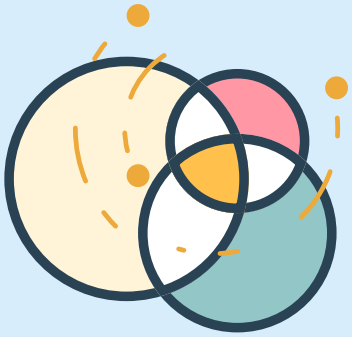
The Advanced AI Buyer's Guide for Behavioral Health

A step-by-step roadmap for choosing the right clinical-grade AI platform

eleos

Most people working in behavioral health are used to making do. Limited budgets, nonstop paperwork, and the threat of burnout are nothing new.

What *is* new is the number of technology companies selling "game-changing" AI tools.



Every vendor arrives with the same promise: that their platform is the answer the field has been looking for—the one that will finally help organizations do more with less.

But behavioral health hasn't come this far by taking every new promise at face value. A certain amount of skepticism is warranted, especially now. Some tools on the market today are little more than generic AI models rebranded with a clinical look and feel. Others don't meet the strict privacy standards required for handling sensitive client data. There are platforms that don't fit well with existing systems, sometimes creating more work than they solve, and others with claims about outcomes that look impressive but don't stand up to scientific research or real-world results.

As outside pressures build, these details matter more than ever. Medicaid rules are shifting and regulations are in flux, so AI investments *need* to show results quickly—without adding new risks to the mix.

It takes more than just a good demo. Leaders must be able to sort out which (if any) company can deliver on their promises, preferably before putting staff satisfaction and client privacy on the line.

This guide is designed to help leaders make confident choices about which AI platforms to trust. We'll cover what capabilities matter most, what questions to ask, and how to weigh the answers—so you can find a solution that works for your clinicians, fits your budget, and protects the people you serve.

Section 1

Evidence:

Does Your Product Do What You Say It Does?

The first rule of shopping in a crowded AI marketplace is simple: don't take anything at face value. The most important skill for any AI buyer is knowing when to ask for proof and how to verify whether that proof stands up to scrutiny. Legitimate vendors will expect these questions—and will have answers at the ready.

To give you a sense of what's realistic, here are some specific benchmarks from publicly available Eleos® case studies and peer-reviewed research. While these examples highlight what one platform has delivered, you can use them as a starting point for challenging any vendor to provide comparable, verifiable evidence of their claims.

Impact Category	**Real-World Benchmark Examples** (Eleos Data)
Return on Investment	Up to **8x ROI** within 12 months Gulf Coast Center
Note Efficiency	**70% less** documentation time Coleman Health Services
Time Savings	**~400 clinician hours** in 6 months GRAND Mental Health
Compliance Coverage	**20x more** notes audited per month without additional staffing Merakey + GRAND Mental Health
Clinician Wellbeing	**90%** of providers report lower stress after adoption Trilogy
Workflow Efficiency	**90% fewer** EHR clicks and **34% faster** group-note completion Large Tennessee-based health system
Clinical Outcomes	**34% greater** depression reduction and **29% greater** anxiety reduction compared to treatment-as-usual Randomized Controlled Trial study

Key Questions to Ask in Every Vendor Call

- ☐ Can you share peer-reviewed or third-party-validated ROI data from organizations like ours?
- ☐ How long did it take those customers to realize efficiency or compliance gains?
- ☐ Do you have any objective data that demonstrates a reduction in provider stress levels as a result of using the product?
- ☐ What metrics have you tracked to validate client outcomes improvements?

Section 2

Clinical Intelligence:

Does the AI "Speak" Behavioral Health?

AI learns from data. During model training, engineers feed the system thousands (even millions) of sample sentences as data. In behavioral health, the safest source for that data is consented, de-identified therapy transcripts that capture real back-and-forth: open-ended questions, reflective listening, risk assessments, and treatment plan updates. If those examples come instead from social media threads or public web sources, the model starts with the wrong dictionary and guesses its way through creating documentation.

Generic public models like GPT-5 are great at summarizing Wikipedia articles, but they can't reliably parse out clinical shorthand. They might translate **SI** as *sports injury* or interpret **DBT** as *data build tool* (a software engineering acronym).

Worse, many AI tools store every keystroke to train future versions, directly violating HIPAA and most BAAs. One popular vendor even warns users not to enter sensitive data because it "may be retained and reviewed." That's an automatic red flag for any covered entity.

Relying on a non-behavioral-health model can lead to costly errors: an incorrect CPT code, a missed suicidal-ideation flag, or a fabricated client quote that never appeared in session. Payer audits, privacy investigations, court subpoenas, and clinical missteps are stressful enough on their own. If a note contains fabrications or omits key details, these scenarios become nightmares.

What Makes a Model *"Built for Behavioral Health"*

When vendors say their AI is designed for behavioral health, that should always indicate:

1 | **Representative clinical corpus**
Training data comes from real, consented therapy sessions that reflect your populations and modalities.

2 | **Clinician–engineer partnership**
Licensed providers help label goals, interventions, and risk cues, then review each release. Otherwise, technical teams miss nuances that drive optimal reimbursement and outcomes.

3 | **Golden-Thread labeling**
Goals, interventions, and progress points stay linked across sessions so any note can survive an audit.

4 | **Grounding against hallucinations**
"Hallucinations" are plausible-sounding outputs the model has made up. Retrieval Augmented Generation (RAG) or similar methods help keep the output tied to the actual session content, which limits invented details.

> ***Quick Definition:*** **Retrieval Augmented Generation (RAG)**
>
> RAG starts by organizing every approved note into a searchable index tagged by client, date, diagnosis, and more. When you ask the AI for a draft, it retrieves only the snippets that match that specific session, then generates text anchored to those facts. That scoped search step keeps the model grounded in facts and protects against hallucinations.

5 | **Privacy-first pipelines**
PHI is encrypted or stripped before training, and raw audio/text is deleted on a set schedule with no data retained "for convenience."

6 | **Quality and rigor**
Every model upgrade should go through structured evaluation before deployment, including clinician-led validation, blinded comparisons, staged releases, and post-deployment monitoring.

Miss one of these pillars, and the platform may generate notes that collapse under payer review—or embed subtle biases against already-marginalized clients.

Key Questions to Ask in Every Vendor Call

- ☐ What percentage of your training data set comes from real behavioral health encounters, and how was consent handled?
- ☐ Who decided what counts as a goal, intervention, or risk cue? Were licensed clinicians involved throughout the definition process?
- ☐ How do you test the model across age, race, gender, and modality? How often is that audit repeated?
- ☐ Can you show an example where the treatment plan, progress notes, and outcomes remain aligned across sessions?
- ☐ Do you use RAG or similar approaches? How often do clinicians manually spot-check outputs to validate the AI's accuracy?
- ☐ Is any raw session data stored after note generation? If so, how is it de-identified, encrypted, and eventually deleted?
- ☐ What methodology do you use to validate model upgrades before deployment (e.g., blinded comparisons, staged releases)?
- ☐ How are clinicians involved in post-deployment monitoring to ensure the model is performing as expected in real-world settings?

Section 3

Future-Proof Architecture:

Can Your Product Keep Up With Changes?

Therapists update intake forms when payers change rules, and leadership rolls out new outcome measures as best practices evolve. Your software has to be able to do the same. Any AI platform that cannot add new data fields, meet changing privacy laws, or handle a larger caseload will slow progress and force an expensive replacement long before the contract ends.

Architecture is the way the tool is wired "under the hood," including details like how it stores data, talks to the EHR, updates the model, and defends against threats. A flexible foundation lets the vendor roll out improvements with little disruption. A brittle system demands workarounds, duplicate data entry, and constant user retraining.

The table below translates key architectural terms into everyday language and shows why each one matters in routine clinical work.

Architectural Feature	Plain-Language Example	Result in Practice
Secure RAG System	The model searches an approved, access-controlled index before it generates note content	Cuts down on made-up text; your clients' information stays in secure storage, never inside the AI's permanent memory
Feedback Loop & Continuous Updates	Users can easily send issues to the vendor's clinical science team for weekly model tuning	Clinical quality and accuracy improves over time
Enterprise Security & Backups	Regular external security checks (HIPAA, SOC 2) are performed; information is secure at rest and in transit, and daily backups are saved in two separate geographic locations	Patient data stays safe and available during outages
Clear Connection Points (Public APIs)	Other secure software systems can send data in or pull data out without waiting for the vendor's engineers	New tools plug in quickly and at low cost
Auto-Scaling Platform	Platform automatically adjusts to changes in processing load and data volume	No slowdown for staff or surprise IT bills

Key Questions to Ask in Every Vendor Call

- ☐ Which EHRs do you already connect to, and how quickly do you fix a broken integration?
- ☐ If a browser plug-in is needed, how is client data protected while it moves, and where is it stored?
- ☐ How do you check new text before it enters the search index, and how do you block harmful prompts?
- ☐ Which outside security audits/reviews have you passed this year, and when is the next one?
- ☐ If servers go down, how long until we are back online, and how much data could be lost?
- ☐ Can we move data in and out via your API, and will that cost extra?

Section 4

Security & Privacy Foundations:

Is My Data Safe?

Behavioral health data is medical information, sure, but it's also much more sensitive, considering that it reflects a client's most personal and private experiences. Any AI tool that touches BH data has to earn trust with clear, verifiable safeguards. And HIPAA is the floor, not the ceiling. Leaders should expect proof of mature security practices today and a plan to meet any new AI standards in the future.

Most security and privacy failures come from basic handling of data, like how it's collected, moved, stored, and deleted. At the same time, AI introduces additional risks: models may retain information from what they process, vendors may route data through third parties, and the rules are still evolving. A production-ready platform should show exactly how it prevents unauthorized access, limits what it collects (data minimization), and adapts as standards tighten.

Privacy & Security Basic Concepts *(and How to Evaluate Them)*

Certifications & Audits

Independent checks of a vendor's security. HIPAA is the U.S. health-privacy rule; SOC 2 is a third-party audit of security controls; ISO 27001 is the leading international standard focused on information security; ISO 42001 is a newer AI governance standard.

Green flag:
Recent SOC 2 Type II, active HIPAA program, ISO 27001, and ISO 42001.

Caution:
"In progress" only, reports older than 12 months, or refusal to share an attestation letter.

Business Associate Agreement (BAA)

A required contract when a vendor handles PHI; it sets allowed uses, retention, and deletion.

Green flag:
Signed BAA spelling out permitted uses, retention, and deletion timelines.

Caution:
Generic template with vague data ownership or no deletion language.

Data Minimization

Collecting only what is needed for care and documentation; deleting or de-identifying the rest on a schedule.

Green flag:
Only collects what's needed; clear deletion or de-identification schedule; option to disable audio/text retention.

Caution:
"We keep everything to improve the model" with no opt-out.

Encryption

Scrambling data so only authorized systems can read it. "In transit" means while moving; "at rest" means while stored.

Green flag:
Data encrypted at rest and in transit; keys rotated and access-controlled.

Caution:
Encryption not specified; broad employee access to keys.

Access Controls

Limits on who can see PHI and what they can do with it (role-based access, multi-factor login, admin audit logs).

Green flag:
Role-based access, MFA, least-privilege by default, admin audit logs.

Caution:
Shared accounts, weak passwords, or no audit trail.

PHI in Model Training

Whether raw session data is used to teach the AI. Safer approaches avoid embedding PHI in model "memory."

Green flag:
No raw PHI embedded; improvements use de-identified data with review.

Caution:
Terms say your data "may be used to train models" without limits.

Incident Response

The plan for finding, containing, and notifying proper parties about security events (and learning from them).

Green flag:
24/7 plan with timelines, notification steps, and routine drills.

Caution:
"We'll handle it if it happens" with no documented process.

Subprocessors & Third-Party Risk

Other companies your vendor uses (cloud, transcription, analytics). Their security becomes your risk, too.

Green flag:
Public vendor list, security reviews, data processing agreements (DPAs), disclosure of geographic locations.

Caution:
Unknown subprocessors or "we use standard cloud" blanket statement with no details.

Trust & Transparency

Clear public documentation, a live status page, and an easy way to report issues.

Green flag:
Readily available privacy/security documentation with specifics; regular status updates; in-product "report an issue" path.

Caution:
Marketing claims with no details.

Retention & Deletion

How long data is kept and how/whether it is removed everywhere (including backups) when time is up or you exit.

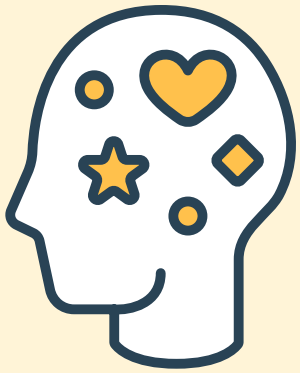
Green flag:
You control retention; automatic deletion after the set period; exit procedures include export + verified purge.

Caution:
Indefinite retention or backups never purged.

Key Questions to Ask in Every Vendor Call

- Do you have HIPAA compliance and a recent SOC 2 report? What is your timeline and plan for ISO 42001?
- What data uses are permitted under your BAA, and how long is PHI retained prior to deletion or de-identification?
- What data do you not collect, and why?
- How is data encrypted in transit and at rest, and who (by role) can access raw session data and notes?
- Is any PHI used to train or fine-tune models? If you use de-identified data, how is de-identification verified, and by whom?
- Which third parties handle PHI, what contracts govern them, and how do you monitor them over time?
- Can you share your incident response playbook? What is your notification process and average time to containment?

Section 5



Quality & Ethics Guardrails:

Will it Do No Harm?

AI should raise care quality and protect clients, not create new risks. Vendors need clear guardrails that reduce bias, prevent harmful output, and keep clinicians in charge of decisions.

Bias can slip in through training data, labels, and day-to-day use. Quality can drift when updates change model behavior without proper testing. Good guardrails make risks visible and fixable, and they should align with standards leaders are already familiar with.

For example, the Coalition for Health AI (CHAI) emphasizes fairness, transparency, and clear risk management, while CARF accreditation highlights person-centered care, cultural competence, and documentation that ties goals, interventions, and outcomes (i.e., the "Golden Thread"). Together, frameworks like these give a shared language for what "do no harm" looks like in practice.

Vendors should be able to show their guardrails across three layers: process, people, and proof.

Layer 1: Bias & Safety Process

Think of this as a loop that a vendor runs before launch and on an ongoing schedule.

Plan

Define fairness goals and what “harm” means in your setting. Identify sensitive attributes to test against.

Collect

Build a representative, consented dataset and document known gaps.

Test

Run bias and accuracy tests across age, race, gender, diagnosis, and modality.
Share the metrics and thresholds.

Validate in practice

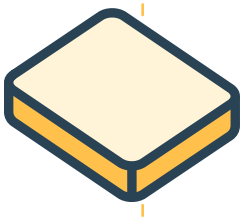
Pilot with human review, track error types, and compare to treatment as usual.

Monitor

Watch for drift after each release.
Re-test the same metrics and publish results on a predictable cadence.

Improve

Retrain or adjust prompts and labels.
Report what changed and why.



Layer 2: Human Oversight & Governance

This layer keeps licensed clinicians in control and makes accountability clear. It also connects the product to recognized frameworks such as CHAI guidance and CARF standards.

Human in the loop

Clinicians review samples, can override outputs, and have a simple way to flag issues.

Incident review

There is a clear playbook for investigating harmful or biased outputs and preventing repeats.

Governance

A documented ethics process aligns with the CHAI pledge for trustworthy health AI and with CARF's quality standards for person-centered, culturally competent care.

Transparent updates

Release notes describe model changes, expected impact, and any new limitations.

Layer 3: Evidence & Validation

This layer demonstrates proof that the tool works and stays safe over time.

Peer-reviewed or third-party studies that show a positive impact on documentation quality, risk identification, or outcomes.

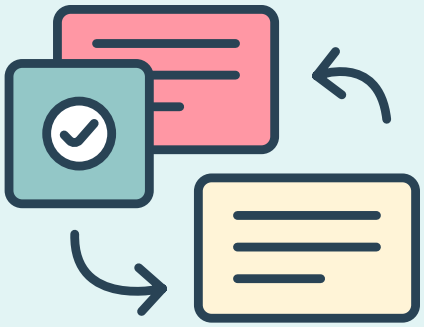
External evaluations for equity and safety (not just vendor-run tests).

Real-world monitoring that continues after go-live, with metrics shared back to customers.

Key Questions to Ask in Every Vendor Call

- ☐ How do you test for bias across demographics and modalities, and what happens when a problem is identified?
- ☐ Who reviews model outputs, and how can clinicians and other users flag and override problematic output?
- ☐ How do your practices align with guidance from standard-setting organizations like CHAI and accreditation bodies like CARF?
- ☐ Before a new product release, what quality checks are run, and do you share the results?
- ☐ If the model produces harmful or false content, what is the investigation process and timeline for remediation?
- ☐ What peer-reviewed or third-party evidence supports improvements in documentation quality, risk detection, or outcomes as a result of using this solution?

Section 6



Real-World Versatility:

Will It Work Across Settings & Use Cases?

Most software demos happen in a perfect test environment. Real behavioral health work does not. Programs run in clinics, schools, homes, shelters, and hospitals—and the teams running them include therapists, psychiatrists, case managers, peer specialists, and supervisors (all of whom document differently). A platform has to show that it can add value across these settings and roles, not just on a programmer's laptop.

A versatile tool will adapt to different care settings, note types, visit formats, and connectivity realities. It should support collaborative documentation, group services, and mobile field work without creating extra steps or privacy risks. Ask vendors to demonstrate these scenarios live with your workflows and documentation requirements.

What “Good” Looks Like

Compliance Management & Audit Readiness

What to look for:
A dashboard that flags missing Golden Thread elements, filters by program and clinician, and exports audit-related data when needed. You should be able to review many more notes with the same staff time.

Proof to ask for:
A real dashboard tour with your templates, plus historical examples from similar organizations.

Individual Therapy

What to look for:
Notes that link goals, interventions, and progress across sessions. Collaborative documentation is available, so the note can be drafted in-session using client-friendly language.

Proof to ask for:
A start-to-finish visit showing a draft that can be reviewed, edited, and signed in the EHR.

Comprehensive Clinical Management

What to look for:
Support for mental status exam elements, medication changes, and risk factors. The note fits E/M coding rules and pulls the med list from the EHR where possible. Vendors should also demonstrate that the system does not filter out or suppress references to suicidal ideation (SI) or homicidal ideation (HI). Notes must capture these risk factors clearly and accurately.

Proof to ask for:
A psychiatry-specific template in action, plus examples of how updates are tracked between visits.

Group Therapy

What to look for:
A roster view for attendance and time in group, shared content for the session, and individualized sections for each participant. Group codes and required modifiers are supported.

Proof to ask for:
A live group note that generates individual documentation for each client without copy-paste.

Mobile Case Management & Community Programs

What to look for:
Reliable usability on tablets or phones, including low-connectivity or offline capture with secure sync later. Short field-friendly templates, location privacy controls, and simple signature capture. A workflow that supports seamless on-the-go data entry (e.g., ability to dictate or briefly summarize the session in the provider's own words and then generate a full clinical note based on that input).

Proof to ask for:
A mobile demo outdoors or over a throttled connection that shows capture, sync, and EHR delivery.

Peer Support Services

What to look for:
Templates and language that match peer roles and recovery plans, with permissions that limit access to sensitive PHI that the role does not need.

Proof to ask for:
A peer note walkthrough and a quick view of role-based access settings.

Supervision & QA

What to look for:
Work queues for supervisors, change history on notes, and sampling tools for routine audits.

Proof to ask for:
A demonstration of how supervisors review, return, and approve notes across programs.

When one platform works across programs, teams don't have to keep switching tools during the day, supervisors get a single view of quality, and IT can manage fewer integrations. Versatility prevents tool sprawl and makes adoption more likely, which reduces training time and ultimately protects your ROI.

Key Questions to Ask in Every Vendor Call

- ☐ Can you run a live walk-through using our forms for individual therapy, psychiatry, group therapy, and field work—not just a demo script?
- ☐ How does collaborative documentation work during a visit, and how do clients review and sign off?
- ☐ How do you create individualized notes from one group session and handle attendance and codes?
- ☐ What happens if internet connection is lost, and how does the system sync securely later?
- ☐ How are templates and permissions set for different roles (e.g., peer support, case managers, and supervisors)?
- ☐ What does your compliance dashboard show, and how does it help us review more notes with the same staff?
- ☐ For each setting, where do the notes and alerts land inside the EHR, and who confirms that mapping before go-live?
- ☐ How does the system handle mentions of suicidal ideation (SI) or homicidal ideation (HI)? Can you show proof that these risk factors are documented consistently and not filtered out due to the model's content restrictions?

Section 7

Partnership & Support:

Who's Got My Back Post-Launch?

Technology is only half the work of AI implementation. The rest is rollout strategy, training, and adoption, because if a vendor cannot get clinicians up and running quickly (or cannot support supervisors after launch), the technology won't ever get used.

Good partners make the path clear by sharing an onboarding plan with owners and dates, offering role-based training for different teams, and meeting with you at regular intervals to review results and make adjustments.

The goal is not a one-time launch. The goal is ongoing value that shows up in documentation quality, time saved, audit readiness, and staff experience.

What Good Support Includes

1 | **Fast start (2–3 hours to first note):**
Ideally, most clinicians can finish training and write their first note in a couple of hours, so operations are not disrupted.

2 | **A named point person (e.g., Customer Success Manager):**
One person you can contact for help, with a backup contact. (Plus a "solutions engineer"—a technical specialist who can adjust settings and connections when necessary.)

3 | **Training you can actually use:**
Live trainings at launch and on an ongoing basis for new hires, weekly office hours for questions, and a searchable resource library for on-demand help.

4 | **Help with change management:**
Templates for staff emails, a "champion" program where early adopters help peers, and an option to roll out in phases instead of all at once.

5 | **A usage and outcomes dashboard:**
Shows adoption by program, average documentation time, and audit-readiness metrics so supervisors can track progress.

6 | **Written support promises (i.e., SLAs):**
Clear targets for response and fix times, plus an escalation path for urgent issues.

7 | **Regular check-ins (e.g., Quarterly Business Reviews):**
Scheduled meetings to review results with your data and strategize on next steps.

Rollout Milestone Map

Use this list to set expectations for what support you should be receiving at each stage of implementation.

Before Launch

What you should see:
A written plan with owners and dates, configuration in a test environment, and short role-based trainings for therapists, psychiatrists, case managers, peer supports, and supervisors. To preview this, ask for a sample project plan, temporary access to the test site, and a training outline.

Launch Week

What you should see:
Most users can learn and complete a first note in under 60 minutes. Live help is available during trainings. A named Customer Success Manager can adjust settings without long ticket chains. To see how this works in practice, ask for a live support schedule and the name and contact info of your CSM.

First 90 Days

What you should see:
Weekly or biweekly office hours, an on-demand training library, and a usage dashboard that shows adoption by program. Early wins are captured and shared. For a concrete example, ask for a sample dashboard, example update emails, and office hour dates.

Ongoing

What you should see:
Quarterly Business Reviews that use your data to check goals and plan changes. There should be a clear path to escalate issues, with standing refresher trainings. To understand the cadence, ask for a QBR agenda template and a training calendar.

A strong partnership keeps adoption high, reduces rework, and protects ROI. Clinicians get help when they need it. Supervisors see quality improve. Leaders have a single place to review progress and adjust the plan.

Key Questions to Ask in Every Vendor Call

- ☐ How long will it take a provider to learn the system and complete a first note, start to finish?
- ☐ Who is our named CSM, and who backs them up? Can we meet them before we sign?
- ☐ What does the training library include, and how often is it updated? Do you offer live office hours?
- ☐ Do you provide communication templates and a peer champion program for early adopters?
- ☐ What adoption and quality metrics will you share with us in the first 90 days?
- ☐ How often will we meet to review results and plan next steps, and who needs to attend?
- ☐ What are your response and resolution time targets, and how do we escalate urgent issues?

Section 8

Quick-Glance Scorecard

Use this side-by-side capabilities matrix during vendor demos to easily compare your options and make an informed decision.

Category	What to Look For	Vendor A	Vendor B	Vendor C	Notes
Evidence	Real data from organizations like ours (ROI, outcomes)				
	How fast customers saw benefits (efficiency, compliance)				
Clinical Fit	Training based on real behavioral health sessions				
	Clinicians helped design, review, and validate the system				

Category	What to Look For	Vendor A	Vendor B	Vendor C	Notes
Technology	Works with our EHR and other systems				
Security & Privacy	HIPAA + SOC 2; clear contract on data use & deletion				
	No client data used to “train the model” & deletion schedule clear				
	Independent security reviews and a backup plan in place				
Quality & Ethics	System tested for bias; easy for clinicians to override				
	Validation methods like blinded comparisons, staged releases, and post-deployment monitoring are used to ensure accuracy				
Versatility	Handles different visit types (individual, group, mobile)				
Support	Easy to start (first note in a couple of hours); named support contact				

Section 9

Buyer's Worksheet:

Questions to Ask

Refer to this tear-out list of questions so every stakeholder scores vendors against the same criteria.

1. **Evidence:** *Does Your Product Do What You Say It Does?*

- ☐ Can you share peer-reviewed or third-party-validated ROI data from organizations like ours?
- ☐ How long did it take those customers to realize efficiency or compliance gains?
- ☐ Do you have any objective data that demonstrates a reduction in provider stress levels as a result of using the product?
- ☐ What metrics have you tracked to validate client outcomes improvements?

2. **Clinical Intelligence:** *Does the AI "Speak" Behavioral Health?*

- ☐ What percentage of your training set comes from real behavioral-health encounters, and how was consent handled?
- ☐ Who decided what counts as a goal, intervention, or risk cue? Were licensed clinicians involved throughout?
- ☐ How do you test the model across age, race, gender, and modality? How often is that audit repeated?
- ☐ Can you show an example where the treatment plan, progress note, and outcomes remain aligned across sessions?
- ☐ Do you use RAG or a similar grounding? How often do clinicians manually spot-check outputs to validate the AI's accuracy?
- ☐ Is any raw session data stored after note generation? If so, how is it de-identified, encrypted, and eventually deleted?
- ☐ What methodology do you use to validate model upgrades before deployment (e.g., blinded comparisons, staged releases)?
- ☐ How are clinicians involved in post-deployment monitoring to ensure the model is performing as expected in real-world settings?

3. **Future-Proof Architecture:** *Can Your Product Keep Up With Changes?*

- ☐ Which EHRs do you already connect to, and how quickly do you fix a broken link?
- ☐ If a browser plug-in is needed, how is client data protected while it moves, and where is it stored?
- ☐ How do you check new text before it enters the search index, and how do you block harmful prompts?
- ☐ Which outside audits have you passed this year, and when is the next one?
- ☐ If servers go down, how long until we are back online, and how much data could be lost?
- ☐ Can we move data in and out through your API, and will that cost extra?

4. **Security & Privacy Foundations:** *Is My Data Safe?*

- ☐ Do you have HIPAA compliance and a recent SOC 2 report? What about ISO 42001?
- ☐ What data uses are permitted under your BAA, and how long is PHI retained before deletion or de-identification?
- ☐ What data do you not collect, and why?
- ☐ How is data encrypted in transit and at rest, and who (by role) can access raw session data and notes?
- ☐ Is any PHI used to train or fine-tune models? If you use de-identified data, how is de-identification verified, and by whom?
- ☐ Which third parties handle PHI, what contracts govern them, and how do you monitor them over time?
- ☐ Can you share your incident response playbook? What is your average time for notification and containment?

5. **Quality & Ethics Guardrails:** *Will It Do No Harm?*

- ☐ How do you test for bias across demographics and modalities, and what happens when an issue is identified?
- ☐ Who reviews model outputs, and how can clinicians and other users flag and override problematic content?
- ☐ How do your practices align with guidance from standard-setting organizations like CHAI and accreditation bodies like CARF?
- ☐ Before a new release, what quality checks are run, and how do you share the results?
- ☐ If the model produces harmful or false content, what is the investigation process and timeline for remediation?
- ☐ What peer-reviewed or third-party evidence supports this product's impact on documentation quality, risk detection, or outcomes?

6. **Real-World Versatility:** *Will It Work Across Settings & Use Cases?*

- ☐ Can you run a live walk-through using our forms for outpatient, psychiatry, group, and field work, not just a demo script?
- ☐ How does collaborative documentation work during a visit, and how do clients review and sign off?
- ☐ How do you create individualized notes from one group session and handle attendance and codes?
- ☐ What happens if internet connection is lost, and how does the system sync securely later?
- ☐ How are templates and permissions set for peer supports, case managers, and supervisors?
- ☐ What does your compliance dashboard show, and how does it help us review more notes with the same staff?
- ☐ For each setting, where do notes and alerts land inside the EHR, and who confirms that mapping before go-live?

7. **Partnership & Support:** *Who's Got My Back Post-Go-Live*

- ☐ How long will it take a provider to learn the system and complete a first note, start to finish?
- ☐ Who is our named CSM, and who backs them up? Can we meet them before we sign?
- ☐ What does the training library include, and how often is it updated? Do you offer live office hours?
- ☐ Do you provide communication templates and a champion program for early adopters?
- ☐ What adoption and quality metrics will you share with us in the first 90 days?
- ☐ How often will we meet to review results and plan next steps, and who needs to attend those meetings?
- ☐ What are your targets for support response and resolution times, and how do we escalate urgent issues?

Section 10

Resource Library:

Dig Deeper

Use these resources to brief stakeholders or explore specific topics in more depth.

Standards & Policy

NIST AI Risk Management Framework (AI RMF 1.0) — core concepts, playbook, and profiles

ISO/IEC 42001 — the AI management-system standard for responsible AI programs

ONC HTI-1: Transparent & Trustworthy AI in Health Care (DSI/FAVES overview)

WHO Regulatory considerations for AI in health (global guidance)

Accreditation & Compliance

CARF: Behavioral Health program standards overview

Security & Safety

OWASP Top 10 for LLM Applications (threats like prompt injection, data leaks)

AHRQ PSNet: Artificial Intelligence and Patient Safety—Promise & Challenges (2024)

Ethics & Equity

Coalition for Health AI (CHAI) — mission and plan for trustworthy health AI

National Academy of Medicine: An AI Code of Conduct for Health & Medicine (2025)

JAMA Viewpoints on responsible AI & equity

Implementation & Governance

AMA Augmented Intelligence policy hub (governance toolkits, CME, transparency)

AMA STEPS Forward Governance for Augmented Intelligence (practical toolkit)

Assistant Secretary for Technology Policy (ASTP)

Behavioral-Health-Specific Learning

National Council for Mental Wellbeing — Evidence Meets Innovation: Clinical Research in AI (webinar recording)

National Council blog — AI & Behavioral Health: The Future Is Here (field perspectives)

About Eleos

Eleos is the leading AI platform for behavioral health, substance use disorder, home health, and hospice. At Eleos, we believe the path to better care is paved with provider-focused technology. Our purpose-built AI platform streamlines documentation, ensures compliance, simplifies revenue cycle management, and surfaces deep care insights to drive better client outcomes. Created using the industry's largest database of real-world sessions and fine-tuned by our in-house clinical experts, our AI tools are scientifically proven to reduce documentation time by more than 70%, boost client engagement by 2x, and improve symptom reduction by 3-4x. With Eleos, providers are free to focus less on administrative tasks and more on what got them into this field in the first place: caring for their clients.