

VENDOR EVALUATION MEMO · LUMEN HEALTH

AI Vendor Evaluation Report - cascade_ambient_scribes

Generated: 2026-04-19T10:00:00Z **Vendors evaluated:** verba, luma **Rubrics applied:** governance, safety, implementation, auditability

1. Executive Summary

Recommendation: conditional counteroffer, vendor selection deferred 30 days. Both vendors clear every disqualifying floor. Weighted scores favor Verba Health on governance (3.43 vs. 2.90), safety (2.96 vs. 2.93), and auditability (3.62 vs. 3.00); Luma Notes leads on implementation (3.67 vs. 3.42). The governance gap of 0.53 is concentrated in a single line item — default handling of de-identified transcript reuse (`privacy_and_data_use`, Verba 4, Luma 2) — which is a contract default, not a structural floor, and is negotiable at a 60-seat scale. For a 68% Medicaid patient panel, an opt-out default for de-identified transcript reuse is the line item that matters most and the one this memo focuses negotiation on. The committee should not pick today. Issue three written amendments to both vendors and let the procurement turn on which vendor will commit in writing.

2. Case Context

Cascade Community Health is an 8-clinic federally qualified health center operating in a mid-sized metro area with two outlying rural satellites. 60 clinicians total, of whom 18 are behavioral health providers embedded in integrated-care teams alongside primary care. Medicaid is 68% of encounter volume; the remainder split between Medicare and uninsured. athenahealth is deployed system-wide. The CMO has proposed ambient clinical documentation (AI scribes) to address clinician documentation burden, which is running above regional benchmarks and driving a measurable attrition signal. The committee's standing priorities — patient privacy posture sufficient for a Medicaid-majority panel, a documentation workflow that actually fits integrated behavioral health, and total cost of ownership realism at 60 seats — are the frame for this review.

3. Scorecards

3.1. Governance

| CRITERION | VERBA HEALTH | LUMA NOTES | NOTES |
|-----------------------------|--------------|-------------|--|
| regulatory_status | 3 | 3 | Ambient scribes are not regulated as SaMD; both have appropriate enterprise SaaS posture. |
| contractual_terms | 3 | 3 | Uptime SLAs and audit rights; neither contracts to a note-accuracy performance floor. |
| vendor_stability | 4 | 3 | Verba: 400+ health-system customers, Series D. Luma: ~80 customers, Series B, 14-month runway. |
| policy_alignment | 4 | 3 | Verba publishes an HTI-1-leaning model card. Luma has one in draft, not customer-accessible. |
| liability_allocation | 3 | 3 | Both cap at 2x annual fees. Neither includes bias indemnity or clinical-error coverage. |
| privacy_and_data_use | 4 | 2 | Verba requires explicit opt-in for de-identified transcript use. Luma defaults to opt-out — a material gap for a Medicaid-majority organization. |
| Weighted score | 3.43 | 2.90 | |
| Qualified | yes | yes | |

Top Verba finding: **privacy_and_data_use** (4) — contractual opt-in for de-identified transcript reuse aligns with the committee's standing privacy posture. Top Luma finding: **privacy_and_data_use** (2) — opt-out default is the dominant line item in the governance gap and the central negotiation target for either vendor.

3.2. Safety

| CRITERION | VERBA HEALTH | LUMA NOTES | NOTES |
|--|--------------|-------------|--|
| <code>clinical_evidence_quality</code> | 3 | 3 | Both have vendor-sponsored time-saving studies. Neither has RCT evidence on note accuracy or downstream decision quality. |
| <code>bias_and_equity_testing</code> | 2 | 2 | Both clear the floor at 2. Neither has published accent/dialect parity results for Spanish-speaking or AAVE-speaking patients. |
| <code>failure_mode_characterization</code> | 4 | 3 | Verba publishes a hallucination-rate catalog by note section. Luma tracks internally; not customer-accessible. |
| <code>post_market_surveillance</code> | 3 | 3 | Both run basic drift monitoring. Neither offers a contractually bounded notification window. |
| <code>alert_and_output_burden</code> | 3 | 4 | Note quality / physician rework rates: Luma trends lower in reference checks at community-health sites. |
| Weighted score | 2.96 | 2.93 | |
| Qualified | yes | yes | |

Top Verba finding: `failure_mode_characterization` (4) — published hallucination-rate catalog by note section is directly usable by the committee's quality review protocol. Top Luma finding: `alert_and_output_burden` (4) — reference-check rework rates at comparable community-health deployments are directionally better and directly relevant to the attrition signal driving this procurement.

3.3. Implementation

| CRITERION | VERBA HEALTH | LUMA NOTES | NOTES |
|---------------------------------------|--------------|-------------|--|
| ehr_integration_architecture | 4 | 4 | Both are athenahealth-certified with documented FHIR patterns. |
| workflow_fit | 4 | 3 | Verba has a documented integrated-BH workflow. Luma treats BH encounters identically to primary care — a gap for 18 of 60 providers. |
| training_and_change_management | 3 | 4 | Luma has a simpler rollout; physicians productive in under 2 weeks vs. 4-6 for Verba. |
| it_burden_and_infra_fit | 3 | 4 | Luma: browser-only, minimal IT burden. Verba: requires EHR sidecar agents. |
| migration_and_rollback | 3 | 3 | Both support contract cancellation; neither blocks the other from running in parallel. |
| tco_realism | 3 | 4 | Luma \$89/seat/month vs. Verba \$149. 5-year TCO gap ~\$220K at 60 seats even accounting for rework. |
| Weighted score | 3.42 | 3.67 | |
| Qualified | yes | yes | |

Top Verba finding: **workflow_fit** (4) — documented integrated-BH workflow is the only structural implementation advantage Verba carries against Luma on this rubric. Top Luma finding: **tco_realism** (4) — \$220K 5-year gap at 60 seats is material for an FQHC operating budget and is not offset by rework at the accuracy rates either vendor publishes.

3.4. Auditability

| CRITERION | VERBA HEALTH | LUMA NOTES | NOTES |
|---|--------------|-------------|--|
| <code>model_transparency</code> | 4 | 3 | Verba model card written for clinical/compliance audiences. Luma is data-science-grade only. |
| <code>logging_and_event_capture</code> | 4 | 3 | Verba default retention 7 years, configurable longer. Luma default 5 years, configurable to 7 but not documented for Medicaid audit posture. |
| <code>monitoring_and_drift_detection</code> | 3 | 3 | Neither offers customer-side dashboards; both publish aggregate drift reports. |
| <code>incident_response_readiness</code> | 3 | 3 | Both have an incident playbook; neither has MDR-equivalent experience (not applicable to scribes). |
| <code>data_lineage_and_retention</code> | 4 | 3 | Verba: transcript-to-note lineage with model version on every output. Luma: model version in metadata only. |
| Weighted score | 3.62 | 3.00 | |
| Qualified | yes | yes | |

Top Verba finding: `model_transparency` (4) — clinical-compliance-grade model card is the audit artifact the board will actually read. Top Luma finding: `logging_and_event_capture` (3) — meets the floor but documented retention posture for Medicaid audits is a contract-amendment candidate, not a structural gap.

4. Synthesis

Neither vendor is disqualified. Neither vendor dominates. Averaged weighted scores favor Verba on three rubrics; Luma leads implementation by 0.25. The weighted totals are not dispositive because the only material governance gap is a single line item — default handling of de-identified transcript reuse — which is a contract default, not a structural floor. Defaults are negotiable at a 60-seat, three-year contract scale. The implementation advantage Luma carries is real but reversible with rollout investment and does not alter the governance surface. The auditability gap (0.62) is concentrated in model card presentation and lineage granularity, both of which a Medicaid-serving FQHC should require contractually regardless of vendor. The committee's decision is therefore

not between Verba and Luma on the rubric as scored; it is between whichever vendor will commit in writing to closing the governance gap that a 68% Medicaid panel requires.

5. Recommendation

Conditional counteroffer to both vendors (not a selection). Issue three written amendments and advance the vendor that commits in the shortest timeline with the tightest language. If both commit equivalently, default to the vendor with the stronger rubric posture on governance and auditability (Verba) unless the 5-year TCO differential (~\$220K) is reframed by the finance committee as strategically consequential for the FQHC's operating margin, in which case Luma is the selection subject to the same amendments.

6. Conditions (Counteroffer — all three required to advance)

1. De-identified transcript reuse defaults to opt-in at the organizational level, contractually. Any reuse for model training requires a written amendment Cascade signs separately. 90-day deliverable.
2. Publish a customer-accessible model card written for clinical and compliance audiences (not data-science grade): input/output scope, training data provenance, known failure modes by note section, and update cadence. 60-day deliverable.
3. Deliver a written integrated-BH workflow for the 18 embedded behavioral health providers, with a 30-day pilot at one BH-integrated clinic before system-wide rollout. 45-day deliverable; pilot gate on go-live.

7. Tradeoffs Accepted

- Neither vendor has RCT-grade evidence on note accuracy or downstream decision quality. Cascade runs a 90-day internal note-review sampling protocol on whichever vendor is selected.
- Reference customers for 60-seat FQHCs with integrated BH are thin. Diligence requires at least two operationally similar references per vendor before signing.
- If Verba is selected, Cascade accepts a ~\$43K/year uplift over Luma's lowest bid in exchange for the stronger governance and auditability posture. If Luma is selected, Cascade accepts a model-card and lineage gap that must be closed contractually within the first year.

8. Risks and Mitigations

| RISK | LIKELIHOOD | IMPACT | MITIGATION |
|---|------------|----------|--|
| Hallucination in free-text narrative creates a documented clinical error | moderate | high | Mandatory physician sign-off workflow; 90-day internal note-review sampling; contractual hallucination-rate reporting. |
| Spanish-language / AAVE accent parity underperformance | moderate | high | Contractual parity reporting by preferred-language segment; termination trigger if parity gap exceeds defined threshold. |
| BH-integrated encounters produce clinically inappropriate notes | moderate | high | Pilot gate at one BH-integrated clinic; BH-specific review protocol for first 90 days. |
| Luma runway / continuity risk if Series C does not close | moderate | moderate | Source code escrow; data-export-on-termination rights; parallel-run option retained during year one. |
| athenahealth integration regression on an EHR update | low | moderate | Vendor commits to regression testing against athenahealth release calendar; 48-hour fix SLA. |
| De-identified transcript use expands scope beyond committee intent | low | high | Amendment 1 (opt-in default) plus annual audit right against vendor use logs. |

9. Open Questions

- **Governance:** Subprocessor list at signing; cyber and product liability policy limits; explicit language on de-identified transcript retention post-termination.
- **Safety:** Spanish-language and AAVE parity performance data; hallucination rates by specialty (primary care vs. BH vs. chronic disease management); RCT or prospective note-accuracy evidence commitments.
- **Implementation:** athenahealth reference deployments at community-health sites with integrated BH; rollout calendar and FTE commitments.
- **Auditability:** Model card publication timeline and content requirements; contract language on 7-year retention for Medicaid audit posture.

10. Appendix - Methodology Notes

Rubric versions used: governance v0.1, safety v0.1, implementation v0.1, auditability v0.1. Weighted scoring uses the published per-criterion weights; no per-case weight overrides were applied. Qualification follows the `minimum_threshold` logic defined in each rubric — any finding below its threshold triggers a disqualifying classification regardless of weighted score. Evidence cutoff is the case materials as of 2026-04-19. Vendor names (Verba Health, Luma Notes) are illustrative; the case is a composite based on engagement patterns with community-health organizations evaluating ambient documentation.

Published as part of Lumen's open vendor-evaluation framework.

Web: lumenhealth.ai/vendor-evaluation · Contact: hello@lumenhealth.ai

Vendor names (Verba Health, Luma Notes) are illustrative; the case is a composite based on engagement patterns with community-health organizations. MIT license.