Kashi — Research Synthesis
Legal Defensibility, Procedural Fairness, and Governance Design

Project-use memo tailored to the current Kashi materials
Version: 2026-04-21

Prepared for	Kashi project team
Purpose	Convert the legal / procedural / governance research into something directly reusable for product design, deck revision, pilot design, and buyer conversations.
Primary question	What can Kashi defensibly claim, what should it avoid claiming, and what concrete governance architecture is required so the concept survives legal, trust, and adoption scrutiny?
Important note	This is a strategy and product-governance memo, not formal legal advice. It is designed to sharpen the product and reduce obvious legal / procedural self-owns before counsel review.

Bottom line

The research does not support Kashi positioning itself as an all-seeing workplace-truth engine. It does support a narrower and stronger claim: Kashi can be framed as a restricted evidentiary workflow for contestable structural interaction signals in recorded meetings. That framing is materially more defensible because it aligns with the strongest parts of the current Kashi build — deterministic structural detectors, review-worthy events rather than legal labels, no emotion inference, no named employee telemetry to managers, and human review before action. The main gap is not detector logic. The main gap is procedure: contest rights, correction workflow, context-window rules, retention under challenge, downstream-use controls, and post-deployment audit / rollback triggers must be specified as operating rules, not left as vibes. [R1][R2][R3][R4][R5][R8][R9][R13]

1. What this memo is for

This memo is not a generic compliance essay. It is a project memo tailored to the actual Kashi materials now in circulation: the current progress-share PDF and the earlier meeting-governance concept note. The goal is to convert the recent research thread into something the team can use immediately for product framing, governance design, investor / partner explanation, pilot scoping, and deck revision. [R1][R2]

Methodologically, the memo does three things at once. First, it extracts the strongest claims already present in the Kashi materials. Second, it stress-tests those claims against current official privacy / AI-governance sources. Third, it translates the result into concrete product and wording implications. This is important because the risk here is not merely that the model could be inaccurate. The bigger risk is that the product could create a procedurally unfair workplace mechanism even if the detectors are technically sound. [R3][R5][R8][R9][R10][R13]

2. Executive findings

- Kashi already has the correct strategic instinct: it is strongest when positioned as governance infrastructure for visible, contestable, pattern-level review objects — not as a harassment classifier, emotion-AI system, or performance-scoring tool. [R1][R2][R3][R4]
- The most defensible part of the current Kashi architecture is the structural-only path: deterministic metrics, baseline-relative analysis, no emotion inference, no employer-facing content classification, and human review before action. [R1][R3][R5]
- The main unresolved weakness is procedural, not algorithmic. The current materials explain what Kashi detects and what it refuses to do, but they still under-specify what happens after a signal exists: who can contest it, how transcript / diarization errors are handled, how much context is shown, when raw data is deleted or preserved, what reviewers must do, and how false positives are audited. [R1][R2][R8][R9][R11][R13]
- The phrase 'not an employee-monitoring tool' is too aggressive and can be used against Kashi. A narrower and stronger description is that Kashi is not a general surveillance or performance-monitoring system; it is a restricted meeting-governance system that processes recorded meeting data under strict procedural and visibility limits. [R1][R8]
- The current deck has a wording contradiction around transcripts and content. It says Kashi uses meeting transcripts, while also saying 'never transcribe for analysis' and 'patterns, not content'. The usable truth is narrower: Kashi ingests transcript-linked meeting records, but employer-facing detection is restricted to structural interaction metadata rather than semantic content classification. That distinction should be stated explicitly. [R1]
- The employee-first trust story and the 'CEO instrument' buyer story are not naturally aligned. Kashi therefore needs an explicit two-lane architecture: a private employee lane and a safeguarded institutional lane. Without this, the product can be attacked as a surveillance tool wearing ethical clothing. [R1][R2][R8]
- If Kashi or a customer deployment drifts into decision support for employment outcomes, the EU AI Act risk profile gets significantly harsher. Employment / worker-management use cases are in Annex III, and deployers of high-risk systems face human-oversight, logging, worker-information, and other obligations. Even where Kashi tries to remain outside that category, the design should assume that functional drift is a real risk. [R4][R5][R6][R7]
- Human review only counts if it is meaningful. Reviewers must have competence, authority, independence, and the ability to reject the system output. Rubber-stamping does not solve the legal or fairness problem; it just launders it. [R9][R10]
- Kashi needs a named post-deployment quality regime. Monitoring, appeals, override, incident response, rollback, and decommissioning are not optional polish. They are part of a serious AI-governance operating model. [R13]

3. What the research supports Kashi claiming with high confidence

The table below distinguishes between claims the research supports, claims that are only partially supportable, and claims that should be removed or rewritten. This is the cleanest way to prevent the deck from over-claiming and getting legally or politically shredded.

Claim type	Claim	Assessment	Project implication
Keep	Kashi surfaces repeated interaction asymmetries as explainable signals for human review.	Strong	This is the core defensible claim. It matches both the current detector design and the official governance logic around transparency, human review, and monitoring. [R1][R2][R9][R11][R13]
Keep	Kashi does not infer emotion, affect, tone, or intent.	Strong	This should remain a hard red line. Workplace emotion inference is explicitly prohibited under the EU AI Act. [R1][R3]
Keep	Kashi should not be used for performance, promotion, discipline, or compensation decisions.	Strong	This is already in the deck and should be elevated from ethics language to use-restriction architecture and contract language. [R1][R4]
Rewrite	Kashi is not an employee-monitoring tool.	Weak / misleading	Too broad. Kashi still processes workplace interaction data. Reframe as 'not a general surveillance or performance-monitoring system'. [R1][R8]
Rewrite	Never transcribe for analysis.	Internally inconsistent	The build uses transcript-linked meeting data. What is true is that employer-facing detection does not classify semantic content. [R1]
Remove or narrow	Kashi is defensible under the EU AI Act because it reduces to counted events from timestamps.	Overconfident	Helpful instinct, but too absolute. Defensibility depends not only on signal type, but on use context, worker information, human oversight, logs, and downstream action. [R4][R5][R6][R7]
Add	Kashi is a controlled evidentiary workflow for contestable structural interaction signals in recorded meetings.	Recommended	This is the best project-level replacement framing found in the research. It is ugly, but strong. [R1][R2][R5][R8][R9]

4. Where the current Kashi materials are already strong

This matters because the research does not require a total conceptual rewrite. Quite the opposite: some of Kashi's strongest design choices are already in place. The problem is that they are not yet tied together into a rigorous procedural architecture. [R1][R2]

4.1 Structural-only detector logic is the right default.

The current progress-share PDF is unusually disciplined in drawing a boundary around what the system measures. It foregrounds deterministic Layer 1 and Layer 2 features; makes clear that the shipped detectors are structural, explainable, and computed from timing / speaker-attribution signals; and explicitly refuses affect, voice stress, facial-expression, keystroke, and employer-facing content-surveillance routes. This is exactly the right instinct. It keeps Kashi away from the worst consumer-surveillance patterns and away from the legally radioactive territory of workplace emotion inference. [R1][R3]

4.2 'Review-worthy event' is a much better object than 'problematic statement'.

The concept note's move from moral labels to review-worthy events is not cosmetic. It is foundational. It lowers overclaim risk, leaves room for transcript error and ambiguity, and is compatible with contestability, human review, and bounded retention. It is one of the clearest signs that Kashi should evolve as a workflow system rather than a truth-judgment system. [R2][R9][R11]

4.3 Role-based visibility is already a core thesis, not an afterthought.

The concept note and the progress-share PDF both repeatedly insist that universal transcript browsing is unacceptable, that employees should see their own data, managers should primarily see their own behavior or aggregate views, and that HR / investigators require procedural justification for drill-down. This is exactly where a trust-preserving governance product must start. [R1][R2]

4.4 Baseline-relative analysis is a strong mitigation against easy false positives.

The progress-share PDF is right to emphasize own-baseline comparison instead of team-average comparison. This directly addresses predictable confounds such as introversion, L2 status, chair role, or local meeting culture. It does not solve fairness by itself, but it is the correct direction and should remain central to the detector story. [R1]

4.5 The refusal of the company-wide 'relationship health' score is strategically correct.

This is one of the most mature decisions in the current materials. A visible headline health score would invite gaming, flatten contested realities into a single number, and create a performative compliance target. The current refusal is therefore not a missed feature; it is a governance asset. [R1]

5. Critical gaps and contradictions that still need fixing

5.1 The product still lacks a procedural-fairness backbone.

The deck explains what Kashi detects and what Kashi refuses. It still does not sufficiently explain what happens after a signal exists. That is the hardest part. In practice, legal defensibility is won or lost in procedure: who can contest, what counts as enough context, which reviewer can override what, how corrections are logged, what survives deletion, and what downstream uses are blocked. The current materials are philosophy-forward but procedure-light. [R1][R2][R9][R11][R13]

5.2 'Not monitoring' is too absolute and therefore fragile.

If a product processes recorded meetings over 30 / 90 / 180-day windows to surface repeated patterns about worker interactions, many audiences will still experience it as a form of monitoring. The real win is not denying that fact; it is proving that the monitoring is narrow, proportionate, purpose-limited, and visibly fenced away from surveillance and performance management. [R1][R8]

5.3 The transcript/content wording is internally contradictory.

Kashi currently says it takes meeting transcripts as input and also says 'never transcribe for analysis'. Those cannot both be read naively as true. This is fixable. The intended meaning appears to be that the organization may provide transcript-linked meeting records, but the employer-facing detector path is restricted to structural interaction metadata rather than semantic content reading. State that directly. [R1]

5.4 Employee trust and CEO-facing authority are in tension.

The project currently says 'mirrors, not microscopes' and also says 'Kashi is the CEO's instrument for seeing the bill before it arrives'. Both statements can be made individually. Together, they create a trust problem unless the system architecture visibly separates private employee visibility from institutional escalation logic. If this separation remains implicit, critics will read the product as surveillance dressed up as empowerment. [R1][R2][R8]

5.5 Manager Mirror is strategically attractive but politically explosive.

The most commercially distinctive Kashi feature — giving powerholders a private mirror of their own interaction patterns — is also the easiest one for managers to interpret as hidden accusation or pre-disciplinary shadow scoring. This means the feature needs stronger procedural guardrails than the current materials provide. [R1]

5.6 The current materials under-specify what happens when the system is wrong.

This is not a minor omission. Transcript error, speaker-attribution error, incomplete context, unusual meeting format, language asymmetry, or legitimate facilitation behavior are not edge cases; they are expected deployment conditions. A product that lacks a structured error and contest path invites both unfairness and adoption failure. [R1][R2][R9][R11][R13]

6. Required procedural-fairness architecture for Kashi

This section is the most important part of the memo. If Kashi does not formalize the following as system behavior, governance posture alone will not be enough.

6.1 A contestability state machine

- Every review-worthy event should move through explicit states: detected → disputed (accuracy) / disputed (speaker) / disputed (context) → under human review → upheld / downgraded / withdrawn → preserved under case hold if needed.
- The rule should be harsh and simple: once disputed, the event stops behaving like settled truth. It must either be excluded from escalation scoring or clearly labeled and down-weighted until review is complete.
- Version history must be immutable: original signal, correction request, reviewer action, timestamp, rationale. No silent overwrite. [R9][R10][R11]

6.2 A bounded-context rule

- A single sentence is procedurally unfair; a full transcript is privacy overexposure. Kashi needs a middle layer.

- Default review should show only a bounded interaction window: for example, a short turn window or time window around the trigger, plus speaker order, interruption timing, and other immediately relevant interaction metadata.
- Expanded context should require role entitlement plus procedural justification, with a complete drill-down audit trail. [R2][R8][R11]

6.3 A retention-under-challenge rule

- The current four-tier retention model is directionally strong, but it needs challenge logic.
- Raw content should auto-delete on the default schedule unless there is an active dispute, case hold, or user-owned encrypted preservation event.
- System logs, event objects, and raw content should be treated as distinct layers with distinct retention reasons.
- Any retention extension should have an owner, reason, timestamp, and expiry review point. [R1][R2][R5][R13]

6.4 A meaningful human-review standard

- Reviewers must have competence, training, authority, and independence to challenge the system output. That is not optional garnish; it is what makes the product 'decision support' rather than pseudo-automated judgment. [R9][R10]
- Kashi should define what a reviewer must check at minimum: signal basis, context window, confounds, transcript quality, repeated-pattern history, and any user challenge or annotation.
- Reviewer performance must itself be auditable. If reviewers merely rubber-stamp outputs, Kashi has a governance failure even if the detectors are technically sound. [R9][R13]

6.5 A downstream-use charter

- Kashi already says 'no HR decisions from the tool'. That should be elevated into a formal use-control layer.
- The system should explicitly forbid use as sole or primary basis for performance assessment, promotion, discipline, compensation, ranking, staffing allocation, or termination.
- The system should also prohibit enterprise-wide health grades, named employee telemetry to managers, and silent re-identification through repeated small-team drill-down.
- These are not merely contract terms. They should be reflected in product behavior, permission design, exports, and auditability. [R1][R4][R5]

6.6 A post-deployment quality and rollback regime

- Kashi needs named measurement and rollback rules, not vague 'we monitor quality'.
- At minimum, the operational quality set should include precision after human review, reversal rate after contest, speaker-attribution error rate, language / meeting-type failure clusters, reviewer disagreement rate, time to resolution, and the proportion of flags that resulted in no action but still burdened participants.
- Kashi also needs rollback triggers: if a rule family overfires, if reversal rates spike, if one meeting type is too noisy, or if a team pattern is systematically over-policed, that detector or deployment mode should be suppressed, narrowed, or reworked.
- This is fully aligned with NIST’s expectation of post-deployment monitoring, appeals / override, incident response, and decommissioning when systems exceed acceptable risk tolerances. [R13]

7. What the current regulatory research means for Kashi specifically

7.1 EU AI Act implications

The most relevant legal lesson is not that the EU AI Act automatically kills the concept. It does not. The lesson is that Kashi must remain disciplined about what the system is and is not used for. The Act explicitly prohibits AI systems that infer emotions in the workplace. Kashi’s no-affect stance therefore remains a hard legal boundary, not just a branding choice. [R3]

At the same time, Annex III and Article 26 matter because AI used in employment or worker-management settings can trigger a significantly heavier compliance posture. Article 26 requires human oversight, monitoring of operation, retention of automatically generated logs for at least six months where under deployer control, and worker / worker-representative information before use of a high-risk system at work. Article 86 also creates a right to clear and meaningful explanation for certain significantly affecting decisions based on high-risk AI outputs. [R4][R5][R6][R7]

The practical implication for Kashi is not 'claim non-high-risk and forget the problem'. The practical implication is 'design as if functional drift toward employment decision support is a real threat and fence it off aggressively'.

7.2 Privacy / worker-monitoring implications

ICO guidance is useful here because it states the obvious without euphemism: worker monitoring can be intrusive, but may be justified if there is a lawful basis and a proportionate reason. It also repeatedly stresses that meaningful human review matters and that human rubber-stamping does not remove the problem. [R8][R9][R10][R11]

This pushes Kashi toward a specific operational posture: narrow purpose, narrow data surface, bounded visibility, auditable drill-down, and strong contest rights. The project should not argue that the system is harmless because it is structural. It should argue that the system is controlled because its purpose, visibility, review, and downstream use are tightly constrained. [R8][R9][R10][R11]

7.3 Consent / lawful-basis caution

The research also supports a caution against leaning too hard on individual worker consent in employment-style contexts. European data-protection guidance repeatedly notes that clear power imbalance makes freely given consent difficult in employer-employee relationships, and ICO employment guidance makes the same point in practical terms. [R14]

For Kashi, the strategic implication is not to build the product story around naive 'everyone opted in, so we are safe' logic. The safer enterprise posture is necessity, proportionality, transparency, worker information, and strict use limitation — with user control features layered on top where appropriate.

7.4 APPI / Japanese data-governance implications

The current Kashi materials are directionally right to treat pattern metadata as personal information when linkable and to think in terms of pseudonymization / limited internal analysis rather than casual free use. PPC materials on pseudonymously processed information are especially relevant because they emphasize use within specified purposes, non-identification, restrictions on third-party provision, and deletion when the information has become unnecessary. [R1][R12]

That does not by itself solve the Kashi problem. But it does support the product direction of narrow internal use, strict role-based access, and separation between user-owned encrypted evidence and employer-facing structural analytics.

8. P0 / P1 project actions

Below is the shortest practical translation of the research into project work.

Priority	Action	Why it matters	Owner shape
P0	Replace 'not an employee-monitoring tool' with a narrower framing.	Stops an easy credibility attack and aligns the language with how regulators	Deck / landing / governance page

		and buyers will actually read the product. [R1][R8]	
P0	Fix the transcript/content wording contradiction.	Prevents the project from sounding evasive or incoherent about what data path is actually used. [R1]	Deck / product copy
P0	Write a formal Procedural Fairness & Decision-Use Controls section.	This is the missing spine. Without it, the project remains conceptually smart but operationally under-specified. [R1][R2][R9][R13]	Core concept doc
P0	Define the dispute / correction state machine.	Essential for transcript error, diarization error, confounds, and reviewer accountability. [R9][R11]	Product + governance spec
P0	Define bounded-context rules and drill-down justification.	Prevents both quote-sniping and universal transcript browsing. [R2][R8][R11]	Product + access-control spec
P1	Formalize Manager Mirror guardrails.	Protects the most distinctive feature from being interpreted as hidden discipline infrastructure. [R1]	Product + sales positioning
P1	Define post-deployment metrics and rollback triggers.	Needed for pilots, credibility, and eventual enterprise review. [R13]	Pilot / MLOps / governance
P1	Separate employer-facing structural analytics from user-owned evidence vault in all wording.	Preserves the core trust story and prevents content-surveillance confusion. [R1]	v2 architecture / copy
P1	Prepare a buyer-facing deployment checklist (worker info, reviewer training, legal review, retention config).	Makes Kashi look operationally serious rather than merely conceptually interesting. [R5][R6][R9][R13]	Sales + implementation

9. Recommended replacement language for the project materials

9.1 One-sentence category description

Recommended replacement:
Kashi is a restricted meeting-governance system that surfaces contestable structural interaction signals from recorded meetings so that repeated workplace asymmetries can be reviewed, challenged, and handled through auditable human procedures.

9.2 What Kashi should say about data

Recommended replacement:
Kashi ingests transcript-linked meeting records for turn structure, timestamps, and speaker attribution. Employer-facing detection is limited to structural interaction metadata rather than semantic content classification. Where user-owned evidentiary preservation is enabled, content snippets remain encrypted for the affected user rather than becoming employer-browsable analysis material.

9.3 What Kashi should say about decisions

Recommended replacement:
Kashi is not an adjudicator and not an employment decision engine. It creates review-worthy event objects that support human review under strict use limits. Kashi outputs must not be used as the sole or primary basis for performance, promotion, discipline, compensation, or termination decisions.

9.4 What Kashi should say about trust

Recommended replacement:
Kashi separates private employee visibility from institutional escalation. The default state is self-visibility and aggregate visibility, not universal visibility. Any drill-down beyond that boundary must be role-limited, justified, and auditable.

10. Open questions that the team should answer before treating the concept as governance-ready

- Exactly which events may create institutional awareness without immediate employee escalation, and under what threshold logic?
- What minimum context window is sufficient for procedural fairness while still respecting third-party privacy?
- What raw-content retention period is workable in practice once challenge handling and evidence-vault logic are considered together?
- Which reviewer role owns final override authority in a pilot: HR, compliance, ombuds, or a mixed panel?
- What is Kashi's position when a manager disputes a signal about themselves but the affected employee does not want immediate escalation?
- How will Kashi communicate uncertainty in the UI without making the system unreadable or useless?
- What deployment checklist is mandatory before a customer can switch on any institutional lane at all?

11. Final judgment

The core research result is not that Kashi should retreat. The result is that Kashi should become stricter. The project is already strongest where it refuses the obvious bad routes: content surveillance, emotion inference, headline organizational health scores, and HR-decision automation. The research supports doubling down on that restraint.

What Kashi still needs is procedural architecture. The project must stop sounding like 'AI sees the pattern and then the organization figures it out'. That gap is exactly where legal fragility, employee distrust, and rollout politics accumulate. Kashi should instead sound and behave like a tightly governed workflow system for contestable pattern evidence in recorded meetings — narrow in scope, explicit in rights, and auditable at every boundary. [R1][R2][R3][R5][R8][R9][R13]

That version is less glamorous than a bold 'AI detects harassment' narrative. It is also far more credible, more deployable, and more likely to survive contact with real organizations.

Appendix A — Reference base used in this memo

The memo uses the following source set. Internal project materials are listed first; official regulatory / governance sources follow.

[R1] Kashi — Progress & Project Overview (2026-04-21). Internal project PDF provided in this conversation.

[R2] meeting_governance_ai_concept_note.docx. Internal concept note provided in this conversation.

[R3] EU AI Act Article 5 (prohibited AI practices): workplace emotion inference prohibition.

[R4] EU AI Act Annex III and Article 26: employment / worker-management relevance, deployer obligations, logs, worker information, human oversight.

[R5] EU AI Act Article 26 (full deployer obligations text) and Article 86 (right to explanation for certain high-impact decisions).

[R6] EU AI Act Recital 92: workers and worker representatives should be informed about planned deployment of high-risk AI systems at work.

[R7] EU AI Act Article 27 / related recitals: fundamental-rights impact assessment obligations in specified high-risk contexts.

[R8] UK ICO guidance on employee monitoring: monitoring may be intrusive, requires lawful basis and justification, and should respect staff privacy.

[R9] UK ICO guidance on human review in AI systems: review must be meaningful; reviewers require competence, authority, and independence.

[R10] UK ICO guidance on AI and data protection / individual rights: rubber-stamping does not turn an automated process into meaningful human decision-making.

[R11] UK ICO guidance on explanations for AI-assisted decisions: rationale, responsibility, data, fairness, safety / performance, and impact explanations.

[R12] Personal Information Protection Commission (Japan) materials on pseudonymously processed information under APPI.

[R13] NIST AI RMF / Playbook: post-deployment monitoring, appeal / override, incident response, decommissioning, acceptable-performance limits.

[R14] EDPB / ICO guidance on power imbalance and the weakness of consent in employment-style settings.

Appendix B — Optional live links for external official sources

- EU AI Act Article 5
- EU AI Act Annex III
- EU AI Act Article 26
- EU AI Act Article 27
- EU AI Act Article 86
- EU AI Act Recital 92
- ICO — Employee monitoring
- ICO — Human review
- ICO — How do we ensure individual rights in our AI systems?
- ICO — What goes into an explanation?
- PPC Japan — APPI English PDF
- NIST AI RMF Playbook — Manage
- NIST AI RMF Playbook — Measure
- EDPB Guidelines 05/2020 on consent