Shumaker

OCTOBER 10, 2025 | PUBLICATION

"AI Related M&A Risks: Acquiring Hidden Liabilities from AI Models"

As AI becomes a cornerstone of business strategy, companies acquiring AI assets must navigate a complex web of legal and operational risks. This article explores how hidden liabilities in AI models can impact merger and acquisition (M&A) transactions and how to mitigate them.

INDUSTRY SECTOR

Technology

SERVICE LINE

Corporate, Tax & Transactions
Merger, Acquisitions & Divestitures
Technology, Data Privacy, Cybersecurity & AI

RELATED PROFESSIONALS

Lloyd J. Wilson

MEDIA CONTACT

Wendy M. Byrne
wbyrne@shumaker.com

Training Data Risks

Training data provenance is increasingly material to M&A deals as a result of recent litigation, including the proposed class action settlement of $3,000 per book for 500,000 pirated works used to train AI models (totaling $1.5 billion)[1] in *Bartz v. Anthropic*[2]. The *Bartz* case underscores an important point: how an AI model is trained can expose a company to outsized liability. Specifically, the allegations in *Bartz* centered on the use and retention of pirated books in model training. *Bartz* shines a light on several important points:

- Sourcing matters: training on lawfully acquired or licensed data is assessed very differently from training on pirated datasets.
- Retention matters: retaining infringing works in centralized repositories can itself be actionable, even if not reproduced verbatim in outputs.
- History matters: discontinuing problematic practices does not erase historical exposure, particularly in class actions.

Legal Considerations

As part of the risk assessment for M&A activity, greater attention should be given to several important questions:

- The acquisition method of training data: did it involve scraping, shadow libraries/torrents, vendor-supplied datasets, or negotiated copyright licenses?
- Storage and retention: was pirated or sensitive data persistently stored, and is there evidence of destruction?
- Identifiability and leakage: are copyrighted or confidential works traceable in outputs or AI model

weights?
- Scale and market impact: what is the volume, nature, and substitutability of the underlying works used to train the AI model(s)?
- Remediation: what steps have been taken to pivot to compliant practices, governance programs, audit trails, and vendor assurances?

Key stakeholders should understand the impact the broader legal landscape has on M&A activity and are advised to seek advice from legal professionals that can leverage their knowledge and understanding of several areas important to a valuation:

- Copyright/Intellectual Property (IP): there are unsettled rights around both training data and AI-generated outputs and open questions on ownership, fair use, text-and-data-mining, and embedded third-party rights.
- Privacy/consumer protection: obtaining consent and implementation of purpose limitations for AI model training and secondary use (regulators have required model destruction when data was improperly processed).
- Contracts with AI vendors: indemnity reversals, user-held IP risk, model training on inputs, data residency, audit rights, and confidentiality carve-outs.
- Regulatory changes: the EU AI Act phases in obligations by risk class; U.S. states and other jurisdictions are rapidly enacting AI and privacy rules with enforcement teeth.

Due Diligence Checklist

The M&A due diligence process needs to go beyond the capitalization table and computer code review. Thorough due diligence should incorporate:

- Data lineage and rights
    - Inventory all training, fine-tuning, eval and benchmarking datasets; trace source, license, terms of use, and opt-outs.
    - Confirm no use of shadow libraries or prohibited scraping; reconcile logs to written policies.
    - Verify destruction of infringing datasets and backups; obtain certificates and preservation exceptions for litigation holds.
- Model behavior and retention
    - Test for outputs that memorize specific details from the training data and reproducing such content in AI model outputs; assess whether model weights embed copyrighted or sensitive material.
    - Review model cards, research and development notes, and prompt/response logs; confirm data retention, redaction, and deletion settings.
- Governance and compliance
    - Evaluate AI governance, privacy, and security programs; AI impact assessments; explainability; and human-in-the-loop controls.
    - Examine privacy policies across time for training/use disclosures and consent; confirm cross-border transfer mechanisms.
    - Assess use of third-party AI tools by employees and vendor compliance (use of unapproved third-party AI tools by employees).
- Vendor and data supplier contracts
    - Scrutinize licenses for training rights, sublicensing, text and data mining carve-outs, restrictions, indemnities, and audit rights.
    - Identify any indemnity obligations owed by the target to platform providers for generated

outputs.
- Litigation and enforcement exposure
    - Map pending/threatened claims, takedown history, peak statutory damages scenarios, and insurance coverage.

Deal Structuring Recommendations

Due diligence recommendations should involve purchase agreement protections tailored for the risk profile:

- Representations and warranties that cover:
    - Clean title/rights in AI-generated assets (with breadth across copyright, trade secrets, database rights).
    - Lawful sourcing and documented rights for all training/fine-tuning datasets and benchmarks.
    - Compliance with privacy and data protection laws; no unauthorized processing; valid consents and disclosures.
    - Conformance with AI vendor terms; no prohibited uses; no undisclosed indemnity assumptions.
- Covenants and closing conditions
    - Pre-close destruction of infringing datasets and archives; disable training on customer/client data; implement guardrails.
    - Delivery of complete dataset inventories, licenses, and governance documentation.
- Indemnities and risk allocation
    - Specific indemnities for copyright/privacy claims and open-source/data license violations.
    - Escrow/holdbacks keyed to identified AI/IP issues; warranty and indemnity insurance where available and scoped to AI risks.
- Ongoing rights and restrictions
    - Post-close access to datasets/tools; vendor assignment/consent mechanics; non-compete/non-solicit where data or models are split.

Post-Close Integration

Precautions should also be instituted for proper AI model integration post-closing of M&A activity. Integration priorities should include:

- Strengthening AI governance: policies on approved tools, prohibited uses (e.g., evidence generation), verification protocols, logging, and audit.
- Remediating data practices: updating privacy notices/consents, implement deletion workflows, and re-papering key data licenses.
- Technical hardening: simulating real-world attacks to find and address security weaknesses, output verification, and implementing controls to prevent AI model memorization and leakage.
- Training and supervision: ensuring teams understand model limits, verification duties, and disclosure expectations.

Conclusion

AI innovation does not excuse weak provenance, privacy gaps, or thin governance. In today's enforcement and litigation environment, training data, retention practices, and vendor terms are core diligence items and not merely footnotes. A thorough diligence process and well-structured deal terms can help manage risk in AI-related acquisitions.

[1] Blake Brittain, *US Judge Preliminarily Approves $1.5 Billion Anthropic Copyright Settlement*, Reuters

(Sept. 25, 2025), https://www.reuters.com/sustainability/boards-policy-regulation/us-judge-approves-15-billion-anthropic-copyright-settlement-with-authors-2025-09-25/.

[2] *Bartz v. Anthropic*, PBC, 3:24-cv-05417, (N.D. Cal.).