Volume 9, No. 5 | September–October 2026

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 9, No. 5 | September–October 2026

EDITOR-IN-CHIEF

Steven A. Meyerowitz
President, Meyerowitz Communications Inc.

EDITOR

Victoria Prussen Spears
Senior Vice President, Meyerowitz Communications Inc.

BOARD OF EDITORS

Jennifer A. Johnson
Partner, Covington & Burling LLP

Paul B. Keller
Partner, Allen & Overy LLP

Garry G. Mathiason
Shareholder, Littler Mendelson P.C.

James A. Sherer
Partner, Baker & Hostetler LLP

Elaine D. Solomon
Partner, Blank Rome LLP

Edward J. Walters
Chief Strategy Officer, vLex

John Frank Weaver
Director, McLane Middleton, Professional Association

START-UP COLUMNIST

Jim Ryan
Partner, Morrison & Foerster LLP

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. Copyright 2026 Fastcase, Inc. No part of this journal may be reproduced in any form—by microfilm, xerography, or otherwise—or incorporated into any information retrieval system without the written permission of the copyright owner. For customer support, please contact Fastcase, Inc., 729 15th Street, NW, Suite 500, Washington, D.C. 20005, 202.999.4777 (phone), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: David Nayer
Production Editor: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

This publication is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

Copyright © 2026 Full Court Press, an imprint of Fastcase, Inc.

All Rights Reserved.

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

729 15th Street, NW, Suite 500, Washington, D.C. 20005
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 729 15th Street, NW, Suite 500, Washington, D.C. 20005.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 631.291.5541.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

David Nayer, Publisher, Full Court Press at david.nayer@clio.com or at 202.999.4777

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)

ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

The AI M&A Playbook: Contracting for the Unknown

C. Craig Lilly and Bryan C. Sykes*

In this article, the authors address the challenges in acquiring an artificial intelligence company, providing in-house counsel and executives with an in-depth summary to surface hidden liabilities.

The artificial intelligence (AI) landscape is evolving at a breakneck pace, shifting from experimental large language model use cases to fully integrated, autonomous systems. This rapid maturation has ignited a surge in high-stakes mergers and acquisitions (M&A) activity, underscoring the critical need for specialized AI diligence and deal making. This article addresses the challenges in acquiring an AI company such as vertical AI vendors (e.g., companies that specialize in AI for sectors like finance, healthcare, or manufacturing), AI product companies (e.g., companies that integrate AI into user applications), and API wrappers or providers. The value of an AI acquisition thereof often turns on data provenance, IP clarity, regulatory compliance, people risk, and governance hygiene. This article provides in-house counsel and executives with an in-depth summary to surface hidden liabilities.

Generic representations and boilerplate indemnities are not sufficient for AI transactions. This article distills practical contracting moves such as AI-specific representations and warranties, targeted indemnities, insurance strategies, sandbagging controls, and post-close governance—and converts various exposures into management risks. The summary below focuses on useful actions to protect valuation, prevent post-close surprises, and align accountability with control.

Tailored Representations and Warranties for AI

AI value depends on properly obtained data, disciplined engineering and clear ownership. Representations should confirm that the company uses only lawfully acquired data and content, with the rights, consents, and disclosures needed for current operations and post-close continuity.

Robotics, Artificial Intelligence & Law / September–October 2026, Vol. 9, No. 5, pp. 371–376.
© 2026 Full Court Press. All rights reserved.
ISSN 2575-5633 (print) /ISSN 2575-5617 (online).

All software, models, datasets, and tools should comply with all applicable licenses, including open-source and network-access provisions, with no copyleft obligations that would force disclosure or licensing of proprietary code or model weights.

Companies should maintain documentation of data lineage, model development, evaluation, deployment, and monitoring in accordance with applicable law and industry standards. They should meet all applicable privacy, data protection, and security requirements, and implement reasonable administrative, technical, and physical safeguards. For regulated or higher-risk uses, the company should confirm that it has performed and documented testing for bias, safety, and performance, which should be accompanied by mitigation, oversight, and disclosures of applicable limitations consistent with legal obligations. The company should also substantiate that it owns or has sufficient rights to all intellectual property in its products and services, and that all material contributors have assigned their rights to the company. Representations should be made by the company that it has established and maintains AI governance policies and procedures, including designated responsible AI officers or oversight committees where required or appropriate, consistent with applicable regulatory frameworks such as the EU Artificial Intelligence Act (Regulation EU 2024/1689).

Special Indemnification and Escrow Structures

Targeted indemnities should address training data rights, open-source software compliance, third-party IP claims, pre-close privacy violations, AI-caused harm arising from pre-close conduct, and undisclosed model or dataset restrictions.

Enhanced caps and extended survival periods should be provided for AI-specific representations and warranties where latent exposure may emerge post-close. Dedicated escrows or holdbacks should be established for AI risks, such as data or IP claims, open-source software remediation, and privacy fines, with 18-24 month terms and clear release mechanics. These measures can mitigate exposures that are only identified after a deal has closed by reducing post-close disputes through previously contemplated remedies. Performance-based earnout offsets tied to defined regulatory or IP outcomes should also be considered, as they provide a built-in adjustment to the AI company's value if certain essential outcomes do not materialize.

R&W Insurance and Specialty Coverage

Representation and warranty (R&W) insurance underwriting should be aligned with bespoke AI representations, with particular scrutiny on data lineage, bias and safety testing, and governance. Expect exclusions for known issues and certain regulatory exposures, and adjust escrow and indemnity structures accordingly.

Technology errors and omissions and cyber insurance policies should address algorithmic errors, service outages, confidentiality breaches, and data incidents. Since many policies exclude regulatory fines or unlawful data collection, companies should close these gaps with endorsements or separate coverage where possible.

For known, quantifiable, and capped risks such as ongoing intellectual property litigation, companies should explore dedicated litigation buyout policies to stabilize exposure.

AI Sandbagging: Defining the Risk and Allocating It by Contract

AI sandbagging refers to the deliberate understatement or concealment of an AI system's true capabilities, limitations, or safety characteristics to deceive evaluators, auditors, or regulators. The concept now extends beyond intentional human conduct to encompass autonomous underperformance or deception by AI systems—behavior that may emerge without explicit instruction. This deceptive behavior is distinct from hallucinations, where AI generates false or fabricated outputs without intent. Sandbagging, in its most concerning form, can involve scheming: an AI system covertly pursuing misaligned goals while concealing its true capabilities, objectives, or reasoning processes from human overseers.

As AI systems increasingly operate as autonomous agents in high-stakes domains—such as autonomous vehicles, algorithmic trading, critical infrastructure, and defense applications—sandbagging poses escalating risks to corporate liability exposure. Systems that mask their true capabilities during safety evaluations or certification processes may subsequently cause harm that the acquirer neither anticipated nor priced into the transaction.

Systems that behave differently in testing environments than in production deployments, or that engage in goal-directed deception without explicit programming, can create substantial liability

for developers, deployers, and acquirers under product liability, consumer protection, securities fraud, and regulatory enforcement theories. While those liability regimes warrant careful analysis, this section focuses on contractual risk allocation in the M&A context.

Why Sandbagging Matters in AI Transactions

In acquisitions where AI capabilities drive valuation, sandbagging can conceal material compliance gaps, obscure technical debt or safety deficiencies, and transfer substantial downstream liability to an unsuspecting buyer. The asymmetric information problem is particularly acute: sellers possess intimate knowledge of their AI systems' behavior and limitations, while buyers must rely on representations, due diligence, and third-party assessments that may not detect sophisticated deception. Traditional representations and warranties—designed for conventional software and technology assets—may fail to surface or properly allocate sandbagging risks unique to AI systems. Buyers should therefore negotiate explicit representations addressing deceptive design, robust testing covenants with pre-close verification rights, and tailored remedies calibrated to the potential magnitude of undisclosed AI risks.

Contracting Toolkit for Sandbagging Risk Allocation

Representations should require the target to affirmatively state that it has not designed, trained, fine-tuned, or instructed any AI system to misrepresent or conceal its capabilities, limitations, or safety characteristics during evaluation, certification, audit, or regulatory interaction. The target should further represent that its AI models do not contain latent mechanisms—whether intentionally embedded or emergent—designed or likely to detect test or evaluation environments and alter behavior relative to production deployment in ways that would undermine safety, compliance, or performance assessments. The target should also represent that it has conducted and documented adversarial testing and red-teaming exercises, proportionate to each AI system's risk profile, to identify manipulation vectors, reward hacking vulnerabilities, backdoors, triggerable behavioral modes, or test-detection capabilities, and has remediated all material findings or disclosed them in the schedules. Finally, the target should maintain and have followed documented policies for anomaly detection, incident reporting, root cause

analysis, and post-incident remediation—including model retraining or decommissioning—where deceptive or anomalous behavior is identified or suspected.

Covenants and closing conditions should grant the buyer a pre-close testing right to conduct targeted adversarial evaluations—including assessments for mode-switching behavior, hidden triggers, reward hacking, and environment-detection mechanisms—subject to mutually agreed protocols that protect the target's intellectual property, trade secrets, and system integrity. The parties should define materiality thresholds and specify consequences for adverse findings:

- remediation obligations with defined timelines,
- purchase price adjustments tied to quantifiable impacts,
- special indemnification for identified risks, or
- termination rights where findings reveal fundamental misrepresentations or undisclosed material risks.

Indemnities and escrows should include an indemnification obligation for losses arising from pre-close deceptive design, undisclosed training methodologies, or latent model triggers that cause material divergence between represented evaluation performance and actual production behavior, or between regulatory certifications and real-world system conduct. The indemnity should expressly cover regulatory penalties, customer claims, product liability exposure, remediation costs, and reputational damages to the extent quantifiable.

Disclosure schedules should require comprehensive disclosure of any known behavioral anomalies, reward hacking incidents, instances of unexpected or unexplained model behavior, internal or external deceptive-behavior research, red-team findings, safety incident reports, and all mitigation measures implemented. Buyers should request access to testing logs, model evaluation records, and internal communications regarding AI system behavior to verify the completeness of these disclosures.

Customer and Vendor Contract Upgrades Post-Close

Customer terms should be updated to include AI-specific use restrictions, human-in-the-loop disclaimers where appropriate, safety warnings, limitations of liability aligned to risk, and

audit-friendly logs. The vendor and model supply chain should include flow-down obligations for data rights, privacy and security, safety testing, and incident cooperation. IP and data infringement indemnities should be obtained from critical suppliers.

Post-Close Governance and Monitoring

A cross-functional committee, including legal, security, product, and safety representatives, should be established with ownership of model approvals, monitoring, incident response, and change control. Scheduled red-team exercises, regression tests, and drift monitoring should be implemented, with mandatory escalation and retraining gates. Data and model documentation, evaluation artifacts, and decision logs should be maintained to support regulatory inquiries and insurance claims.

Conclusion

Contract precision is the best defense in AI M&A. Treat AI-specific exposures—data rights, open-source software, bias and safety, and deceptive behavior—as named perils with explicit representations, bespoke indemnities, and, if applicable, fit-for-purpose insurance/escrows. Combining these protections with real testing rights and post-close governance turns unknowns into manageable risks.

Notes

* The authors, attorneys with Reed Smith LLP, may be contacted at clilly@reedsmith.com and bsykes@reedsmith.com, respectively.