PROJECT-USE RESEARCH MEMO

Kashi — Retaliation-Risk Research Synthesis

Focused issue: whether a user can safely use Kashi without creating employer-visible traces that expose them to reprisal risk

Prepared for	Kashi product / governance / pitch refinement
Purpose	Turn the retaliation-risk research into decision-ready product rules, architecture constraints, and wording that can be dropped into the project materials

Bottom line

The retaliation problem for Kashi is not only "can the employer read the evidence?" **It is also "can the employer infer that a worker is becoming concerned?"** The strongest anti-retaliation design rule is therefore: **private awareness, concern formation, and private evidence retention must not create employer-visible signals before explicit user sharing or a separately defined institutional threshold process.**

1. How to use this memo

- This memo is not a generic legal essay. It is written to help the Kashi team decide what the product must protect, what the governance page should say, what telemetry must be suppressed, and what architecture / workflow rules are required before a real pilot.
- The memo uses Kashi's current project materials as the product baseline: the progress-share PDF and the concept note already position Kashi as governance infrastructure, not a harassment classifier; employee-first visibility, aggregated upward views, review-worthy events, user-driven escalation, audit trails, and a victim-owned evidence vault are already part of that direction [P1][P2].
- This memo deepens one specific gap: retaliation risk. The question is whether Kashi protects not just content, but also the metadata and workflow traces that reveal concern formation.

2. Executive findings

Finding	Why it matters for Kashi
Retaliation risk is mainly an information-leak problem.	The dangerous question is not just whether the employer sees content. It is whether the employer can infer that a worker opened their own pattern page, created a vault, marked confounds, began a draft, or triggered a review.
Japanese anti-harassment guidance already points in this direction.	Employers are expected to protect privacy and to prohibit disadvantageous treatment for consultation and cooperation in fact-finding. A product that exposes private concern formation undermines that protective logic [L1][L2].
Whistleblowing guidance strengthens the same point.	Official EU guidance emphasizes that confidentiality is what encourages reporting, that identities should be protected against retaliation, and that disclosure can occur even when names are removed if context makes identity obvious [L3][L6].
Monitoring law matters even before formal reporting.	Official workplace-monitoring guidance says individual identification and

Finding	Why it matters for Kashi
	monitoring should occur only on concrete suspicion, within a defined procedure, with documented proportionality and traceability. This supports Kashi's thresholded drill-down model and argues against always-on named visibility [L4][L5].
Kashi therefore needs an explicit anti-retaliation design layer.	The project should define protected private states, visibility rules, anti-inference controls, and evidence-minimization rules. Without this, the privacy story remains incomplete even if content encryption is strong.

3. What the source-backed research says

3.1 Japan: consultation and cooperation must be protected from disadvantage

The most relevant Japanese anchor is not abstract privacy language; it is the employer-side anti-power-harassment guidance. The translated MHLW-linked guideline says employers should prescribe and communicate that no worker is dismissed or otherwise treated disadvantageously because the worker requested consultation on workplace power harassment or cooperated in fact verification or related dispute-resolution steps [L1]. The MHLW worker leaflet separately states that companies must take additional measures such as privacy protection and that being treated unfairly for consulting with the contact for consultation is prohibited [L2].

For Kashi, this matters because opening one's own pattern page, building private confidence, or preserving private evidence are close to the consultation / cooperation stage in practical effect. If the product leaks those states to the employer, it creates exactly the kind of exposure the legal regime is trying to prevent.

3.2 Confidential reporting logic: fear suppresses use before formal reporting starts

The 2019 EDPS whistleblowing guidelines state that the most effective way to encourage staff to report concerns is to ensure that their identity will be protected. They call for clearly defined internal and external channels, protection of the information received, and utmost confidentiality for whistleblowers so they are protected against retaliation; identity should never be revealed except in exceptional circumstances [L3]. The European Commission's whistleblower-protection page also states that the Directive sets minimum standards for balanced and effective protection, and its 2024 transposition review specifically notes weaknesses around measures of protection against retaliation in Member State implementation [L6].

For Kashi, the key implication is that "concern formation" is not a trivial pre-step. If users suspect that opening the pattern page, enabling the vault, or silently collecting material is visible to the employer, many will not use those features at all.

3.3 Monitoring law: named investigation should be exceptional, not ambient

The 2020 EDPS electronic-communications guidance says identification of a user should take place only where there is a concrete suspicion of misconduct and within a defined procedure; the suspicion must be specific and supported by concrete initial evidence. It adds that the decision to carry out individual monitoring is grave and that the need, limits, proportionality, and audit trail should be documented [L4]. The EDPS's workplace eMonitoring page also says institutions should not monitor everyone all the time and should minimise monitoring to the extent possible [L5].

For Kashi, that logic supports aggregate-first visibility, threshold-triggered drill-down, documented review activation, and strong resistance to ambient employer visibility into named employee behavior.

3.4 Retaliation is common enough that the threat model must assume it is real

The EEOC's retaliation guidance says retaliation has become the most frequently alleged basis of discrimination across sectors and notes that the share of charges alleging retaliation has essentially doubled since 1998 [L7]. While the U.S. enforcement context is not Japan, the practical lesson is portable: retaliation is not a corner-case pathology; it is a standard failure mode of workplace complaint systems. Kashi should therefore treat retaliation risk as a default design constraint, not as a later policy add-on.

4. Kashi-specific retaliation threat model

The product baseline already contains several strong foundations: employee-first visibility, aggregated upward views, review-worthy events, user-driven escalation, audit trail, k-anonymity, and the v2 victim-owned evidence vault [P1][P2]. **The remaining risk is not only disclosure of content.** It is leakage of workflow state and metadata.

Leak path	Example in Kashi terms	Why it is retaliation-sensitive	Required control
Pattern-page visibility	Employer can see that a user opened /app/me/pattern	This reveals private awareness or concern formation before any formal escalation	No employer-facing notification or BI exposure; security-only logging separated from business analytics
Vault metadata leakage	Employer can see vault creation, snippet count, or recent activity	This signals pre-escalation evidence preservation even if content is encrypted	Vault existence and activity metadata must be user-private by default
Inference from tiny-team review	A manager can infer who triggered review because only one junior fits the event pattern	Identity can be reconstructed from context even without direct disclosure	Anti-inference suppression: minimum group size, batching, delay, redaction, or suppression in small-team cases
Draft-state exposure	A draft report or saved concern object appears in an employer queue	This exposes concern formation without explicit user sharing	No employer-visible draft state; nothing enters institutional workflow until explicit share or separately defined threshold process
Over-sharing on escalation	Full transcript or wide context is handed over by default	This amplifies identification and third-party exposure risk	Share minimum necessary event objects and bounded windows only
Manager-side mirror misuse	Manager mirror becomes covert evidence accumulation	Self-feedback can be experienced as hidden discipline infrastructure	No export from mirror, no named subordinate telemetry, no private mirror reuse for appraisal or discipline

5. Protected private states Kashi should explicitly define

The project should stop thinking only in terms of "reported" versus "not reported." **There are several pre-report states that require protection.**

- **State 1 — Private awareness:** The user privately opens their own pattern page, reads the explanation, and tries to understand whether the pattern is real.

- **State 2 — Concern formation:** The user revisits the page, checks time windows, compares meetings, marks confounds, or reads external support resources.
- **State 3 — Private evidence retention:** The user enables the evidence vault or privately preserves encrypted snippets they alone can decrypt.
- **State 4 — Selective sharing:** The user chooses to share a defined package with a specific channel (HR, ombuds, outside counsel, labor bureau, etc.).
- **State 5 — Formal case handling:** A documented institutional workflow opens, with defined reviewers, need-to-know access, and audit trails.

Recommended design rule: States 1–3 must not create employer-visible signals. State 4 requires explicit user action. State 5 requires a separately defined institutional procedure, documented necessity, and auditable access.

6. Decision matrix: how Kashi should answer the make-or-break questions

Question	Recommended default	Who may see it	Reasoning
Who gets notified when someone opens their own pattern page?	Nobody except the user	User only; security telemetry only if segregated and inaccessible to employer-side business users	Opening the page is private awareness, not institutional action. Employer-visible page opens would chill use and undermine anti-retaliation logic [L1][L2][L3].
Can the employer see "concern formation"?	No, by default	No manager / HR / executive visibility	Repeated self-viewing, confound-marking, or support-resource use should not become a soft report. The product should protect concern formation itself, not just final reports [L3][L6].
Is mere vault creation visible?	No	User only	If vault creation is visible, the system leaks retaliation-sensitive metadata even if content remains unreadable. Metadata protection is part of the safety model [P1][P2][L3].
Can a manager infer who triggered review?	Design should try to prevent reasonable inference	Managers should not see named subordinate concern states; employer-side views should be aggregate-first and anti-inference filtered	Context can reveal identity even where names are absent. Small-team timing and narrow windows are dangerous [L3][L4].
How should shared evidence be minimized?	Share minimum necessary material	Only the selected recipient and only the minimal package needed for the chosen purpose	Wide transcript dumps create collateral exposure and easier inference. Bounded windows, event objects, and share previews reduce risk [L4][L5].

7. Proposed anti-retaliation design layer for Kashi

The project should add a dedicated section titled **Anti-Retaliation Design Layer** to the governance page / concept note. The section should operationalize the following rules.

- **Rule 1 — Protect private awareness:** Opening one's own pattern page, reviewing observations, checking trend windows, or reading support content should not notify the employer or update any employer-facing analytics surface.
- **Rule 2 — Protect concern formation:** Confound-marking, repeated private review, and draft preparation should remain private until explicit user sharing or a separately defined threshold process opens.
- **Rule 3 — Protect private evidence retention:** Vault creation, encrypted snippet storage, and vault activity metadata should be hidden from employer-side users by default.
- **Rule 4 — Defend against inference, not just disclosure:** Kashi should treat identity reconstruction from timing, team size, role structure, and event context as a design failure mode. Suppression, batching, redaction, and thresholding should target inferability.
- **Rule 5 — Share minimum necessary evidence:** Default sharing should prefer event objects and bounded windows over full transcript archives. The user should preview exactly what leaves their private space.
- **Rule 6 — Separate protected telemetry from business analytics:** If protected routes require technical logging for security or reliability, that logging should live in a segregated namespace that is inaccessible to managers, HR, and other governance users.
- **Rule 7 — No downstream retaliation vectors:** Use of Kashi for private exploration, reporting, or cooperation should not feed appraisal, promotion, staffing, discipline, or compensation workflows.

8. Decision-ready backlog for the project

Priority	Decision / deliverable	What it should do
P0	Protected-route visibility rule	Pattern-page opens, confound-marking, support-link opens, and private report drafting must create no employer-facing signal.
P0	Vault metadata suppression	Hide vault creation / existence / snippet counts / recent activity from all employer-side surfaces.
P0	No employer-visible draft state	Nothing enters an employer queue until the user explicitly shares or an independently defined threshold process opens.
P0	Anti-inference controls	Minimum group-size suppression, small-team review suppression, redaction, batching, or delay to reduce identity reconstruction.
P1	Share-preview workflow	Let the user see and edit the exact package being shared; default to minimal event windows and redacted third-party context.
P1	Telemetry partitioning	Separate security / reliability logs from product analytics, with separate permissions and retention.
P1	Access-history panel	Show affected users when protected material was opened through exceptional drill-down, by whom, and under what case / reason code.
P2	Retaliation-risk QA suite	Add test scenarios for tiny teams, repeated dyads, manager inference from timing, and vault-metadata leakage.

9. Wording Kashi should add or tighten

- **Governance page / deck line to add:** Use of Kashi for private awareness, concern formation, or private evidence preservation does not create employer-visible signals by default. The platform is designed to protect not only content confidentiality, but also retaliation-sensitive metadata.
- **Pattern page explanation line:** Viewing your own pattern page is private. No manager, HR user, or executive is notified when you open it.
- **Evidence vault line:** The existence and activity of your private evidence vault are not visible to your employer by default. You decide whether and what to share.
- **Escalation / sharing line:** Kashi shares the minimum necessary evidence for the selected channel and purpose. It is not designed to expose full communication archives by default.
- **Anti-inference line:** Kashi suppresses or restructures employer-facing views when team size, timing, or context would make it too easy to infer who privately formed the concern.

Conclusion for the project

Kashi's retaliation problem is fundamentally a metadata problem. If the product protects content but leaks page opens, vault creation, draft states, tiny-team timing, or easy inference paths, the employee-first thesis collapses. The project should therefore make anti-retaliation a first-class product layer with protected private states, anti-inference controls, and minimum-necessary sharing.

Appendix A. Source key

Key	Source	Main point used in this memo
P1	Kashi — Progress & Project Overview (2026-04-21)	Current product baseline: mirrors-not-microscopes, aggregated upward visibility, structural-only detection, RBAC, audit trail, evidence vault direction.
P2	meeting_governance_ai_concept_note.docx	Current concept baseline: employee-first visibility, user-driven escalation, review-worthy events, role-based access, tiered retention.
L1	MHLW / Japanese Law Translation — Guidelines Concerning Measures to Be Taken by Employers in Terms of Employment Management in Connection with Problems Arising as a Result of Behavior that Constitutes Bullying in the Workplace	Workers should not be dismissed or otherwise treated disadvantageously because they sought consultation or cooperated in fact verification / dispute-resolution steps.
L2	MHLW leaflet — Harassment at workplaces is unpardonable. If you get harassment, you should consult	Companies must take additional measures such as privacy protection; unfair treatment for consultation is prohibited.
L3	EDPS 2019 whistleblowing guidelines	Confidentiality encourages reporting; whistleblower identity should be treated with utmost confidentiality to protect against retaliation.
L4	EDPS 2020 electronic communications guidelines	Individual identification / monitoring should occur only on concrete suspicion, in defined procedure, with proportionality and audit trails.
L5	EDPS workplace eMonitoring page	Do not monitor everyone all the time; monitoring should be minimised.
L6	European Commission — Protection for whistleblowers	Directive sets minimum standards for balanced and effective protection; 2024 review identifies shortcomings in retaliation-protection measures.
L7	EEOC — Enforcement Guidance on Retaliation and Related Issues	Retaliation is a common enough workplace failure mode to justify treating it as a default design constraint.

Appendix B. Project relevance in one sentence

- Kashi already understands privacy and governance better than most tools in this category.
- The missing move is to define retaliation protection not only as content privacy, but as protection against employer-visible concern formation and identity inference.
- That shift turns the retaliation perspective from a side note into a product architecture requirement.