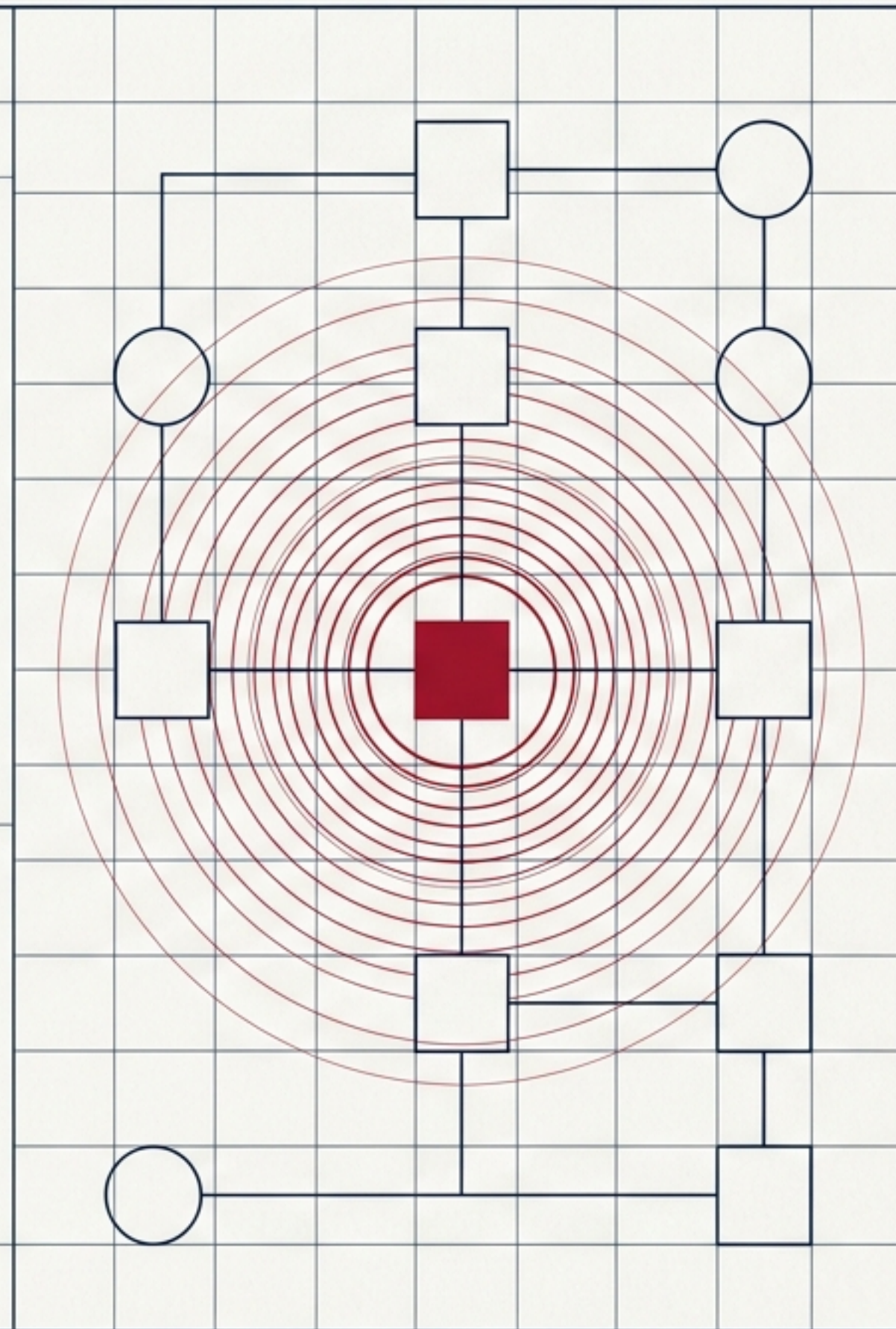


EXECUTIVE BRIEFING & STRATEGIC RISK ALERT

データ来歴の代償： Anthropic和解が暴く 「AIセーフハーバー」の幻想

15億ドルの米国著作権訴訟が
突きつける、日本企業への戦略的
警告と実務プレイブック



危険な誤解 (The Myth)

「Anthropicは15億ドル（約2400億円）を支払い、AI学習の免罪符を買った。和解金を払えば、今後もウェブ上のあらゆるデータをスクレイピングして学習させることができる」



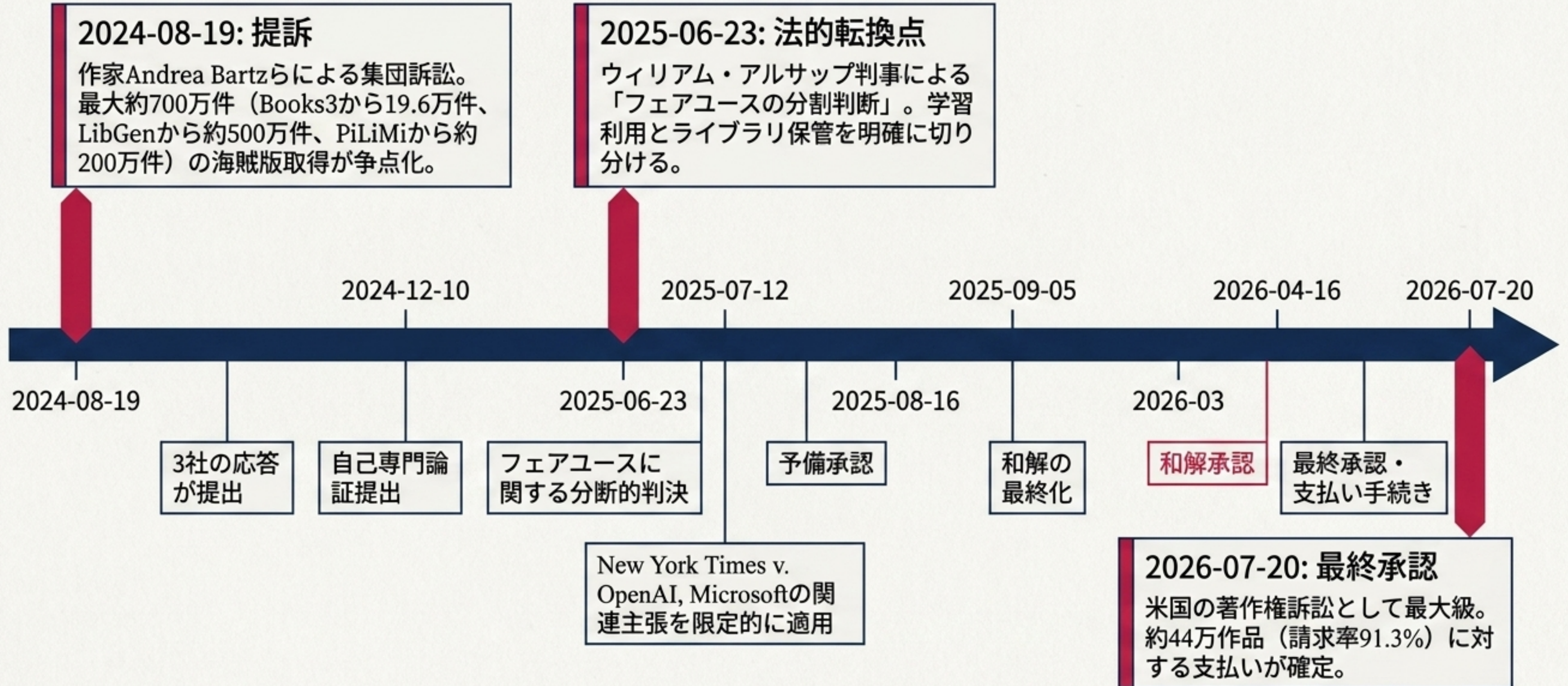
新たな現実 (The Reality)

「本件和解は、AI学習そのものを違法としたわけではない。出所不明・海賊版データの『違法法取得と恒久保管』に対する巨額の制裁である。データ来歴 (Provenance) を証明できないAI開発は、致命的な財務・事業リスクに直面する」



和解は「AI学習はすべて適法」と宣言したものではなく、違法入手・違法保管が最大の争点であることを可視化した（エグゼクティブサマリーより）。

Bartz v. Anthropic の主要経過



Anatomy of a Settlement

\$1.5 Billion

(約2400億円)

非返還型和解基金
(Non-reversionary Fund)

[財務的ペナルティ]

1作品あたりの純受取見込額は約3,127ドル。対象作品数が50万作を超えた場合は、1作あたり3000ドルを追加で支払う義務（キャップなしの青天井リスク）。

[将来免責の拒絶 (Critical Insight)]

過去（2025年8月25日以前）の行為のみを処理。将来の複製・二次的著作物作成、および「AI出力に基づく請求」は放棄対象外。継続的なライセンス付与ではない。

[データの物理的破棄]

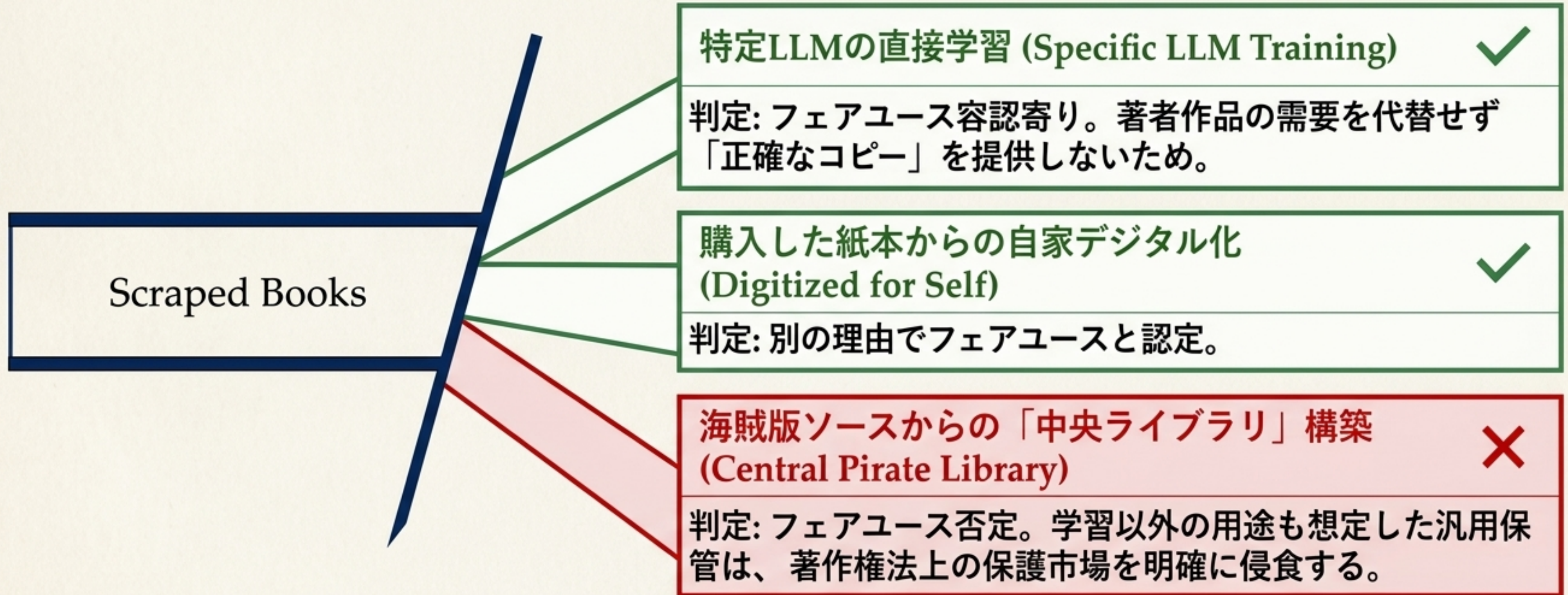
LibGen/PiLiMi由来の原本ファイルと、その派生コピーの完全破棄義務。

[商用モデルへの非混入表明]

Anthropicは、これら海賊版データセットが商用リリース済みLLM（Claude等）の学習コーパスに入っていないことを法的に表明。

法的スカルペル: データパイプラインの法的分断

(Legal Scalpel: Legal Segmentation of the Data Pipeline)



法的含意: 「AI学習そのもの」が裁かれたのではない。「データ取得経路」と「保管態様（汎用ライブラリ化）」が著作権侵害のトリガーとなる。

The Hidden Bottleneck: Japanese Legal Realities

日本の著作権法 第30条の4（法的入口）

「享受目的がない利用（情報解析等）は原則適法」

一見すると、米国フェアユースよりもAI学習に広く、明確な入口（セーフハーバー）を与えているように見える。



文化庁 2024年ガイダンス（実務上の隠れた罠）

条文上は広くても、実務ガイドラインが厳格なフィルターとして機能。

「ただし書（権利者の利益を不当に害する場合）」に抵触する
具体的なレッドラインを設定。



1. AI学習用に有償提供されているデータベースの無許諾利用



2. 既存の創作的表現を出力させる目的の追加学習 / RAG



3. 特定クリエイターの少量作品に対する集中学習（スタイル模倣）



結論: 日本は決して「スクレイピングの完全な逃避地」ではない。

比較：米国 vs 日本の実務リスク構造

[適法性のゲートウェイ]	
フェアユース (17 U.S.C. § 107)	第30条の4
多要素衡量型。市場代替性（第4要素）が極めて重視される。	明示的な権利制限だが、「享受目的の併存」や「市場阻害（ただし書）」で突如として適用外になる柔軟性のジレンマ。
[損害賠償の構造 (財務リスク)]	
法定損害賠償 (17 U.S.C. § 504)	逸失利益・ライセンス料相当額の推定 (第114条)
1作品あたり最大15万ドル（故意の場合）。天井のない作品数連動型の「 <u>爆発的リスク</u> 」。	爆発的な法定損害はないが、海賊版や無許諾学習が常態化すれば、推定額は膨れ上がる。
[企業への実存的脅威]	
集団訴訟 (Class Action) による巨額和解	差止め請求 (第112条) と証拠開示
<u>本件の15億ドルのような直接的なキャッシュアウト。</u> 	学習済みモデルからのデータ除去、最悪の場合は「 <u>モデル自体の廃棄請求</u> 」という <u>事業停止リスク</u> 。 

The Financial Stakes: Economic Scale & Legal Risk

約1.5兆円 (1兆5462億円)

2025年 日本の出版市場規模 (紙・電子合計)

⚠
海賊版の被害額：経産省推計で
日本の出版分野のオンライン海賊
版被害は年間2.6兆円にのぼる。

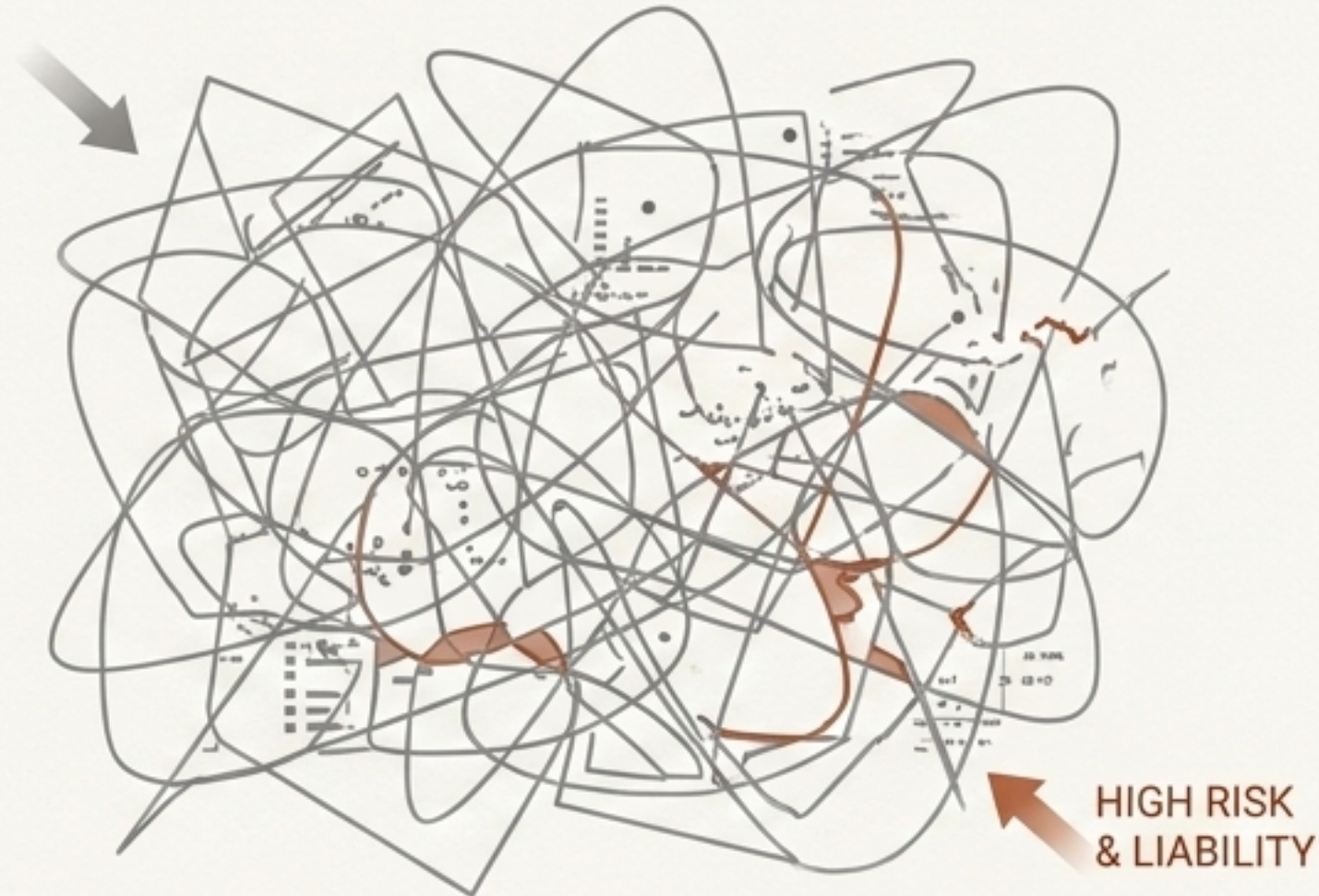
約2,400億円 (\$1.5 Billion)

Bartz v. Anthropic 訴訟の和解総額

「 和解額は日本の年間出版市場の極めて大きな割合を占める。
コンテンツ産業（翻訳、電子、コミック等）にとって、無許諾学習の放置は死活問題で
であり、今後は集団的な法的措置や交渉が激化することが確実視される。 」

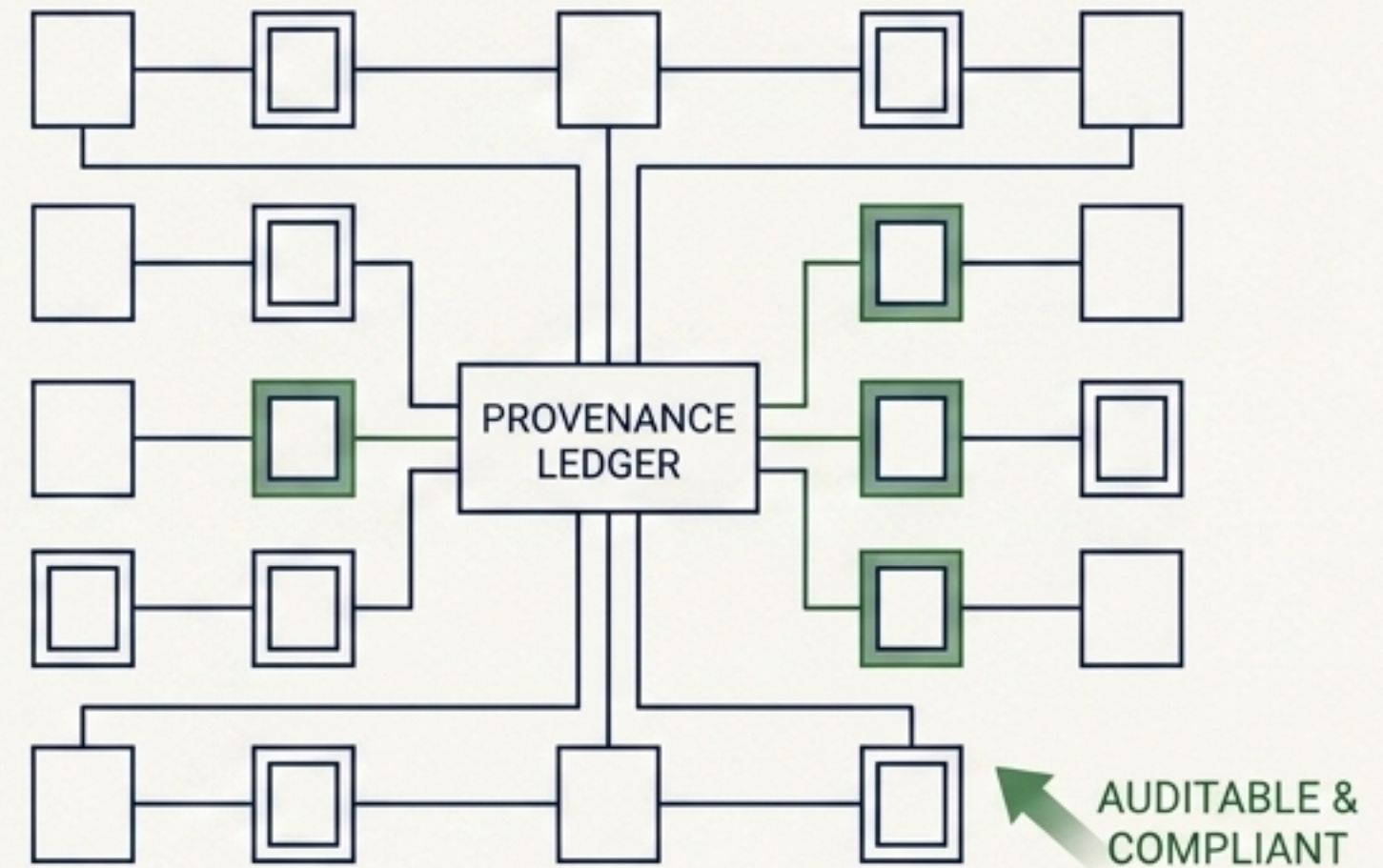
Before & After: Architectural Shift in Data Acquisition

過去の慣行：暗闇でのスクレイピング



- ・匿名データ収集（Books3, 海賊版サイト）
- ・学習と保管の未分離（巨大なデータレイクの放置）
- ・「とりあえずすべて集める」アプローチ

今後の要請：データ来歴の監査（Provenance Ledger）



- ・取得源・権限・利用目的の厳格な台帳管理
- ・有償ライセンス市場の積極的活用
- ・学習用途限定・短期保管・削除証跡の実装

米国著作権局も「市場ベースのライセンス形成」を勧告。違法入手の高コスト化により、データは「拾うもの」から「監査し、ライセンスを受けるもの」へ移行する。

アクション・プラン：AI開発企業向け（Defensive Auditing）

「学習フェーズ」だけでなく、データのサプライチェーン全体を監査せよ。

1

データ来歴台帳（Provenance Ledger）の構築 **【優先度：最優先】**

各データの取得源、取得日、権限、削除可否を完全にトレース可能にする。不明ソース・海賊版（LibGen等）は学習が適法でも「入手違法」の火種となるため即時パージする。

2

「学習用コピー」と「保管ライブラリ」の物理的分離 **【優先度：最優先】**

本件の敗因を回避するため、用途を限定し、保存期間を短縮。削除証跡を残し、裁判所に「中央ライブラリ」と認定される隙を与えない。

3

ベンダー・利用規約におけるAI保証条項 **【優先度：高】**

データプロバイダーに対して、適法性の保証、補償、および削除への協力義務を契約に明記させる。

4

RAGと追加学習の内部審査 **【優先度：高】**

日本法下でも、既存表現の近似出力を目的とするRAG設定は差止めリスクが極めて高い。設計段階でのフィルタリングを義務化する。

アクション・プラン：出版社・クリエイター向け（Offensive Fencing）

単なる「学習禁止」の声明は無力。「どう意思表示し、どう売るか」の設計へ。

「
1
」

AI学習用ライセンスの積極的商品化 **【優先度：最優先】**

「代替可能な有償ライセンス市場が存在する」という事実そのものが、無許諾スクレイピングを法的に不利（米国第4要素・日本ただし書）にする最強の武器となる。

「
2
」

出版・エージェント契約の全面見直し **【優先度：最優先】**

学習、埋め込み、要約、RAGへの利用権限と、収益配分（本件では著者・出版社間で原則50:50）を契約書に明記し、権利の所在を整理する。

「
3
」

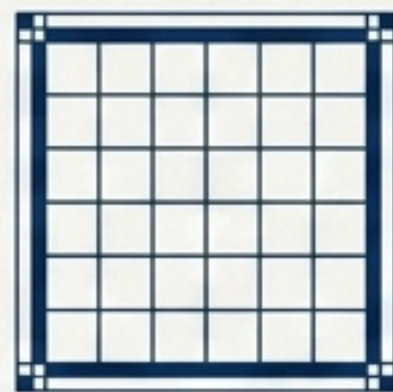
機械可読なオプトアウトとログ保全 **【優先度：高】**

robots.txtや認証ゲートウェイによるアクセス制御を実施し、「明確な拒絶の意思表示」と「不正アクセスの証拠（ログ）」を同時に確保する。

「
4
」

著作者団体による標準化 **【優先度：高】**

個別の交渉コストを下げるため、団体主導で「標準条項」と「標準料金表」を整備し、市場形成を加速させる。



不確実性を放置するコスト

「日本法では広く学習が許されるはず」という社内の短絡的な法解釈は、もはや通用しない。Anthropicの15億ドル和解が示した真の教訓は、AI開発における勝負の分水嶺が「学習アルゴリズムの適法性」から、「データ来歴の証明と管理」へと完全にシフトしたという事実である。