Ethical AI Use in ABA Supervision and Performance

Purpose:

This brief summarizes how the ten ethical guideline areas from the AIC-ABA (2025) Guidelines apply to AI use in ABA staff supervision and performance management. Each area includes a concise supervisory obligation and a concrete applied example drawn from ABA service delivery.

Key Framing Questions for Discussion:

- Shadow AI use and the capacity for developing AI tools at the clinician levels creates some ethical concerns for the field of ABA.
- Which of these guideline areas represent the greatest risk or gap in current practice and literature?
- What organizational structures (policies, training, committees) would need to exist to implement these guidelines? What controls need to be developed at the organizational and individual level?
- How should frontline staff (RBTs, BCBAs) be involved in AI governance decisions that affect their supervision?
- How can we focus on personalizing feedback and preserving relationships in the age of AI?

Please note that Claude was used as a thought partner to synthesize information for this document.

AIC Guideline	Supervisor Obligation	Applied Example in ABA
1. Human-Led Oversight	Supervisors retain full professional responsibility for all AI-generated performance evaluations and training outputs. AI provides a starting point, not a final determination	*BCBAs should review AI-scored treatment integrity data before delivering staff feedback. Flagged sessions should also initiate human review, not automatic corrective action without human review.*
2. Transparency & Explainability	Staff must be explicitly informed when AI evaluates their performance, what data is collected, and how findings influence supervisory decisions. Appeal processes must be available and clearly communicated to staff.	*Written AI disclosures should be provided to staff during onboarding. Supervisors should also communicate when and how AI data contributed to feedback (e.g., 'The system flagged this session, which led me to review it.').*
3. Privacy & Data Security	Client and employee performance data in AI systems require rigorous protection. PHI/PII must not be entered into open AI systems. Supervisors model responsible data hygiene for staff.	*Organization should adopt a closed, HIPAA-compliant platform for performance tracking. BCBAs and RBTs should also be provided training on what data cannot be entered into open AI tools and systems.*
4. Accountability	Written policies must define AI's role in personnel decisions before deployment. Error-reporting processes must exist. Collaborators including frontline staff, should have input in AI governance.	*Standard operating processes should specify AI cannot be sole basis for disciplinary action. Staff must have a formal grievance process for AI-influenced performance decisions as well as AI Ethics Review Committee oversight.*

Please note that Claude was used as a thought partner to synthesize information for this document.

AIC Guideline	Supervisor Obligation	Applied Example in ABA
5. Safety, Quality & Accuracy	AI performance data must be verified for accuracy. Supervisors must have a protocol to pause AI tool use when discrepancies or errors arise, defaulting to direct human observation.	*Quarterly audits should compare AI session scores with supervisor direct observation IOA data. Organizations should also adopt defined 'pause protocol' procedures when AI accuracy falls below acceptable threshold.*
6. Equity & Minimizing Bias	AI performance tools may disadvantage staff from diverse backgrounds due to limited training data and algorithm bias. Supervisors must actively monitor for demographic bias patterns and scrutinize information on model training.	*Model training information should be thoroughly reviewed before procurement of AI tools and systems. If clinicians suspect that AI tools and system's outputs are biased, a report should be made to the systems administrator immediately. AI performance scores should also be analyzed semi-annually for patterns by staff demographic group. Additionally, training scenarios reviewed for cultural relevance of training data to application contexts.*
7. Evaluation & Continuous Improvement	AI supervision tools require structured ongoing evaluation. Baselines must be set before adoption. Staff feedback must be collected and genuinely used to inform decisions about AI tools.	*Semi-annual staff reviews should be examined for accuracy, staff satisfaction, and client outcome trends. Consequently, frontline staff surveys should inform continued use or discontinuation decisions on how these tools are used in performance management.*

Please note that Claude was used as a thought partner to synthesize information for this document.

AIC Guideline	Supervisor Obligation	Applied Example in ABA
8. Mitigating Dual Use Risks	Performance AI must not become punitive surveillance. Organizational policy must explicitly limit AI data use to quality improvement and professional development, and not as a discipline tool. AI performance feedback should be personalized to the staff and outputs should be reviewed and adapted to the individual while focusing on preserving the relationship.	*AI session observation tools should be scored scoped to clinical sessions only not other outside session interactions. Organization policy should also prohibit use of AI performance data as standalone basis for termination or discipline of staff. Additionally, vendor contracts should restrict third-party data sharing to protect staff privacy.* *AI tools and performance feedback should be adapted to the individual.*
9. AI Literacy & Training	Supervisors must achieve competence in AI tools before deploying them with supervisees. AI literacy including critical evaluation of AI outputs must be built into staff training at all levels. Please note, metrics for AI competence and literacy in behavior analysis have yet to be established (for information on general AI literacy components please see Stanford University Teaching Commons, n.d.)	*Supervisors must complete AI literacy training before using AI performance data in staff evaluation. Also, RBT onboarding should include AI literacy modules covering data rights and limitations of AI measurement.*
10. AI & Client Outcomes	AI supervision tools must ultimately demonstrate benefit to client service quality. Parallel tracking of staff performance and client outcome data is required. Additionally, gains in efficiency must translate to clinical value for clients to justify the use of AI tools and systems.	*Baseline client and staff metrics should be documented before AI tool launch to evaluate the effects of the tool. Additionally, the freed supervisor time should be reinvested in direct observation and coaching to support the quality of clinical care. Also, client outcome data should be included in AI tool evaluation.*

Please note that Claude was used as a thought partner to synthesize information for this document.

Three Overarching Principles for Discussion

Human Oversight is non-negotiable	**Staff and Client Rights Must Be Actively Protected**	**Client Outcomes Are the Ultimate Metric**
No AI system can substitute for the professional judgment of a qualified behavior analyst in personnel and clinical decisions. AI informs; humans decide.	*Transparency, accountability, equity monitoring, and anti-surveillance safeguards ensure AI performance tools serve client and staff development, not punitive control.*	*All AI use in supervision must be evaluated against its impact on the quality of services delivered to clients; not only organizational efficiency.*

Please note that Claude was used as a thought partner to synthesize information for this document.

References:

Jennings, A. M., Para-Cremer, M., Garrison, B., Britwum, K., King, K., Del Rio, M., ... Kazemi, E. (2025, December 12). Artificial Intelligence and Behavior Analysis: Guidelines for Ethical Use. https://doi.org/10.31234/osf.io/axurd_v1

Stanford University Teaching Commons. (n.d.). Understanding AI literacy. Stanford University. https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide/understanding-ai-literacy