

What the research library changed 可視

Kashi started as a 23-hour hackathon build. Then two research libraries landed — 24 business PDFs + 18 technical-dev memos, 42 documents total. Below is a honest table of what changed, why, and which memo drove it.

2026-04-21 · Companion to </business.html> · Source docs at </research/>

42

total research documents integrated

11

risks now in register (was 6)

5

moat layers (was 4)

6

success metric layers (was 4)

01 Positioning & language

strikethrough = what we used to say · regular text = what we say now

AREA	BEFORE	AFTER	DRIVER MEMO
Hero positioning	"Governance infrastructure for workplace-power dynamics"	"Privacy-bounded meeting governance that surfaces repeated interaction asymmetries under uncertainty"	category-entry · measurement-science
Thesis line	"The pattern is the harm"	"The pattern may constitute evidence consistent with uneven conversational treatment over time"	measurement-science · legal-procedural-fairness

AREA	BEFORE	AFTER	DRIVER MEMO
Detection claim	"Detect harmful team dynamics earlier"	"Surface repeated interaction asymmetries earlier, as contestable structural signals for human review"	measurement-science · product-wedge
Buyer framing	"The CEO's instrument for seeing the bill before it arrives" (front-door)	"A bounded governance instrument" (front-door) · "CEO's instrument" demoted to closed-sponsor context only	category-entry · product-wedge · trust · founder-narrative
Anti-surveillance line	"Not an employee-monitoring tool" (overclaim — we DO create person-linked telemetry)	"Not a general-purpose employee-surveillance product. A tightly restricted governance system with structural metadata under explicit anti-misuse limits"	labor-relations · trust
Detector purity claim	"All deterministic, metadata-only, no content reading" (contradicted by embedding-based detectors)	Three lanes: 3 structural-only (default employer-facing) · 4 hybrid text-informed (tenant opt-in) · refused. Machine-readable at <code>src/lib/pipeline/detector-registry.ts</code>	detector-boundary · model-boundary · measurement-science · legal-max-hardening

02 Market & ROI

AREA	BEFORE	AFTER	DRIVER
Hero cost metric	"¥7.9M per leave case × N cases = big savings"	Hidden-loss portfolio (presenteeism + regrettable attrition + formal-escalation premium + leave tail). ¥7.3T of Japan's ¥7.6T loss is presenteeism, not absence	rebuilt-ROI · research-synthesis

AREA	BEFORE	AFTER	DRIVER
ROI math	8x / 3.3x / 1.3x by company size (overstated, assumed full leave-cost capture)	A × C × I × E discount framework: Strict 0.54% = ¥0.69M · Mid 1.75% = ¥2.22M · Upper 4.80% = ¥6.10M captured per 30-case year	rebuilt-ROI
Pricing defense	Enterprise ¥10M base justified by leave-cost savings	Enterprise base = governance-service fee (consultation + CAIQ pack + SOC 2 + CSM + compliance artifacts); leave-cost capture alone requires 94.4% which is indefensible	rebuilt-ROI · procurement-security
SAM sizing	"JP mid-market, ~45k companies"	Same ~45k SAM but segmented by knowledge-work filter, blended ACV ¥4.2M, revenue ceiling ~¥190B/yr	research-synthesis · funding-pathways

03 Go-to-market & expansion

AREA	BEFORE	AFTER	DRIVER
Expansion order	EU-first (DE → NL → UK), US second	Japan-first doctrine · Singapore second (PSG/EDG grants) · NL/UK year 4 · DE only after works-council playbook matures · US year 5+ after SOC 2	cross-cultural · funding-pathways · labor-politics

AREA	BEFORE	AFTER	DRIVER
Buyer model	One buyer: "the CEO"	Three buyers with distinct rhetorical registers: Economic sponsor (CEO/CFO) · Operational gatekeeper (HR/Legal/Compliance) · Trust counterparty (worker rep / union / works council)	founder-narrative · labor-politics · trust
Sales framing	"Intro to the CEO at a 50-500-person JP company"	"Requires executive sponsor + legal review + worker-representation alignment before pilot launch. All three, or none."	labor-relations · labor-politics · rollout
Channel stack	5 channels (CEO outbound, labor-law, academic, JP VC, content)	8 channels (+ worker-rep/union engagement, + occupational-health channel, + ombuds/neutral-review) with priority weighting	labor-politics · rollout · ally-path
Financing	Pure equity (pre-seed ¥50M → seed ¥500M → Series A ¥3B)	Equity + non-dilutive track in parallel: JP AI Adoption Subsidy 2026 (P1), Workstyle Reform Subsidy (P1), JOHAS routes, Singapore PSG/EDG. Customer-side subsidy is the enterprise pricing unlock	funding-pathways

04 Risk register

Expanded from 6 risks to 11. The 5 new risks are not edge cases — they are category-killers that v1 silently dismissed.

RISK #	NAME	WHAT KILLS US IF WE IGNORE IT	DRIVER
R1-R3	Pilot lawsuit · MS/Zoom copies · Works-council consent	Present in v1. Mitigations refined.	all

RISK #	NAME	WHAT KILLS US IF WE IGNORE IT	DRIVER
R4 · NEW	Adversarial adaptation / metric gaming	Managers reroute dominance into channels Kashi doesn't see (agenda control, 1:1s, async, hierarchical laundering). Metric improves, harm doesn't.	adversarial · adversarial-for-devs
R5 · NEW	Institutional capture / selective deployment	Org buys Kashi, carves leadership out of scope, surveils ICs only. "Kashi surveils workers, protects leaders" becomes the public story.	anti-capture
R6 · NEW	Metadata leakage / inference reconstruction	Employer infers worker opened pattern page / created vault / marked confounds — directly actionable retaliation risk under MHLW guidelines. No content needed.	retaliation · privacy-retention
R7 · NEW	False-negative laundering	Clean dashboard cited as exoneration ("we ran Kashi, no signal, therefore no harm"). EEOC base rate: 3 of 4 targets never speak up.	false-negative
R8 · UPGRADED	ASR + diarization input contamination	46.8% tcpWER single-channel, 54.9% overlap-heavy, 2x racial WER disparity. Demographic bias could become the scandal.	scientific-attack-surface · speaker-identity
R9-R10	Regulator walkback · Scandal	Present in v1. Mitigations refined.	all
R11 · NEW	Goodhart within Manager Mirror	Even private self-view is gameable. Keigo/interruption improve on the measurable surface while dominance expresses in untouched channels.	measurement-science

05 Architecture & moat

AREA	BEFORE	AFTER	DRIVER
Moat layers	4 layers: product (refusals) → process (consultation) → data (baselines) → regulatory (compliance)	5 layers (+ Layer 5: anti-inference architecture — retaliation defense via telemetry partitioning, min-group-size suppression, batching/redaction) · Layer 1 upgraded to procedural-fairness spine (contestability state machine, bounded-context rule, retention-under-challenge)	retaliation · legal-max-hardening · legal-procedural-fairness
Detector output	Scalar intensity enum (calm/watch/concern)	<code>ConfidenceBundle</code> with 5 sub-components + reason codes + abstention state + evidence grade. Types declared in <code>src/lib/types.ts</code>	confidence-abstention · measurement-science
Detector taxonomy	All 7 detectors treated as equivalent "structural"	Three lanes declared at <code>src/lib/pipeline/detector-registry.ts</code> . Employer-facing default = structural-only. Tenant feature flag required for hybrid lane.	detector-boundary · model-boundary
Meeting-type handling	Universal baseline, cross-type pooling	<code>MeetingType</code> + <code>scoringMode</code> as first-class fields. MVP scoreable = weekly_sync, standup, design_critique. Rest = observation-only. Cross-type pooling blocked.	meeting-type-normalization
Governance "will not do"	10 items	12 items (+ no decryption of evidence server-side, + no company-wide health score) +	manager-adoption · labor-

AREA	BEFORE	AFTER	DRIVER
		3-lane accountability model (self-correction / governed remediation / formal review) + anti-laundering rule + responsibility-boundary CAN/CANNOT table	relations · measurement-science

06 Measurement & success

AREA	BEFORE	AFTER	DRIVER
Success model	4 layers: adoption · behavior · responsiveness · strategic	6 layers (+ remediation quality — was the handling actually fair? · + human recovery — did the affected person recover?) · resists Goodhart by requiring worker-side fairness rating	remediation-outcomes
North-star metric	Detection pass rate (3/3 on seed, 0 FP on control)	Paying JP companies with signed 就業規則 amendment AND completed Lane-B remediation cycle with worker-rated fairness $\geq 4/5$	remediation-outcomes · labor-relations
Evidence quality	Binary: signal detected / not detected	4-state dashboard: signal detected / no qualifying signal in observed data / insufficient observation / out-of-scope. Absence of signal \neq absence of problem.	false-negative

07 What was REMOVED

Worth calling out explicitly. These five items were in v1 as strengths; the research made them liabilities:

- **"The pattern is the harm"** — philosophically satisfying but legally indefensible. Replaced with "evidence consistent with uneven conversational treatment."
- **"CEO's instrument for seeing the bill"** as top-of-document identity — reads as surveillance-for-executives. Demoted to closed-sponsor context.
- **"Metadata only, no content"** blanket claim — falsified by our own embedding-based detectors. Replaced with explicit lane taxonomy.
- **EU-first expansion** — DE requires works-council co-determination on monitoring tech; we don't have the playbook yet. Japan is actually the easier first market.
- **Single-CEO buyer** — reads as top-down imposition to worker reps, who will block deployment. Three-buyer model required.

08 What the research did NOT change

Worth also naming. Several architectural choices held up across all 42 documents and required no revision:

- **Structural-only detector core** — 3 detectors (intrusive-interruption, chilling-delta, Gini) remain the default employer-facing lane. Endorsed by every relevant memo.
- **Per-speaker 90-day baseline** — defended explicitly by baseline-calibration + measurement-science memos.
- **Refusal of company-wide health bar** — multiple memos independently endorsed the refusal.
- **Role-based visibility with audit trail** — endorsed by trust + legal-procedural-fairness.
- **Victim-owned E2E evidence vault** — endorsed by retaliation + legal-max-hardening as one of the strongest architectural defenses.
- **"Review-worthy event" language** (not "harassment detected") — endorsed throughout.
- **EU AI Act Art. 5 no-emotion-inference alignment** — endorsed as architecturally required.

The one-sentence summary: Kashi's *product architecture* was mostly right from v1. Its *language, ROI math, buyer model, and risk register* were not. The research didn't require a rewrite of the product; it required a rewrite of how the product presents itself and defends its claims.

09 v2 → v3 delta (2026-04-21 post-GitHub-push)

Small but important shipped-delta after the wave-3 docs pass. This is the bridge commit: UX commitment from the Ideas_wave3 research lands in the live product, not just in types.

AREA	V2 (DECLARED)	V3 (SHIPPED)	DRIVER
ConfidenceBundle emission	Types declared in <code>types.ts</code> ; not yet emitted. Detectors still return legacy scalar outputs.	Backward-compat wrapper at <code>confidence-bundle.ts</code> . Every <code>ManagerMirror</code> + <code>PatternSummary</code> now carries <code>overallGrade</code> + <code>overallAbstention</code> + <code>overallReasonCodes[]</code> . UI shows grade badge with hover rationale on <code>/demo/mirror</code> + <code>/demo/ceo</code> .	confidence-abstention · measurement-science
Weakest-link grade logic	No aggregation logic specified	Headline grade = MIN of (inputQuality, exposureSupport, contextSupport, detectorConfidence, aggregationSupport). "A signal is only as strong as its weakest component."	confidence-abstention
Tech-doc §7.5	§7 listed 23 improvements, ranked but not sequenced	New §7.5 "Next 3 builds" critical review — picks Japan data residency + meeting_type gating + RLS CI tests as the first post-funding Sprint 1. Also flags 3 items the original §7 missed (webhook ingest, feature-flag admin UI, detector cost telemetry).	internal critical review

AREA	V2 (DECLARED)	V3 (SHIPPED)	DRIVER
Code shipped	0 lines of ConfidenceBundle consumption	~250 lines: wrapper + type extensions + 2 seed updates + 2 UI components + 3 doc updates	this pass

Companion docs: [/business.html](#) · [/governance](#) · [/deck.html](#) · [/tour.html](#) · [research library](#)

v1 → v2 diff · integrated 2026-04-21 · 24 business PDFs + 18 technical-dev memos