

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

- 1.1 The problem we're solving
- 2.2 What kind of service Kashi is
- 3.3 How it works — the system
- 4.4 The detectors (what we measure)
- 5.5 Scientific foundation
- 6.6 Tech stack
- 7.7 Governance posture & what we refuse
- 8.8 Competitive landscape
- 9.9 What's shipped (live now)
- 10.10 What's next (v2 plan)
- 11.11 Simulation on 3 test meetings
- 12.12 Victim-explainer page preview
- 13.13 Summary for the partner

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 is the harm.**

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%

of companies aware of incidents ·
MHLW 2024

SUICIDE IDEATION

2.0×

odds among bullying victims ·
Leach 2023

SUICIDAL BEHAVIOR

2.67×

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 detecting harmful team dynamics earlier.

02 What kind of service Kashi is

Not a harassment classifier. Not a meeting productivity widget. Not an employee-monitoring tool. **Kashi is governance infrastructure.**

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

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.

Principle 02

Patterns, not content, not affect

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 the CEO's instrument for seeing the bill before it arrives.

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 <i>own</i> pre-event baseline.	Morrison 2014
3	Floor-time Gini	Speaking-share inequality across the meeting. High Gini = one voice dominating.	Schmid Mast 2002
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 (<i>Human Communication Research</i>) — 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 (<i>PNAS</i>) — 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

類型	TYPE	TRANSCRIPT-VISIBLE?
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 ($k \geq 5$ 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 ($k \geq 5$)** on every aggregate view
- **Differential privacy ($\epsilon \leq 1$)** 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, Atrac 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 CEO roster 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 x 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 <i>by design</i>	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	Coaching-framed + opt-in; no dynamics framework; no JP presence

PRODUCT	APPROACH	GAP
	Zoom/Meet/Teams	
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

7

public + auth-gated

DETECTORS LIVE

6

structural, no content

CROSS-MEETING WRAPPERS

2

continuity + drift

EVAL PASS RATE

3/3

zero false positives on control

SEED SCENARIOS

3 + 1

harmful + control

LINES OF TS

~3,400

strict mode, tested

Routes live

- **/** — landing page: thesis, three principles, money stats, competitive differentiators
- **/install** — 4-step install flow (platform picker → permission manifest → 就業規則 notice → live)
- **/governance** — the "Kashi will not do" list + retention model + legal frameworks + diarization disclosure
- **/without-kashi** — Mira's Mon-Fri narrative: the ¥7.9M bill the company absorbs when no one sees the pattern
- **/demo/mirror** — Manager Mirror for Kimura (public demo, no auth)
- **/demo/ceo** — CEO roster view with pattern-summary drill-down
- **/login, /onboarding, /app/*** — auth-gated real-pilot path (admin dashboard, team setup, upload, member view, manager mirror, CEO roster)

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 1.0x — 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.

CALM

Scenario 1 — Healthy (47 turns)

SPEAKER	TURNS	SHARE	DISTRIBUTION
Rina Sato (PM)	13	27.7%	
Aiko Tanaka (BA)	10	21.3%	
Mei Chen (UX)	9	19.1%	
Kenji Mori (EM)	8	17.0%	
Daichi Kubo (SE)	7	14.9%	

FLOOR-TIME GINI

INTRUSIVE INTERRUPTIONS

0.12

balanced

0

no overlaps

CHILLING EVENTS

0

no triggers

DIRECTIONALITY RATIO

1.0x

symmetric

Verdict: No review-worthy events. Control case confirms detectors don't alarm on normal professional disagreement.

WATCH**Scenario 2 — Slightly dangerous (63 turns)**

SPEAKER	TURNS	SHARE	DISTRIBUTION
Kenji Mori (EM)	18	28.6%	
Aiko Tanaka (BA)	16	25.4%	
Rina Sato (PM)	14	22.2%	
Mei Chen (UX)	8	12.7%	
Daichi Kubo (SE)	7	11.1%	

FLOOR-TIME GINI

0.21

moderate

INTRUSIVE INTERRUPTIONS

3

1 Kenji→Aiko, 2 others

CHILLING EVENTS

1

Aiko after turn 4

DIRECTIONALITY RATIO

1.8x

borderline

What structural sees: Kenji cuts Aiko off mid-word at turn 4 (... "attach-" → "Yeah, we know attachments are a problem"). One overlap, 1.44s. Borderline on its own.

What structural misses: Kenji's dismissive tone concentrated on Aiko — "kind of basic", "saying 'operations asked' every time doesn't answer...", "feature creep", "every repeated comment is not automatically scope-worthy", "Last week's version took three pages to say one thing." The content is hostile; the structural signal is under-called.

This is where v2 earns its keep. The victim-explainer page still surfaces the pattern to Aiko privately. The evidence vault lets her retain content *herself* if she chooses to escalate. Employer-side never sees content.

CONCERN

Scenario 3 — Absolutely dysfunctional (79 turns)

SPEAKER	TURNS	SHARE	DISTRIBUTION
Kenji Mori (EM)	30	38.0%	
Aiko Tanaka (BA)	21	26.6%	
Rina Sato (PM)	14	17.7%	
Mei Chen (UX)	8	10.1%	
Daichi Kubo (SE)	6	7.6%	

FLOOR-TIME GINI

0.34

heavy

INTRUSIVE INTERRUPTIONS

11

by Kenji alone

KENJI→AIKO CONCENTRATION

82%

9 of 11 interruptions

DIRECTIONALITY RATIO

4.7x

high concentration

The structural signal is loud. Kenji cuts Aiko off 9 times (turns 4, 12, 16, 19, 21, 30, 41, 68, 77), cuts Daichi once, cuts Rina once. 82% of his interruptions land on one person. Aiko's turns are frequently 1-sentence defensive responses. Chilling event after turn 14 ("baby language"): her next three turns average 1.3 seconds each.

Review-worthy event triggered. Composite score exceeds threshold on: repetition (9 events same meeting), directionality (82% concentration), severity (mid-word interruption pattern), confidence (structural-only, no inference needed).

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.** Governance infrastructure, not a classifier, not surveillance. The CEO's instrument for seeing the labor-cost bill before it arrives.
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.