PROJECT USE ONLY

Kashi Research Memo

Employee trust, bounded visibility, and anti-surveillance positioning

Evidence-grounded implications for product design, governance, messaging, and rollout

Prepared	21 April 2026
Prepared for	Kashi / 可視 project team
Purpose	Turn the trust / anti-surveillance research into project-usable judgments, product implications, and paste-ready wording.
Scope note	This is a product and governance memo, not legal advice. It synthesizes current Kashi materials plus external governance sources that are relevant to worker monitoring, AI governance, and contestability.

Core conclusion

Kashi is strongest when it is framed and built as worker-protective governance infrastructure: the institution is deliberately allowed to see less than the technology could show, while the affected individual is allowed to see more than they normally can and to control escalation. **The trust architecture is therefore not a side risk section. It is part of whether the product is legitimate, adoptable, and defensible at all.**

1. Executive judgment

The research result is not that Kashi merely needs kinder messaging. The result is that worker trust, constrained institutional visibility, contestability, and procedural access control are first-order product architecture. Kashi already has unusually strong instincts in this direction — structural-only analytics, no content-reading, no affect inference, no HR decisions from the tool, k-anonymized upward visibility, auditable drill-downs, and a victim-owned evidence vault concept [K1]. However, the current materials still understate how central these features are, and some lines — especially the more buyer-centric or CEO-centric framing — can weaken the employee-trust story if left unbalanced [K1][K2].

Judgment	What the evidence supports	Project meaning
Trust is a deployment condition	Worker monitoring is intrusive and must be justified, proportionate, and tightly governed [S1][S3].	The trust architecture belongs in the product thesis, not in a late-stage risk appendix.
Consent is weak in employment	Employers usually cannot rely on consent where power imbalance is real [S2].	Kashi should not sell its legitimacy story as 'employees agreed'.
Emotion AI is a red line	Emotion recognition in the workplace is prohibited in the EU except narrow medical/safety cases [S5][S6].	Kashi must stay structural, non-affective, and non-psychologizing.
Human review must be meaningful	Oversight, override logging, and review discipline matter; logging alone is not enough [S4][S5].	Every drill-down and every consequential output needs procedural accountability.
Worker consultation matters	OECD work shows consultation can mitigate risks and improve acceptance of workplace algorithms [S8][S9].	Rollout governance should include worker or representative consultation, not management-only deployment.
The buyer and the protected person are not the same	Institutions buy the tool, but workers bear much of the surveillance risk [K1][S8].	Kashi must deliberately withhold some power from the buyer to remain credible.
Kashi is directionally right but under-articulated	Current Kashi materials already encode strong anti-surveillance design moves [K1][K2].	The next step is not invention; it is sharpening, surfacing, and operationalizing what is already there.

2. What Kashi already gets right

Kashi already positions itself as governance infrastructure rather than as a harassment classifier or generic meeting productivity tool [K1][K2]. That is the correct conceptual lane. It directs attention toward longitudinal pattern visibility, accountable human review, and bounded outputs instead of pseudo-certainty about intent or legality.

The structural-only design is a major legitimacy asset. Kashi's current materials explicitly reject content surveillance, emotion inference, tone/voice stress analysis, facial analysis, keystroke logging, screen capture, browsing capture, and direct employer access to message body text [K1]. That is not only ethically cleaner. It also aligns the product with the strongest external legal and governance constraints that currently exist for workplace AI [S1][S5][S6].

The permission model is also directionally strong. Kashi already states: mirrors, not microscopes; primary visibility is personal; managers see their own behavior; upward visibility is aggregated and k-anonymized; audit trails are viewable by the affected individual; review-worthy events are retained separately from raw data; and the server-side evidence-vault concept is intentionally unreadable by the employer [K1]. Those are not cosmetic features. They are the beginning of a real anti-surveillance architecture.

Kashi also already refuses several poisoned feature paths that would make the product materially worse: binary abuse labels, performance scoring, promotion/discipline/compensation use, future-behavior prediction, company-wide relationship health bars, and employer-readable content analysis [K1]. This is strategically important because many adjacent products lost trust precisely by turning visibility into scorekeeping or managerial ranking.

3. Where the current story is still weak

- The current framing still risks sounding more buyer-driven than worker-protective. Kashi says it is 'the CEO's instrument for seeing the bill before it arrives' [K1]. That line is commercially sharp, but psychologically it can undo part of the trust story. A worker reading it can easily conclude: the real product is executive visibility, not worker protection.
- The phrase 'not an employee-monitoring tool' is directionally understandable but too absolute if left unqualified [K1]. From an employee's perspective, Kashi still ingests workplace interaction data and produces institutional signals. The safer claim is not that Kashi does no monitoring whatsoever, but that it is deliberately designed to minimize institutional visibility, reject content surveillance, and avoid conventional employee-monitoring logic.
- Contestability is present but not yet fully specified. Kashi has user-marked confounds, a private victim-explainer concept, and a worker-owned evidence vault [K1]. But the current materials still under-specify a full challenge path covering transcript accuracy, speaker attribution, context-window fairness, summary wording, thresholding, escalation disposition, and correction or suppression rights.
- Worker consultation is not yet prominent enough in the rollout model. The current materials are strong on technical governance and legal posture, but weaker on the socio-political governance of deployment. For a workplace-facing system, that is a real hole. Consultation is not simply a nice-to-have; it is one of the strongest available ways to reduce backlash, improve acceptability, and surface bad assumptions early [S8][S9].
- The exception path for deeper access is still more implicit than explicit. Kashi already has RBAC, audit trails, legal-hold logic, and investigator roles [K1][K2]. But it should say more bluntly that raw-context access is not a normal operating mode; it is an exceptional, procedurally justified path.

4. External research and governance findings

4.1 Worker monitoring and data protection

The ICO treats surveillance systems as intrinsically intrusive and expects necessity, proportionality, purpose limitation, data minimization, restricted access, and retention discipline [S1]. This matters even though ICO guidance is UK-specific, because it is one of the clearest operational benchmarks for turning abstract privacy principles into product design decisions.

The ICO's DPIA guidance states that a DPIA is required where processing is likely to present high risk to rights and freedoms, especially when using new technologies [S3]. For Kashi, the practical implication is simple: even where exact legal classification varies by jurisdiction, the team should behave as though high-scrutiny impact assessment is the default, not an optional extra.

The ICO also states that consent is usually inappropriate where there is a clear imbalance of power, and it identifies employers as a particularly important example [S2]. That point is directly relevant to Kashi because it means legitimacy cannot rest on a thin story of nominal worker agreement. The defensible basis has to come from bounded purpose, constrained access, transparency, and procedural safeguards.

4.2 EU AI Act and EU-facing governance constraints

The European Commission's current materials make two points especially relevant for Kashi. First, emotion recognition in the workplace is prohibited except for narrow medical or safety reasons [S5][S6]. Second, deployers of high-risk AI systems in the workplace must inform affected employees and workers' representatives beforehand, and affected persons can have a right to a clear and meaningful explanation when high-risk output is used in a decision producing legal effects [S5].

The exact legal classification of Kashi's final deployment posture still depends on how the product is marketed, configured, integrated, and actually used. The memo therefore does not claim that all Kashi deployments are automatically high-risk in the same way. What the evidence supports is a more practical rule: because Kashi operates in the employment context and changes what an institution can see about workers, the product should be designed and governed to high-risk discipline even where formal legal classification remains counsel-dependent.

The Commission's AI literacy materials also matter for deployment. They stress that deployers and providers must ensure staff understand how the system works, what risks exist, and how oversight is supposed to function [S7]. For Kashi, this implies that rollout cannot stop at giving administrators an interface. The institution must also train the humans who interpret, review, and act on the system.

4.3 OECD evidence on algorithmic management and consultation

OECD work on algorithmic management repeatedly points to both promise and risk: productivity and managerial consistency on one side, and digital-surveillance stress, unclear accountability, low explainability, and worker-health concerns on the other [S8]. This is directly relevant because Kashi sits adjacent to algorithmic management even if it is narrower and more governance-focused than many management tools.

OECD materials also say worker consultation can mitigate risks and boost engagement and acceptance, and later OECD work reports that worker, manager, and representative consultation can produce designs that stakeholders believe preserve productivity while improving job quality [S8][S9]. That finding is unusually useful for Kashi because it bridges ethics and buyer logic: consultation is not anti-business; it can improve adoption quality without throwing away employer value.

4.4 Japan-specific governance signals relevant to rollout

MHLW materials require employers to establish consultation systems, respond appropriately and promptly, protect privacy, and prohibit disadvantageous treatment of people who consult or cooperate in fact-finding [S10]. This matters because Kashi's escalation, access, and anti-retaliation posture should not merely sound globally thoughtful; it should align with the practical logic of Japanese workplace-harassment response obligations.

The Personal Information Protection Commission's guidance on 仮名加工情報 is relevant to Kashi's current privacy-design ambition [S12]. That does not settle all legal questions by itself, but it supports the broader internal project direction that analytics layers, role-based access, and pseudonymization discipline should be treated as substantive design work rather than as afterthoughts.

MHLW’s harassment survey materials provide part of the empirical problem context already used in Kashi’s project materials [S11][K1]. For project use, the important point is less the exact percentage itself and more the structural pattern: harassment is common, under-acted-on, and often not converted into timely institutional response.

5. Synthesis: what the evidence means for Kashi

The deepest synthesis is this: Kashi’s legitimacy comes from disciplined non-seeing. The product should win not by maximizing what management can know, but by deliberately limiting casual institutional sightlines while still creating enough traceable visibility for accountable action [K1][K2][S1].

That leads to a sharper project doctrine:

> **Kashi should be framed as worker-protective governance infrastructure built from constrained institutional visibility and worker-controlled escalation power.**

This is stronger than a generic 'AI for healthier meetings' story, and it is also more credible than a pure CEO-risk dashboard story. It explains why Kashi keeps structural-only analytics, why it refuses affect inference, why it refuses performance and discipline uses, why it limits upward visibility, and why the worker-owned evidence-vault concept is so important [K1].

It also clarifies a key tension: the institution buys the product, but the protected person is often the worker. That means Kashi cannot simply mirror the buyer’s maximum information appetite. If it does, it stops being accountability infrastructure and becomes a surveillance archive. The product therefore has to withhold some power from the buyer by design. That is not a flaw; it is the trust mechanism.

5.1 Proposed Kashi doctrine set

Doctrine	Why the evidence supports it	What Kashi already has	What still needs to be added or surfaced
Constrained visibility by default	Surveillance guidance favors least intrusive, purpose-limited, access-restricted systems [S1][S3].	K-anonymity, DP, aggregate upward views, no browsing of others' data [K1].	Say explicitly that raw-context access is an exception path requiring justification, logging, and review.
Employee-facing value first	Worker trust collapses if the product is mainly experienced as employer visibility [K2][S8].	Private dashboard, confounds, evidence-vault concept [K1].	Make this a front-stage pillar, not an implied feature.
Every consequential representation is challengeable	Meaningful human review and explanation require contestability beyond a final flag [S4][S5].	Human approval for events; explainability via turn IDs [K1].	Specify challenge rights for transcript, speaker attribution, context window, summary, thresholds, and disposition.
Logs must expose power, not just record it	Human-review guidance stresses override logging and	Audit trail viewable by affected individual [K1].	Add review process, abnormal-access alerts, reason codes,

Doctrine	Why the evidence supports it	What Kashi already has	What still needs to be added or surfaced
	accountability [S4].		and periodic audit of access behavior.
No affect, no psychologizing, no employer-readable content	EU materials make workplace emotion inference a legal red line; worker-monitoring concerns intensify when personal communications are mined [S5][S6][S8].	Kashi already refuses affect inference and employer-readable content [K1].	Keep the red line absolute in product scope, marketing, and architecture.
Consultation before rollout	OECD says consultation can improve acceptance and reduce risk; EU-facing materials can require informing workers and representatives for certain deployments [S5][S9].	Current materials hint at governance seriousness [K1].	Add a deployment requirement: consultation, notice, and anti-retaliation policy before pilot or rollout.

6. Project decisions that follow from the research

1. Decision 1 — Strengthen the framing, not by softening it but by making the asymmetry explicit. The product should state that institutional visibility is deliberately weaker, slower, narrower, and more procedural than the technology would technically permit. That is the cleanest way to differentiate Kashi from surveillance-adjacent products.
2. Decision 2 — Reframe the trust section as a product pillar. Do not leave anti-surveillance logic buried under 'risks and tensions'. Move it into the front-stage architecture and principles. A useful title would be: 'Worker trust, bounded visibility, and contestability.'
3. Decision 3 — Build a real contestability workflow. The victim-explainer concept is good, but it is still more recognition than full contestability. Kashi needs a way for users to mark confounds, dispute representation errors, request review, see access history, and understand outcomes.
4. Decision 4 — Treat rollout governance as part of the product. For pilots and enterprise deployment, prepare a basic governance pack: purpose statement, what Kashi will not do, retention model, visibility matrix, anti-retaliation statement, worker/representative notice template, access-review procedure, and reviewer training note.
5. Decision 5 — Keep the worker-owned evidence-vault direction and treat it as strategic, not optional. Of all current ideas, this is one of the strongest ways to prove that Kashi is not merely a better employer archive. It materially changes the power structure around preserved context.
6. Decision 6 — Do not expand into content surveillance or generic relationship scoring. The current refusal of a company-wide relationship-health bar is correct and should remain a competitive advantage, not a temporary omission [K1].

6.1 Immediate project backlog implied by this memo

Priority	Change	Why	Concrete deliverable
P0	Rewrite trust/anti-surveillance logic as a top-level principle set.	Current materials are substantively strong but rhetorically underpowered.	New deck/site copy; one dedicated governance slide; one principle block on the product page.
P0	Soften or rebalance buyer-centric phrasing such as 'CEO's instrument'.	Commercial clarity is useful, but it can weaken adoption narrative among workers.	Copy revision in deck, landing page, and governance page.
P0	Specify exception-path access rules.	Without explicit access doctrine, trust defaults to fear.	Reason codes, approval logic, worker-visible access log, access-review cadence.
P1	Design the contestability workflow.	Challenge rights need to extend beyond a final flag.	UX spec for dispute, annotate, request review, resolve, and suppress flows.
P1	Formalize pilot governance pack.	Deployment legitimacy cannot rely on ad hoc explanation.	PDF/slide one-pager for admins, workers, and works council or representative audience.
P1	Preserve evidence-vault roadmap priority.	This is one of Kashi's strongest anti-surveillance differentiators.	Threat model, UX flow, recovery model, and escalation sharing flow.
P2	Define trust metrics, not just detection metrics.	Success is not only detection accuracy; it is also whether the system is trusted and challengeable.	Metrics for dashboard use, confound usage, appeals, access-log review, and misuse incidents.

7. Claim discipline for Kashi

This section is included because a large part of project risk is not technical; it is representational. Certain claims are supportable now, some should be softened, and some should not be made at all.

Safe to say	Better if softened	Do not say
Kashi surfaces repeated structural interaction asymmetries for human review [K1].	Kashi is not an employee-monitoring tool.	Kashi detects harassment / intent / illegality.

Safe to say	Better if softened	Do not say
Kashi deliberately limits institutional visibility and rejects content surveillance [K1][S1].	Kashi is compliant by design with all relevant law.	Kashi proves abuse or identifies abusers.
Kashi uses deterministic, explainable structural signals and keeps AI in an assistive role [K1][K2].	Kashi eliminates retaliation risk.	Kashi causally reduces harassment.
Kashi does not infer emotions, affect, or voice stress [K1][S5][S6].	Kashi is purely for CEOs.	Employee consent is sufficient legitimacy for deployment.
Kashi is designed so that workers can see their own pattern context before institutions get investigative depth [proposed positioning].	Kashi is not surveillance at all.	A company-wide relationship-health score would measure workplace safety well.

8. Paste-ready wording for Kashi materials

8.1 A cleaner positioning paragraph

Kashi is governance infrastructure for surfacing repeated workplace power-asymmetry patterns without turning the workplace into a searchable surveillance archive. It uses structural interaction signals — not content-reading, not affect inference, and not productivity scoring — to make recurring suppression patterns visible enough for accountable human review. The design principle is deliberate bounded visibility: workers see their own pattern context first, while institutional users see aggregated or procedurally justified views only.

8.2 A cleaner trust / anti-surveillance paragraph

Employee trust is not a communications concern around Kashi; it is a product requirement. If workers believe the platform primarily expands management visibility, the system will be interpreted as surveillance infrastructure rather than accountability infrastructure. Kashi therefore limits casual institutional access, keeps drill-downs auditable, preserves challenge paths for consequential outputs, and rejects affect inference, content surveillance, and performance-use cases by design.

8.3 Principle block proposal

- Principle 01 — Mirrors before microscopes: The primary view is the individual's own pattern context. Managers see self-mirrors. Upward visibility begins in aggregate, not in named-individual browsing.
- Principle 02 — Patterns, not content, not affect: Kashi measures structural interaction signals and refuses employer-readable content analysis, emotion inference, and psychologizing.
- Principle 03 — Contestability, not capture: Every consequential representation must remain challengeable. Users can mark confounds, dispute errors, request review, and see when protected drill-downs occur.

- Principle 04 — No HR decisions from the tool: Kashi is not for performance scoring, promotion, discipline, or compensation. It is a governance input for accountable human review, not an automated personnel authority.

9. Open questions the project should resolve next

- How, exactly, will a worker challenge a structural event: transcript error, speaker error, context-window unfairness, summary wording, or threshold logic?
- What reason codes and approval logic will govern deeper drill-down into protected context?
- What is the minimum deployment governance pack required before any pilot: notice, consultation, training, policy sign-off, and audit cadence?
- How will Kashi distinguish legitimate chairing/facilitation from dominance without overfitting to one meeting culture?
- Which trust metrics will be tracked alongside detector metrics?
- How will the product phrase uncertainty without becoming mushy or useless?

Appendix A. Source register

The body text uses the source codes below. Internal project materials are marked K; external sources are marked S.

Code	Institution	Source	Use in this memo
K1	Internal project document	Kashi — Progress & Project Overview (2026-04-21)	Internal project PDF used for current product posture, architecture, refusals, and roadmap context.
K2	Internal concept note	Transparency That Drives Institutional Accountability (meeting_governance_ai_concept_note.docx)	Internal concept note used for generic concept framing and its earlier trust/governance language.
S1	ICO	Guidance on video surveillance / surveillance systems	https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/cctv-and-video-surveillance/guidance-on-video-surveillance-including-cctv/how-can-we-comply-with-the-data-protection-principles-when-using-surveillance-systems/
S2	ICO	When is consent appropriate?	https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/consent/when-is-consent-appropriate/
S3	ICO	When do we need to do a DPIA?	https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/data-protection-impact-assessments-dpias/when-do-we-need-to-do-a-dpia/
S4	ICO	Human review	https://ico.org.uk/for-organisations/advice-and-services/audits/data-protection-audit-framework/toolkits/artificial-intelligence/human-review/
S5	European Commission	Navigating the AI Act (FAQs)	https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act
S6	AI Act Service Desk / European Commission	Guidelines on prohibited AI practices / Article 5 materials	https://ai-act-service-desk.ec.europa.eu/en/faq
S7	European Commis	AI Literacy — Questions & Answers	https://digital-strategy.ec.europa.eu/en/faqs/ai-literacy-questions-answers

Code	Institution	Source	Use in this memo
	sion		
S8	OECD	Algorithmic management in the workplace (2025)	https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/algorithmic-management-in-the-workplace_3c84ed6d/287c13c4-en.pdf
S9	OECD	AI and work — consultation and adoption materials	https://www.oecd.org/en/topics/ai-and-work.html
S10	MHLW	職場におけるハラスメント防止のために	https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/koyou_roudou/koyoukintou/seisaku06/index.html
S11	MHLW	職場のハラスメントに関する実態調査（令和 5 年度）	https://www.mhlw.go.jp/content/11909000/001259093.pdf
S12	Personal Information Protection Commission (Japan)	仮名加工情報・匿名加工情報ガイドライン	https://www.ppc.go.jp/personalinfo/legal/guidelines_anonymous/

Important note: several cited governance pages explicitly state that some guidance is under review because of post-2025 legal changes or staggered AI Act implementation. This memo uses those materials as current design and governance benchmarks as of 21 April 2026, but any external-facing legal claims should still be checked with counsel before commercial deployment.

Appendix B. Final one-line project stance

Kashi should present itself as a system that makes institutions see enough to become accountable, while refusing to let them see so much that the system itself becomes the next source of workplace fear.