AI Voice and Workflow Companies in Healthcare: A Diligence Framework for Growth Investors

Underwriting Clinical Fit, Data Rights, Model Performance, Integration Burden and Customer Economics

Authored by Chennakeshav Adya, Independent Researcher

August 2026

Abstract

Healthcare AI voice and workflow companies are moving from transcription into ambient documentation, coding support, inbox management, order preparation, prior-authorisation workflows, patient communication and operational orchestration. The category can create genuine value by reducing documentation burden and improving the flow of information. It can also conceal material risk. A polished demonstration may depend on ideal acoustics, narrow specialties, permissive data access, extensive human correction, subsidised implementation or workflow assumptions that do not survive production deployment.

This paper develops a growth-investor diligence framework for healthcare AI voice and workflow companies. It connects intended use, workflow fit, data permissions, model architecture, clinical evaluation, human oversight, privacy, cybersecurity, regulatory classification, interoperability, implementation cost, customer cohort economics, revenue quality, gross margin, intellectual property, organisational capability, valuation and transaction protection. Five original figures and five implementation tables translate the method into a workflow map, data-rights ledger, model-evaluation scorecard, integration-cost curve and cohort-economics bridge.

The framework uses primary and authoritative sources as control boundaries. WHO guidance identifies governance, safety, privacy, bias, automation and human-rights considerations for AI and large multi-modal models in health.[1][2] FDA guidance distinguishes non-device clinical decision support from device functions and addresses lifecycle, change control and cybersecurity for regulated software.[3][4][5] ONC's HTI-1 rule establishes algorithm-transparency requirements for predictive decision support in certified health IT.[6] HHS guidance makes contractual and security responsibilities for protected health information relevant to cloud and vendor relationships.[7][8] The EU AI Act, European Health Data Space, UK medical-software guidance and NHS clinical-risk standards create additional jurisdiction-specific workstreams.[9][10][11][12] NIST and IMDRF provide risk-management and clinical-evaluation disciplines.[13][14]

Recent clinical evidence supports a measured approach. A 2025 rapid review found sparse real-world evidence, variable effects and gaps against standardised evaluation frameworks.[15] A pragmatic randomised trial across 238 outpatient physicians provides stronger evidence on documentation-time and user outcomes while retaining application-specific results and limitations.[16] Other studies show that performance, edit burden and productivity effects can differ by setting and comparator.[17][18] Every adoption rate, price, cost, retention figure, time saving, gross margin, market share, valuation multiple and investment return used below is a hypothetical modelling assumption for demonstrating the framework. It is not a forecast, benchmark, clinical recommendation or investment conclusion.

JEL Classification: G24, G32, I11, L86, O33

Keywords: healthcare AI, ambient AI, AI scribe, clinical workflow, growth equity, technology diligence, data rights, model evaluation, healthcare software, cohort economics

1. Define the investment thesis around a bounded job to be done

Healthcare AI voice and workflow companies should be underwritten around a bounded job to be done. The investor should state which user performs which task, in which care setting, using which input, producing which output, with which review obligation and measurable economic result. “AI for healthcare” is too broad to support diligence. Ambient documentation for outpatient clinicians, autonomous patient calls, inpatient handover summaries, coding suggestions and prior-authorisation preparation occupy different clinical, regulatory and economic positions.

The product perimeter should distinguish capture, transcription, extraction, summarisation, recommendation, action preparation and action execution. Capture records or streams speech. Transcription converts audio to text. Extraction identifies specified facts. Summarisation compresses information. Recommendation proposes a judgement or next step. Action preparation creates an order, message, claim or task for review. Action execution changes a record, initiates communication or affects care. Risk rises as the product moves from representation toward judgement and execution.

The thesis should identify the buyer, budget, beneficiary and veto holder. A chief medical information officer may sponsor the clinical case. Operations may own rollout. Security and privacy teams may approve data flows. Finance may require a productivity or retention case. Clinicians and patients determine practical adoption. Procurement may impose integration, insurance and contractual conditions. A company can have enthusiastic users and a blocked enterprise sale when these roles are not aligned.

The diligence question is therefore causal: does the product create a repeatable improvement in a high-value workflow, under the controls required by customers, at a deployment and support cost that permits durable unit economics? The answer should be tested through workflow evidence, production performance, rights, cohort data and customer references. Market enthusiasm, demonstration quality and model sophistication remain supporting context.

Table 1. Product-perimeter and intended-use map

Product function	Primary output	Accountable reviewer	Principal diligence question
ambient documentation	draft encounter note	treating clinician	does the draft reduce total work while preserving accuracy?
coding support	suggested code or evidence	qualified coding or clinical owner	does the suggestion improve compliant yield without upcoding?
inbox workflow	prioritised message and draft response	authorised care-team member	are urgency, routing and escalation reliable?
order preparation	draft order or referral	authorised clinician	can the user review the basis before action?
patient voice agent	conversation and task	designated clinical or service owner	are identity, consent, boundaries and escalation controlled?
prior-authorisation workflow	evidence packet and status	revenue-cycle owner	does it reduce cycle time and rework across payer rules?

Classification and obligations require jurisdiction-specific legal and regulatory advice.

2. Map the complete workflow before evaluating the model

Model accuracy has meaning only inside a workflow. Diligence should begin with direct observation, screen evidence, system logs and user interviews that reconstruct the current and proposed process. The map starts before the model receives data and ends after the final user action, downstream reconciliation and exception handling. It should show identities, systems, queues, handoffs, review points, wait states and failure recovery.

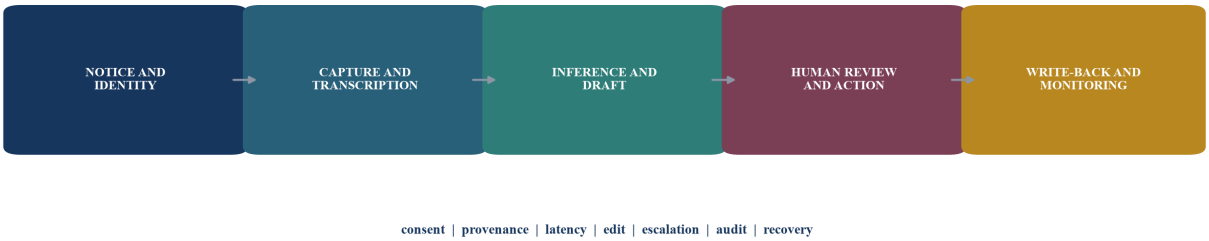
For ambient documentation, the workflow can include patient notification or consent, microphone selection, session start, speaker separation, audio transmission, transcription, structured extraction, note generation, clinician review, edit, attestation, coding, record write-back and deletion or retention. Each stage can fail independently. A strong transcript can still produce a poor note. A good note can still increase work when EHR navigation, template selection, attribution or signing is awkward.

The map should record time and quality for both the old and new process. Time saved during dictation may be offset by editing, searching, workflow interruptions or after-hours corrections. A product may transfer work from clinicians to medical assistants, coders, support teams or implementation staff. The investor should quantify total system work and identify who absorbs it.

Workflow fit also varies by specialty and encounter. A short primary-care visit, a paediatric consultation with several speakers, a mental-health encounter, an emergency department visit and a surgical consent discussion have different vocabulary, turn-taking, privacy and documentation structures. The diligence sample must represent the commercial expansion plan rather than the easiest current use case.

Figure 1. Healthcare AI voice and workflow map

THE PRODUCT IS THE COMPLETE CONTROLLED WORKFLOW



The map identifies the full control chain; actual workflows vary by product, care setting and jurisdiction.

3. Build the evidence room around production reality

The diligence room should be organised around claims that matter to customers and investors. For each claim, the company should provide the definition, measurement population, period, comparator, source system, exclusions, statistical method, owner and reproducible output. Evidence should distinguish internal tests, customer pilots, production deployments, peer-reviewed studies and marketing estimates.

Core materials include product specifications, intended-use statements, model cards, validation plans, test datasets, release histories, incident logs, clinical-safety files, privacy assessments, security reports, architecture diagrams, data-flow maps, customer contracts, business-associate agreements, subprocessor lists, implementation plans, support tickets, uptime records, EHR integration evidence, cohort revenue, invoices, cloud bills and customer-reference permissions. The request list should connect each document to an investment question.

Production logs deserve special attention. They can show encounter attempts, successful captures, abandonment, latency, output generation, review time, edit distance, acceptance, write-back, error, override and support events. The investigator should reconcile event definitions across versions. A metric may improve because a failed or difficult encounter disappeared from the denominator. A product release may change logging before it changes performance.

The evidence room should preserve negative evidence. Failed pilots, low-adoption cohorts, safety concerns, data-rights restrictions, unprofitable integrations and customer churn can reveal the limits of repeatability. Management should explain whether the issue arose from product capability, workflow design, customer readiness, pricing, implementation, governance or competition. The investor can then test whether the current plan addresses the cause.

Table 2. Evidence inventory for growth-investor diligence

Evidence domain	Minimum evidence	Reconciliation test	Decision use
product	specifications, versions, release notes	claims match shipped functions	scope and roadmap risk
model	datasets, evaluations, failures, monitoring	test population matches use	performance confidence
workflow	logs, observations, user references	total work and exceptions reconcile	adoption and value
rights	contracts, notices, consents, licences	every use has an identified basis	data and IP risk
deployment	statements of work, tickets, integrations	effort matches reported cost	margin and scalability
commercial	contracts, invoices, cohorts, churn	bookings bridge to cash and retention	revenue quality

Required depth should reflect the product's intended use, customer footprint and transaction materiality.

4. Create a data-rights ledger for every processing purpose

A healthcare AI company can possess data without holding every right required by its product and growth plan. Diligence should separate the right to receive data from the right to process it for a specific purpose. Recording, transcription, note generation, storage, quality assurance, support, model evaluation, product improvement, model training, benchmarking and customer export can require different contractual, privacy, consent and regulatory analyses.

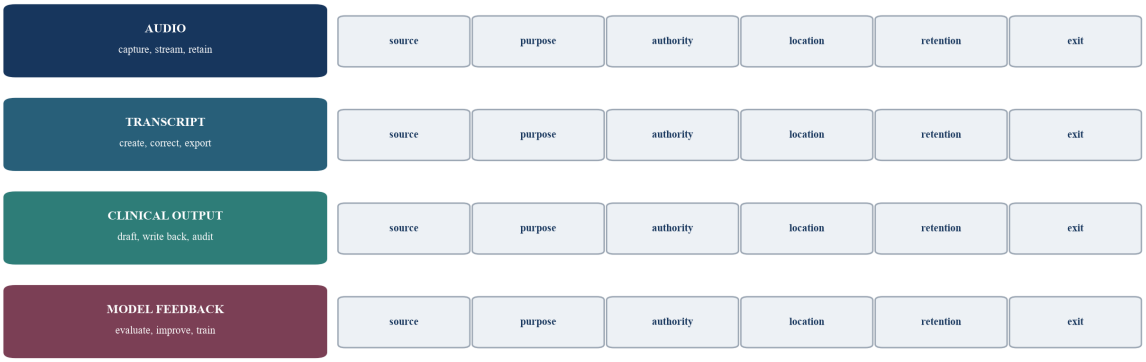
The ledger should identify data category, subject, source, controller or covered entity, processor or business associate, subprocessor, location, purpose, legal or contractual basis, retention, deletion, export, training use, audit right and incident responsibility. Audio deserves separate treatment from transcript and generated text. Audio can contain voices of patients, relatives, clinicians and bystanders. Derived features, embeddings, labels and model feedback may retain sensitivity or contractual restrictions.

HHS states that covered entities using business associates require written arrangements specifying permitted functions and safeguards for protected health information.[7] HHS cloud guidance also makes a business-associate agreement relevant when a cloud service creates, receives, maintains or transmits electronic protected health information on behalf of a covered entity or business associate.[8] Products outside HIPAA's covered-entity and business-associate perimeter may still face other privacy and breach obligations. The FTC's Health Breach Notification Rule is one US example for covered personal-health-record vendors and related entities.[19]

Rights should be tested through termination and change of control. The investor should determine whether a customer can revoke improvement rights, require deletion, prohibit model training, restrict cross-border processing or demand migration assistance. The company should demonstrate how data lineage supports those obligations. Contract language and system capability must agree.

Figure 2. Data-rights ledger across the healthcare AI lifecycle

DATA RIGHTS ARE PURPOSE-SPECIFIC AND MUST SURVIVE CHANGE



Each processing purpose requires its own evidence; the figure does not determine a legal basis.

5. Understand the model and software supply chain

The product may combine speech recognition, speaker diarisation, language detection, specialty vocabularies, retrieval, large language models, rules, templates, coding engines and workflow orchestration. Diligence should identify which components are proprietary, licensed, open source, customer-specific or provided by a third party. The architecture should show where each component runs, which data it receives and which output it controls.

Dependency analysis should cover foundation-model terms, speech-model terms, hosting concentration, rate limits, regional availability, price changes, deprecation, safety filters and rights to generated output. A company can own the interface and workflow logic while relying on a small number of upstream providers for core inference. This may still be an attractive model, but the investor should price substitution cost and negotiating leverage.

The model register should record purpose, version, input, output, training or adaptation method, evaluation population, known limitations, approved use, monitoring and release authority. Deterministic rules, statistical models and generative systems should remain distinct because their errors and test methods differ. A single product version may contain several model versions and prompt or retrieval configurations.

Change control is central. FDA's guidance on predetermined change control plans and lifecycle management shows why planned modifications, validation and monitoring matter for AI-enabled device functions.[4][5] The precise regulatory application depends on intended use and jurisdiction. Even when the software is outside device regulation, customers still need evidence that material changes are evaluated before release and that regressions can be detected, contained and reversed.

The investor should test the emergency substitution path. Can the company change an upstream model without losing function, rights, latency, geography or unit economics? Does it have regression tests and customer-notification rules? Architecture optionality is valuable only when it

has been exercised or evidenced.

6. Evaluate performance at the workflow decision unit

Aggregate transcription accuracy is an incomplete measure. The investor should test the unit that affects workflow and care: a medication, allergy, symptom, diagnosis, duration, negation, speaker attribution, task, code, order element, urgency classification or patient instruction. A note can read fluently while omitting a clinically material fact or assigning it to the wrong speaker.

The evaluation plan should define intended use, population, setting, reference standard, adjudication, metrics, thresholds and failure response. It should include known positives, known negatives, difficult edge cases and out-of-distribution conditions. Samples should cover specialties, encounter types, accents, languages, noise, telehealth, interruptions, several speakers and relevant patient groups. The commercial roadmap should determine the test matrix.

Performance dimensions include factual completeness, factual correctness, unsupported content, negation, temporality, speaker attribution, structure, action extraction, latency and abstention. Workflow dimensions include review time, edit distance, acceptance, override, abandonment, support use and downstream correction. Safety dimensions focus on clinically material errors and whether controls detect them before action.

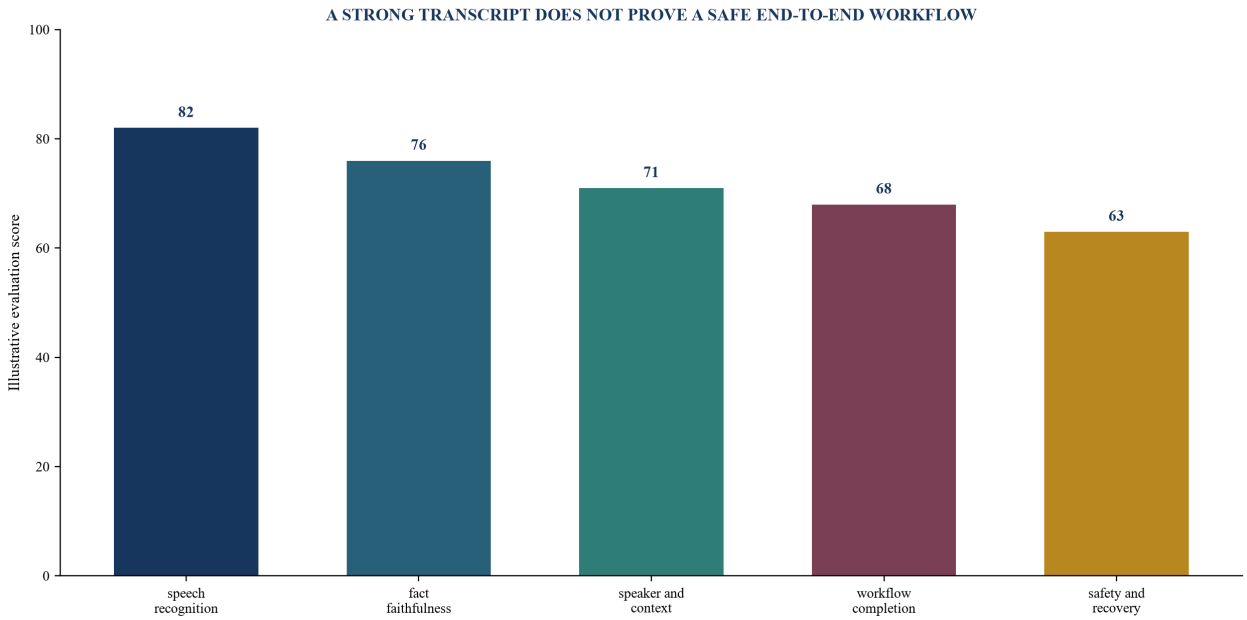
The company should preserve versioned results. A global score can improve while a material subgroup regresses. Thresholds should therefore be set by use case and harm, not by marketing convenience. Evaluation data should be sufficiently independent from development data, and customer-specific tuning should be separated from core-product performance.

Table 3. Model-evaluation design by decision unit

Decision unit	Example failure	Evaluation measure	Required review
medication	name, dose or negation error	exact and clinically weighted error	clinician or qualified reviewer
speaker	symptom assigned to wrong person	attribution precision and recall	encounter-level adjudication
summary fact	omitted or unsupported statement	factual completeness and support	blinded reference review
task	missed urgency or wrong routing	sensitivity, specificity and escalation	operational and clinical owner
workflow action	unauthorised or incorrect write-back	controlled-action success and recovery	safety and system owner

Metrics and thresholds require product-specific clinical and statistical design.

Figure 3. Model-evaluation scorecard from speech to safe workflow action



Scores are illustrative dimensions; investors should use product-specific metrics and thresholds.

7. Test clinical safety and human oversight as operating systems

Human oversight should be specified as an operating control. “Human in the loop” does not explain who reviews, what they see, how long they have, which errors they can detect, what happens under workload pressure or how the system responds when review fails. Diligence should map the accountable role, review interface, escalation route and evidence retained for each material output.

The review design should expose source context and uncertainty where useful. A user may need the relevant transcript span, speaker, timestamp, confidence, supporting record and reason for a suggested action. A long fluent draft can make errors harder to find. Interface design, alert burden and time pressure therefore affect the practical reliability of human oversight.

WHO guidance on large multi-modal models highlights risks including inaccurate outputs, bias, automation bias, privacy and cybersecurity.[2] NHS DCB0129 requires clinical-risk management for health-IT manufacturers, while DCB0160 addresses deployment and use by care organisations.[12] These standards illustrate the shared responsibility between product and adopter. The company should show a hazard log, clinical-safety case or equivalent risk file appropriate to its markets and use.

Incident management should connect product telemetry, customer reports, clinical review, engineering investigation and executive decision rights. Severity definitions should capture potential harm and near misses, not only confirmed injury. Material incidents should trigger containment, customer communication, root-cause analysis, corrective action, regression testing and monitoring.

The investor should test the degraded mode. If transcription, inference, EHR connectivity or identity resolution fails, can the user continue care safely? The workflow needs clear status, no silent write-back, retry rules and a manual route. A safe failure path can matter more than average model performance.

8. Examine subgroup performance, language and accessibility

Voice products operate across accents, languages, code-switching, speech impairments, age groups, microphone quality and clinical environments. An aggregate result can conceal weak performance in commercially important or vulnerable populations. Diligence should align subgroup testing with the company's deployed footprint and expansion plan.

The company should explain how it defines groups, obtains lawful evaluation data, establishes reference truth and avoids unstable conclusions from small samples. Performance differences should be assessed at clinically and operationally meaningful units. A modest word-error difference may become material when it affects medication names, negation, numbers or urgent symptoms.

Language support should distinguish user-interface translation, speech recognition, clinical vocabulary, note generation and workflow execution. A product may recognise conversational speech in a language while lacking validated specialty terms, templates or downstream EHR fields. Mixed-language encounters require separate evidence. Marketing labels such as multilingual should be decomposed into tested functions and settings.

Accessibility includes hearing, speech, cognitive, visual and motor needs as well as digital literacy. Patient-facing voice agents should handle repetition, interruption, uncertainty, interpreter involvement and transfer to a human. They should avoid creating a barrier for people who cannot or do not wish to use voice automation.

WHO's broader AI governance principles place inclusiveness and equity alongside autonomy, safety, transparency, responsibility and sustainability.[1] The investor should translate those principles into versioned subgroup evaluation, customer deployment criteria, monitoring and remediation. The commercial model should include the cost of supporting the populations it promises to serve.

9. Audit privacy, consent, security and resilience

The data-flow map should follow audio, transcript, generated content, identifiers, metadata, logs, support access, backups and model feedback across devices, networks, clouds, subprocessors and customer systems. Each transfer needs an owner, purpose, protection, retention rule and incident route. Security certificates and penetration tests are useful evidence, but they do not replace an architecture-specific threat model.

Voice creates particular risks. Always-on or ambient capture can include unintended speakers and sensitive discussion. Session start and stop controls, visible status, consent or notice, microphone permissions, local buffering and deletion behaviour require testing. The product should prevent one patient's content from entering another record and should surface identity uncertainty before write-back.

Security diligence should cover authentication, role access, least privilege, encryption, key management, secrets, tenant isolation, software supply chain, vulnerability management, secure development, logging, anomaly detection, backup, recovery and customer notification. Generative systems add prompt injection, data leakage, malicious content, retrieval poisoning and model-service dependency. NIST's AI RMF and Generative AI Profile provide a structured risk-management reference.[13]

For regulated device functions, FDA's 2026 cybersecurity guidance addresses design, quality-system and submission considerations.[5] Application again depends on the product and intended use. All companies should demonstrate resilience through tested recovery objectives, dependency failures and incident exercises.

Contractual promises should reconcile with operations. If the agreement promises regional processing, rapid deletion, subprocessor notice or a recovery objective, the company should show system evidence. A gap between sales language and production capability can become a renewal, liability and valuation risk.

10. Classify regulatory exposure by function and jurisdiction

Regulatory classification follows intended use and function, not the company's preferred category. A documentation tool that records and drafts a note can occupy a different position from software that recommends diagnosis, prioritises urgent symptoms or executes an order. Product modules and claims should therefore be classified separately.

In the United States, FDA's January 2026 clinical-decision-support guidance explains criteria for certain non-device CDS functions and gives examples distinguishing non-device from device software.[3] FDA's other digital-health guidance addresses AI-enabled device software, change control and cybersecurity.[4][5] ONC's HTI-1 rule adds algorithm-transparency requirements for predictive decision support supplied through certified health IT.[6] The diligence team should map whether the company is a certified health-IT developer, integrates with one or supplies functionality that triggers contractual flow-down.

In the European Union, the AI Act creates obligations based on system role and risk classification, including high-risk categories connected with regulated products and specified uses.[9] The European Health Data Space adds staged requirements for electronic health data and EHR systems while preserving the application of GDPR, medical-device and AI rules.[10] The effective dates and product impact require current jurisdiction-specific analysis.

In the United Kingdom, MHRA guidance and its software and AI as a medical device programme frame product classification and lifecycle expectations.[11] NHS deployments may also require clinical-risk, data-protection, technical-security, interoperability and usability evidence through DCB standards and assessment criteria.[12]

The regulatory matrix should record module, claim, user, action, jurisdiction, classification view, evidence, registration or certification, post-market duties and change consequence. Management's conclusion should be supported by qualified advice where material. A label such as administrative software does not settle the analysis.

11. Underwrite integration architecture and interoperability

Integration is part of the product. Enterprise value can depend on reliable identity, context launch, encounter matching, schedule access, documentation write-back, task routing, coding, consent, audit and analytics across customer systems. A vendor with strong models and fragile integration can face long sales cycles, high services cost and variable outcomes.

The architecture review should identify each interface, protocol, authentication method, data object, write permission, polling or event mechanism, latency, retry, error queue, observability

and support owner. Standards such as HL7 FHIR can improve portability, but customer implementation, versions, profiles and proprietary constraints still matter.[20] A claimed integration should be evidenced in production rather than inferred from an API specification.

The investor should distinguish reusable connectors from customer-specific work. Reusable configuration can support scale. Bespoke transformations, template mapping, workflow redesign, security review and testing can create recurring labour even when the technical interface is standard. Implementation data should show hours by role, elapsed time, dependency waits, defects and post-go-live support.

Write-back deserves a separate control path. The system should confirm patient, encounter, document type, author, version and final approval. It should prevent duplicate or stale writes and preserve an audit trail. Failed writes need a visible queue and accountable recovery. Clinical content should not disappear into a generic integration log.

The diligence model should value interoperability through evidence: shorter time to live, lower implementation cost, fewer defects, higher adoption, lower support load and easier expansion. Connector count alone can overstate readiness when integrations are old, lightly used or maintained manually.

12. Convert implementation burden into a total cost to live

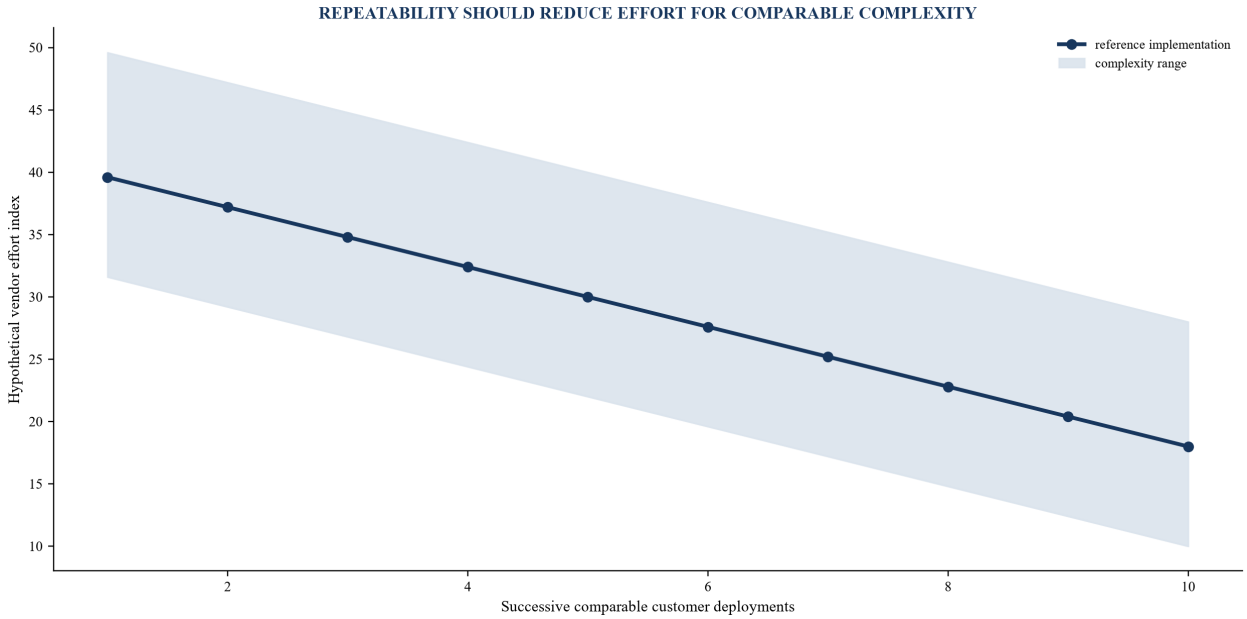
Customers buy a working change, not an isolated model. The total cost to live includes security and legal review, clinical governance, procurement, integration, configuration, workflow design, training, communications, support, measurement and dual running. These costs can sit with the vendor, customer, implementation partner or EHR provider. The investor should capture the complete system cost and identify who pays.

Implementation should be decomposed into repeatable stages: qualification, contracting, data and security approval, technical build, configuration, clinical-safety review, pilot, training, go-live, monitoring and scale. For each stage, the company should provide median and distribution of elapsed time, internal hours, customer hours, blockers, rework and abandonment. Averages can conceal a long tail that consumes cash and leadership attention.

The cost curve should separate fixed platform investment, reusable product work and customer-specific labour. Early cohorts may be expensive while the company learns. Later cohorts should show evidence of declining effort for comparable complexity. Growth created by adding implementation headcount at the same rate as revenue has different economics from product-led repeatability.

Customer-side burden affects adoption and renewal even when it does not appear in vendor cost of revenue. Clinicians may spend time reviewing, correcting and learning. Informatics teams may maintain templates and governance. IT teams may operate interfaces. The customer reference should test whether realised benefit exceeds this burden.

Figure 4. Hypothetical integration cost curve by customer complexity



Values are hypothetical; live diligence requires customer-level hours, cash costs and elapsed-time evidence.

Table 4. Integration responsibility and cost matrix

Workstream	Vendor responsibility	Customer responsibility	Evidence for repeatability
security and privacy	architecture, controls, responses	risk review and approval	standard pack and cycle time
interface	connector and monitoring	access, testing and change control	reusable code and defect history
workflow	product configuration	operating design and ownership	configuration range and adoption
clinical safety	product hazards and controls	deployment hazards and governance	approved files and incidents
training and rollout	materials, enablement, support	attendance, local champions, policy	time to competent use
value measurement	telemetry and definitions	baseline, comparator, outcome review	reconciled benefit evidence

The matrix should be completed for each material customer segment and EHR environment.

13. Rebuild customer economics from cohorts and usage

Contracted annual value does not prove durable economics. The investor should build cohorts by start period, segment, product, geography, EHR, deployment complexity and contract type. Each cohort should show contracted value, live value, usage, expansion, contraction, churn, implementation cost, support cost, cloud and model cost, gross margin, cash collection and customer concentration.

The bridge begins with eligible users or encounters, activated users, active use, successful workflow completion, reviewed output and sustained use. Seats can overstate value when usage is low. Encounter pricing can overstate predictability when volume or specialty mix changes. Enterprise pricing can conceal unprofitable heavy users. The commercial model should reconcile pricing unit to value and cost drivers.

Retention should be measured on several bases. Logo retention captures relationship continuity. Gross revenue retention excludes expansion. Net revenue retention includes expansion and contraction. Usage retention shows whether the product remains embedded. Margin retention tests whether renewed revenue remains economical after inference, support and integration. Definitions and currency treatment should be consistent.

Expansion requires causal evidence. A customer may add clinicians because the pilot worked, because a separate budget opened or because the vendor discounted an enterprise conversion. Expansion into coding, inbox or patient voice can have a different buyer, risk review and support model. The investor should separate same-product scale from cross-product sales.

Customer references should test counterfactual behaviour. Would the customer renew at current price? What measurable workflow changed? Which users benefit and which do not? What burden remains? Which product failure would cause removal? Reference enthusiasm should be reconciled to invoices, usage and renewal decisions.

14. Test revenue quality, pricing power and gross margin

Revenue quality depends on enforceable contracts, deployment status, acceptance, usage, renewal, collection and delivery cost. Bookings can include pilots, cancellation rights, future modules or contingent volumes. Diligence should bridge signed contracts to recognised revenue, invoices, cash and live utilisation by customer.

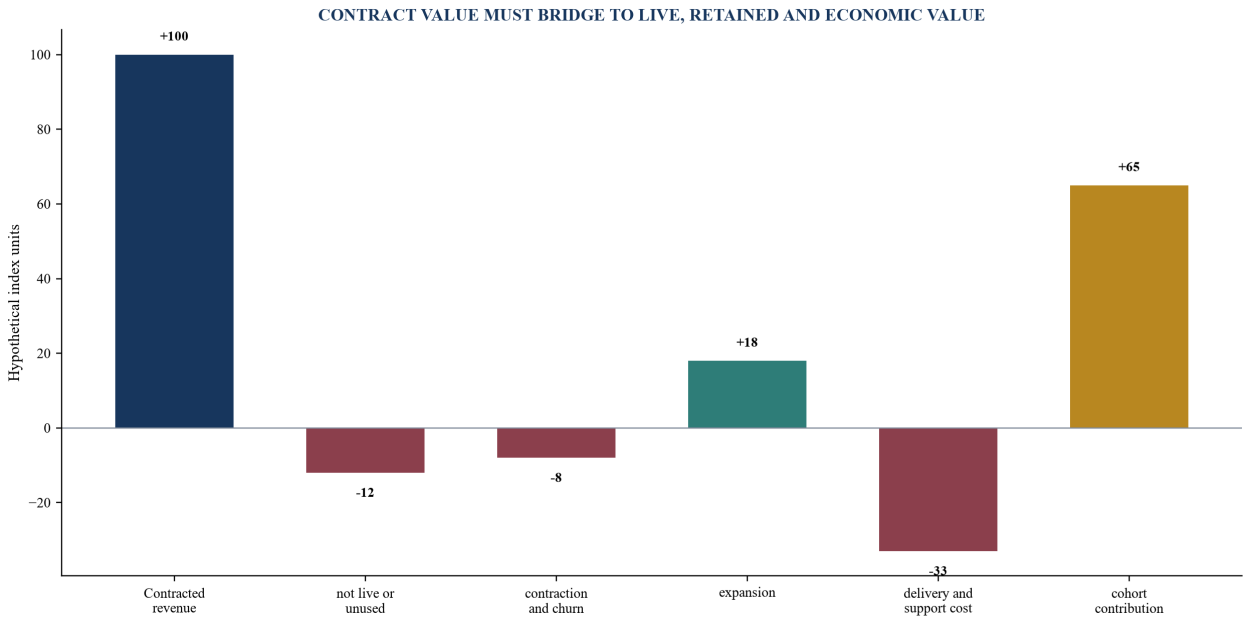
Pricing should reflect the customer's value unit while remaining observable and governable. Per-clinician pricing is simple but can detach from encounter volume. Per-encounter pricing aligns usage but creates volume variability. Enterprise pricing supports budget predictability while requiring controls for heavy usage. Outcome-linked pricing can be attractive where the measure, baseline and attribution are credible. Hybrid structures may fit different workflows.

Pricing power should be tested through renewal history, discounts, competitive replacements, procurement concessions and willingness to expand. A high list price with extensive implementation credits, free pilots or bundled modules may create a weaker net price. The investor should calculate effective price after incentives and service commitments.

Gross margin should include speech, foundation-model, storage, observability, EHR fees, support, clinical operations, implementation, customer success and third-party royalties where attributable to delivery. Capitalised development and allocation policy should be understood separately. Upstream inference price declines can improve margin, while usage growth, larger context, richer outputs or redundancy can offset the benefit.

The company should demonstrate margin by cohort and product. Mature customers should reveal the steady-state support and compute profile. New products may have lower margin while evidence and workflow mature. The valuation model should reflect the time and investment required to reach target margin rather than applying a software benchmark by label.

Figure 5. Hypothetical customer cohort economics bridge



Values are hypothetical index units and do not represent a market benchmark or forecast.

15. Separate clinical evidence from marketing claims

Evidence strength depends on design. A vendor case study can establish feasibility and customer experience. A pre-post study can describe change while remaining exposed to selection, timing and concurrent initiatives. A matched observational study can improve comparison while retaining confounding. A randomised trial can strengthen causal inference for its tested product, users, period and outcomes. No single study proves every specialty, customer or product version.

The investor should build a claim-to-evidence matrix. Each clinical, workflow and economic claim receives an intended population, endpoint, study design, version, comparator, result, limitation and permitted commercial wording. Evidence for documentation time does not automatically prove reduced burnout, higher patient capacity, better note quality or improved revenue.

The 2025 rapid review of ambient AI scribes found only six eligible real-world studies from a much larger search set and reported sparse evidence, mixed standardisation and limited productivity effects.[15] A later systematic review likewise identified implementation benefits alongside accuracy variation, manual edits and evaluation gaps.[18] These reviews support disciplined evidence collection rather than a category-wide conclusion.

A pragmatic randomised trial across 238 outpatient physicians tested two ambient-scribe applications against usual care and reported application-specific outcomes.[16] Another emergency-department comparison found similar or lower note quality and greater clinician contribution for AI-generated notes versus human scribes in that setting.[17] These studies show why comparator, setting, user and metric matter.

The company should maintain post-deployment evidence. Version changes, new specialties, languages, EHRs and workflow actions can alter performance. Monitoring should include adoption, edits, errors, incidents, subgroups, support and customer outcomes. Commercial claims should follow the approved evidence set.

16. Examine intellectual property, data advantage and defensibility

Defensibility can arise from workflow design, distribution, integrations, evaluation systems, proprietary data rights, clinical trust, regulatory capability and operating evidence. A model alone may be replicable or sourced from an upstream provider. The investor should identify which assets remain valuable if a foundation model improves, prices fall or customers gain similar features from an incumbent platform.

The IP review should cover employee and contractor assignment, background IP, open-source software, model and dataset licences, customer feedback, patents, trademarks, domain names, code provenance and restrictions in commercial partnerships. Generated code and content should be assessed under applicable law and contract. Material components should have a clear owner and permitted commercial use.

A data advantage requires lawful, contractually permitted and technically usable data. Large volume is insufficient when rights prohibit training, labels are weak, populations are narrow or versions cannot be linked to outcomes. The company should explain how data improve a defined capability and how the improvement is measured. Customer-specific data may create switching cost without supporting a transferable model.

Evaluation infrastructure can be a durable asset. Curated test sets, error taxonomies, clinical adjudication, subgroup analysis, production monitoring and release gates can accelerate safe improvement. Their value depends on representativeness, rights, freshness and reproducibility.

The investor should run a substitution test. If the upstream speech or language model changed, what remains? If an EHR vendor launched a competing feature, what workflow, evidence, integration or distribution advantage would protect renewal? Defensibility should be described as a tested economic mechanism.

17. Assess organisation, governance and deployment capability

The organisation should match the risk and operating model of the product. Relevant capabilities can include machine learning, speech engineering, clinical informatics, product management, security, privacy, regulatory affairs, clinical safety, integrations, implementation, customer success and revenue-cycle expertise. The required mix depends on intended use and customer segment.

Key-person analysis should identify who owns model architecture, clinical judgement, regulated quality systems, major integrations, customer relationships and incident response. Knowledge should be evidenced in documentation, review processes and succession rather than concentrated in a founder or a few engineers. Compensation, retention and hiring plans should align with the expansion case.

Governance should connect product release, clinical safety, privacy, security and commercial claims. A release committee can require completed evaluation, hazard review, security testing, documentation, support readiness and customer communication. Exceptions should have a named approver, rationale, expiry and monitoring plan.

Sales incentives require review. Teams may be rewarded for contracted value before security approval, go-live or sustained use. Product claims can outrun evidence when competitive pressure

is high. Compensation and approval gates should support qualified revenue and safe deployment.

Board reporting should combine product performance, incidents, rights, regulatory status, customer deployment, retention, margin and roadmap risk. A growth board needs early-warning indicators: stalled implementations, low active use, rising edits, subgroup regressions, support concentration, inference-cost changes and renewal exposure. Governance becomes an asset when it produces timely decisions and credible evidence for customers.

18. Translate findings into valuation and financing

Each confirmed finding should change cash flow, timing, risk, optionality or transaction structure. A slower implementation curve delays revenue and consumes services capacity. Higher review burden can weaken adoption. Restricted data rights can limit model improvement or geography. Upstream concentration can affect gross margin and resilience. Regulatory reclassification can require evidence, quality systems and longer sales cycles.

The valuation model should bridge pipeline, contracted value, go-live, active use, retention, expansion and contribution margin. Scenario variables should include security and procurement cycle, implementation capacity, inference cost, customer concentration, renewal, price, product mix and evidence investment. Management's base case should be reconciled to historical cohort behaviour.

Double counting should be prevented. A slower go-live may already reduce revenue and cash. The same issue should not receive an additional multiple discount unless residual uncertainty remains. A higher support cost may already reduce margin. Finance and product reviewers should approve the causal bridge together.

Financing needs can rise through delayed collections, implementation labour, clinical studies, security remediation, regulated quality systems, EHR partnerships and geographic expansion. The sources-and-uses model should include these investments. A growth-equity round sized only to sales hiring can underfund evidence and deployment capability.

Valuation comparables require careful normalization. Healthcare software companies differ by regulated status, services content, customer concentration, growth, gross margin, retention and cash consumption. A headline revenue multiple can obscure product and cohort quality. The investment committee should receive a transparent bridge from evidence to assumptions and value.

19. Convert diligence into transaction protections and the ownership plan

Transaction protection should follow the causal risk. A closing condition can require delivery of a customer consent, security remediation, key contract, regulatory filing or validated release. Price and proceeds can reflect a measurable revenue, cash or liability gap. Escrow, retention, warranty, indemnity or covenant can allocate a defined residual exposure, subject to governing law and qualified advice.

Representations should address ownership, licences, data rights, privacy, security, regulatory status, open source, customer contracts, model dependencies, incidents and claims. Their scope should match the evidence and business model. A generic compliance warranty can be weaker than a schedule of known processing purposes, subprocessors, incidents and limitations.

The ownership plan should exist before signing. Initial priorities can include freezing product and model versions, preserving data and logs, confirming customer obligations, securing key staff, validating monitoring, reviewing high-risk workflows, reconciling cohort economics and closing evidence gaps. Integration with a parent platform should proceed through clinical, security and customer gates.

The board should distinguish accepted risk from unfinished diligence. Each accepted item receives an owner, budget, milestone, evidence and decision right. Material unresolved items may require further diligence, a condition, a valuation change or rejection.

Table 5. Investment-committee gates and failure responses

Gate	Required evidence	Decision owner	Failure response
workflow	observed process, user burden and exception map	product diligence lead	narrow thesis or extend evidence
data	purpose-specific rights and lineage	privacy and legal leads	restrict use, remediate or stop
model	versioned representative evaluation	clinical and technical leads	limit use or require validation
deployment	repeatable integration and cost evidence	operating lead	reprice growth and margin case
economics	cohorts, usage, retention, margin and cash	investment committee	revalue, restructure or stop

The gate design should reflect transaction materiality, intended use and delegated authority.

20. Establish rejection gates and a repeatable diligence system

Some findings should stop or narrow the investment. Examples include inability to establish rights for a core processing purpose, clinically material errors without an effective detection path, unsupported regulated claims, unresolved tenant separation, unreliable identity matching, non-reproducible evaluation, customer concentration without credible retention evidence or unit economics that remain negative at mature comparable deployments.

Rejection gates should be explicit before management presentation. This reduces confirmation bias and prevents timetable pressure from turning missing evidence into assumed comfort. A failed gate can lead to rejection, a narrower product perimeter, a lower valuation, a staged investment, a condition or a funded remediation plan. The response should reflect control and materiality.

The diligence system should preserve a question register, evidence index, data-rights ledger, model register, evaluation results, workflow maps, cohort model, finding log and decision record. Each finding should include source, version, causality, materiality, owner, reviewer, economic effect and transaction response. Outputs should remain traceable to evidence.

After investment, actual go-live time, adoption, edits, incidents, retention, expansion, gross margin and cash can be compared with diligence assumptions. This feedback reveals which indicators predicted performance and which controls were weak. Portfolio learning should retain product type, care setting, buyer, EHR, workflow risk, geography and version so that evidence is not generalised beyond its context.

Research gaps remain substantial. Future work should examine long-term patient and clinician outcomes, subgroup performance, downstream coding and claim effects, safe autonomy, cross-language deployment, model drift, customer switching and the economics of EHR-native competition. Growth investors can contribute by requiring transparent evaluation and supporting evidence infrastructure as part of the value-creation plan.

The decision test is straightforward: can the investor reproduce how the product turns authorised data into a controlled workflow improvement, how that improvement becomes retained customer value, and how the company earns durable margin while meeting its clinical and regulatory responsibilities? Evidence that survives this chain supports an investable thesis. Breaks in the chain identify the next diligence question or the reason to stop.

References

[1] World Health Organization, Ethics and governance of artificial intelligence for health, 28 June 2021, https://www.who.int/publications/i/item/9789240029200

[2] World Health Organization, Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models, 25 March 2025, https://www.who.int/publications/i/item/9789240084759

[3] US Food and Drug Administration, Clinical Decision Support Software, Final Guidance, January 2026, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software

[4] US Food and Drug Administration, Guidances with Digital Health Content, https://www.fda.gov/medical-devices/digital-health-center-excellence/guidances-digital-health-content

[5] US Food and Drug Administration, Cybersecurity in Medical Devices: Quality Management System Considerations and Content of Premarket Submissions, February 2026, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/cybersecurity-medical-devices-quality-management-system-considerations-and-content-premarket

[6] Office of the National Coordinator for Health Information Technology, HTI-1 Final Rule, https://healthit.gov/regulations/hti-rules/hti-1-final-rule/

[7] US Department of Health and Human Services, Covered Entities and Business Associates, https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html

[8] US Department of Health and Human Services, Cloud Computing and electronic protected health information, https://www.hhs.gov/hipaa/for-professionals/faq/2075/may-a-hipaa-covered-entity-or-business-associate-use-cloud-service-to-store-or-process-ephi/index.html

[9] European Union, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[10] European Union, Regulation (EU) 2025/327 on the European Health Data Space, https://eur-lex.europa.eu/eli/reg/2025/327/oj/

[11] UK Medicines and Healthcare products Regulatory Agency, Software and artificial intelligence as a medical device, https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device

[12] NHS England, DCB0129 and DCB0160 clinical risk management standards, https://digital.nhs.uk/data-and-information/information-standards/governance/latest-activity/standards-and-collections/dcb0129-clinical-risk-management-its-application-in-the-manufacture-of-health-it-systems

[13] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework and Generative AI Profile, https://www.nist.gov/itl/ai-risk-management-framework

[14] International Medical Device Regulators Forum, Software as a Medical Device: Clinical Evaluation, 21 September 2017, https://www.imdrf.org/documents/software-medical-device-samd-clinical-evaluation

[15] Kanaparthy NS et al., Real-World Evidence Synthesis of Digital Scribes Using Ambient Listening and Generative Artificial Intelligence for Clinician Documentation Workflows: Rapid Review, JMIR AI, 2025, https://pubmed.ncbi.nlm.nih.gov/41071988/

[16] Shah SJ et al., Ambient AI Scribes in Clinical Practice: A Randomized Trial, 2025, https://pubmed.ncbi.nlm.nih.gov/41497288/

[17] Morey J et al., Ambient Artificial Intelligence Versus Human Scribes in the Emergency Department, Annals of Emergency Medicine, 2025, https://pubmed.ncbi.nlm.nih.gov/41251650/

[18] Clinical Implementation of Artificial Intelligence Scribes in Health Care: A Systematic Review, 2025, https://pubmed.ncbi.nlm.nih.gov/40306686/

[19] US Federal Trade Commission, Health Breach Notification Rule, Final Rule, 30 May 2024, https://www.ftc.gov/legal-library/browse/rules/health-breach-notification-rule

[20] HL7 International, FHIR Overview, https://www.hl7.org/fhir/overview.html

About the Author

Authored by Chennakeshav Adya
Independent Researcher

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds; bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.