams / Google Meet, triggered when the platform finishes transcribing.

- **Zoom:** Marketplace app (Server-to-Server OAuth) → webhook `recording.transcript_completed` → `POST /api/webhooks/zoom` (HMAC-SHA256 verify via `x-zm-signature`, 5-minute replay window, URL-validation handshake implemented). Requires org-level "Cloud Recording + Audio Transcript" toggle on.
- **Microsoft Teams:** Azure AD app + app permission `OnlineMeetingTranscript.Read.All` → Graph change-notification subscription → `POST /api/webhooks/teams`. Requires Teams Premium OR tenant-enforced `AllowTranscription` policy.
- **Google Meet:** Workspace OAuth + domain-wide delegation → Workspace Events API + Cloud Pub/Sub push → `POST /api/webhooks/meet`. Transcript entries arrive structured; no VTT parsing needed.

All three paths normalize through `src/lib/pipeline/transcript-parser.ts` → `Turn[]` → detector pipeline below. The /demo/ingest surface exposes the same normalizer + detectors to a stateless paste for hands-on demos. Production stores structural metrics only; transcript body text is discarded after detector run.

Lane 1 — structural-only (3 detectors, default employer-facing)

1. INTRUSIVE-INTERRUPTION

Detects when speaker B starts while speaker A is still speaking AND A's turn ends within a threshold.

```
// src/lib/pipeline/layer1-deterministic.ts
const OVERLAP_THRESHOLD_MS = 500;      // overlap window
const TRUNCATION_WINDOW_MS = 500;   // A's turn must end within this

export function detectIntrusiveInterruptions(turns: Turn[]): IntrusiveInterruption[] {
  // O(n) scan: for each adjacent pair (B, A+1), check if B.startMs < A.endMs
  // AND A.endMs − B.startMs < OVERLAP_THRESHOLD_MS
  // AND A's turn ends within TRUNCATION_WINDOW_MS → count it
}
```

Research anchor: Anderson & Leaper 1998 (meta-analysis, 43 studies, *d*=0.33).

Known confounds: facilitator/chair role · incident-bridge meeting type · overlap-heavy audio. Handled by caveat surface on the Mirror UI; should be handled by `meetingType` gating (§7 gap).

2. CHILLING-DELTA

Per-speaker participation drop in a 5-minute window after a trigger turn, compared to that speaker's own rolling baseline (not team avg).

```
const CHILLING_DELTA_WINDOW_MS = 300000;  // 5 min
const CHILLING_DELTA_THRESHOLD = 0.4;    // 40% drop
const COLD_START_MIN_MEETINGS = 5;
```

Research anchor: Morrison 2014 (organizational silence); Detert & Burris 2007.

Cold-start rule: no baseline signal emitted if the speaker has fewer than 5 prior meetings. Today this rule lives in `perSpeakerBaseline()` at `layer1-deterministic.ts:118` — it returns null when comparable meetings < 5. See §7 gap: should be lifted to global abstention doctrine.

3. FLOOR-TIME GINI

Standard Gini coefficient over speaker durations within a meeting. Gini 0 = perfectly equal, 1 = one person dominates.

Research anchor: Schmid Mast 2002 (Human Communication Research meta-analysis).

Lane 2 — hybrid text-informed (4 detectors, tenant opt-in required)

Why these are lane 2, not lane 1. All four of these detectors read transcript text — either via lexical pattern matching, embedding similarity, or surface-grammar classification. The v1 deck's blanket "metadata only, no content" claim was falsified by the existence of these detectors. The registry makes the lane explicit and gates them behind a tenant feature flag that defaults off.

4. UNANSWERED-QUESTION RATE

A turn ending in lexical question markers (「か？」/「?」/「かな」) that receives no substantive response within N turns. Short acknowledgment responses (<2s or lexical "yes/なるほど/そうですね") don't count as a response for this detector.

Research anchor: Stivers et al. 2009 (PNAS cross-linguistic turn-taking).

5. TOPIC-CREDIT IGNORED-TURNS

A substantive turn (>3s) from speaker A, followed either by (a) silence ≥ 2× the meeting's median inter-turn gap, or (b) a topic redirect by speaker B that captures credit for a similar proposal later in the meeting. Similarity is computed via embedding distance.

Research anchor: Sacks / Schegloff / Jefferson 1974; maps to MHLW パワハラ 類型 3 + 5.

6. AGREEMENT ASYMMETRY (同調圧力)

Directional rate at which position-statements shift toward a specific speaker. Position-statements detected via lexical cues; shift detected by comparing turn-order positional claims.

7. KEIGO (敬語) PEER-ADDRESSEE ASYMMETRY

Per-speaker-per-addressee politeness-register score from surface grammar. Classifies each turn into `sonkeigo` / `kenjougo` / `teineigo` / `plain` / `imperative` / `mixed`. Detects when the same speaker uses meaningfully lower register toward one addressee than toward peers.

```
// src/lib/pipeline/keigo.ts — excerpt
const SONKEIGO_PATTERNS = [/いらっしゃ/, /おっしゃ/, /なさる/, /お[一-鿐]+になる/];
const PLAIN_PATTERNS = [/だよ/, /だね/, /でしょ/, /よね/, /じゃん/];

const PEER_ASYMMETRY_THRESHOLD = 0.25;
const MIN_TURNS_PER_TARGET = 2;
```

On our Kimura/Mira seed: Kimura scores **0.88 toward Nakamura** (honorific) vs **0.38 toward Mira** (plain form). Gap = 0.50. The largest gap in our control team (Sato's): 0.0 — zero asymmetries.

Research anchor: Cook 2011 (*J. Pragmatics* 43(15)); Saito 2011 (*J. Pragmatics* 43(6)); Pizziconi 2003; Ide wakimae framework. **Unique to Kashi.** No Western product has this.

Cross-meeting wrappers (in `continuity.ts`)

- **Dyadic interruption continuity** — MHLW 継続性 3要素 test: is an (interrupter, interruptee) pair appearing across ≥3 meetings? Emits a persistence score (0-1).
- **Speaker baseline drift** — detects the longest sustained drop window where a speaker is below their own 90-day baseline. Example seed output: Mira down 68% for 63 continuous days.

Aggregation thresholds (live in `aggregate.ts`)

```
const WINDOW_DAYS = 90;
const CONCERN_DIRECTIONALITY = 3.0;  // 3× peer rate → concern
const CONCERN_BASELINE_DROP = 0.4;   // 40% drop → concern
const WATCH_DIRECTIONALITY = 2.0;
const WATCH_BASELINE_DROP = 0.2;
```

```
const COST_PER_CASE_MIN_YEN = 3_000_000;
const COST_PER_CASE_MAX_YEN = 8_000_000;
```

Known simplification: These thresholds are universal. The meeting-type-normalization research says they should be *per-meeting-type*. Fix in the §7 gap list.

05 Data model

Core types at `src/lib/types.ts`:

```
// The atomic unit of input.
export type Turn = {
  speakerId: string;
  startMs: number;
  endMs: number;
  text: string;
};

export type Meeting = {
  id: string;
  dateIso: string;
  teamId: string;
  title: string;
  turns: Turn[];
  // Declared (not yet threaded through detectors):
  meetingType?: MeetingType;
  meetingTypeConfidence?: number;
  scoringMode?: ScoringMode;
};

// New types committed 2026-04-21 · awaiting end-to-end threading:
export type DetectorClass =
  | "STRUCTURAL_ONLY"
  | "TEXT_DERIVED_DETERMINISTIC"
  | "HYBRID_TEXT_INFORMED"
  | "GENERATIVE_ASSIST"
  | "REFUSED";

export type EvidenceGrade =
  | "BLOCKED"           // input quality below gate
  | "INSUFFICIENT"      // thin exposure
  | "WEAK"
  | "EMERGING"
  | "STABLE"
```

```
  | "HIGH_CONFIDENCE_STABLE";

export type AbstentionState =
  | "NO_COMPUTE"
  | "COMPUTE_NO_INTERPRETATION"
  | "WATCH_ONLY"
  | "INTERPRETABLE_PRIVATE_ONLY"
  | "INTERPRETABLE_ROLE_BOUNDED";

export type ConfidenceBundle = {
  inputQuality: number;
  contextSupport: number;
  exposureSupport: number;
  detectorConfidence: number;
  aggregationSupport: number;
  reasonCodes: ReasonCode[];
  abstention: AbstentionState;
  grade: EvidenceGrade;
};
```

Database schemas at `src/lib/db/types.ts` — multi-tenant with `org_id` scoped RLS on every table. Raw transcript text is stripped before DB write; only `length_chars` is persisted.

SUPABASE SCHEMA (ABBREVIATED)

```
orgs(id, name, created_at)
profiles(id, org_id, email, role, created_at)
  -- role enum: admin | ceo | member (see §7 gap: needs 6)
teams(id, org_id, name)
meetings(id, org_id, team_id, title, date_iso, turns_metadata jsonb)
  -- turns_metadata = TurnMetadata[] with text STRIPPED
meeting_metrics(meeting_id, speaking_share jsonb, intrusive_interruptions jsonb, ...)
manager_mirrors(manager_profile_id, week_ending_iso, ...)
pattern_summaries(manager_profile_id, pattern_intensity enum, ...)
user_keys(user_id, public_key_jwk jsonb)  -- evidence vault (planned)
evidence_vault(id, user_id, event_id, encrypted_snippet, encrypted_data_key, iv)
```

MIGRATIONS APPLIED

- `0001_init.sql` — schema + initial RLS policies
- `0002_rls_hardening.sql` — replaced v1 policies with isolation-verified v2 versions
- `0003_indexes_and_metrics.sql` — query-performance indexes + metric materialization

06 Security posture (honest snapshot)

What we've actually built vs what we've only declared. Naming the gap is the first step to closing it.

GUARANTEE	BUILT IN CODE?	DECLARED IN DOCS?	GAP (IF ANY)
Tenant isolation via RLS	YES	YES	No automated RLS isolation tests in CI yet — fix in Sprint 1
Magic-link OTP auth	YES	YES	—
Raw transcript stripping (length only stored)	YES	YES	—
Japan data residency (Supabase Tokyo)	PARTIAL	YES	Vercel primary-compute region is US. Regulated content flowing through Vercel runtime contradicts the claim. Fix in Sprint 1.
ConfidenceBundle on every event	YES (v3)	YES	Emitted end-to-end as of 2026-04-21. Headline `EvidenceGrade` + `AbstentionState` + `ReasonCode[]` visible on every Manager Mirror + Executive Brief event. Per-detector bundles remain Sprint-1.
k-anonymity (k≥5) on aggregates	DECLARED	YES	Enforced at UI layer today, not at query layer. Sprint 2: move enforcement into SECURITY DEFINER RPCs.
Differential privacy (ε≤1) on exec dashboards	NOT YET	YES	Math declared in governance page; not implemented. Sprint 3.
Audit log on every drill-down	PARTIAL	YES	Individual drill-downs logged; app-layer audit event schema with reason codes + SIEM streaming is Sprint 2.
4-tier retention (14d / 24mo / 12mo / legal-	PARTIAL	YES	Retention bands defined in schema, purge jobs exist for raw layer. Tombstones +

GUARANTEE	BUILT IN CODE?	DECLARED IN DOCS?	GAP (IF ANY)
hold)			restore-reconciliation: Sprint 3.
No admin-level content access	YES	YES	Architecturally: Kashi staff cannot read customer transcripts. Break-glass procedures documented; tabletop exercise Sprint 2.
SOC 2 Type II / ISO 27001	PRE-CERT	honestly disclosed	Year-1 target. Architecture ready for audit; controls documentation is the gap.

The Japan data-residency gap is the single most urgent fix before any Enterprise-tier JP sale. Solvable without rearchitecture: pin Vercel function regions to Tokyo for sensitive routes, ensure no transcript or vault content flows through Vercel logs/previews, update subprocessor map to disclose Vercel-US honestly.

07 Gap analysis — where we can improve

This section is the honest answer to "what's not yet done?" All items below are surfaced directly from the Ideas_wave3 technical-dev research library (17 memos) + the business research memos' technical implications. Every item has a path to code.

P0 Critical — pilot-blockers

1. ConfidenceBundle · HEADLINE SHIPPED (2026-04-21)

Backward-compatible wrapper shipped at `src/lib/pipeline/confidence-bundle.ts`. Every ManagerMirrorData + PatternSummary now carries `overallGrade` + `overallAbstention` + `overallReasonCodes`. UI renders grade badge with hover-tooltip rationale on `/demo/mirror` + `/demo/ceo`.

Remaining Sprint-1 scope: refactor each of the 7 detectors to emit their OWN bundle natively (not just at the aggregation boundary). The UX commitment is landed; the per-detector audit trail is next.

1b. Per-detector ConfidenceBundle emission (Sprint-1)

Currently the bundle is computed at aggregate.ts boundary. Research wants each detector to emit its own bundle so reviewers can see which detector is weak and why. Full signature refactor of 7 detectors + threading through aggregate + eval harness re-verification. ~1 week.

2. Meeting-type normalization missing

Universal thresholds (`CONCERN_DIRECTIONALITY=3.0`) pool across all meeting types. A weekly sync and an incident bridge and a 1:1 and a training session are treated identically. The meeting-type-normalization research says this is the largest red-team attack surface.

Fix: add `meetingType` to `Meeting` (declared, not used); gate detector execution on `scoringMode` ; block cross-type baseline pooling in `aggregate.ts` . ~2h.

3. Role enum too narrow (3 values, research requires 6)

Current: `admin | ceo | member` . Research requires: Individual, Manager, HR/Compliance, Executive, Restricted Investigator, System Admin. Current enum structurally cannot satisfy the §3 visibility matrix in the role-and-visibility architecture memo.

Fix: migration 0004 · expand `profiles.role` enum · update RLS policies · split `admin` into support-admin vs behavioral-data investigator. ~3h including RLS policy tests.

4. Contestability state machine doesn't exist

Research (legal-procedural-fairness memo §5) is explicit: challenge / dispute / correction workflow is the single biggest product gap. Without it, every review-worthy event is uncontestable — which breaks the entire fairness story and fails EU AI Act Annex III §4 meaningful-human-review requirement.

Fix: 8 new tables (disputable_object, dispute_ticket, dispute_evidence, review_decision, correction_patch, recompute_job, aggregation_exclusion, access_history) · dispute API endpoints · server-enforced DRAFT privacy. Sprint 2, ~2 weeks.

5. Japan data residency gap (Vercel primary-US)

See §6. Marketing claim is false until fixed. Fix is not architectural — it's configuration + disclosure. Sprint 1, ~1 day.

6. No RLS isolation tests in CI

RLS policies exist. We have no automated test asserting that role X cannot SELECT from view Y. Today the isolation is verified manually. An RLS regression could ship without anyone noticing.

Fix: `test/rls-isolation.test.ts` · test harness that logs in as each role, attempts forbidden queries, asserts they fail. Gated in CI. ~1 day.

P1 High priority — next sprint

7. Semantic-detector quarantine in code (not just registry)

The registry declares which detectors are lane-2 hybrid. But the employer-facing output endpoint doesn't filter by `detectorsAllowedForTenant(flags)`. Today the tenant flag is decorative.

Fix: wrap detector invocation in `aggregate.ts` with `detectorsAllowedForTenant()`. Employer-facing default = lane-1-only. Semantic lane requires explicit per-tenant flag flip in `orgs.feature_flags` jsonb. ~4h.

8. Speaker-ID provenance chain

Today `Turn.speakerId` is a flat string. Research wants 5-layer provenance: `UtteranceSegment → DiarizationCluster → MeetingParticipantInstance → CanonicalPerson`, with status enum (RESOLVED / UNKNOWN / WRONG / SPLIT_SUSPECTED / MERGE_SUSPECTED / OVERLAP_AMBIGUOUS).

Fix: refactor `Turn` type + migration + identity-mapping tables with versioning. Metric-eligibility gate suppresses person-level output when `unknown_speaker_duration > 15%`. ~1 week.

9. Telemetry partitioning (anti-retaliation)

Employer must not be able to infer that an employee opened their pattern page, created a vault, marked a confound, or filed a dispute. Today these events could in theory appear in business-analytics. Research (retaliation-risk memo) flags this as directly actionable under MHLW retaliation-prohibition.

Fix: split telemetry namespace · protected-route events go to a separate store (not exposed to tenant BI) · small-team (<5 user) inference suppression · batching + delay on employee-facing events. ~1 week.

10. Per-meeting influence cap + leave-one-out fragility

Current aggregation: any single meeting can dominate the 90-day signal. A noisy meeting creates a fake trend. Research wants: cap single-meeting contribution at 20% of weighted evidence; run leave-one-out to confirm signal survives without any single meeting.

Fix: weight normalization in `aggregate.ts`. Degrade evidence grade if signal disappears leave-one-out. ~3 days.

11. 7-gate input-quality pipeline

Research wants every meeting to pass 7 sequential quality gates BEFORE any detector runs: substrate presence · parser integrity · speaker attribution · transcript text · language regime · meeting type · sample sufficiency. Gates produce a `detector_eligibility_map` telling each detector if it's allowed to run.

Fix: new `InputQualityGate` layer before Layer 2. Meetings failing Gate N get `scoringMode="observation_only"`. ~3-4 days.

12. Vault metadata suppression

E2E evidence vault is planned (ciphertext employer cannot decrypt). But the *existence* of a vault, its snippet count, last-activity timestamp, and draft state still leak via database metadata. Research: metadata leakage is as dangerous as content leakage.

Fix: vault tables in a separate Postgres schema with no cross-schema joins from tenant BI · admin access requires break-glass · metadata counts suppressed on exports. ~1 week.

P2 Medium — this quarter

13. Adaptation-watch layer (anti-gaming)

Once managers know the detector surface, some will game it (channel-shift dominance into async/1:1). Research: flag suspiciously clean metric improvement as a signal, not success. Multi-metric corroboration required before declaring an "improvement."

14. Multi-metric corroboration rule

No single-metric victory claims. An "improvement" requires ≥2 detector families moving in the right direction + no adjacent metric worsening. Enforce in aggregation layer before event emission.

15. Remediation-quality + human-recovery measurement

Business plan §11 declares the 6-layer success model. Layers 5 (remediation quality) and 6 (human recovery) aren't measured in code yet. Requires post-Lane-B fairness-rating capture from the affected employee + speaking-share-recovery tracking at 30/60/90 days.

16. Dispute API endpoints

The contestability state machine (§P0) needs paired API: `POST /disputes` · `POST /disputes/:id/submit|triage|resolve` · `GET /objects/:id/explanation` · `GET /access-history/:object_id`. Build once the state machine lands.

17. 3-lane accountability in code (currently only in docs)

The Lane-A (private self-correction) → Lane-B (governed remediation) → Lane-C (formal review) model is declared on governance page. The automatic transition trigger (pattern persists after notice → escalates to Lane B) is not implemented. Today it's a policy, not a product feature.

`Strategic` 6-12 month horizon

18. NAQ-R outcome validation study

Partnership with 津野香奈美 lab (Kanagawa U. of Human Services). 24-month validation: do Kashi's structural signals predict NAQ-R self-report outcomes at 6-month follow-up in treatment vs matched controls? Publishable result. Cohort of 5 JP companies.

19. Cross-platform ASR matrix

Platform × language × feature support matrix. Teams vs Meet vs Zoom, each with different transcription quality. Japanese-specific: the 46.8% single-channel tcpWER baseline is dangerous. Overlap flags, L2 caution surfaces, per-platform eligibility maps.

20. Multilingual locale packs

Universal detector core + per-locale calibration layers. Locale registry: country × language × mixed-language flag × platform × transcript-confidence band × calibration pack × legal pack × UX copy pack. Enables Singapore / NL / UK expansion.

21. Victim-explainer page

Research (manager-adoption + trust memos) endorses. When a review-worthy event fires, the affected individual sees a private explainer page. What happened, what it probably means, what their options are, who they can contact, how to enable the evidence vault.

22. Victim-owned E2E evidence vault

Planned. WebCrypto RSA-OAEP-2048 + AES-256-GCM envelope encryption. Employer stores ciphertext; victim holds private key. Vault metadata suppression (§12) is a prerequisite.

23. SOC 2 Type II + ISO 27001

Year-1 SOC 2 target. Year-2 ISO 27001. Architecture is ready; what's missing is control documentation, quarterly tabletops, third-party pen-test cycle.

07·5 Critical review — if you can only ship 3 things before pilot

§7 lists 23 improvements ranked by priority. But priority lists ≠ sequencing decisions. This section re-ranks the §7 items by **pitch impact × pilot necessity × feasibility with current team size**, and picks the three that should be Sprint-1 scope after pre-seed funding.

Pick 1 · Japan data residency fix (§7 item 5)

Smallest implementation cost, biggest procurement-unblock. Configuring Vercel function regions + auditing data flow is ~1 day. Without it, the Japan-data-residency claim is technically false and the first Enterprise-tier JP prospect will catch it in their security review. **Must land before any Enterprise conversation starts.**

Pick 2 · meeting_type gating (§7 item 2)

Universal thresholds pool across meeting types — a known validity hole that a hostile technical reviewer will notice in the first pilot post-mortem. Adding `meetingType` + `scoringMode` to the Meeting schema + gating detector execution in `aggregate.ts` is ~2h. It closes the biggest red-team attack surface and enables honest observation-only mode for unsupported formats (incident bridges, 1:1s, exec reviews).

Pick 3 · RLS isolation tests in CI (§7 item 6)

One engineer-day. Writes a test harness that logs in as each role, attempts forbidden queries, asserts they fail. Gated in GitHub Actions on every PR. Without this, the tenant-isolation claim is "verified by faith." With it, every PR is automatically checked. **The cheapest credibility upgrade in the entire roadmap.**

What NOT to build first (despite being P0)

- **Role enum expansion (§7 item 3)** — important but 3h of migration + RLS rework. Not the first blocker.
- **Contestability state machine (§7 item 4)** — 2-week Sprint-2 scope. Too big to jam into pre-pilot prep.
- **Per-detector ConfidenceBundle (§7 item 1b)** — UX commitment is already landed via the wrapper. Full refactor can wait.

Principle: the first three builds post-funding should be ones where shipping pays for itself in *unblocked sales conversations*, not code hygiene. Data residency, meeting-type gating, and RLS tests all unblock specific investor/customer objections in the first technical-review meeting.

What the gap analysis missed (v3 addition)

- **Webhook ingestion path for Zoom/Teams/Meet.** Today Kashi runs on uploaded transcripts. For production, platforms push transcripts via webhook. Missing from §7 because it's plumbing, not product — but it's the integration gate for any real pilot. Add as P1 item 24.
- **Per-tenant feature-flag admin UI.** The detector-registry defines flags (`semantic_lane_enabled`, etc.). No admin surface to flip them per tenant. A pilot customer can't opt into lane-2 detectors because there's no way to say yes. Add as P1 item 25.

- **Cost-of-detector telemetry.** We don't instrument how long each detector takes per meeting. At scale that matters for the pricing model defensibility. Add as P2 item 26.

08 Roadmap by sprint

Tied to the financing plan (business.html §6).

SPRINT	WEEKS	FUNDED BY	DELIVERABLES
Sprint 1	1–4	Pre-seed	§7 items 1 (ConfidenceBundle threading) · 2 (meeting-type gating) · 3 (role enum expansion) · 5 (Japan data residency fix) · 6 (RLS isolation tests in CI)
Sprint 2	5–8	Pre-seed	§7 items 4 (contestability state machine) · 7 (semantic-detector quarantine enforced) · 9 (telemetry partitioning) · start 16 (dispute API)
Sprint 3	9–12	Seed	§7 items 8 (speaker provenance) · 10 (influence cap) · 11 (input-quality gates) · 12 (vault metadata suppression) · start 22 (evidence vault MVP)
Sprint 4	13–16	Seed	§7 items 13 (adaptation-watch) · 14 (multi-metric corroboration) · 15 (remediation-quality measurement) · 17 (3-lane in code) · 21 (victim-explainer page)
Sprint 5-6	17–24	Seed	§7 items 18 (NAQ-R study kickoff) · 19 (ASR matrix) · SOC 2 control documentation · third-party pen-test #1
Sprint 7+	25–52	Series A	§7 items 20 (multilingual locale packs) · 23 (SOC 2 Type II achieved · ISO 27001 in progress) · Singapore regional rollout

Velocity assumption

Sprints 1-2 deliverable on the 2-person founding team. Sprint 3+ assumes 1 additional engineer (the security/DevSecOps hire from business.html §5). Sprint 5+ assumes 2 additional engineers + CS lead from the seed round.

09 What we will NOT build

The companion to governance's "Kashi will not do" list, at the technology level. These are architectural refusals — not "we haven't gotten around to it yet," but "we will actively not ship this."

- **Audio / prosody / voice-stress analysis.** EU AI Act Art. 5(1)(f) + Commission Guidelines C(2025) 884 final. Plus the science is chance-level (Damphousse 2008, Eriksson & Lacerda 2007).
- **Emotion / affect / sentiment inference.** Same Art. 5 prohibition. Full stop.
- **Generative summaries of individual behavior.** No "Kashi thinks X about Y." The tool describes what happened; humans interpret.
- **Company-wide health score.** Refused on evidence. See governance page §11 and business plan §7.
- **HR-decision export path.** No API, no data format, no workflow feeds performance / promotion / discipline / compensation systems.
- **Keystroke, screen, or browser monitoring.** Out of scope. Always.
- **Real-time in-meeting intervention.** No "please stop interrupting" popups. Kashi is asynchronous, weekly, private.
- **Identity-based classifications.** No gender / race / age / religion inference. The detectors are about structure, not demographics.
- **Pulse-survey features.** Not our category. Wevox, Peakon, Culture Amp cover this and have failed to move harassment outcomes. Adding a survey tab would dilute our positioning.
- **A "score your manager" public leaderboard.** Goodhart's law predicts this fails in 6 months. Not built.

> *Every refusal above is simultaneously a technology decision and a market-positioning decision. A competitor who wants to copy Kashi's safety story has to reproduce all 10 refusals. They can't cherry-pick the detectors and skip the discipline.*

Companion docs: business plan · governance · integration log · investor tour · research library (42 docs)

Technology architecture + improvement roadmap · v1 · 2026-04-21 · every file path resolvable in the live codebase