UC BERKELEY
CENTER FOR LONG-TERM CYBERSECURITY

CLTC WHITE PAPER SERIES

AI Risk Governance

A STRUCTURED APPROACH FOR INVESTORS AND CORPORATE BOARDS

OUMOU LY

CLTC WHITE PAPER SERIES

AI Risk Governance

A STRUCTURED APPROACH FOR INVESTORS AND CORPORATE BOARDS

OUMOU LY

August 2026

This report is provided for general informational and educational purposes only. It does not constitute legal, regulatory, investment, or financial advice, and should not be relied upon as a substitute for consultation with qualified legal, compliance, or investment professionals. The frameworks and analysis presented reflect the author's assessment as of August 2026.

Contents

Terms and Definitions

AI	Artificial Intelligence
API	Application Programming Interface
ESG	Environmental, Social, And Governance
Inference process	The process by which an AI model receives inputs and produces outputs in operation, potentially exposing it to manipulation intended to compromise its outputs, security, or integrity.
M&A	Mergers And Acquisitions
Model Drift	Model drift is the decline in a machine learning model's predictive accuracy over time as the statistical patterns in real-world data diverge from those the model was originally trained on.
ML	Machine Learning
PE	Private Equity
Risk Management	The process of identifying, assessing, and mitigating risks to reduce potential harm and optimize outcomes under uncertainty
VC	Venture Capital

Foreword

As global spending on artificial intelligence approaches nearly $2.6 trillion (Gartner, 2026), investors are seeking strategies to optimize returns and demonstrate that capital deployed into AI will continue to generate value.

The stakes are highly consequential. Increasing concentration in a small number of AI infrastructure providers has made markets especially sensitive to poor monetization and repeated valuation correction. It has also given rise to new risk exposures that existing investor frameworks cannot sufficiently address.

In this environment, investors need methods to determine which companies can successfully scale AI-native businesses, provide or capture efficiency gains from AI integration, and avoid value-harming risks, such as reputational impacts, cybersecurity and privacy incidents, and poor financial performance. This report provides an initial approach to help investors better understand value-harming risks in the context of AI, as well as how to partner with companies to mitigate them.

In 2025, the project team conducted interviews with dozens of investors and investment advisors to gather the insights used to develop the approach outlined in this document. One of our most salient takeaways from those interviews is that, for many firms, the interrelation between harm to customers, society, and the environment and harm to returns is often undefined or underestimated. This report will be of assistance to investors looking to better understand these relationships conceptually, as well as those looking to develop quantitative approaches to represent them.

A second important reflection is that investor risk oversight can influence the norms that take shape within companies as they scale, and this can foster a high standard for risk diligence in a given sector or jurisdiction. In the aggregate, this produces better outcomes for customers and for society. This reflection was the primary impetus for undertaking this report.

Importantly, engaging in AI oversight also benefits investors themselves, including by helping widen exit pathways and by helping investors to exercise the influence of their capital in a way that contributes to returns. For investors with ESG mandates, AI oversight offers the added benefit of serving as a complementary mechanism for meeting responsible investment objectives.

While this report is primarily intended for investors, it may also be of benefit to corporate boards of AI-intensive and AI-native enterprises, to those designing oversight frameworks around AI, or to those working in sectors that rely on mechanisms for ongoing visibility into emerging risks.

We hope this work contributes to a broader shift in how the investment community understands their role in the innovation economy, and we are grateful to all who contributed their time and perspective to help shape this report.

Executive Summary

AUDIENCE AND SCOPE

This paper offers a structured approach for assessing AI risk in investment processes. It is written for practitioners, including fund and asset managers; investment, stewardship, and diligence teams; and the advisors who support them.

The approach is particularly applicable in the following contexts:

- **Early- and growth-stage investment in AI-native companies**. In these contexts, governance structures are still forming, making diligence an especially important input to investment decisions and a natural point of engagement on risk practices with founding teams.
- **Corporate venture capital**. In these contexts, investment decisions have financial and operational risk implications for the parent organization, which prompts a need for diligence approaches that account for downstream risks to the parent's operations and reputation.
- **Portfolio monitoring and governance**. For investors with existing AI-related exposures, ongoing visibility into system performance is essential for accurate valuation and risk management. This report supports stewardship and monitoring teams in developing frameworks for tracking how AI systems perform in real-world contexts and identifying corresponding material changes in a portfolio company's AI risk profile.

This approach is designed to serve two types of users:

1. Investors working in one or more of the above contexts who need information to ask the right questions of target companies and determine whether deeper diligence is warranted. Those investors should focus on the risk exposure proxy table on pages 14–16, which provides a practical starting point.
2. Investors working in one or more of the above contexts with the capacity and intent to engage in more comprehensive AI risk oversight, who may benefit from both the proxy table and diligence questions library on pages 17–28.

We hope this approach becomes a working tool in deal rooms and portfolio reviews, and that it is adapted for fund strategies and sector contexts. We are particularly hopeful that it can help smaller funds and first-time investors in AI-intensive sectors develop risk oversight practices that are proportionate to their resources but impactful in their value added.

OVERVIEW OF THE APPROACH

The approach focuses on three categories of AI-related risk that are most directly material to investment outcomes:

1. **Financial performance risks** arise when AI systems fail to deliver expected revenue or monetization outcomes, require unexpected capital expenditures for remediation, or create exposures that impair company valuations.
2. **Cybersecurity and privacy risks** arise when AI systems introduce vulnerabilities that expose sensitive data, expand attack surfaces for adversarial exploitation, or create exposure under applicable data protection frameworks.
3. **Reputational loss risks** emerge when AI systems produce harmful or incorrect outputs or fail to meet customer quality expectations, or through other means, damage the brand value of portfolio companies or their investors.

Although these risks are organized into separate categories, together they all have potential to affect a firm's financial performance, making them interconnected.

The approach is organized around five distinct phases of the investment lifecycle, with particular focus on establishing a governance trajectory that mitigates risk across the investment lifecycle and corresponding AI value chain.

- **Deal sourcing:** At the earliest stages of an investment, firms encounter potential portfolio companies and conduct preliminary assessments. Their priorities center on understanding a company's trajectory and financial viability before committing significant diligence resources.
- **Due diligence:** During deep evaluation of investment targets, investors require detailed visibility into AI-related exposures to identify risks early, uncover hidden liabilities, and validate the business model before committing capital.

- **Post investment:** During this phase, investors' priorities shift to governance structures and risk monitoring. This can include establishing structures like board seats, reporting requirements, and key performance indicators (KPIs).
- **Growth:** As portfolio companies scale operations, investors continuously track how risks surfaced during due diligence evolve and intensify. Their priorities include providing companies with resources, expertise, and guidance to scale.
- **Exit.** At the final stage of the lifecycle, investors are focused on preparing the company for a liquidity event such as an acquisition, merger, public offering, or secondary sale. Priorities include resolving any outstanding governance or liability issues and protecting the valuation of the company during the transaction.

This approach equips investors with the structure and tools to move from ad hoc awareness of AI-related risks to systematic risk oversight over the investment lifecycle.

Introduction

Investors typically sit several layers removed from the parts of the value chain that give rise to AI-related risk. Without a structured approach, they lack the tools to surface and manage the exposures that most directly threaten their returns.

The consequences of this gap are immediate and material. A target company in a merger or acquisition (M&A) transaction may have deployed AI systems that violate emerging regulatory requirements, which can create substantial compliance costs and legal liabilities that surface only after acquisition, when remediation is far more expensive.

Without AI-specific screening integrated into existing diligence processes, these value-destroying risks can go unpriced and unmanaged until the point at which they are most difficult and expensive to remediate.

This report addresses this critical gap by enabling investors to:

- **Identify and evaluate AI-specific risks** across their portfolios using structured assessment criteria.
- **Develop screening questions and evaluation protocols** that can be integrated into existing diligence processes for deal evaluation, portfolio monitoring, and exit preparation.
- **Categorize and prioritize risks** according to three categories that capture the exposures most material to investment outcomes.

AI Risk in Context

As it is used in this report, the term "AI-related risk" encompasses the range of negative outcomes to customers, companies, the environment, society, and investment performance that can arise from the development and deployment of AI systems.

The most significant investor risks can be categorized into three overarching groups: **(1) financial performance risks, (2) cybersecurity and privacy risks, and (3) reputational and trust risks**.

1. Financial performance risks

Financial performance risks are materially connected to investment outcomes, making them central to an investor's fiduciary oversight responsibilities.

These risks can include exit-phase complications, where AI governance issues surface during due diligence and impair valuation or deal completion; escalating insurance costs driven by AI liability exposure; and unexpected operational costs from AI infrastructure requirements, compute expenses, incident remediation activities, and more.

Specific financial performance risk exposures include:

- **Model performance degradation:** Over time, this reduces the value proposition of AI-enabled products or services.
- **Third-party dependency risks:** Pricing changes or service disruptions from foundation model or infrastructure providers affect unit economics, particularly during scaling.
- **Technical debt:** Costly refactoring requirements from AI implementations accumulate over time.
- **Valuation impairment:** Undisclosed or unidentified AI-related exposures that surface post-investment or post-acquisition may materially erode enterprise value.
- **Exit-phase complications:** AI governance issues that surface during acquisition due diligence can impair valuation or deal completion.
- **Insurance and liability costs:** AI liability exposure could drive up insurance premiums in ways not reflected in current operating cost structures.

A company that has not stress-tested its financial model against scenarios of degraded AI performance, increased compute costs, or third-party provider disruption may be carrying significantly more financial risk than its headline metrics suggest.

Investors who engage in AI oversight can better price these risks, influence AI risk management within portfolio companies, and reduce AI-related risk across the broader portfolio.

2. Cybersecurity and Privacy Risks

Cybersecurity and privacy risks encompass the technical vulnerabilities and process failures that arise across the full lifecycle of an AI system, including through development, deployment, and ongoing operation. These include risks such as unethical or undocumented data sourcing during the training of AI models, adversarial attacks on deployed models, and model drift that degrades performance over time. Left unaddressed, these exposures can induce regulatory penalties, hurt enterprise value, or expose sensitive and proprietary information.

Importantly, conventional cybersecurity does not directly apply to AI systems and is often insufficient to address the distinct threats they introduce. For instance, AI systems are distinguished by their dependence on large datasets for training; their integration with complex supply chains of third-party models, data vendors, and cloud infrastructure; and their susceptibility to a class of advanced attacks that include adversarial inputs, model extraction, data poisoning, and more.

Cybersecurity and privacy risk exposures include:

- **Data breaches:** Unauthorized access of AI systems exposing sensitive information including customer records, training data, model weights, or proprietary datasets, leading to regulatory penalties, costly remediation, and lasting damage to competitive positioning.
- **Third-party dependency risks:** Vulnerabilities introduced through third-party APIs, pre-trained models, or data vendors that can compromise dependent systems across the AI supply chain.
- **Adversarial attacks:** Manipulation of an AI system's inputs or outputs to extract sensitive information, degrade performance, trigger unintended behavior, or otherwise undermine the model's inference process.

- **Operational disruptions:** System outages that interrupt critical business processes and directly impact revenue, particularly for AI-as-a-service business models in which uptime is tied to commercial performance.

Without particular attention to these risks in the unique context of AI, investors are susceptible to significant financial and reputational costs.

3. Reputational and Trust Risks

Reputational and trust risks capture the potential for AI-related incidents to damage company credibility and stakeholder relationships. The primary mechanism of harm in this category is the erosion of the trust and confidence upon which franchise value, customer relationships, and market positioning depend.

Reputational risks in the context of AI are also distinct from those associated with conventional software product failures. AI systems can produce harmful, discriminatory, or factually incorrect outputs in ways that are difficult to predict or explain. Conversely, conventional software product failures often have traceable remediation pathways.

An AI-related incident often has no simple explanation, leaving companies unable to credibly account for their system's behavior. In the eyes of customers, partners, and regulators, that inability signals a deeper failure of governance and control.

Specific reputational and trust risk exposures include:

- **Public incidents:** When AI systems generate offensive or factually incorrect outputs that attract media scrutiny and public backlash.
- **Erosion of customer trust:** When AI systems fail to meet customer reliability or product quality expectations, leading to churn and contract terminations.
- **Association risks:** Reputational damage from connection to harmful AI applications or incidents, including for corporate venture investors when portfolio company failures reflect negatively on the parent organization. Moreover, an investor's association with poorly governed AI companies can limit access to future capital, deal flow, and acquirer interest, directly impairing long-term investment performance.

- **Competitive disadvantage:** Negative reputational impact from AI-related issues can hinder portfolio companies' commercial prospects, particularly in enterprise markets where vendors are increasingly evaluated on AI risk posture in certain jurisdictions.
- **Privacy and regulatory non-compliance:** Failure to meet obligations under applicable privacy regulations, including GDPR, CCPA, and a growing body of AI-specific legislation, exposing companies to enforcement action, significant financial penalties, and reputational harm from public disclosure of compliance failures.
- Talent/customer attrition due to perceived AI risk exposure.

Reputational risk is also highly asymmetric: the reputational benefit of AI systems functioning as intended is typically modest and largely expected by customers, while the reputational damage from a high-profile failure can be disproportionately large and persistent. Investors who do not assess reputational risk exposure as part of their AI risk oversight are therefore likely to underestimate the full downside risk profile of their AI-related investments.

INTERACTION BETWEEN RISK CATEGORIES

While these three risk categories are analytically distinct, they are interconnected in practice. For example, a cybersecurity incident will typically carry both direct financial cost in remediation, regulatory penalties, and litigation, as well as reputational consequences that affect customer retention and future commercial prospects.

Similarly, a public incident arising from harmful AI system outputs may trigger regulatory scrutiny that introduces compliance costs and, if the underlying failure reflects inadequate security or governance practices, surface financial liabilities that were not previously visible.

Investors who assess the risk categories in isolation, or who focus exclusively on the most immediately quantifiable exposures, are likely to underestimate their aggregate exposure. An integrated approach to AI risk oversight considers how risks in each category can propagate and amplify risks in the other categories and is essential for investors seeking to accurately price AI-related exposure.

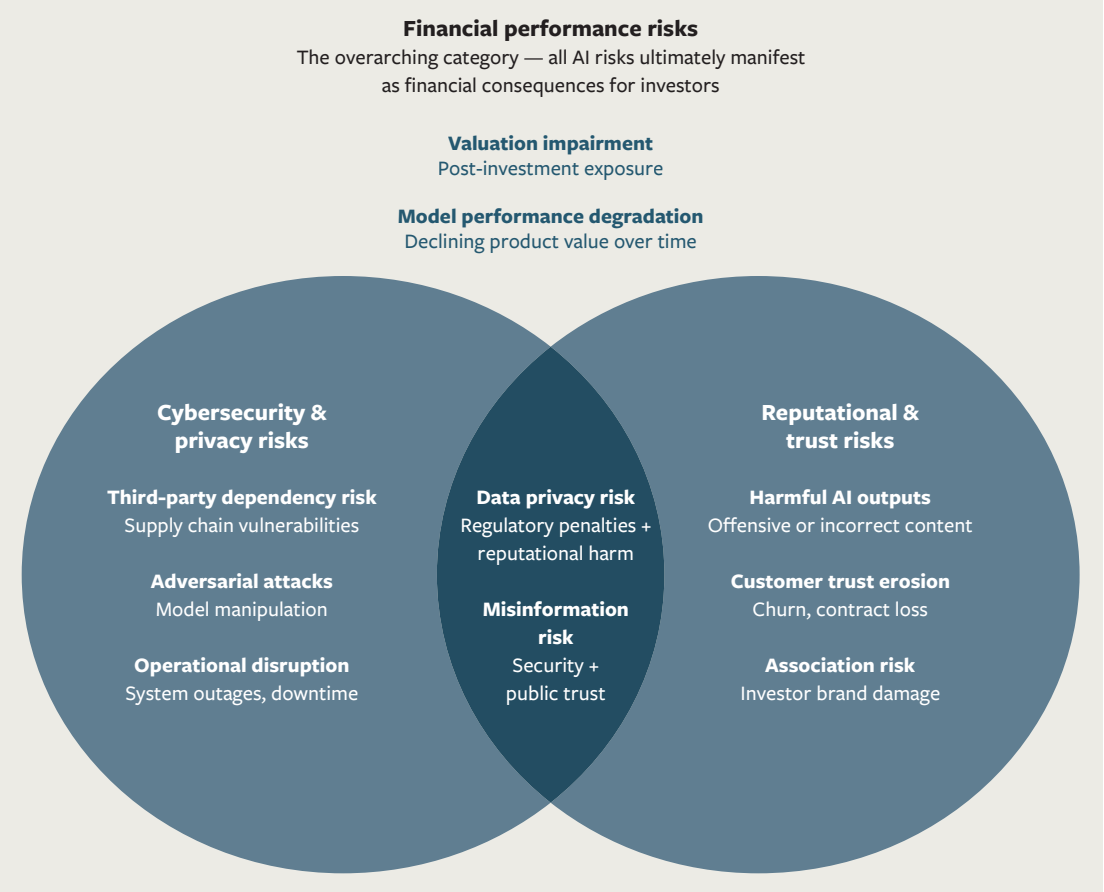

Figure 1: Risk Categorization Diagram.
This figure shows the different types of AI-related risks that can arise for investors and how they frequently overlap.

EXPOSURE VECTORS IN SCOPE

This paper is concerned with two primary sources of risk exposure for investors:

(1) portfolio risks from AI development and deployment within companies across investor holdings, and

(2) systemic risks from concentration in AI infrastructure, whereby a small number of providers, platforms, or capital flows underpin the broader industry, such that a shock to any one of them (or a broader repricing of AI valuations) could propagate losses across otherwise unrelated holdings simultaneously.

While risks from these vectors overlap and interact, they require distinct oversight mechanisms and mitigation strategies. To surface potential risk exposures, the structured approach offers two complementary oversight tools. The first is a proxy table that presents positive and negative risk signals across the three risk categories at each stage of the investment lifecycle, designed for investors who need efficient screening without extensive time commitment.

The second is a detailed questionnaire that enables more comprehensive assessment of financial, reputational, and cybersecurity risks at each lifecycle stage. This resource is intended for investors with the capacity for deeper engagement.

Together, these tools are supported by the AI Risk and Investor Oversight Matrix, which provides a framework for prioritizing risks by their potential impact on returns and the degree of investor influence over their outcome.

Methodology

From March 2025 to October 2025, the project team conducted interviews with more than 20 corporate, venture capital (VC), and equity investors to assess current AI investment and risk management practices, identify gaps in existing approaches, and determine where additional guidance is needed.

Interviews were structured to allow for flexibility to explore firm-specific practices in depth. Each interview followed a common discussion guide organized around four thematic areas: (1) internal AI use and governance, (2) visibility into AI use within portfolio companies, (3) existing risk oversight and control mechanisms, and (4) perceived gaps, constraints, and priorities related to AI risk management. Interviews typically lasted between 50 to 75 minutes and were conducted under conditions that encouraged candid discussion of internal processes and challenges.

Core prompts included, but were not limited to:

- How AI systems are used within the investor's own operations (e.g., investment decision-making, due diligence, portfolio monitoring, compliance, or internal productivity).
- Whether and how AI use is identified, documented, or monitored across portfolio companies, including reliance on third-party or foundation models.
- The extent to which AI-related risks (e.g., security, reliability, misuse, regulatory exposure, liability, or concentration risk) are considered during investment, post-investment oversight, or exit planning.
- What formal or informal controls exist to manage AI risk, such as policies, reporting requirements, escalation processes, or board-level engagement.

Follow-up questions were used to probe the depth and consistency of practices described by respondents. Responses were analyzed using a qualitative coding framework. Each interview was assessed across three core dimensions:

- Operational maturity: the degree to which AI-related policies, controls, and responsibilities are formalized, repeatable, and integrated into existing governance, diligence, and stewardship functions.

- Visibility into AI use cases: the extent of the investor's awareness of how AI is used within portfolio companies, including clarity around use cases, model types (where applicable), data dependencies, and third-party providers.
- Capacity for AI risk oversight: the investor's ability to engage meaningfully in AI risk governance, including access to relevant information, internal expertise, escalation pathways, and leverage mechanisms with portfolio companies.

For each dimension, responses were scored on an ordinal scale reflecting increasing levels of maturity and capability. Scoring was based on evidence provided by respondents regarding documented practices, consistency across teams or investments, and demonstrated ability to act on identified risks.

VALIDATION AND SYNTHESIS

To reduce interpretive bias, preliminary findings and scoring assumptions were refined iteratively. Aggregate insights were further validated through structured feedback sessions with a broader group of stakeholders, including selected investors, AI security researchers, and AI developers.

December 2025 Convening

In December 2025, the team from the UC Berkeley Center for Long-Term Cybersecurity's Artificial Intelligence Security Initiative (AISI) convened a private roundtable of more than 50 asset owners and allocators, AI security researchers, and representatives from leading AI development labs to discuss an early version of the approach and help investors conceptualize the range of potential actions they could take to begin engaging in AI governance. The discussion helped the project team form the approach outlined in this report.

Explanation of the Approach

This paper addresses risks arising from two primary sources:

(1) AI development and deployment within portfolio companies, and

(2) systemic risks from concentration in AI infrastructure across portfolios.

While these risk categories overlap and interact, they require distinct oversight mechanisms and mitigation strategies. The structured approach is organized around the investment lifecycle, enabling investors to identify financial, reputational, and cybersecurity exposures from both sources at each phase.

PHASE 1: DEAL SOURCING

At this stage, investors encounter one or more potential portfolio companies and conduct preliminary assessments. The structured approach provides guidance on:

- Initial flagging of risk based on the three risk categories (financial performance, reputational loss, and cybersecurity/privacy).
- Identifying whether target companies develop, deploy, or critically depend on AI systems.
- Determining whether AI-related risks warrant deeper technical due diligence.
- Establishing a picture of AI risk governance maturity.

PHASE 2: DUE DILIGENCE

During deep evaluation of investment targets, investors require detailed visibility into AI-related exposures. The structured approach includes information that may enable investors to:

- Assess companies' internal model development practices and data provenance.
- Evaluate third-party dependencies, regulatory compliance posture, and liability exposures.
- Identify AI systems embedded in target companies' operations and price associated risks.
- Make information requests from target companies (e.g., model cards, data lineage records, security assessments, or incident logs) that enable investors to assess potential risk exposures in selected risk domains.

PHASE 3: POST-INVESTMENT

Immediately following investment, investors need to establish governance structures and baseline risk monitoring. The structured approach specifies:

- Key performance indicators and metrics for tracking AI system performance and risk exposure.
- Integration of AI risk oversight into existing portfolio monitoring processes.
- Initial risk assessment and documentation of AI systems.
- Board-level reporting requirements for AI-related incidents.

PHASE 4: GROWTH

As portfolio companies scale operations and AI deployment, risks evolve and intensify. The structured approach addresses:

- Triggers for enhanced oversight when portfolio companies deploy new AI systems or increase dependency on existing ones.
- Monitoring protocols for tracking changes in third-party AI dependencies as companies scale.
- Assessment frameworks for evaluating new AI use cases and their associated risk profiles.
- Mechanisms for sharing risk intelligence across portfolio companies facing similar exposures.
- Escalation procedures for when risk metrics exceed predetermined thresholds.

PHASE 5: EXIT

As investors prepare to exit positions, AI-related risks affect exit valuations and transaction success. The structured approach includes:

- Pre-exit assessments to identify and remediate AI-related issues that could impair valuations or deter acquirers.
- Documentation requirements for demonstrating AI governance maturity to potential buyers.

- Disclosure considerations for communicating AI-related risks and risk management practices to acquirers.

- Post-exit liability considerations, particularly for investors with board seats or governance roles.

The proposed structured approach offers a two-tiered method for gaining visibility into AI-related risks across priority risk domains. The first tier, depicted in Table 1, presents a set of observable proxies (i.e., signals) that allow investors to gauge the magnitude of risk exposures throughout the investment lifecycle.

The second tier, outlined in Table 2, provides a more detailed set of considerations, evaluation criteria, and engagement prompts designed for investors with the capacity and intent to assess risk more thoroughly.

Table 1. Investor Oversight Risk Exposure Proxies.
This table provides signals and flags that enable investors to build a picture of a company's risk maturity across three risk domains.

Phase	Financial performance risks	Cybersecurity & privacy risks	Reputational loss risk
Deal flow	✖ **Negative signals** • No AI-specific unit economics disclosed CRITICAL • Unvalidated revenue assumptions HIGH • Hidden dependencies on subsidized compute or data access CRITICAL • Claimed margins incompatible with comparable AI companies HIGH	✖ **Negative signals** • Use of scraped or unlicensed training data CRITICAL • Unclear ownership of customer prompts/outputs CRITICAL • Prior undocumented incidents or vulnerabilities HIGH • No third-party risk controls MEDIUM	✖ **Negative signals** • Founder controversies or red flags HIGH • Overpromising capabilities in sales materials CRITICAL • Early customer complaints about AI accuracy HIGH • No disclosure of model limitations to customers HIGH
	✓ **Positive signals** • Clear cost model for training, inference, and compute HIGH • Plausible path to cost efficiency at scale HIGH • Benchmarked unit economics against AI-native peers MEDIUM	✓ **Positive signals** • Documented data provenance with licensing CRITICAL • Clear data processing agreements in place HIGH • Basic access controls and monitoring MEDIUM • Appropriate certifications available or in progress MEDIUM	✓ **Positive signals** • Transparent claims about capabilities and limits HIGH • Neutral to positive early market perception MEDIUM • Customer testimonials validate AI value proposition MEDIUM

Phase	Financial performance risks	Cybersecurity & privacy risks	Reputational loss risk
Due diligence	✖ **Negative signals** • Revenue model assumes linear scaling without compute adjustments CRITICAL • Customer acquisition cost (CAC) ignores per-customer fine-tuning costs HIGH • No benchmarking against AI-native business models HIGH • Dependence on a single foundation model provider HIGH	✖ **Negative signals** • No data lineage tracking or removal capabilities CRITICAL • Vague answers about data retention and versioning CRITICAL • Customer data used for training without explicit consent CRITICAL • No model versioning/rollback capability HIGH	✖ **Negative signals** • History of overstating AI capabilities CRITICAL • Demonstrable bias in outputs without mitigation CRITICAL • Customer churn due to performance vs. promise gaps HIGH • Marketing claims versus peer-reviewed benchmarks misaligned HIGH
	✓ **Positive signals** • Itemized R&D vs. production inference costs CRITICAL • Evidence of model efficiency improvements over time HIGH • Clear break-even inference volume identified HIGH • Multi-model or multi-provider strategy MEDIUM	✓ **Positive signals** • Complete data provenance documentation CRITICAL • SOC 2 Type II in progress or completed CRITICAL • Data processing agreements (DPAs) and tenant isolation architecture in place HIGH • Completed pen testing with remediation plan HIGH	✓ **Positive signals** • Transparent model cards with capabilities and limitations HIGH • Customer case studies with realistic metrics HIGH • Published incident response plan for model failures MEDIUM
Post-investment	✖ **Negative signals** • Inference costs growing faster than revenue CRITICAL • Margin compression due to model complexity creep HIGH • Over-reliance on single cloud provider with no negotiated rates MEDIUM	✖ **Negative signals** • Security incidents not disclosed to board within 24-48 hours CRITICAL • Delayed or incomplete compliance certifications HIGH • Growing technical debt in security infrastructure HIGH	✖ **Negative signals** • Negative press about AI outputs (bias, errors, harm) CRITICAL • Customer complaints escalating to social media HIGH • Growing complaints about AI accuracy in production HIGH • Regulatory scrutiny about AI practices HIGH
	✓ **Positive signals** • Quarterly improvement in cost-per-inference metrics CRITICAL • Implementation of caching, quantization, distillation strategies HIGH • Diversified compute strategy (multi-cloud, on-prem) MEDIUM	✓ **Positive signals** • Zero customer data breaches or prompt injection incidents CRITICAL • Regular third-party security audits with clean findings HIGH • Bug bounty program actively maintained HIGH • Automated compliance monitoring dashboards MEDIUM	✓ **Positive signals** • Proactive stakeholder communication about guardrails HIGH • Advisory board with domain experts engaged MEDIUM • Positive case studies from marquee customers MEDIUM

Phase	Financial performance risks	Cybersecurity & privacy risks	Reputational loss risk
Growth	✖ **Negative signals** • Unit economics deteriorating as customer base scales CRITICAL • High-touch custom deployments preventing standardization HIGH • Inability to pass compute costs to customers via pricing HIGH	✖ **Negative signals** • Compliance certifications not keeping pace with new markets CRITICAL • Inadequate data residency controls for international expansion CRITICAL • Shadow AI usage by customers creating liability exposure HIGH • Supply chain attacks on ML pipeline HIGH	✖ **Negative signals** • Brand association with AI incidents industry-wide HIGH • Competitive win/loss analysis showing trust concerns HIGH • G2/Gartner reviews mentioning AI reliability errors HIGH • Legislative proposals targeting business model MEDIUM
	✓ **Positive signals** • Model serving infrastructure scales sub-linearly with users CRITICAL • Pricing model aligned with value and compute consumption HIGH • Evidence of economies of scale in training and inference HIGH	✓ **Positive signals** • ISO 27001, GDPR, HIPAA readiness as needed CRITICAL • Regional data sovereignty solutions in place HIGH • Enterprise customer security reviews passing consistently HIGH • Red-team exercises identifying and fixing vulnerabilities MEDIUM	✓ **Positive signals** • Industry leadership in AI governance and standards-setting HIGH • Thought leadership establishing company as responsible actor MEDIUM • Customer retention despite AI market commoditization HIGH
Exit	✖ **Negative signals** • Acquirer discovers undisclosed compute liabilities or contracts CRITICAL • Financial model based on outdated or soon obsolete AI architecture CRITICAL • Revenue multiples don't account for AI margin structure HIGH	✖ **Negative signals** • Data licensing issues surface during acquirer due diligence CRITICAL • Customer data processing agreements block acquisition CRITICAL • Unresolved security debt or technical vulnerabilities HIGH	✖ **Negative signals** • Pending litigation or IP disputes related to training data CRITICAL • Brand damage from AI incidents reduces valuation multiple HIGH • Customer concentration risk due to reputational constraints MEDIUM
	✓ **Positive signals** • Clean, auditable separation of AI vs. non-AI costs CRITICAL • Demonstrable moat via proprietary data or model performance CRITICAL • Scalable margins validated across customer cohorts HIGH	✓ **Positive signals** • Complete data lineage documentation ready for review CRITICAL • All compliance certifications current and transferable CRITICAL • Security posture meets or exceeds acquirer standards HIGH • No outstanding regulatory inquiries or liabilities HIGH	✓ **Positive signals** • Clean IP and no material legal or regulatory exposure CRITICAL • Strong brand equity as a trusted, responsible AI provider HIGH • Customer references validate AI practices MEDIUM

Proxies are observable indicators that offer investors preliminary insight into AI-related risks without requiring specialized technical expertise or extensive investigation. Proxies act as early warning signals, highlighting where potential exposures may exist, the relative magnitude of the exposures, and the likelihood that they could propagate further. While proxies do not provide a definitive measure of risk, they enable investors to allocate resources strategically to mitigate potential harms.

For each investment phase, the table provides a risk signal in response to each proxy. For example, at the deal flow stage, proxies help distinguish companies that merit deeper scrutiny. During due diligence, proxies inform the collection of concrete evidence, including incident histories, contractual obligations, financial stress tests, regulatory compliance documentation, and security assessments. During the post-investment phase, proxies support the establishment of governance structures, reporting protocols, and ongoing monitoring mechanisms aligned with a company's operational complexity and risk profile.

As portfolio companies scale, using the table can facilitate dynamic oversight by helping to identify emerging risks related to new AI deployments, increased reliance on third-party providers, and more. At the exit phase of a deal, the table functions as a comprehensive checklist to ensure that AI-related exposures have been addressed and documented.

Firm-specific risk proxies can be uncovered using a diligence tool currently in development. For investors with the ability to engage in deeper risk oversight, a more complete table is included below.

Table 2: Recommended Questions and Actions for Expanded Investor Oversight.
This table provides an expanded set of diligence questions and a task inventory for investors with capacity and intent to engage in deeper screening.

Deal Flow	
Financial Performance Risks	**Capital Efficiency & Economics** • How do your unit economics (cost per inference/prediction) trend as you scale? • What percentage of revenue is dependent on third-party AI services (e.g., AI model developers, cloud storage services), and what are the pricing risks? • Have you experienced any AI-related incidents that resulted in customer refunds, contract terminations, or other types of financial losses? • What is your customer acquisition cost and how does AI performance impact conversion and retention?

Deal Flow	
	Infrastructure & Operational Costs • What cloud infrastructure commitments have you made and what are the penalty terms? • What is your current burn rate on AI R&D and infrastructure versus revenue generation? • Have you stress-tested your financial model against scenarios of degraded AI performance or increased compute costs?
	Insurance & Liability Exposure • What insurance coverage do you carry (e.g., errors and omissions, cyber, product liability) and what are the premiums? • Are current and projected insurance costs accounted for? • Have you received any insurance claims or faced difficulty obtaining coverage due to AI-related risks? • What indemnification obligations have you accepted in customer contracts for AI outputs or decisions? • Are there any pending or threatened legal claims related to AI performance, bias, or harm?
	Exit Readiness • Have previous investors flagged any AI-related concerns that could impact future fundraising or exit? • What percentage of customers have contractual provisions that could be triggered by ownership changes? • Are there any regulatory approvals or compliance certifications that would need to transfer to an acquirer?
Reputational Loss Risks	**Public Perception & Controversies** • Has your company or its AI technology been featured in negative media coverage or public controversies? • Have you faced backlash from customers, employees, or advocacy groups regarding your AI applications? • What is your positioning on controversial use cases (e.g., surveillance, weapons, political campaigns)? • Have you withdrawn from or declined to pursue any markets or customers due to ethical concerns?
	Brand & Market Positioning • Have you received any recognition or certification for responsible AI practices (e.g., SOC 2, ISO certifications)? • What percentage of enterprise deals require AI governance assurances or audits before contract signature? • Have you lost deals specifically due to AI trust, transparency, or governance concerns?
Cybersecurity and Privacy Risks	**Data Security & Breach History** • Have you experienced any data breaches, security incidents, or unauthorized access in the past three years? • What data security certifications do you hold (e.g., SOC 2 Type II, ISO 27001) and when were they last audited? • How is training data secured, and who has access to raw datasets versus processed/anonymized versions? • What is your encryption approach for data at rest, in transit, and during model training? • Do you keep a log of AI-related incidents? • How are model weights being protected?

Deal Flow	
	Privacy & Data Governance • What types of personal data do you collect, and under what legal basis (e.g., consent, legitimate interest, etc.)? • How do you handle data subject rights requests (e.g., access, deletion, portability) at scale? • What is your data retention policy, and how do you ensure deletion cascades through all systems, including AI models? • Have you had to face any regulatory penalties?
	Supply Chain & Third-Party Dependencies • What third-party AI services, APIs, or models do you depend on (e.g., cloud providers, foundation models, data vendors)? • What happens to your service if a critical third-party provider has an outage or terminates your access? • How do you vet and monitor the security posture of your AI supply chain? • Do you have contractual commitments regarding data handling with third-party providers? • Have you experienced any security incidents originating from a vendor or partner? • Are models deployed as open-source or via an API?

Due Diligence	
Financial Performance Risks	**Technical Due Diligence - Cost Structure** • What are your cloud provider contract terms, committed use discounts, and flexibility to switch providers? • What is your roadmap for cost optimization (e.g., model compression, efficient architectures, custom hardware)? • Have you conducted sensitivity analysis on how margin varies with model performance requirements?
	Incident Cost Analysis • For each major incident, what was the financial impact (e.g., revenue loss, credits, legal costs, remediation expenses)? • What are your insurance deductibles and coverage limits for different incident types? • Have insurance premiums increased, and what factors drove the changes? • What are the anticipated insurance premium increases?
	Cash Runway & Capital Requirements • What are your projected cash needs for the next 12-24 months, broken down by category? • What assumptions about AI R&D spending, compute costs, and headcount are embedded in your model? • When do you expect to reach profitability? • Under what scenarios would you need to raise additional capital before reaching profitability?
	Exit Scenarios & Value Preservation • What percentage of enterprise value is attributable to proprietary AI/ML capabilities versus other assets? • Are there any regulatory approvals required for change of control in your key markets? • What technical dependencies or key personnel would an acquirer need to retain? • Have you identified any IP, compliance, or technical issues that could surface in a buyer's due diligence?

	Due Diligence
Reputational Loss Risks	**Incident & Controversy History** • Provide a complete timeline of any public controversies, media incidents, or customer complaints related to AI. • For these incidents, what were the root causes, how were they addressed, and what preventive measures were implemented? • Have you issued public statements or apologies related to AI incidents? If so, provide copies. • What customer communication occurred following incidents, and what was the sentiment? • Were the incidents reported in public AI incident databases?[1] • Were the incidents reported to any oversight bodies (e.g., the EU AI office)?
	Internal Culture & Ethics • What is your employee turnover rate, particularly among AI/ML staff? • Are there required employee trainings around responsible AI?
	Governance & Accountability • Are roles and responsibilities regarding the safety framework established and clear?
	Stakeholder Feedback Mechanisms • How do you solicit and incorporate feedback from affected communities or users? • What external advisory relationships do you maintain (e.g., with academics, ethicists, domain experts)? • Describe any participatory design or inclusive development processes for AI features.
	Public Commitments & Accountability • Do you publish transparency reports, AI incident disclosures, or model cards? • What industry pledges, frameworks, or standards have you committed to follow? • Have you submitted a report to the G7 Reporting Framework (OECD n.d)?
Cybersecurity and Privacy Risks	**Security Architecture & Controls** • Provide your security architecture documentation, including network diagrams, data flows, and access controls. • Walk through your identity and access management approach for employees, contractors, and systems. • What security controls protect your training data, model weights, and inference infrastructure? • How do you segregate customer data, and can one customer's data leak into another's results? • What penetration testing, vulnerability scanning, and red-team exercises have you conducted? • What percentage of enterprise RFPs include cybersecurity testing requirements?
	Incident Response Capabilities • Provide your incident response plan and evidence of tabletop exercises or simulations. • Walk through your most recent security incident, from detection through resolution. • What is your mean time to detect (MTTD) and mean time to respond (MTTR) for different incident types? • Who is on your incident response team, and what external resources (e.g., forensics, legal) are pre-arranged? • How do you determine notification obligations and communicate with affected parties?

1 Examples include the AI Incident Database (AIID n.d.), the ATLAS AI Incidents (MITRE n.d.), the MIT AI Incident Tracker (MIT 2025), and the AI Incidents and Hazards Monitor (OECD.AI n.d.).

Due Diligence	
	Privacy Program Maturity • Provide documentation of your privacy program, including policies, procedures, and training materials. • Walk through your data inventory: what data you collect, legal basis, retention periods, and deletion processes. • How do you conduct Privacy Impact Assessments (PIAs) or Data Protection Impact Assessments (DPIAs)? • Provide evidence of data subject rights request handling (e.g., volume, response times, escalations).
	Regulatory Compliance Status • Provide all compliance certifications, audit reports, and regulatory correspondence from the past three years. • Have you received any regulatory inquiries, complaints, or enforcement actions? Provide documentation. • What privacy regulations apply to your business (e.g., GDPR, CCPA, HIPAA, etc.) and how do you ensure compliance? • Do you conduct Data Protection Impact Assessments (DPIAs) for high-risk AI processing? • What cross-border data transfer mechanisms do you rely on (e.g., Standard Contractual Clauses, adequacy decisions)?
	Supply Chain Security • Provide a complete inventory of third-party AI dependencies (e.g., cloud, APIs, models, data vendors, tools, etc.). • What security assessments or audits do you conduct of critical vendors? • Provide copies of data processing agreements (DPAs) and security exhibits with key vendors. • What is your process for monitoring vendor security posture and responding to their incidents? • How do you manage open-source dependencies, vulnerability scanning, and patching? • What happens if a critical vendor (e.g., cloud provider, API) experiences an outage or security incident?
	Business Continuity & Resilience • Provide your business continuity and disaster recovery plans. • What is your RTO/RPO for critical AI systems, and what redundancy exists? • How quickly can you failover to backup infrastructure, and what degradation is acceptable? • Walk through your model deployment pipeline and rollback procedures. • What monitoring and alerting exists for model performance degradation, data drift, or anomalous behavior?
Post-Investment	
Financial Performance Risks	**Baseline Metrics & Monitoring** • Establish monthly reporting on: revenue, gross margin, compute costs as percentage of revenue, customer churn, and cash runway. • What are your quarterly targets for unit economics improvement and cost optimization? • Set thresholds for early warning indicators: What level of cost increase or margin compression triggers board escalation? • What financial impact would a significant AI incident have, and what reserves or insurance exist?

Post-Investment	
	Cost Management & Optimization • What initiatives are underway to improve AI infrastructure efficiency and reduce costs? • How are you tracking actual compute usage versus budget, and what variance is acceptable? • What experiments or pilots are running to test alternative models, providers, or architectures? • Are you considering custom hardware, edge deployment, or other cost reduction strategies?
	Insurance & Risk Transfer • Have you obtained or increased errors and emissions E&O, cyber, and product liability insurance as discussed during diligence? • What is the timeline for obtaining any missing insurance coverage identified during diligence? • Are there any customer contracts requiring insurance coverage increases or additional indemnification?
	Crisis Response Planning • What financial reserves or credit facilities exist to respond to a major AI incident? • Who has authority to approve emergency spending for incident response or remediation? • What customer communication and service credit policies are in place for AI performance failures?
Reputational Loss Risks	**Baseline Metrics & Monitoring** • Establish monthly reporting on: revenue, gross margin, compute costs as percentage of revenue, customer churn, and cash runway. • What are your quarterly targets for unit economics improvement and cost optimization? • Set thresholds for early warning indicators: What level of cost increase or margin compression triggers board escalation? • What financial impact would a significant AI incident have, and what reserves or insurance exist? • Are AI incidents reported in public AI incident databases?[2]
	Cost Management & Optimization • What initiatives are underway to improve AI infrastructure efficiency and reduce costs? • How are you tracking actual resource usage (e.g., compute) versus budget, and what variance is acceptable? • What experiments or pilots are running to test alternative models, providers, or architectures? • Are you considering custom hardware, edge deployment, or other cost reduction strategies?
	Insurance & Risk Transfer • Have you obtained or increased E&O, cyber, and product liability insurance as discussed during diligence? • What is the timeline for obtaining any missing insurance coverage identified during diligence? • Are there any customer contracts requiring insurance coverage increases or additional indemnification?
	Crisis Response Planning • What financial reserves or credit facilities exist to respond to a major AI incident? • Who has authority to approve emergency spending for incident response or remediation? • What customer communication and service credit policies are in place for AI performance failures?

2 Examples include the AI Incident Database (AIID n.d.), the ATLAS AI Incidents (MITRE n.d.), the MIT AI Incident Tracker (MIT 2025), and the AI Incidents and Hazards Monitor (OECD.AI n.d.).

Post-Investment	
Cybersecurity and Privacy Risks	**Security Posture Baseline** • Establish risk tolerance levels, or risk tiers,[3] including responses to each risk level if reached. • Establish quarterly security metrics reporting: incidents, vulnerabilities, penetration test findings, compliance status. • Set remediation timelines for any gaps identified during due diligence. • What is the budget and timeline for achieving target security certifications (SOC 2, ISO 27001)? • Assign executive accountability for information security (CISO or equivalent).
	Incident Response Readiness • Validate that incident response plan is documented and tested, and that stakeholders are trained. • Conduct tabletop exercise within first 90 days covering AI-specific scenarios (e.g., model poisoning, data leak, inference manipulation). • Establish notification protocols to board and investors for different incident severity levels. • Confirm cyber insurance coverage is in place and all stakeholders understand the claims process.
	Privacy Program Implementation • Establish quarterly privacy metrics: data subject requests, processing activities, DPIA completions, training completion rates. • Set timeline for completing any outstanding DPIAs for high-risk AI processing. • Implement or validate data retention and deletion automation. • Establish a privacy review process for new data collection or AI feature deployment.
	Supply Chain Risk Management • Implement a vendor risk management program that covers inventory, risk assessments, monitoring, and incident protocols. • Establish service-level agreements (SLAs) with critical AI vendors, including notification requirements for their security incidents. • Set timeline for achieving contractual protections (DPAs, security exhibits) with all critical vendors. • Implement monitoring for open-source dependency vulnerabilities and patching processes.
	Resilience & Business Continuity • Validate that disaster recovery (DR) and business continuity plans are documented and tested. • Conduct DR test within first six months and establish annual testing cadence. • Establish monitoring and alerting for critical AI systems with defined escalation paths. • Develop a runbook for common AI system failures (e.g., model degradation, data pipeline breaks, API failures). • How much is the company investing in responsible AI research and implementation?

Growth	
Financial Performance Risks	**Scaling Economics** • As you enter new markets or launch new products, how are unit economics trending versus projections? • What investments in AI infrastructure are required to support growth (e.g., 2x–3x), and how will this impact margins? • Are you achieving economies of scale in compute costs, or are costs scaling linearly with volume? • What percentage of new revenue comes from higher-risk AI use cases, and how does this affect your risk profile?

3 AI companies may already have risk tiers and risk tolerance outlined in their safety frameworks. For example, the OpenAI Preparedness Framework (OpenAI 2025) describes the "critical" capability thresholds for cybersecurity risk, and states that if this threshold is reached a halt in development may be necessary.

Growth	
	Geographic & Use Case Expansion • What additional regulatory compliance costs arise from new jurisdictions (e.g., AI Act obligations)? • Do new markets require data localization, which increases infrastructure costs? • Have you stress-tested financial models against potential regulatory enforcement or product restrictions? • What insurance cost increases result from expansion into new geographies or higher-risk verticals?
	Exit Value Optimization • What strategic initiatives would maximize valuation in an exit scenario (e.g., customer concentration reduction, margin improvement, IP strengthening)? • Are you proactively addressing any "red flags" that surfaced in prior fundraising or M&A discussions? • What financial metrics are most valued by strategic acquirers in your space? • Are ESG materials prioritized, collected, and stored?
Reputational Loss Risks	**Controversy & Incident Tracking** • Have there been any AI-related incidents, customer complaints, or media coverage in the past quarter? • What is trending in customer feedback regarding AI transparency, fairness, or reliability? • Are you tracking sentiment among different stakeholder groups (e.g., customers, employees, regulators, advocacy groups)? • Have any new use cases or customer types introduced reputational risk?
	Geographic & Cultural Considerations • As you expand to new geographies, how are you adapting AI models and practices for local contexts? • Have you conducted fairness testing for new demographic groups or cultural contexts? • What local partnerships or advisory relationships exist to guide responsible AI in new markets? • Are you monitoring for cultural insensitivity or bias in AI outputs across different regions?
	Stakeholder Engagement • How are you engaging with stakeholders proactively in new jurisdictions? • What industry associations, standard-setting bodies, or policy initiatives are you participating in? • Are you building relationships with academic researchers, civil society groups, or domain experts in expansion areas? • How are you communicating AI risk management practices to an increasingly diverse stakeholder base?
	Brand & Market Position • What competitive intelligence exists on how peers are positioning on AI ethics and governance? • Are you seeing enterprise RFPs with increasingly stringent AI governance requirements? • How are you publicly demonstrating commitment to responsible AI (e.g., transparency reports, certifications, external audits)?
Cybersecurity and Privacy Risks	**Security Scaling Challenges** • As you scale infrastructure and enter new markets, how is your attack surface changing? • What new security risks arise from geographic expansion (e.g., data localization, jurisdiction-specific threats)? • Are security resources and controls scaling proportionally with business growth? • What new third-party dependencies have been introduced, and how are they vetted?

Growth	
	Privacy Complexity • Under how many different privacy regimes do you now operate, and how do you ensure compliance across all? • Are you conducting DPIAs for each new market or use case as required? • How do you handle conflicting privacy requirements across jurisdictions (e.g., data localization vs. centralized processing)? • What privacy-enhancing and privacy-protecting measures are you deploying to minimize data collection or enable compliant cross-border transfer?
	Incident Response at Scale • Have there been any security incidents or breaches in the past quarter? What was learned, and have these incidents resulted in a change in the company's risk landscape? • Are incident detection and response capabilities (including detection and response timing) keeping pace with organizational complexity? • What tabletop exercises or simulations have been conducted recently?
	Supply Chain Evolution • What new vendors, APIs, or third-party AI services have been onboarded? • How are you managing the expanding web of data processing agreements and security commitments? • Have any vendors experienced security incidents that could affect your systems? • Are you conducting regular vendor risk assessments as the supply chain grows?
	Resilience & Availability • As system complexity increases, are you maintaining or improving uptime SLAs? • What redundancy and failover capabilities exist across multiple regions? • How quickly can you detect and respond to model degradation or data quality issues at scale? • Are you stress-testing disaster recovery plans against realistic growth scenarios?
	AI-Specific Security Risks • Are you monitoring for adversarial attacks, model extraction, or data poisoning attempts? • What red teaming or security testing is conducted on AI systems specifically? • How are you evaluating the system for emerging or novel threats and vulnerabilities? • Are you monitoring and submitting discovered vulnerabilities to public databases such as the AI Vulnerability Database (AIVD n.d.)? • How do you protect model weights and proprietary training data as the organization scales? • Are you implementing AI-specific security controls (e.g., input validation, output filtering, anomaly detection)?
Exit	
Financial Performance Risks	**Financial Cleanup & Optimization** • Review and normalize financial statements to remove one-time AI incidents or investments from operating metrics. • What is the clean earnings before interest, taxes, depreciation, and amortization (EBITDA) or revenue run-rate, excluding non-recurring AI remediation costs? • Have all AI-related contingent liabilities been quantified and disclosed (e.g., pending claims, warranty reserves, insurance deductibles)? • What is the expected financial impact of any outstanding AI compliance gaps that must be remediated pre-exit?

Exit	
	Cost Structure Validation • Can you defend current compute costs and demonstrate a path to continued margin improvement? • What cloud or infrastructure commitments exist that could transfer to the acquirer or create liabilities? • Are there any upcoming contract renewals with AI vendors that could materially impact costs? • What cost optimization initiatives are in flight that could improve valuation?
	Revenue Quality & Sustainability • What percentage of revenue comes from customers with AI-specific performance guarantees or SLAs? • Are there any customers at risk of churn due to AI performance, competitive alternatives, or pricing pressure? • Prepare analysis of revenue retention, expansion, and churn drivers with AI performance as a variable.
	Insurance & Liability Documentation • Provide complete insurance history: coverage types, limits, premium trends, claims, and denials. • What tail coverage or extended reporting period provisions exist that could transfer to the acquirer? • Compile all customer indemnification obligations related to AI into a single reference document. • What reserves exist for potential AI-related liabilities (e.g., litigation, regulatory fines, warranty claims)?
	Exit Economics • Model exit scenarios across different valuation multiples showing impact of any outstanding AI risks. • What earn-out or escrow provisions might buyers require based on AI performance or compliance uncertainties? • Prepare waterfall analysis showing how AI-related contingencies affect proceeds to investors.
Reputational Loss Risks	**Stakeholder Reference Preparation** • Prepare case studies demonstrating successful AI deployments and customer satisfaction. • Document any awards, certifications, or third-party validations of responsible AI practices. • Prepare testimonials or reference letters from key partners, advisors, or independent experts.
	Governance & Culture Documentation • Document the AI governance framework, including policies, frameworks, committee charters, independent third-party evaluators or auditors, escalation protocols, and decision-making authorities. • Prepare materials demonstrating response. • Compile complete chronology of all AI-related incidents, controversies, media coverage, and customer complaints. • For each incident, document the root cause, financial impact, remediation steps, and lessons learned. • Identify any unresolved or recurring issues that could concern buyers. • Compile evidence of proactive risk management, which may include safety testing, fairness audits, red teaming, bug bounties, and external reviews. • Document stakeholder engagement activities through advisory boards, community consultations, and regulatory interactions.

Exit	
	Media & Public Perception Audit • Conduct media analysis identifying positive and negative coverage, overall sentiment, and emerging narratives. • Prepare response briefs for any negative coverage or controversies that may surface in buyer diligence. • Document all public commitments, statements, or pledges related to AI ethics that buyer would inherit.
	ESG & Responsible AI Positioning • Prepare ESG materials highlighting responsible AI practices, governance, and social impact. • Does the company have an AI safety and security framework? • Does the framework align with regulatory requirements (e.g., SB53, the EU AI Act)? • Does the framework align with best practices and standards (i.e., ISO42001, EU GPAI CoP, NIST AI RMF)? • Document alignment with emerging AI regulations (e.g., EU AI Act, state laws) and readiness for compliance. • Compile diversity and inclusion data for AI teams and fairness testing across demographic groups. • Prepare a narrative on how the company has evolved responsible AI practices and what the buyer is acquiring.
Cybersecurity and Privacy Risks	• Provide current security audit reports (e.g., SOCS, ISO 27001, ISO 42001) with no significant findings outstanding. • Compile penetration test results, vulnerability assessments, and remediation evidence from the past 12-24 months. • Document security architecture, controls inventory, and evidence of continuous monitoring. • Prepare summary of security incidents from past 3 years with resolution and ongoing impacts.
	Privacy Compliance Evidence • Compile all privacy certifications, audit reports, and regulatory correspondence demonstrating clean compliance records. • Provide evidence of DPIA completion for all high-risk AI processing activities. • Document data inventory, retention policies, and deletion automation capabilities. • Prepare data subject rights request log showing timely and complete responses. • Compile all data processing agreements (DPAs) with customers, vendors, and partners.
	Incident History Transparency • Prepare a complete incident log with evidence that all incidents were properly contained, remediated, and disclosed. • Document notification made to affected parties, regulators, and insurers where required. • Provide post-incident reviews demonstrating root cause analysis and preventive measures. • Identify any incidents with ongoing monitoring requirements or potential tail risk.
	Regulatory & Compliance Status • Confirm no outstanding regulatory inquiries, investigations, or enforcement actions. • Compile evidence of compliance with all applicable privacy and AI regulations across operating jurisdictions. • Document any regulatory exemptions, waivers, or special arrangements that may not transfer. • Prepare a summary of the regulatory landscape and a compliance roadmap that the buyer would inherit.

Exit	
	Supply Chain & Dependency Audit • Provide complete vendor inventory with risk assessments, contract terms, and security documentation. • Identify any vendors with security concerns, contract renewal risks, or dependencies that could affect transition. • Document all third-party data processing relationships and confirm DPAs are in place and transferable. • Assess which vendor relationships are critical and may require buyer consent or renegotiation.
	Business Continuity & Resilience • Provide current disaster recovery and business continuity plan (DR/BCP) documentation with evidence of successful testing within the past 12 months. • Document uptime history, incident resolution times, and SLA performance. • Demonstrate redundancy, failover capabilities, and resilience across critical AI systems. • Identify any single points of failure or resilience gaps that should be addressed pre-exit.
	Data Room Preparation • Organize comprehensive, logically structured documentation for security, privacy, and compliance. • Prepare an executive summary highlighting security posture, compliance status, and risk management maturity. • Anticipate buyer questions and prepare FAQ or supplemental materials. • Identify any sensitive information requiring additional NDAs or restricted access during diligence.
	Transition Planning • Document critical security personnel, knowledge, and capabilities that must transfer to the buyer. • Identify any security tool licenses, vendor contracts, or certifications that need buyer assignment or renewal. • Prepare an onboarding plan for the buyer's security team covering systems, processes, and ongoing obligations. • Plan for post-close security monitoring to ensure continuity and address any transition risks.

AI RISK AND INVESTOR OVERSIGHT MATRIX

To supplement the proxy table and expanded set of diligence questions, the AI Risk and Investor Oversight Matrix offers broad strategies to mitigate the impact of risks identified through the structured approach. It compares the level of influence an investor has over the evolving risk situation with the risk's impact on potential returns.

The x-axis of the matrix gauges how much leverage an investor has over the risk situation. High leverage corresponds to a risk situation within an investor's direct governance scope, while an investor with low leverage is not able to significantly influence the situation with their current resources and position.

The y-axis of the matrix measures the severity of the risk situation's impact on an investor's returns. A risk situation placed higher on this axis has a greater financial impact on the portfolio, while risks placed lower are not as financially costly.

The resulting four quadrants of the matrix are as follows:

Engage Externally (Industry Response)
Risks in this quadrant typically cannot be directly addressed by investors, but they have a high impact on portfolio performance. These risks must still be addressed through industry standards, regulation, engagement, and continued monitoring.

Priority Action (Direct Oversight)
Risks that are within an investor's direct governance scope and have a high impact on returns should be immediately addressed via engagement, diligence protocols, and ongoing monitoring.

Monitor Only (Acknowledge)
These are risks that are low in severity and over which investors have little leverage. Investors should acknowledge these risks, but may not find it impactful to dedicate extensive oversight resources to address them.

Best Practice (Portfolio Support)
This quadrant encompasses lower-severity risks over which investors have direct influence. In order to avoid the risks that fall under this category, investors can engage in best practices that offer portfolio companies support, shared resources, or guidance.

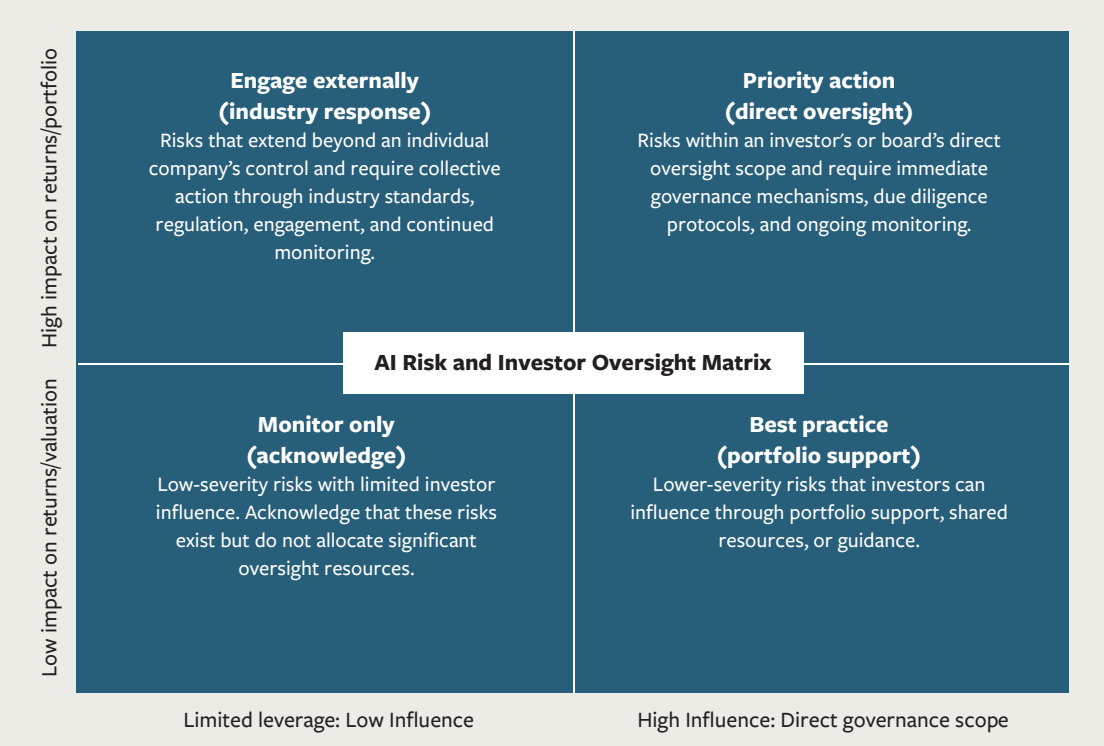

Figure 2: AI Risk and Investor Oversight Matrix.
This figure provides a framework for sorting AI risks facing investors' firms into four quadrants, organized by leverage over the risk and its potential impact on portfolio returns.

CASE STUDY

Below is a scenario in which a fictional company attempts to manage a rapidly escalating risk. Read the scenario and identify the risks for which an investor or prospective investor might consider creating mitigation strategies.

Scenario

DataDrive AI is a growth-stage company that developed a proprietary large language model for enterprise contract analysis. To train the model, the company ingested data through a range of third-party APIs, including legal databases, document repositories, and contract management platforms, without implementing a process for documenting data provenance or verifying the licensing terms governing the data's use.

Eighteen months after a Series B investment, a competing legal technology firm files suit alleging that DataDrive's training dataset included a substantial portion of their proprietary

contract database, which had been accessed without authorization through a misconfigured API. During discovery, DataDrive's legal team is unable to produce documentation accounting for the origins of approximately 40% of its training data.

As the litigation progresses, a data protection regulator opens a parallel inquiry into the company's data collection practices, citing potential violations of applicable privacy laws. Several of DataDrive AI's largest enterprise customers, including law firms and financial institutions with their own regulatory obligations, begin asking whether their previously submitted documents may have been ingested into the training dataset without consent.

Two customers suspend their contracts pending clarification, and a third initiates termination proceedings. The company's legal costs escalate rapidly. Its engineering team is pulled into remediation efforts, delaying a major product release. A prospective acquirer that had been in preliminary discussions quietly withdraws. The Series B investors, who hold board seats, are now managing an issue that was not identified during diligence and for which no governance protocols were in place.

Case Study Discussion

This scenario elucidates a variety of risks including, but not limited to, the cybersecurity, financial, and reputational risks covered in the approach. Additional risks from the NIST AI Framework, the International AI Safety Report, and the MIT AI Risk Repository are relevant to investors as well.

- **Cybersecurity risk:** The misconfigured API through which DataDrive took in third-party training data enabled unauthorized access to a competitor's proprietary database and remained undetected until it surfaced through litigation.
- **Financial performance risk:** DataDrive's data governance failures cause escalating legal costs, two suspended contracts, a third customer initiating termination, a delayed product release, and the withdrawal of a prospective acquirer, all of which represent material impairment to the company's revenue trajectory and exit prospects.
- **Reputational risk:** Enterprise customers, including law firms and financial institutions with their own regulatory obligations, are actively questioning whether their previously submitted documents were ingested into DataDrive's training dataset without consent.
- **Data privacy risk:** DataDrive's failure to verify the licensing terms governing its API-sourced training data has triggered a parallel regulatory inquiry into potential violations of applicable privacy laws.

- **Intellectual property risk:** A competing firm alleges that DataDrive accessed their proprietary database without authorization. Legal and technical due diligence at the Series B stage could have identified this exposure by reviewing the company's data acquisition practices and licensing agreements.
- **Copyright infringement:** The absence of enforceable industry standards governing the use of third-party data in AI model training meant that DataDrive's practices were not unusual at the time of investment. Resolving this exposure at a market level requires regulatory and legislative action that sits outside the scope of individual investor oversight.
- **Information integrity:** Because DataDrive cannot account for the origins of approximately 40% of its training data, the reliability and accuracy of its model outputs are fundamentally uncertain, creating the potential spread of dis- and misinformation.

As an exercise, use the blank table below to sort the risks identified above into the quadrants of the AI Risk and Investor Oversight Matrix. The following page includes an "answer key" which shows how identified risks could be sorted into the appropriate quadrants.

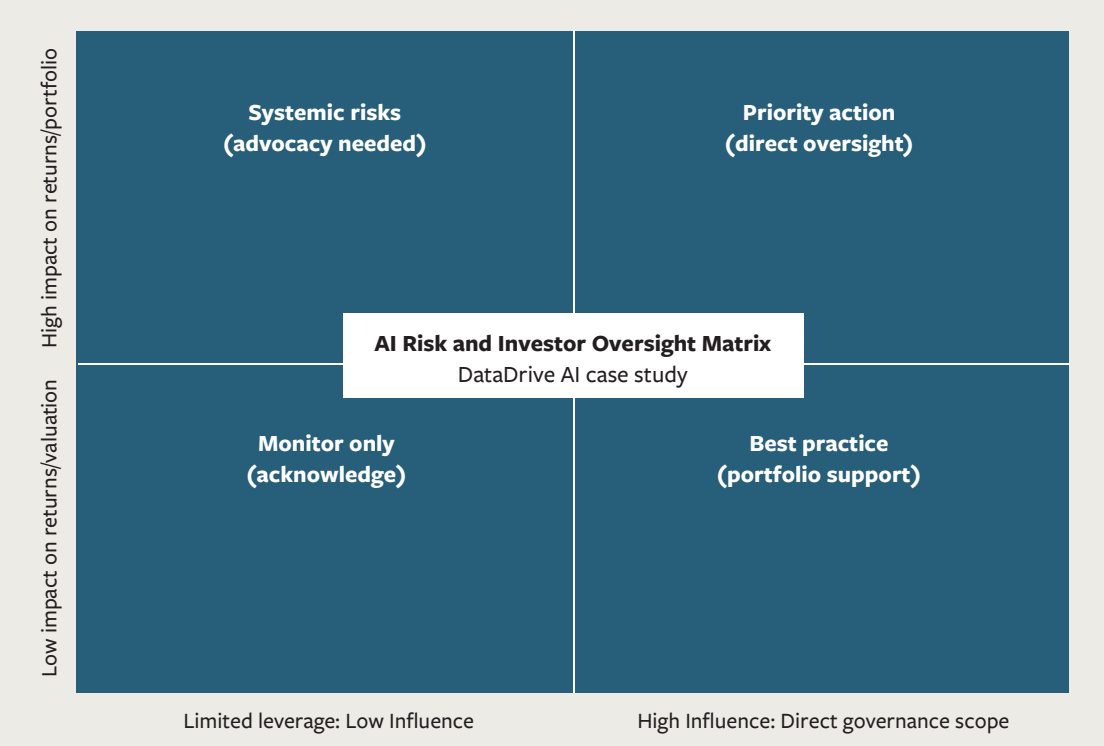

Figure 3: Blank AI Risk and Investor Oversight Matrix.
This figure is an empty version of Figure 2 into which investors can sort risks identified from the DataDrive AI case study.

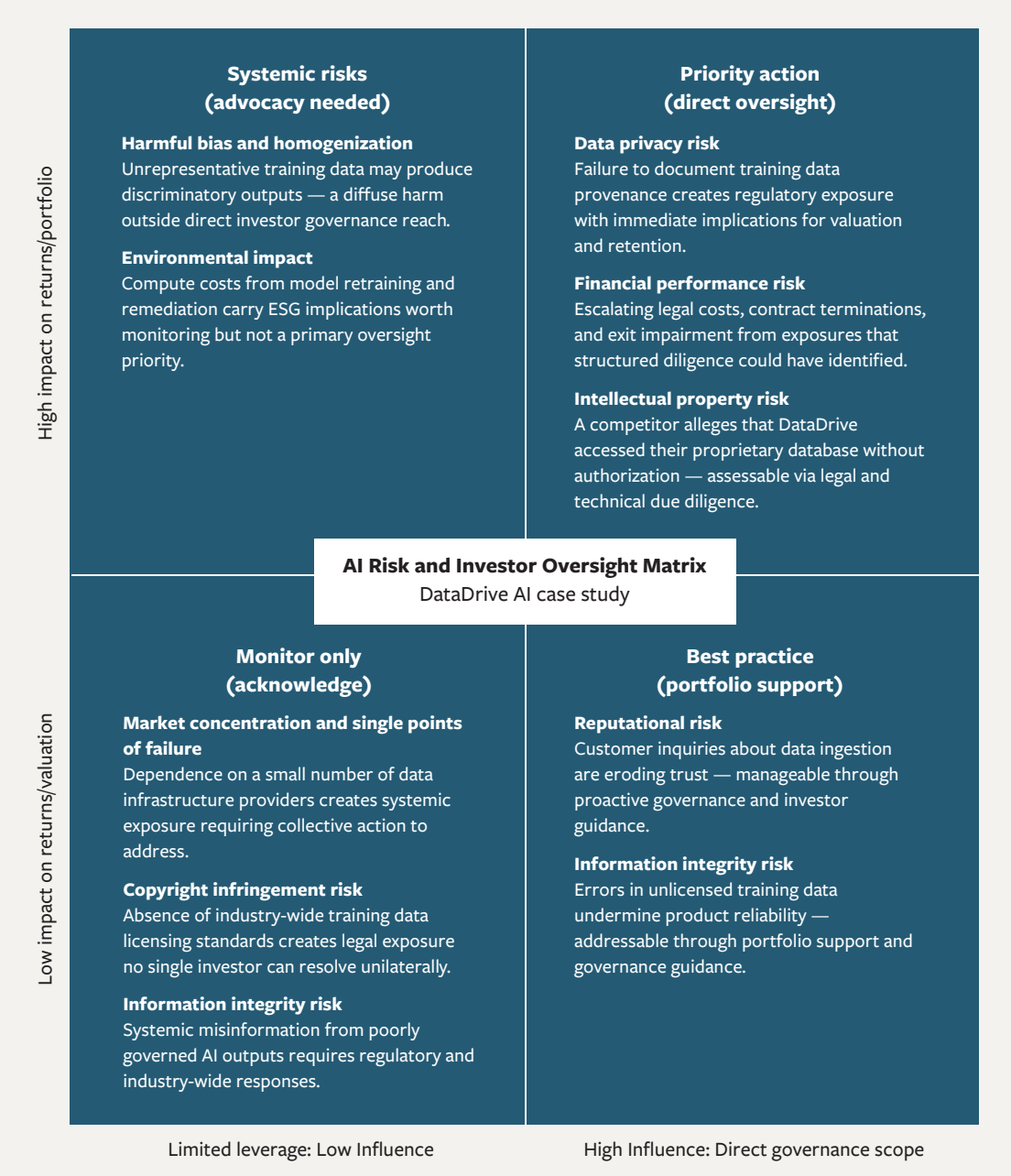

Figure 4: AI Risk and Investor Oversight Matrix - DataDrive AI Case Study.
This figure sorts the AI risks facing DataDrive's investors into four quadrants, based on their degree of leverage over each risk and its potential impact on portfolio returns.

Limitations and Areas for Further Study

While the structured approach outlined in this paper provides a method for surfacing and assessing AI-related risks, several limitations should be acknowledged. First, the approach focuses primarily on corporate, equity, and venture investors and may not fully generalize to other investor classes, such as hedge funds or mutual funds, each of which has distinct risk considerations.

Second, the approach is designed to provide visibility into three priority risk domains, and does not capture the full spectrum of AI-related risks that may indirectly affect investment outcomes. The three categories of risk illustrated here represent the most accessible entry point for investors beginning to assess how their AI-related decisions affect portfolio performance. However, future research may extend the framework to incorporate other risks.

Third, while the two-tiered structure enables both preliminary signal-based assessment and deeper engagement for investors with greater oversight capacity, it does not prescribe or evaluate specific mitigation actions. Investors must interpret the guidance provided in the context of their portfolio, resources, and risk appetite.

Fourth, the approach relies on observable proxies and self-reported information from portfolio companies and third-party disclosures, which may be incomplete, inconsistent, or subject to reporting bias.

Fifth, this framework is built around the full five stages of the investment lifecycle, and therefore may not be best suited for compressed investment timelines.

Finally, the structured approach has not yet been extensively validated through diverse investment scenarios. Its utility may vary depending on investment stage, sector, jurisdictional regulatory requirements, and the maturity of portfolio companies' AI governance practices. Consequently, this work should be understood as a foundational contribution, intended to inform and evolve through continued research and practice.

Future Work

The structured approach introduced in this report represents an initial step toward systematic AI risk oversight in investment contexts. While it provides foundational guidance, several opportunities for refinement and expansion exist.

Subsequent versions of this work may introduce risk-scoring mechanisms that enable more systematic assessment and portfolio-level analysis. Developing weighted scoring methods across the three primary risk domains could allow investors to make more data-driven resource allocation decisions and identify concentration risks across their portfolios. Future iterations of this work could also introduce tiered mitigation guidance calibrated to risk severity, which could enable investors to undertake proportionate responses to risks that manifest at various levels of severity.

More broadly, future development could extend the framework in several directions. For example, it could include broader coverage of investor types and asset classes, deeper treatment of systemic risk from AI infrastructure concentration, and more granular guidance for specific sectors and deal structures,. Moreover, the next phase of work could include designing standardized reporting and information sharing mechanisms for investors and boards to enable better shared visibility into AI risk for those with governance responsibilities in both groups.

In the coming months, a tool will become available to help investors put this approach into practice. By analyzing portfolio company responses to a structured set of diligence questions, the tool surfaces use-case-specific risk insights, giving investment, diligence, and stewardship teams visibility a company's AI risk profile through the investment process, along with a clear basis for directing governance resources to where they carry the greatest impact across the investment lifecycle. Readers interested in learning more are welcome to reach out directly.

We welcome collaboration and contributions from investment professionals and security researchers and practitioners who share an interest in making this framework more useful and complete.

Acknowledgments

A special thanks to all who contributed to this report, including Alex Alben, Luca Belli, Tess Buckley, Ryan Carrier, Gwenn Cujdik, Shu Dar, Jordan Famularo, Rokas Gipiskis, Olaf Groth, Andrew Harrison, Stefan Heuser, Jessica Ji, Prakaah Krishnan, Devin Lynch, Chris McClean, Debbie Taylor More, Evan Murphy, Yakaira Nunez, Caraline Pellatt, Karine Perset, Carole Piovesan, Christabel Randolph, Sekhar Sarukkai, Raymond Sheh, Joyce Shen, Peter Slatter, Aparna Surendra, Jennifer Tang, Miriami Tkeshelashvili, Jim Williams, Cherry Wu, Emily Yang, and Polina Zvyagina.

A special thanks to Amaya Blevins, Nada Madkour, Deeksha Vaidyanathan, and Michael Benedict Yamoah for their special contributions to this publication.

References

The following sources were referenced and/or were used as background to develop the approach:

AIID (n.d.): "AI Incident Database," *AI Incident Database,* https://incidentdatabase.ai/.

Anthropic (2025a): "Responsible Scaling Policy, Version 2.2," *Anthropic,* https://www-cdn.anthropic.com/872c653b2d0501d6ab44cf87f43e1dc4853e4d37.pdf.

Anthropic (2025b): "Proposed Frontier Model Transparency Framework," *Anthropic,* https://www-cdn.anthropic.com/19cc4bf9eb6a94f9762ac67368f3322cf82b09fe.pdf.

BIS (2024): "Artificial Intelligence and the Economy: Implications for Central Banks," *Bank for International Settlements,* https://www.bis.org/publ/arpdf/ar2024e3.htm.

Barrett, A. M., J. Newman, B. Nonnecke, N. Madkour, D. Hendrycks, E. R. Murphy, K. Jackson, and D. Raman (2025): "AI Risk-Management Standards Profile for General-Purpose AI (GPAI) and Foundation Models," *arXiv,* https://arxiv.org/abs/2506.23949.

Bengio, Y., S. Mindermann, D. Privitera, T. Besiroglu, R. Bommasani, S. Casper, Y. Choi, P. Fox, B. Garfinkel, D. Goldfarb, H. Heidari, A. Ho, S. Kapoor, et al. (2025): "International AI Safety Report," *arXiv,* https://arxiv.org/abs/2501.17805.

Castillo, A, and C. Rosales (2025): "Building Technology with Purpose: The Role of VCs in the Future of AI," *IDB Lab,* https://bidlab.org/en/news/building-technology-purpose-role-vcs-future-ai.

CFA Institute (2025): "New CFA Institute Report Urges Financial Sector to Prioritize Explainability in AI Adoption," *CFA Institute,* https://www.cfainstitute.org/about/press-room/2025/explainable-ai-in-finance-2025.

CRFM (2025): "The Foundation Model Transparency Index," *The Stanford Center for Research on Foundation Models,* https://crfm.stanford.edu/fmti/December-2025/index.html.

EC (2025): "The General-Purpose AI Code of Practice," *European Commission,* https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

EU (2024): "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)," *European Union,* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

GAO (2025): "Artificial Intelligence: Use and Oversight in Financial Services," *U.S. Government Accountability Office,* https://www.gao.gov/products/gao-25-107197.

Gartner (2026): "Gartner Says Worldwide AI Spending Will Total $2.5 Trillion in 2026," *Gartner*, https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026.

Google (2025): "Frontier Safety Framework, Version 2.0," *Google,* https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/updating-the-frontier-safety-framework/Frontier%20Safety%20Framework%202.0.pdf.

ILPA (2023): "Responsible AI Quick Guide for Asset Owners," *Institutional Limited Partners Association,* https://ilpa.org/wp-content/uploads/2023/10/Responsible-AI-Quick-Guide-for-Asset-Owners.pdf.

ISO (2023): "ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management," *International Organization for Standardization*, https://www.iso.org/standard/77304.html.

METR (n.d.): "Frontier AI Safety Policies," *METR,* https://metr.org/faisc

MIT (n.d.): "What are The Risks of Artificial Intelligence?" *MIT AI Risk Initiative,* https://airisk.mit.edu/.

MIT (2025): "How is AI Harming us?" *MIT AI Risk Initiative,* https://airisk.mit.edu/ai-incident-tracker.

MITRE (n.d.): "MITRE ATLAS AI Incidents," *MITRE,* https://ai-incidents.mitre.org/.

Morris, L. (2023): "Responsible AI Startup (RAIS) Framework," *Radical Ventures,* https://github.com/radicalventures/RAIS-Framework.

NIST (n.d.): "NIST AI RMF Playbook," *National Institute of Standards and Technology,* https://airc.nist.gov/airmf-resources/playbook/.

NIST (2023): "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," *National Institute of Standards and Technology,* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

NIST (2024): "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," *National Institute of Standards and Technology,* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

OECD (2024): "Regulatory Approaches to Artificial Intelligence in Finance," *OECD Publishing,* https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/09/regulatory-approaches-to-artificial-intelligence-in-finance_43d082c3/f1498c02-en.pdf.

OECD.AI (2025): "G7 Reporting Framework – Hiroshima AI Process (HAIP) International Code of Conduct for Organizations Developing Advanced AI Systems," *OECD,* https://transparency.oecd.ai/.

OECD.AI (n.d.): "AIM: AI Incidents and Hazards Monitor," *OECD,* https://oecd.ai/en/incidents.

OpenAI (2025): "Preparedness Framework, Version 2," *OpenAI,* https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf.

Primack, D. (2025): "AI is Eating Venture Capital, or at Least its Dollars," Axios, https://www.axios.com/2025/07/03/ai-startups-vc-investments.

RIAA (2024): "Artificial Intelligence and Human Rights Investor Toolkit," *Responsible Investment Association Australia,* https://cdn.prod.website-files.com/67a17dcc2362afe3723b9c97/6850b57133a141d30d4c6b4b_AI-and-Human-Rights_Investor-Toolkit-FINAL-27.4.2024%20%282%29_compressed.pdf.

Wan, A., K. Klyman, S. Kapoor, N. Maslej, S. Longpre, B. Xiong, P. Liang, and R. Bommasani (2025): "The 2025 Foundation Model Transparency Index," *arXiv,* https://arxiv.org/abs/2512.10169.

WEF (2025): "Responsible AI Playbook for Investors," *World Economic Forum,* https://www3.weforum.org/docs/WEF_Responsible_AI_Playbook_for_Investors_2024.pdf.

About the Author

Oumou Ly is a Nonresident Research Fellow with the AI Security Initiative at UC Berkeley's Center for Long-Term Cybersecurity, where she conducts applied research on the systemic risks that emerging technologies pose to global financial systems. Oumou is a technologist who shapes how global governments and industries confront the risks and opportunities of artificial intelligence and other emerging technologies.

From 2023 to January 2025, Oumou served as Senior Advisor for Technology and Ecosystem Security at the White House, where she was an author of the 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Oumou also played a key role in developing national strategies to improve global technology security outcomes, including the National Cybersecurity Strategy, the National Cyber Workforce and Education Strategy, and the U.S. Standards Strategy for Critical and Emerging Technologies. She also served in a Presidentially appointed role at the Cybersecurity and Infrastructure Security Agency, where she led efforts to advance data-driven methods for reducing cybersecurity risk.

Prior to serving in the executive branch, Oumou was a researcher at the Berkman Klein Center for Internet & Society at Harvard University, where she co-authored the concept paper for BlueSky Social, a public benefit corporation that hosts one of the first decentralized social networking platforms. Earlier, Oumou was an advisor to the United States Senate Democratic Leader Charles Schumer, where she counseled the Leader and U.S. Senators on national security and technology policy.

Oumou holds a B.A. from Notre Dame de Namur University in Belmont, CA and an MSc from the London School of Economics and Political Science. She is a member of the Board of Trustees at Notre Dame de Namur University.


CLTC
Center for Long-Term
Cybersecurity
UC Berkeley