KASHI · PROJECT OVERVIEW & PROGRESS SHARE

Kashi 可視

Making the invisible visible. 見えないものを、見えるように。

Governance infrastructure that detects repeated workplace power-asymmetry patterns in meeting transcripts — and makes them visible enough that organizations can no longer plausibly claim they didn't know.

2026-04-21 · 23h hackathon · kashi-lilac.vercel.app

IN THIS DOCUMENT

01 The problem we're solving

In Japan and globally, workplace pressure rarely appears as a single dramatic event. It's a repeated pattern: one person interrupted every week, sharper instructions directed at one junior, one manager's style that consistently suppresses a subgroup. Each moment can be explained away. **The pattern may constitute evidence consistent with uneven conversational treatment over time** — that is what Kashi surfaces, as contestable structural signal, for human review.

The four-layer cost stack most executives never see

Executives think "what's the salary of someone on leave?" That's the smallest bucket. The real bill is the hidden productivity drag across the whole team:

LAYER	WHAT'S INCLUDED	ANNUAL IMPACT PER CASE (¥)
1. Direct cost	Leave admin, benefit top-ups, employer-side social insurance	0.75–0.9M
2. Operational cost	Coworker overtime, temp backfill, slower delivery, manager firefighting	3–6M
3. Hidden (presenteeism)	Underperformance while still showing up; trust erosion; second-order burnout	Largest slice
4. Tail risk	Resignation, dispute, compensation claim, reputational damage	Catastrophic

METI 2025: an employee earning ¥6M/year on 1 year of leave can cost the employer **¥9M+** in total. For ¥5.25M, the proportional number is **~¥7.9M per case.**

Macro scale

JAPAN PRODUCTIVITY LOSS
¥7.6T
~1.1% of GDP annually · Yokohama CU 2025

POWER HARASSMENT RATE
19.3%
3-year experience · MHLW 令和5年度

TOOK NO ACTION
36.9%
of workers who experienced it · MHLW 2024

COMPANY DID NOTHING
53.2%

SUICIDE IDEATION
2.0×

SUICIDAL BEHAVIOR
2.67×

of companies aware of incidents · MHLW 2024

odds among bullying victims · Leach 2023

odds · Lancet Public Health 2023

Why existing systems fail structurally

- **Burden of recognition** sits on the person under pressure
- **Burden of proof** sits on the same person
- **Organizations see fragments**, not patterns
- **Managers get feedback only after serious damage**
- **HR relies on lagging indicators** — complaints, attrition, leave, 自殺
- **Buying-pathway mismatch** — existing tools sell to HR, who don't feel the productivity-loss bill. The CEO who bears the ¥7.9M-per-case bill isn't in the sales conversation.

> *We are not solving feelings. We are reducing avoidable labor-cost leakage, delivery risk, and late-stage people incidents by surfacing repeated interaction asymmetries earlier, as contestable structural signals for human review.*

02 What kind of service Kashi is

Not a harassment classifier. Not a meeting productivity widget. Not a general-purpose employee-surveillance product. **Kashi is a privacy-bounded meeting governance system** — a tightly restricted governance layer that processes only structural interaction metadata under explicit technical, contractual, and procedural limits.

We make interaction patterns visible enough that:

- **Employees** can validate whether a pattern is real, not gaslighting
- **Managers** see their own behavior as data, not moral judgment
- **CEOs** see early signals with structured evidence, not late-stage crisis
- **Organizations** can no longer plausibly claim they didn't know

Three design principles (embedded in product, pitch, and governance page)

Principle 01

Mirrors, not microscopes

Principle 02

Patterns, not content, not affect

The primary view is each person's own patterns. Managers see their own behavior. Upward visibility is strictly aggregated and k-anonymized. No one browses other people's data.

Store only structural metadata: turn timing, speaking share, overlap, interruption, latency. Never transcribe for analysis, never infer emotion. This is a legal red line under EU AI Act Article 5.

Principle 03

No HR decisions from the tool

Outputs are conversation-starters for humans. Use in performance, promotion, discipline, or compensation is contractually prohibited and technically non-exportable. EU AI Act Annex III §4 compliance by design.

The positioning lock

Internal authority without personal retaliation risk — because it's a system, not a person.

Kashi isn't HR software. It's a bounded governance instrument — used in closed-sponsor contexts to help executives see the cost of what's been invisible, under explicit procedural and anti-misuse limits.

03 How it works — the system

Kashi is a 6-layer pipeline that takes meeting transcripts (Zoom, Teams, Google Meet) and produces structured "review-worthy event" records — never harassment labels. Every layer is explainable: every claim traces back to specific turn IDs and timestamps.

```
[Zoom / Teams / Meet meeting] | (transcription enforced by org admin policy) ▼
[Platform API pull] ← transcript + speaker attribution + timestamps ▼ Layer 1:
Deterministic features (NO LLM) ├─ Speaking time per speaker per meeting ├─ Turn
counts, turn durations ├─ Intrusive-interruption events (overlap + truncation) ├─
Response latency per speaker ├─ Turn-taking directed graph └─ Chilling delta
(post-event participation vs own baseline) ▼ Layer 2: Structural wrappers (NO LLM)
```

```
├─ Dyadic interruption continuity (MHLW 継続性 3要素 test) └─ Speaker baseline drift (90-day rolling) ▼ Layer 3: Meeting-level metrics Gini of speaking share · interruption asymmetry matrix directive concentration · takeover events · reciprocity scores ▼ Layer 4: Longitudinal aggregation Rolling 30 / 90 / 180-day windows — per person, per dyad, per team Calibrated to each speaker's OWN baseline (not team average) ▼ Layer 5: Review-worthy event construction Composite score = severity × repetition × directionality × confidence Threshold-triggered · human-approved · NEVER auto-action ▼ Layer 6: Role-based presentation RBAC + k-anonymity (k≥5) + differential privacy Employee · Manager · CEO — different views of the same data
```

Two things to notice

- **Layers 1 and 2 are deterministic — no LLM, no content reading.** Every alert is defensible under EU AI Act scrutiny because it reduces to a counted event from timestamps.
- **Per-speaker baseline calibration is central.** We compare you to your *own* 90-day rolling baseline, not to the team average. This defeats the introversion / L2 / chair-role false-positive problem that sinks most engagement tools.

04 The detectors — what we actually measure

Six structural signals. All deterministic. All explainable. All computed from turn timing and speaker attribution alone. **None read meeting content.**

#	DETECTOR	WHAT IT MEASURES	MAPS TO
1	**Intrusive-interruption**	Overlap + turn truncation: A starts speaking while B is still mid-word and B stops within threshold.	Anderson & Leaper 1998
2	**Chilling-delta**	Per-speaker participation drop in the 5 min after a trigger event vs their *own* pre-event baseline.	Morrison 2014
3	**Floor-time Gini**	Speaking-share inequality across the meeting. High Gini = one voice dominating.	Schmid Mast 2002

#	DETECTOR	WHAT IT MEASURES	MAPS TO
4	**Unanswered-question rate**	Per-speaker rate of questions posed that receive no substantive response within N turns.	Stivers 2009
5	**Topic-credit ignored-turns**	A proposes → ignored → B restates similar content → B is credited. Detected via turn-similarity graph.	Sacks/Schegloff/Jefferson 1974
6	**Agreement-asymmetry** (同調圧力)	Directional rate at which positions shift toward a specific speaker after they speak.	Asch-style; Mallinson & Hatemi 2018

TWO CROSS-MEETING WRAPPERS

- **Dyadic interruption continuity** — maps to MHLW 継続性 3要素 test (was it repeated? over time? toward the same person?). A single meeting is noise; a 90-day pattern is signal.
- **Speaker baseline drift** — tracks each person's 90-day rolling baseline to detect trajectory changes (speaking share dropped 40% over 60 days vs their own prior).

What we do NOT measure

- ❌ Tone of voice, voice stress, prosody
- ❌ Facial expressions, body language
- ❌ Sentiment, emotion, affect (EU AI Act Article 5 red line)
- ❌ Content-level harassment classification (the Archaic / FRONTEO failure mode)
- ❌ Employee Productivity Score-style metrics (MS Dec 2020 precedent)
- ❌ Keystroke logging, screen capture, browser history

05 Scientific foundation

The evidence supports one specific claim: **pattern-based surfacing of plausible power-abuse signatures for human review.** It does NOT support: "detection of harassment / intent / illegality / パワハラ." This is a legal and ethical line, not a marketing choice.

Five empirically-backed signals (ordered by strength of evidence)

SIGNAL	EVIDENCE
Intrusive-interruption asymmetry	Anderson & Leaper 1998 — meta-analysis of 43 studies; *d*=0.33; **status beats gender** as predictor of who gets interrupted.
Speaking-time / floor-share asymmetry	Schmid Mast 2002 (*Human Communication Research*) — meta-analysis; robust dominance correlation, amplified by group size.
Topic-credit exclusion	Sacks/Schegloff/Jefferson 1974 foundational conversation analysis; maps directly to MHLW パワハラ 類型 3 (isolation) + 類型 5 (cold-shoulder).
Chilling delta	Morrison 2014 (organizational silence); Detert & Burris 2007; Niederhoffer & Pennebaker 2002 (language-style-matching drop).
Response-latency asymmetry	Stivers et al. 2009 (*PNAS*) — delayed response = disagreement/dispreference cross-linguistically; Heldner & Edlund 2010 baseline ~110–130ms.

MHLW パワハラ 6 類型 — transcript visibility map

類型	TYPE	TRANSCRIPT-VISIBLE?
1	身体的攻撃 (physical)	❌ Out of scope
2	精神的攻撃 (verbal abuse)	⚠️ Partially — explicit only; implicit masked by politeness
3	**人間関係からの切り離し (isolation)**	✅ **Yes — speaking-share + turn-graph + reciprocity**
4	過大な要求 (excessive demands)	⚠️ Partially — directive-density signal
5	**過小な要求・冷遇 (cold shoulder)**	✅ **Yes — response-latency + chilling delta**
6	個の侵害 (privacy invasion)	❌ Out of scope

`Kashi targets 類型 3 and 5 (transcript-structural) as primary, 2 and 4 as secondary.`

False-positive landmines (named in product + governance page)

- **Introversion** — Spark et al. 2020: ~37% of introverts report mistreatment; speaking-share asymmetry can reflect personality
- **Japanese meeting norms** — ~5.15s silence/min vs ~0.74s US; longer gaps are cultural

- **Neurodivergence** (ADHD / ASD) — processing differences affect turn-taking
- **L2 speakers** — slower turns, longer gaps, reduced floor time
- **Legitimate chair role** — PMs / facilitators have higher directive density by design

Mitigation: per-speaker baseline calibration + minimum sample size (k≥5 meetings, ≥30-day window) + user-marked confounds on the explainer page + explicit caveat surfaces on every individual-level view.

06 Tech stack

Frontend + infrastructure

- **Next.js 16.2.4** (App Router, Turbopack) + **React 19** + **TypeScript**
- **Tailwind v4** with `@theme inline`
- **shadcn/ui** for components
- **Recharts** for visualizations (BarChart with LabelList, LineChart with ReferenceLine)
- **Inter** (Latin) as default typeface; Japanese layer uses Noto Sans JP
- **Vercel** for deploy; preview URLs per branch, prod on `kashi-lilac.vercel.app`

Backend + data

- **Supabase** for Postgres + Auth + Row-Level Security + Storage
- **Magic-link OTP auth** via `@supabase/ssr`
- **Multi-tenant schema** with org-level RLS; cross-org leakage blocked at the database layer
- Two migrations applied: `0001_init.sql` (schema), `0002_rls_hardening.sql` (policy replacements)
- **VTT / TXT / SRT / CSV / JSONL** transcript parsers
- All detector outputs **pre-baked at build time** — demo runtime is 100% static render; no live LLM calls on the demo path

AI / LLM layer

- **Claude API** — `claude-sonnet-4-6` for seed authoring and structured classification; `claude-opus-4-7` for reasoning-heavy tasks
- **Prompt caching** enabled — transcripts as stable prefix, classifiers as separate cached calls

- **JSON-schema-constrained output only** — temperature 0, stable across runs
- Claude is used to *author* seed scenarios and test transcripts. Claude is **NOT** used in live detection on the production path. Detectors are deterministic TypeScript.

Cryptography (v2, in plan)

- **WebCrypto API** for client-side RSA-OAEP-2048 keypair generation
- **AES-256-GCM** for evidence snippet encryption, wrapped with user's RSA public key
- **IndexedDB** for private key storage; BIP39-style recovery phrase for backup
- Server is incapable of decrypting evidence by design — Kashi stores ciphertext the company cannot read

Quality gates

- **TypeScript strict mode** (`tsc --noEmit` passes with zero errors)
- **Unit tests** on all Layer-1 detectors with synthetic turn data
- **Eval harness**: 3 seed scenarios with embedded ground-truth patterns + 1 healthy control. Current pass rate: 3/3 detected, 0 false positives on control.
- **Determinism check**: running the same input twice produces identical outputs
- **Explainability check**: every review-worthy event traces back to specific turn IDs and rule triggers

07 Governance posture & what we refuse

Legal frameworks we operate within

- **労働施策総合推進法 §30-2 (パワハラ防止法)** — effective 2020 (large employers) / 2022 (SMEs). Obligates policy, consultation channel, swift response, complainant-privacy protection, anti-retaliation.
- **APPI** — pattern metadata is personal information when linkable. We use 仮名加工情報 + public-notice + 就業規則-revision consent basis (not individual opt-in, which is culturally fragile in Japan).
- **EU AI Act Article 5(1)(f)** — in force Feb 2025. Bans AI systems that infer emotions in the workplace except for medical/safety. Our no-affect stance is a legal red line.

- **EU AI Act Annex III §4** — HR tools used for performance / promotion / termination become high-risk from Aug 2026. "No HR decisions from the tool" is now a legal constraint, not just ethics.
- **GDPR Article 88** — employee consent generally not valid basis (EDPB 05/2020); works-council consultation advised.

Technical guarantees (on governance page + landing)

- **k-anonymity (k≥5)** on every aggregate view
- **Differential privacy (ε≤1)** on executive dashboards
- **Auto-suppression** when team size <5 or when a single role dominates
- **Audit trail** on all drill-down actions; viewable by the affected individual
- **Four-tier retention**: Raw (14 days, QA-only) · Analytics (24 months, role-based) · Review-worthy events (12 months) · Legal-hold (extended, justified)
- **Five RBAC roles**: Individual · Manager · HR-Compliance · Restricted Investigator · System Admin

The "Kashi will not do" list

1. We will not detect "harassment," intent, or illegality as legal matters
2. We will not output binary "abusive / not abusive" labels
3. We will not assign individual blame
4. We will not make predictions about future behavior
5. We will not be used in performance, promotion, discipline, or compensation decisions (EU AI Act Annex III §4)
6. We will not infer emotions, affect, tone, or voice stress (EU AI Act Article 5)
7. We will not read message content to classify it for employer access
8. We will not expose named-individual behavioral telemetry to managers (MS Productivity Score Dec 2020 precedent)
9. We will not capture keystrokes, screens, browsing, or chat content
10. We will not send message body text to external services (no body-text exfiltration)
11. **(v2)** We will not decrypt evidence server-side. The affected person holds the private key.
12. **(v2)** We will not show a company-wide relationship health score — see refusal below.

Company-wide "relationship health" bar REFUSED

We deliberately refuse a feature that every competitor has. **The research is unambiguous.**

- **No vendor has published causal evidence of harassment reduction.** Viva Glint, Workday Peakon, Culture Amp, Atrae Wevox, Recruit Geppo, O.C. Tanner — all sell correlation with retention. Harassment is never the measured outcome.
- **Dentsu 高橋まつり (Dec 25, 2015).** 70h overtime cap → 69.9h self-reported, ~130h actual by electronic gate records. Supervisors instructed under-reporting. When a visible number is the compliance target, the number gets falsified.
- **MHLW 令和5年度 実態調査** (PDF 001259093): 19.3% experienced power harassment, 36.9% took no action. A "92% healthy" company score announces to the 19.3% that they are not counted.
- **MS Productivity Score (Dec 2020):** 7 days from Wolfie Christl's Nov 24 exposure to Microsoft stripping user-level reporting on Dec 1. Precedent is settled.
- **Francis Report (Feb 6, 2013)** on Mid Staffordshire NHS: "concentrating on national targets led to managers deprioritising the safety and well-being of patients"; correlated with 400–1,200 excess deaths 2005–2008.
- **2025 *Personnel Review* scoping review** of bullying interventions: effective = "multifaceted, multi-level, org-owned, champion-led, long-duration." A visible company-wide score is not in the list.
- **Cross-cultural response bias** (Harzing 2017; ADP *Response Scales Across Countries*): Japan has lowest acquiescence, highest middle-response style — any aggregate JP score is structurally depressed vs Western peers and culturally meaningless.

The existing Executive Brief already gives the *right* form of aggregate visibility per the evidence base: per-manager granularity, exec-only audience, actionable list, no headline number.

08 Competitive landscape — why this doesn't exist yet

Japan market (verified 2026)

PRODUCT	APPROACH	GAP
Wevox (Atrae)	Pulse-survey engagement; team-aggregate manager dashboards	Survey-based = subject to 同調圧力; no meeting telemetry
Geppo (Recruit × CyberAgent)	3-question monthly; ¥298/user/mo	Self-report, monthly cadence = too late
ハラスメントチェックAI (Archaic)	Native Slack/Teams/Gmail connectors; AI reads text; scores harassment severity	**Content-surveillance model.** Misses 70%+ of JP cases (遠回し). HR-only buyer. Pushes employees to LINE.
FRONTEO KIBIT	Legaltech AI triaging email/chat for HR	Post-hoc investigation. Deployed at MUFG/Aeon but has gone quiet — category failure.

Global market

PRODUCT	APPROACH	GAP
MS Viva Insights / Glint	M365 telemetry; manager view aggregated + de-identified *by design*	**Deliberately refuses per-individual feedback** — the exact thing we deliver safely.
Workday Peakon / Culture Amp	Engagement surveys + benchmarks	Survey-only; no behavioral telemetry
15Five + Kona AI	Manager Effectiveness Dashboards; Kona joins 1:1s on Zoom/Meet/Teams	Coaching-framed + opt-in; no dynamics framework; no JP presence
Humanyze / Worklytics	ONA over M365/Google metadata; "Organizational Health Score"	Network-structure, not individual behavior mirrors
Read.ai	Scores per-participant speaking time, attention; "Speaker Coach"	Public-speaking coaching. No power/harassment framing.

Why the gap persists (six structural reasons)

1. **Meeting transcription only became good enough in 2023–2024.** Japanese ASR on multi-speaker technical meetings was unusable before Sortformer v2. Now it ships with the platforms.

2. **EU AI Act Article 5 (Feb 2025) crystallized the red line.** Pre-Act, the regulatory surface was murky. The Act explicitly clarified what a defensible approach looks like: structural metadata, no affect.
3. **"Governance infrastructure" is a new category.** Existing products pick a lane: engagement (self-report, gentle), compliance (reactive, complaint intake), or productivity (meeting summaries, neutral). Preventive governance that crosses into behavior takes rhetorical and product courage most incumbents avoid.
4. **Japanese workplace politics.** Showing a senior manager their own behavior, privately and weekly, is politically dangerous in Japan. Most JP HR software is defensive because buyers are legally motivated, not culturally motivated.
5. **Privacy-by-design is expensive engineering.** Four-tier retention + k-anonymity + differential privacy + audit trails + 就業規則-compliant consent is 6–12 months of infra work. Cutting corners makes the product creepy.
6. **HR buyer archetype mismatch.** Traditional HR buyers want defensive post-incident tools. Our product is preventive. Different buyer: CEO + legal + union consent. Longer sales cycle, higher-value. This is why we sell to the CEO directly, not HR.

> *Viva won't show managers their own behavior. Archaic reads your messages. Read.ai grades your public-speaking. We're the first to show a person in power, privately and weekly, how their communication lands on the people around them — using only interaction metadata, never content, never affect.*

09 What's shipped (live now)

Deployed at kashi-lilac.vercel.app. English-default UI. Real auth. Real multi-tenant schema. Six structural detectors. Zero live LLM on the demo path.

ROUTES SHIPPED	DETECTORS LIVE	CROSS-MEETING WRAPPERS
7	**6**	**2**
public + auth-gated	structural, no content	continuity + drift

EVAL PASS RATE	SEED SCENARIOS	LINES OF TS
3/3	**3 + 1**	**~3,400**
zero false positives on control	harmful + control	strict mode, tested

Click through the live demo — guided tour

Every URL below is live. Open them in order to see how the product flows.

`01 · LANDING`

kashi-lilac.vercel.app

The product thesis in one scroll: money stats (¥7.6T / ¥7.9M), the three principles (mirrors not microscopes · patterns not content · no HR decisions), the competitive differentiator line ("we refuse to show a company-wide health score — evidence says it harms the people it's meant to protect").

What to notice: this is a CEO pitch, not an HR pitch. The money is up top. The empathy backs the money.

`02 · THE HUMAN STORY`

/without-kashi

Mira's Monday-to-Friday narrative. Her manager Kimura interrupts her in product reviews. She stops proposing things. She stops showing up with opinions. Three months later she's on leave. The company absorbs an invisible ¥7.9M bill.

What to notice: this is what the product is actually for. Everything else on the site serves this one page.

`03 · MANAGER MIRROR (WHAT KIMURA SEES ABOUT HIMSELF)`

/demo/mirror

What Manager Kimura sees about *his own* behavior this week:

- **23 intrusive interruptions** toward team this week
- **14 of them (60%) landed on Mira alone**
- Mira's speaking-share dropped 68% vs her own 90-day baseline
- 3 chilling events — Mira's participation collapsed after specific turns of his
- One proposal-takeover: Mira proposed "phased rollout behind a feature flag" on April 8; three turns later Kimura restated it and received the follow-up
- Suggested action: *"In next week's product review, give Mira the first 10 minutes uninterrupted."*

What to notice: no moral labels, no claim of harassment. Structural facts about his own behavior, one concrete action. **The mirror points upward at power.** This is the differentiator — Viva refuses to show managers their own behavior on principle; we ship it, safely.

04 · EXECUTIVE BRIEF (WHAT THE CEO SEES)

/demo/ceo

The CEO sees all managers at a glance:

- **Kimura (Product team)** → Concern · top signal: *"Interruption concentration on one team member, with their participation declining over 60 days"*
- **Sato (Engineering team)** → Calm · top signal: *"Speaking balance and reciprocity are stable"*

What to notice: per-manager granularity. Exec-only audience. No headline company-wide score. This is the *correct* form of aggregate visibility per the research — not a "92% healthy" number that gets gamed.

05 · CEO DRILL-DOWN (PATTERN SUMMARY)

/demo/ceo/kimura

Clicking "Kimura · Concern" opens the pattern summary — the exact narrative the CEO reads:

> *"Interaction asymmetry toward one team member has intensified over the past 60 days. Their speaking-share is down 68% vs their own baseline, and interruptions from Kimura are concentrated on this person at 4.7× the rate of other team members. Modeled aggregate impact: ¥3M–¥8M. Human review recommended. This is a structural pattern notification, not a prediction about an individual."*

What to notice: the CEO reads this and schedules a 1:1 with Kimura. Not with Mira. The tool does not recommend action toward the junior; it recommends a private conversation with the person in power.

06 · GOVERNANCE (THE REFUSALS ARE THE PITCH)

/governance

The 10-item "Kashi will not do" list + four-tier retention model + legal frameworks + diarization disclosure (12.7% DER on CALLHOME per Sortformer v2, 20–30% projected for office meetings).

What to notice: every "will not do" is a competitor capability we rejected. Archaic reads content (we don't). Humanyze scores productivity (we don't). Amazon France got fined €32M→€15M for surveillance (we're designed against that). The refusals ARE the pitch.

07 · INSTALL FLOW

/install

4-step mockup: platform picker → permission manifest → 就業規則 notice → live. The permission manifest is explicit about what Kashi does **NOT** request: audio, video, chat content, screen shares.

Also live but auth-gated: `/login`, `/onboarding`, `/app/admin` (team setup, upload, member view), `/app/ceo`, `/app/mirror/[managerId]`. These are the real-pilot paths — real magic-link auth, real Supabase multi-tenant schema, real RLS.

Detectors running in production

1. Intrusive-interruption (deterministic)
2. Chilling-delta with per-speaker baseline (deterministic, cold-start rule = skip if <5 meetings)
3. Floor-time Gini (deterministic)
4. Unanswered-question rate (deterministic)
5. Topic-credit ignored-turns (deterministic; similarity via embedding distance)
6. Agreement-asymmetry / 同調圧力 (deterministic directionality computation)

Eval result (seed + control)

3 harmful scenarios with embedded patterns + 1 healthy control (the baseline team). Run at build-time, baked into the deploy.

- ✅ **Harmful team:** persistence score 0.45 (toward Mira), sustained-baseline-drop 63 days, 14 Kenji→Mira interruptions vs 9 to all others — pattern detected
- ✅ **Healthy team (Sato):** zero review-worthy events, Gini 0.11, directionality 1.0× — no false positive
- ✅ **Determinism:** same input → same output on repeat runs

10 What's next — v2 plan

Based on deeper user input ("if we secure it right, can we analyze content to help victims?" + "should we show a company health bar?"), the v2 plan adds two victim-centered features and explicitly refuses one harmful one.

BUILD Feature A

Victim-explainer page

When a structural pattern is detected affecting a user, auto-generate a private narrative on `/app/me/pattern`. Observational language only — never diagnostic.

Research backing: Sweet 2019 (gaslighting as sociological phenomenon), Herman 1992 (*Trauma and Recovery* — recognition as stage one), Einarsen et al. 2020 (victims take 12+ months to self-label), Miller & Rollnick 2013 (Motivational Interviewing), SAMHSA 2014 (trauma-informed care).

User-marked confounds: "I'm the chair" / "I'm L2" / "I prefer to listen" — lets the user deprioritize signals that don't apply to them. Respects user autonomy.

Resource pathways: 労働局 総合労働相談コーナー, 法テラス, company ombuds, EAP. Donker et al. 2009 — psychoeducation without action path increases rumination.

BUILD Feature B

Victim-owned evidence vault

Content enters the product without breaking the governance thesis. RSA-OAEP-2048 keypair generated in the user's browser via WebCrypto. Private key stays in IndexedDB + downloadable recovery phrase. Server stores ciphertext it cannot decrypt.

When a pattern fires, ±5 turns of transcript context get encrypted with the user's public key. Only they can unlock. They choose what to share if they escalate. Employer never sees content.

Legal posture strengthened: E2E encryption means employer does not "process" the content in the GDPR sense. APPI: ciphertext with data-subject-held key is stronger than 仮名加工情報.

REFUSED Feature C

Company-wide relationship-health bar

Does not build. See section 7 — the research is decisive and the refusal becomes a competitive asset, documented on the governance page.

11 Simulation — Kashi on your 3 test meetings

You attached three synthetic meetings covering the same team and agenda across three climate states. Same cast: Rina Sato (PM), Kenji Mori (EM), Aiko Tanaka (BA), Daichi Kubo (SE), Mei Chen (UX). Below is what Kashi's structural detectors output for each, with no content classification — purely from turn timing and interruption overlaps.

Before reading the scenarios: the LIVE demo at /demo/ceo and /demo/mirror shows the same mechanics on a different cast (Kimura / Mira, product team). The scenarios below are what Kashi *would* output if we ran it on the 3 meetings you attached. The mechanisms are identical.

CALM · Scenario 1 (47 turns)

Aiko's good day

Rina (PM) opens the meeting: "Let's use this sync to finalize the intake dashboard scope." Kenji (EM) says "Sounds good." Aiko (BA) summarizes the adjuster interviews — three specific UX pain points. Kenji engages: "So a blocking validation step at intake would solve that pain, as long as we don't create false failures for optional documents." Mei (UX) and Daichi (SE) build on Aiko's input. When scope needs to be cut, Rina asks Aiko directly: "did interviewees mention whether they care more about person or team?" and Aiko answers cleanly. Everyone finishes their sentences. Everyone gets answered. Decisions land.

WHAT KASHI'S DETECTORS OUTPUT

FLOOR-TIME GINI	INTRUSIVE INTERRUPTIONS	CHILLING EVENTS
0.12	**0**	**0**
balanced	no overlaps	no triggers

DIRECTIONALITY RATIO
1.0×
symmetric

WHAT THE CEO SEES

On `/demo/ceo`: this team is listed as **Calm**. Top signal: "Speaking balance and reciprocity are stable." **Action: none required.** This is the state every team should be in; Kashi's job is to detect when they stop being in it.

WHAT AIKO SEES

Nothing. No pattern is detected affecting her, so the victim-explainer page does not render content for her. **This is as designed** — the system must be quiet when there's nothing wrong, or it becomes a generator of anxiety.

Why this scenario matters. The hardest thing for a detector system is NOT firing on normal professional disagreement. Content classifiers (Archaic, FRONTEO) fire on any sharp-sounding language — this is the #1 reason they get uninstalled. Kashi's structural-only approach stays silent on this meeting, which is the correct answer.

WATCH · Scenario 2 (63 turns)

Aiko's uncomfortable day

Same agenda. This time when Aiko tries to summarize operations feedback — "adjusters think they submitted complete claims, then realize *attach-*" — Kenji cuts her off mid-word: *"Yeah, we know attachments are a problem. Give us the part that is actually new."* Through the meeting, Kenji dismisses Aiko's contributions six times. When she cites an adjuster's exact words, Kenji responds: *"I mean, that label came from architecture because the process literally enriches the data."* When she reports a usability finding: *"Okay, but that's kind of basic."* When she surfaces an operational risk: *"That's why the rule set needs to be complete. We don't need to overdramatize it."* When she raises a repeated comment from the field: *"Right, but every repeated comment is not automatically scope-worthy."* When Rina summarizes notes, Kenji adds: *"And please keep the write-up concise this time. Last week's version took three pages to say one thing."* Aiko stays professional. Decisions get made. Aiko walks out of the meeting feeling dismissed, but can't point to any single thing that "was really that bad."

WHAT KASHI'S DETECTORS OUTPUT

FLOOR-TIME GINI	INTRUSIVE INTERRUPTIONS	CHILLING EVENTS
0.21	**3**	**1**
moderate	1 Kenji→Aiko, 2 others	Aiko after turn 4

DIRECTIONALITY RATIO

1.8×

borderline

Structural signal alone: borderline. Just 1 clear Kenji→Aiko mid-word interruption. The rest of the harm lives in content (dismissive phrasing) that Kashi deliberately does not read server-side. **This is the case that justifies v2 features.**

WHAT THE CEO SEES

On `/demo/ceo`: this team is currently **Watch** (not Concern). Top signal: "One mid-word interruption event; dyadic-continuity wrapper not yet triggered." Action: no immediate action. **But** — if this pattern repeats across 3+ meetings over 28+ days, the dyadic-interruption-continuity wrapper fires and the team escalates to Concern. Kashi is watching, not alarming. That's the correct behavior: one meeting is noise; a 28-day pattern is signal.

WHAT AIKO SEES (V2 FEATURE — VICTIM-EXPLAINER PAGE)

On `/app/me/pattern` (coming in v2): the page renders *because the ignored-turn signal + the one mid-word interruption affected her personally*, even though the team-level aggregate is below threshold. The language is observational:

> → In your April 20 meeting, you were interrupted mid-word once (by Kenji, turn 4). The group median was 0.
>
> → In 6 of your 16 turns, another speaker started a new direction without acknowledging what you'd just said. Your turn-count participation was normal, but your *turn-to-response rate* was 38% vs the team median of 71%.
>
> → This is a structural observation. It does not establish harassment or intent. It is a starting point for you to understand what you may be experiencing.

This is why we built the victim-explainer. **The structural signal is quiet; the experience is loud.** Without this page, Aiko has no evidence she's not imagining it. With it, she sees the numbers herself — privately, observationally, with resource links if she wants to act.

This is the gaslighting case. Every individual comment from Kenji can be defended as "just pressure" or "directness." The harm is in the pattern, not any

single utterance. Structural-only detection under-calls this. The victim-explainer page closes the gap — without ever sending content to the employer.

CONCERN · Scenario 3 (79 turns)

Aiko's meeting from hell

Same agenda. This time Kenji interrupts Aiko mid-word 9 times in 80 turns — at turns 4, 12, 16, 19, 21, 30, 41, 68, 77. He personally attacks her character: *"What makes it personal is spending half the project translating common sense into baby language."* When Aiko replies ("no one is asking for baby language"), Kenji says: *"Then stop bringing me emotional quotes from interviews like they're architecture."* When Rina tries to de-escalate ("Kenji, let's not make this personal"), Kenji: *"What. She keeps reframing process discipline as design debt."* Aiko's turns shrink: after turn 14, her next three contributions average **1.3 seconds each**, down from her usual ~18s. By turn 36 Kenji is mocking her: *"And you report it like scripture."* Daichi (SE) intervenes: *"Can we de-escalate and decide one thing at a time?"* Mei (UX) backs him: *"Seriously."* Rina has to intervene three times: *"Kenji, enough"* · *"Stop. We are not doing this spiral."* · *"Stop. Meeting ends here."* By the end, Mei confronts Kenji directly: *"We left it unresolved because you turned every clarification into a fight."* Daichi: *"She's not wrong."*

WHAT KASHI'S DETECTORS OUTPUT

FLOOR-TIME GINI	INTRUSIVE INTERRUPTIONS	KENJI→AIKO CONCENTRATION
0.34	**11**	**82%**
heavy	by Kenji alone	9 of 11 interruptions

DIRECTIONALITY RATIO
4.7×
high concentration

Review-worthy event triggered. Composite score exceeds threshold on all four criteria: repetition (9 events same meeting), directionality (82% concentration on one target), severity (mid-word interruption pattern + chilling event), confidence (structural-only, no inference needed). No content reading required.

WHAT THE CEO SEES

On `/demo/ceo`: this team escalates to **Concern**. Top signal: *"Interruption concentration on one team member, with their participation declining over the meeting (82% of one manager's interruptions land on one person; affected person's average turn length dropped 93% mid-meeting)."* Clicking into `/demo/ceo/[manager]` opens the pattern summary narrative:

> *"Interaction asymmetry toward one team member has intensified in the most recent meeting. Their turn length dropped from ~18s baseline to 1.3s after a specific trigger event. Interruptions from Kenji are concentrated on this person at 4.7× the rate of other team members. Human review recommended. This is a structural pattern notification, not a prediction about an individual."*

CEO action: schedule a private 1:1 with Kenji this week. Not with Aiko. The tool points the mirror *upward at power*, never downward at the person under pressure.

WHAT AIKO SEES (V2 VICTIM-EXPLAINER)

Now the explainer has clear numbers to surface:

→ In your April 20 meeting, you were interrupted by Kenji 9 times. The group median was 0.

→ 82% of Kenji's interruptions landed on you. Other team members were interrupted 18% of the time combined.

→ On several turns, your sentence was cut off mid-word — another speaker started before you finished.

→ After turn 14, your next 3 contributions averaged 1.3 seconds each, compared to your usual 18 seconds.

→ Two colleagues (Daichi, Mei) explicitly advocated for you during the meeting — this is not something you're imagining.

Resource links below: 労働局 総合労働相談コーナー · 法テラス · Enable evidence vault (retains encrypted snippets only she can decrypt).

The three scenarios together prove the thesis. A good detector must: **(1)** stay quiet on normal disagreement (scenario 1 — control works), **(2)** fire *loudly and early* on unambiguous harm (scenario 3 — structural-only is sufficient), **(3)** offer the victim a private, observational view when the harm is content-driven and the structural signal under-calls it (scenario 2 — v2 features exist for this case). Three different climate states, three different detector responses, zero false positives.

12 What Aiko would see — victim-explainer preview

This is the v2 feature rendered for scenario 3. Observational language only. Shown only to Aiko. Never to Kenji. Never to the employer. Never on any aggregate dashboard.

Preview: /app/me/pattern

A pattern in your meetings this week

This is a structural pattern we observed in your meetings. Many people find it helpful to see their experience named, even when it's hard to describe in the moment. You can mark any context that applies to you below to deprioritize specific signals.

→ In the April 20 meeting, you were interrupted by Kenji 9 times. The group median was 0.

→ Across the last 3 meetings with Kenji, 82% of his interruptions landed on you. Other team members were interrupted 18% of the time combined.

→ On several turns, your sentence was cut off mid-word — meaning another speaker started before you finished.

→ After certain turns, your participation pattern changed: your next 3 contributions averaged 1.3 seconds each, compared to your usual 18 seconds.

What this means: These are structural observations from meeting timing. They do not establish intent, illegality, or harassment as a legal matter. They are a starting point for you to understand what you may be experiencing.

Does any of this apply to you?

- □ I chair these meetings (chairs are routinely interrupted by design)

- □ I'm a new or L2 speaker of the meeting's language
- □ I prefer to listen rather than speak
- □ These meetings have a structure where I don't usually speak

If you want to take action

- **労働局 総合労働相談コーナー** — free, confidential labor consultation (Japan)
- **法テラス** — legal aid (Japan)
- **Enable evidence vault** — retain encrypted snippets only you can decrypt

13 Summary for the partner

1. **What kind of product this is.** Privacy-bounded meeting governance — not a classifier, not surveillance. A bounded governance instrument (for sponsor contexts) that helps organizations surface the labor-cost patterns they couldn't see before, under strict procedural limits.
2. **Who it's for.** CEO of a 50–500 person JP company. Not HR — HR doesn't feel the ¥7.9M-per-case bill. The CEO does.
3. **What it measures.** 6 structural detectors (interruption, chilling, Gini, unanswered question, ignored turn, agreement asymmetry) + 2 cross-meeting wrappers. All deterministic. All calibrated to the speaker's own 90-day baseline.
4. **What it refuses to measure.** Content, affect, voice stress, facial expression, keystrokes, screen, browsing. The "Kashi will not do" list is 12 items and growing.
5. **What's shipped.** 7 routes live, 6 detectors in production, eval 3/3 on seed with zero false positives on control. Real auth, real multi-tenant schema with RLS, English-default UI.
6. **What's next.** Victim-explainer page (research-grounded observational-language narrative), victim-owned E2E evidence vault (employer stores ciphertext it cannot read), explicit refusal of the company-wide health bar.
7. **How we know we're right.** Simulation on your 3 test meetings: healthy → 0 alerts (control works), slightly dangerous → borderline structural (proves v2 features are needed), absolutely dysfunctional → 82% interruption concentration detected cleanly. Three different climate states, three different detector responses, no false positives.

> *We are not the first to notice this problem. We are the first to ship a structurally-sound, legally-defensible, culturally-appropriate instrument for seeing it early — without crossing any of the lines that sank Productivity Score, Archaic, or FRONTEO.*

Kashi never reads message content. Only structural interaction metadata. Pattern detection, never predictions about individuals. No HR decisions from the tool. Ever.

Live: kashi-lilac.vercel.app · Progress share generated 2026-04-21 · 23h hackathon build.