



ALLIANCE
FOR ARTIFICIAL
INTELLIGENCE
IN HEALTHCARE

Powered By  CONNECT

Critical Considerations for AI Model Licensing Agreements in Healthcare

INSIGHTS FROM THE AAIH
WEBINAR: "CRITICAL
CONSIDERATIONS FOR AI MODEL
LICENSING AGREEMENTS AND
DATA OWNERSHIP"

Webinar date: February 10, 2026

AUTHORS:

ELAINE HAMM, PHD, ALLIANCE FOR ARTIFICIAL INTELLIGENCE IN HEALTHCARE;
EXECUTIVE IN RESIDENCE, TULANE UNIVERSITY SCHOOL OF MEDICINE

BROOKE FRITZ, PARTNER, FOLEY HOAG LLP

JOHN D. LANZA, PARTNER, FOLEY HOAG LLP

BRANDON ALLGOOD, PHD, PARABILIS MEDICINES

ANAND KUMAR, INTUITIVE.CLOUD

Executive Summary

AI licensing in healthcare is rarely “just a license.” It is a package of decisions about (1) what the asset actually is, (2) who controls it in practice, and (3) who remains accountable as models, data, and regulatory expectations evolve.

In a recent webinar sponsored by Foley Hoag and the AAIH, a consistent message surfaced across operator, cloud, and legal perspectives: most negotiations break down when contracts assume static assets and perfect control. Data is often messy and perishable, model artifacts can unintentionally reveal information, and “return or delete” clauses do not map cleanly to how machine learning works.

The most useful agreements are the ones that (a) start with an honest valuation of the dataset or model for a specific use case, (b) translate “ownership” into a precise bundle of rights (access, training, derivatives, outputs), and (c) specify the lifecycle obligations that make the deal workable over time: versioning, auditability, update governance, and security controls that match the deployment reality.

What readers should take away

- Do not start with royalties and reach-through. Start with a use case, access pattern, and a rights bundle that matches how ML actually behaves.
- Assume “toothpaste” dynamics: once data or model leaves a controlled environment, you should plan as if you cannot put it back.
- Make lifecycle obligations first-class: provenance, model versioning, drift monitoring, update approval, rollback, and inspection readiness.

Panel Context

AAIH and Foley Hoag convened a cross-functional panel to discuss contracting nuances that matter when AI models and health data intersect, including data ownership and restrictions, the right to train in-house models, bias and dataset understanding, field-of-use restrictions, and the practical challenges of non-returnable data and models. The discussion was designed for biopharma, medtech, provider, and payer leaders negotiating AI partnerships where patient privacy, regulatory expectations, and commercial strategy must align.

Key Themes and Insights

1) Start with asset triage: when data is valuable versus expensive

Organizations frequently overestimate the standalone value of their data. The panel emphasized that the commercially relevant “asset” is often not the raw dataset, but the governed, cleaned, and well-described resource that is fit for a specific use case. Raw data is rarely in the final usable state.

A practical first step is customer discovery. If you believe you have valuable data, talk to the likely buyers early, identify what is missing, and learn what level of curation, labeling, or harmonization is required before the dataset becomes useful.

Practical valuation questions to ask before negotiating terms

- What is the intended use and decision it will support (research exploration, clinical validation, operational workflow, regulated decision support)? How valuable is the decision?
- What work is required to make the dataset usable (cleaning, normalization, labeling, missingness handling, governance documentation)?
- Is the value time-sensitive (will similar datasets emerge elsewhere, will clinical practice change, will codes and ontologies shift)?
- Can the buyer evaluate the dataset’s distributions and gaps before signing (exploration tools, summary statistics, sample schemas)?

2) Replace “ownership” with a rights bundle that matches technical reality

The panel repeatedly returned to a core point: “ownership” is too blunt for AI and data deals. Contracts need to define distinct assets and rights, including raw data, labels, derived data (such as embeddings), model artifacts (weights), and outputs.

In practice, the most contentious negotiations cluster around the right to train, the scope of derivative works (including derivative models), and the control of outputs that could leak sensitive information.

Rights bundle checklist

- Access rights: where data is accessed (inside a controlled environment, via API, via data export).
- Training rights: whether training and fine-tuning are permitted, and for what purpose (product improvement vs competitive model development).
- Derived data and outputs: what is permitted to be retained, reused, published, or shared.
- Residuals and termination: what happens to model weights and artifacts that may retain traces of the training dataset.

Key Themes and Insights

3) Trade secret protection meets model leakage and adversarial querying

A recurring tension is the collision between trade secret language and the ways data can leak once incorporated into a model. Even with confidentiality obligations, model-mediated leakage can occur through prompt attacks, memorization, or regurgitation behaviors. An example of such a leak was exemplified by the recent accusation by OpenAI of companies distilling their model to train competitor models.

The practical implication is not that contracts are useless, but that they must be drafted with a clear threat model and paired with concrete technical controls. Agreements should anticipate model inversion-style risks and define what “reasonable safeguards” means for the specific deployment.

Controls that should be discussed alongside confidentiality clauses

- Access pattern: keep sensitive datasets in a controlled environment when possible (secure enclave, behind-firewall compute).
- Output controls: logging, rate limiting, and review processes for high-risk exports (models, embeddings, large result sets).
- Security testing: adversarial testing and red teaming aligned with the model’s interface and threat vectors.
- Contractual prohibitions: prohibit training a second model on the licensed model where that could recreate the distribution of the underlying data (distillation-style leakage).

4) Deal structures must reflect the “temporal” nature of data

Traditional IP license patterns (upfront fees, long-tail royalties) often misfit data transactions because data can depreciate quickly. Data may be valuable today but less differentiated tomorrow as comparable datasets emerge or clinical practice evolves.

The panel discussed several approaches that better match this reality: time-limited royalties, milestone payments tied to verifiable events, subscription-style access fees for ongoing data updates, and co-development structures where value exchange is not purely monetary.

Negotiation framing that reduced friction

- Separate ‘data as input’ transactions from ‘model as product’ transactions. They carry different risk profiles and should not share a one-size contract.
- For high-uncertainty exploration, consider simpler economics (access fees, time-limited terms) rather than broad reach-through across unknown future products.
- If the dataset is strategic and foundational, ensure the agreement addresses future data updates and refresh cadence.

Key Themes and Insights

5) Field of use: sometimes essential, sometimes a proxy fight

Field-of-use restrictions are standard in patent licenses, but their role in data licenses is more nuanced. In many cases, the dataset itself constrains what can realistically be done, making aggressive field-of-use negotiation less valuable than parties assume.

Where field-of-use still matters is competitive positioning. It can be used to prevent a counterparty from using the relationship to enter a sensitive adjacent market, or to prevent productization beyond the intended scope.

6) The “toothpaste problem”: return and deletion clauses need honesty

When training occurs, data cannot always be meaningfully ‘returned’ because its influence may be embedded in model parameters and derived artifacts. The panel emphasized that the most reliable control mechanism is to avoid uncontrolled export in the first place.

When export is necessary, contracts should be explicit about what must be deleted (raw data, identifiers, working copies), what may be retained (aggregated statistics, audit logs), and what obligations survive termination (confidentiality, restrictions on derivative models, audit rights).

Drafting approaches that align with ML reality

- Define confidential information to include embeddings and other derived artifacts where appropriate.
- Clarify ‘residual rights’ in model weights and set restrictions on reuse for competitive training.
- Consider controlled access plus inspected exports: models and artifacts leaving an enclave are reviewed through a defined process (IRB or governance committee, as applicable).
- If “right to be forgotten” obligations apply, discuss traceability and versioning early.

Key Themes and Insights

7) Lifecycle obligations: versioning, drift, updates, and inspection readiness

- Define confidential information to include embeddings and other derived artifacts where appropriate.
- Clarify 'residual rights' in model weights and set restrictions on reuse for competitive training.
- Consider controlled access plus inspected exports: models and artifacts leaving an enclave are reviewed through a defined process (IRB or governance committee, as applicable).
- If "right to be forgotten" obligations apply, discuss traceability and versioning early.

Minimum lifecycle capabilities to address explicitly

- Model versioning and change logs.
- Data lineage and provenance documentation.
- Drift monitoring and performance surveillance in the target environment.
- Update governance: validation steps, customer notification, rollback procedures.
- Audit and inspection readiness: who can inspect what, and under what process.

8) Vendor selection and differentiation: demand objective proof

The panel closed with a market reality: the vendor landscape is noisy, and many claims are lookalike. Buyers should demand objective proof of differentiation and clarity about what is actually being licensed.

Useful diligence questions include reproducibility on external datasets, benchmarking against public baselines, clarity on unique data access or moats, security and integration maturity, and a roadmap tied to measurable outcomes.

A practical checklist for your next AI data or model agreement

- Name the asset precisely: raw data, labels, derived data, embeddings, model weights, outputs, and operational logs.
- Write the intended use and deployment pattern first. Then write the rights bundle to match it.
- Decide whether this is 'data as input' or 'model as product' (or both), and separate the terms accordingly.
- Assume non-returnability after training. Draft deletion and termination terms that distinguish raw data from model artifacts.
- Add lifecycle obligations: versioning, lineage, drift monitoring, update governance, rollback, and inspection readiness.
- Pressure test trade secret language against model leakage threat models, and pair it with concrete controls.
- Keep economics aligned with temporal value: consider time-limited royalties, milestone payments, or subscription access for ongoing updates.
- Treat vendor selection as part of risk management: require objective evidence, external validation, and security maturity.

Conclusion

Healthcare AI partnerships are accelerating, but the most common deal failures still come from the basics: vague asset definitions, unrealistic control promises, and contracts that ignore how models evolve.

The path to better agreements is not a longer contract. It is a more precise contract: one that translates ownership into rights, translates confidentiality into controls, and translates "future proof" into lifecycle obligations that can be executed.

AAIH will continue to convene cross-functional practitioners to share field-tested contracting patterns and operational practices that support responsible scaling of AI in healthcare.

Disclaimer: This document is for informational purposes only and does not constitute legal advice. Readers should consult counsel for advice regarding their specific circumstances.