Building a Plan for Monetizing Your Data in an AI Ecosystem

_____

Celesq® AttorneysEd Center
www.celesq.com

5301 North Federal Highway, Suite 150, Boca Raton, FL 33487
Phone 561-241-1919

Building a Plan for Monetizing Your Data in an AI Ecosystem

Charles R. Macedo

March 5, 2025

DISCLAIMER

The following presentation reflects the opinion of its human author and does not necessarily represent the views of his respective clients, partners, employers, or of Amster, Rothstein & Ebenstein LLP, nor CelEsq or any of its officers, or employees.

Additionally, the following content is presented solely for the purposes of discussion and illustration and does not comprise, nor is to be considered, as legal or business advice.

No humans were hurt in making this presentation, but AI assisted in its preparation.

About our Speakers

Charles R. Macedo is a seasoned intellectual property attorney with deep expertise at the intersection of emerging technologies, data-driven businesses, and commercialization. He is recognized as an author, thought leader and frequent lecturer in intellectual property, blockchain, artificial intelligence/software, and data monetization. With a technical background in physics and decades of experience guiding Unicorns, startups, financial services firms, and technology innovators, he has developed IP strategies as they launch and monetize new product lines resulting in collections of hundreds of millions of dollars of royalty revenue.

J.D. 1989, Columbia Law School; B.S./M.S., Physics, 1986; former Law Clerk to Hon. Daniel M. Friedman at U.S. Court of Appeals for the Federal Circuit.
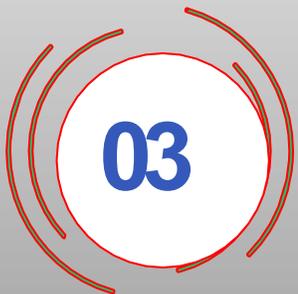
Amster
Rothstein &
Ebenstein LLP

Key Topics:

01   **Trump Administration policies and implementations to monetizing Data Sets**

02   **Fair Use as it relates to utilizing data sets for training and outputs of generative AI**

03   **How to set up and implement policies within an organization to generate monetizable Data Sets**

Topic 1:
Trump Administration policies and implementations to monetizing Data Sets

March, 5, 2026

Executive Order 14179 of January 23, 2025
Removing Barriers to American Leadership in Artificial Intelligence

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered:

Section 1. Purpose. The United States has long been at the forefront of artificial intelligence (AI) innovation, driven by the strength of our free markets, world-class research institutions, and entrepreneurial spirit. To maintain this leadership, we must develop AI systems that are free from ideological bias or engineered social agendas. With the right Government policies, we can solidify our position as the global leader in AI and secure a brighter future for all Americans. This order revokes certain existing AI policies and directives that act as barriers to American AI innovation, clearing a path for the United States to act

Sec. 2. Policy. It is the policy of t
dominance in order to promote h

Sec. 3. Definition. For the purpos
forth in 15 U.S.C. 9401(3).

> The term "artificial intell
> human-defined objective
> or virtual environments.
> to- (A) perceive real and
> through analysis in an a
> for information or action

Sec. 4. Developing an Artificial Intelligence Action Plan.
(a) Within 180 days of this order, the Assistant to the President for Science and Technology (APST), the Special Advisor for AI and Crypto, and the Assistant to the President for National Security Affairs (APNSA), in coordination with the Assistant to the President for Economic Policy, the Assistant to the President for Domestic Policy, the Director of the Office of Management and Budget (OMB Director), and the heads of such executive departments and agencies (agencies) as the APST and APNSA deem relevant, shall develop and submit to the President an action plan to achieve the policy set forth in section 2 of this order.

Sec. 5. Implementation of Order Revocation.
(a) The APST, the Special Advisor for AI and Crypto, and the APNSA shall immediately review, in coordination with the heads of all agencies as they deem relevant, all policies, directives, regulations, orders, and other actions taken pursuant to the revoked Executive Order 14110 of October 30, 2023 (Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence). The APST, the Special Advisor for AI and Crypto, and the APNSA shall, in coordination with the heads of relevant agencies, identify any actions taken pursuant to Executive Order 14110 that are or may be inconsistent with, or present obstacles to, the policy set forth in section 2 of this order. For any such agency actions identified, the heads of agencies shall, as appropriate and consistent with applicable law, suspend, revise, or rescind such actions, or propose suspending, revising, or rescinding

(c) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

THE WHITE HOUSE,

January 23, 2025.

Jan. 23, 2025, President Trump **revokes** Biden Era AI Policy, and directs his Cabinet "shall develop and submit .. an action plan to achieve policy" to "**sustain America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security.**"

Winning the Race: America's Action Plan, July 2025

THE WHITE HOUSE

Winning the Race

AMERICA'S AI ACTION PLAN

JULY 2025
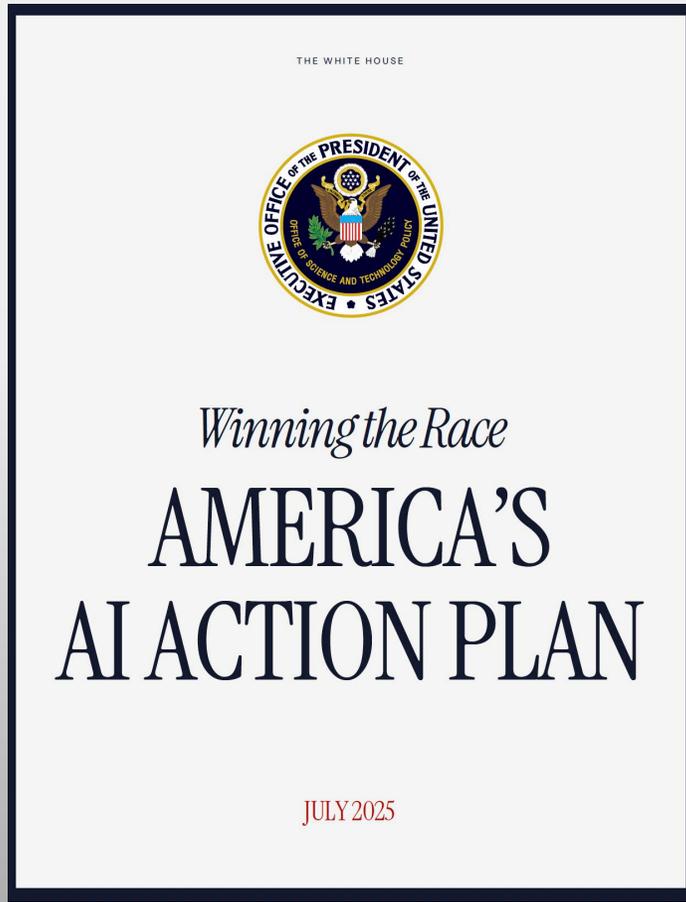
The United States is in a race to achieve global dominance in artificial intelligence (AI). Whoever has the largest AI ecosystem will set global AI standards and reap broad economic and military benefits. Just like we won the space race, it is imperative that the United States and its allies win this race.

America's AI Action Plan is a blueprint to achieve global dominance in AI for the U.S., and has 3 pillars:
- **innovation,**
- **infrastructure, and**
- **International AI diplomacy and security**

America's AI Action Plan has three pillars: innovation, infrastructure, and international diplomacy and security.

Pillar 1: Accelerate AI Innovation

Pillar 2: Build AI Infrastructure

Pillar 3: Lead in International AI Diplomacy and security

March, 5, 2026

https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.p

Winning the Race: America's Action Plan, July 2025

Encourage Open-Source and Open-Weight AI

Open-source and open-weight AI models are made freely available by developers for anyone in the world to download and modify. Models distributed this way have unique value for innovation because startups can use them flexibly without being dependent on a closed model provider. They also benefit commercial and government adoption of AI because many businesses and governments have sensitive data that they cannot send to closed model vendors. And they are essential for academic research, which often relies on access to the weights and training data of a model to perform scientifically rigorous experiments.

We need to ensure America has leading open models founded on American values. Open-source and open-weight models could become global standards in some areas of business and in academic research worldwide. For that reason, they also have geostrategic value. While the decision of whether and how to release an open or closed model is fundamentally up to the developer, the Federal government should create a supportive environment for open models.
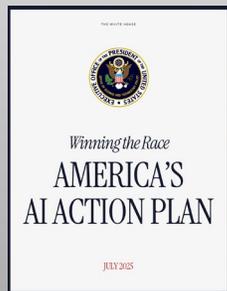
Build World-Class Scientific Datasets

High-quality data has become a national strategic asset as governments pursue AI innovation goals and capitalize on the technology's economic benefits. Other countries, including our adversaries, have raced ahead of us in amassing vast troves of scientific data. The United States must lead the creation of the world's largest and highest quality AI-ready scientific datasets, while maintaining respect for individual rights and ensuring civil liberties, privacy, and confidentiality protections.

Amster Rothstein & Ebenstein LLP

March, 5, 2026

8

> In this pivotal moment, the challenges we face require a historic national effort, comparable in urgency and ambition to the Manhattan Project that was instrumental to our victory in World War II and was a critical basis for the foundation of the Department of Energy (DOE) and its national laboratories.

American Science and Security Platform – Secretary of Energy led platform to integrate Federal scientific data with cloud-based AI computing environment to support large-scale model training, simulation, and inference;.

Goal: train **scientific foundation models** and create **AI agents** to automate research workflows.

Targets at least **20 National Science & Technology Challenges including:** advanced manufacturing, biotechnology, critical materials, nuclear fission and fusion energy, quantum information science; and semiconductors and microelectronics.

March, 5, 2026

https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/

NIH Public Access is already in effect

NIH GRANTS & FUNDING

NIH Public Access Policy Overview

⚠ This page provides information about the **2024 NIH Public Access Policy**, which went into effect on **July 1, 2025**. For those with questions during this transition, please see the 2024 NIH Public Access Policy's Frequently Asked Questions.

Public Access Policy Details

Increasing access to publications resulting from National Institutes of Health (NIH) funding offers many benefits to the scientific community and the public who funded the underlying research. When patients, families, and healthcare providers can access published findings resulting from NIH funding, they are able to better understand and address the most critical health concerns facing their communities. It also allows researchers, students, and members of the public in all communities to have equitable access to such content.

To achieve these goals, the 2024 NIH Public Access Policy requires Author Accepted Manuscripts accepted for publication in a journal, on or after July 1, 2025, to be submitted to PubMed Central upon acceptance for publication, for public availability without embargo upon the Official Date of Publication.

Public Access Policy Implementation accelerated from Dec. 2025 to July 1, 2025

March, 5, 2026

"Bulk Data Rule"

PRESS RELEASE

Justice Department Implements Critical National Security Program to Protect Americans' Sensitive Data from Foreign Adversaries

Friday, April 11, 2025

Today, the Justice Department took significant steps to move forward with implementing a critical program to prevent China, Russia, Iran, and other foreign adversaries from using commercial activities to access and exploit U.S. government-related data and Americans' sensitive personal data to commit espionage and economic espionage, conduct surveillance and counterintelligence activities, develop AI and military capabilities, and otherwise undermine our national security.

The Data Security Program implemented by the National Security Division (NSD) under **Executive Order 14117** addresses this "unusual and extraordinary threat...to the national security and foreign policy of the United States" that has been repeatedly recognized across political parties and by all three branches of government.

The Justice Department's continued prioritization of the Data Security Program delivers on promises made by President Trump in his **America First Investment Policy** and **NSPM-2 on Imposing Maximum Pressure on Iran,** addresses threats identified in the 2025 Annual Threat **Assessment of the U.S. Intelligence Community** and President Trump's 2017 National Security Strategy, and responds to the national emergency President Trump declared in **Executive Order 13873.**

Amster Rothstein & Ebenstein LLP

"Bulk Data Rule"

PRESS RELEASE

Justice Department Implements Critical National Security Program to Protect Americans' Sensitive Data from Foreign Adversaries

Friday, April 11, 2025

Today, the Justice Department took significant steps to move forward with implementing a critical program to prevent China, Russia, Iran, and other foreign adversaries from using data and Americans' sensitive personal data to com... counterintelligence activities, develop AI and milita...

The Data Security Program implemented by the Nati... "unusual and extraordinary threat...to the national s... recognized across political parties and by all three b...

The Justice Department's continued prioritization of... Trump in his America First Investment Policy and NS... threats identified in the 2025 Annual Threat Assessm... National Security Strategy, and responds to the natio...

"If you're a foreign adversary, why would you go through the trouble of complicated cyber intrusions and theft to get Americans' data when you can just buy it on the open market or force a company under your jurisdiction to give you access?" said Deputy Attorney General Todd Blanche. "The Data Security Program makes getting that data a lot harder."

To address this urgent threat, the Data Security Program establishes what are effectively export controls that prevent foreign adversaries, and those subject to their control, jurisdiction, ownership, and direction, from accessing U.S. government-related data and bulk genomic, geolocation, biometric, health, financial, and other sensitive personal data.

Amster Rothstein & Ebenstein LLP

March, 5, 2026

12

"Bulk Data Rule" (28 C.F.R. pt. 202)

Bulk Data

U.S. Sensitive Personal Data	Threshold
Human genomic data	100 U.S. persons
Human 'omic data	1,000 U.S. persons
Biometric Identifier	1,000 U.S. persons
Precise geolocation data	1,000 U.S. persons
Personal health data	10,000 U.S. persons
Personal financial data	10,000 U.S. persons
Covered personal identifiers	100,000 U.S. persons
Combined data	Lowest applicable number

Restricted Countries

- China (including Hong Kong and Macau) (NOT Taiwan)
- Cuba
- Iran
- North Korea
- Russia
- Venezuela

Key Takeaways

- **Covered Persons** (28 CFR § 202.211.) – apply to employees, consultants, and anyone who has access to Bulk Data. This includes parties involved in transfer and conditioning of data. For example, third party processors and employees/consultants/contractors who gain temporary data access.

- **Exemptions** - certain exemptions exist for therapeutics and diagnostics provided for regulatory approval from clinical trials and post surveillance market access.

- **Personal Health** - non biometric and non 'omic data points like x-rays and blood tests qualify as Bulk Data under the personal health category in volumes of 10,000 U.S. persons.

- **Know your customer** - customers who gain access to Bulk Data must be identified and verified to ensure they are not covered people or countries.

- **Track Rejections** - rejections of prohibited data brokerage transactions must be tracked and reported if the transaction is denied.

Understanding the Ecosystem

Amster
Rothstein &
Ebenstein LLP

Example of How LLMs Answer Queries

How do I treat Parkinson's Disease?

Input

RAG

Real Time Data Sets

Large Language Model (LLM)

Historical Training Sets

Output

Agents / Orchestrators

How do I treat Parkinson's Disease?

I need to be clear up front: I cannot provide personalized medical advice. Parkinson's disease (PD) is a serious, progressive neurological condition, and treatment should always be guided by a neurologist or movement disorder specialist.
That said, here's an overview of **current standard approaches to treating Parkinson's disease**:

1. Medications
The mainstay of treatment is medication to restore or mimic dopamine in the brain:
•**Levodopa (with carbidopa)** – the most effective drug for improving motor symptoms (tremor, rigidity, slowness).
•**Dopamine agonists** (e.g., pramipexole, ropinirole, rotigotine) – mimic dopamine effects, sometimes used earlier or alongside levodopa.

Interface – Asking Query

Retrieval Augmented Generation (RAG) enables retrieval of real time data to supplement query formation for LLM

A **Large Language Model** is a type of artificial intelligence that can **understand and generate human language**.

Agents and **Orchestrators** help generate and select the optimum insight to deliver in response to the query.

Interface – Responding with output

02 Analysis of ongoing training set litigation and recent fair use rulings

March, 5, 2026
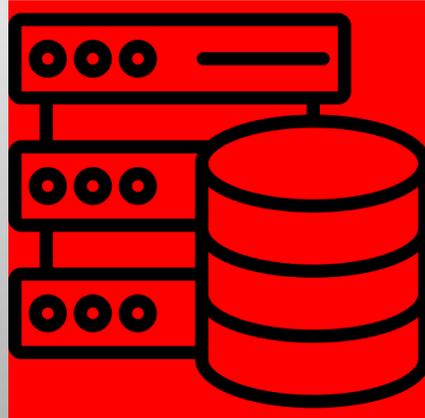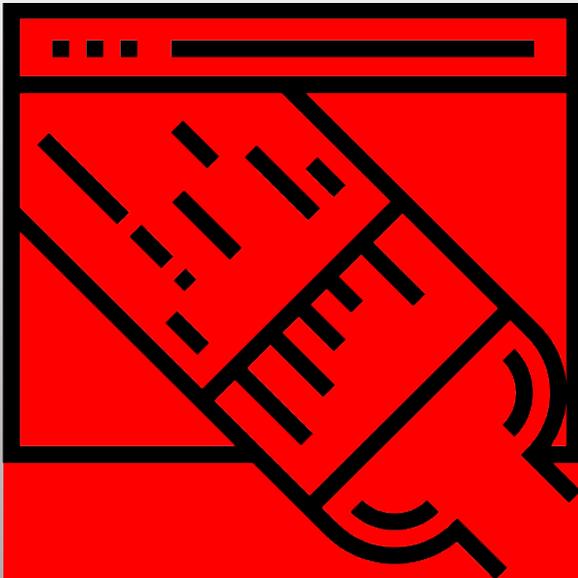
Fair Use Factors

107. Limitations on exclusive rights: Fair use

Notwithstanding the provisions of sections <u>106</u> and <u>106A</u>, ==the fair use of a copyrighted work==, including such use by reproduction in copies or phonorecords or by any other means specified by that section==, for purposes such as== criticism, comment, news reporting, teaching (including multiple copies for classroom use), ==scholarship, or research==, ==is not an infringement of copyright==. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—

(1==) the purpose and character of the use==, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) ==the nature== of the copyrighted work;

(3) ==the amount and substantiality of the portion== used in relation to the copyrighted work as a whole; and

(4) ==the effect of the use upon the potential market for or value== of the copyrighted work.

The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.

March, 5, 2026

17 U.S.C. 107

17

Cases Relevant to Fair Use in the AI Context In the Circuit Courts

Reverse Engineering Cases:

- *Sega Enters. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992)
- *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir.
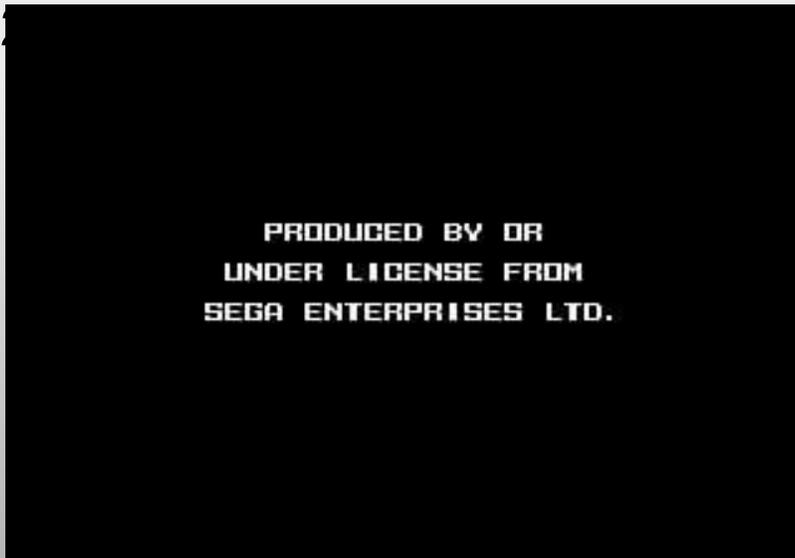
Sample screen shown by use of SEGA header in Sega Enters.

Copying of Entirety of Works:
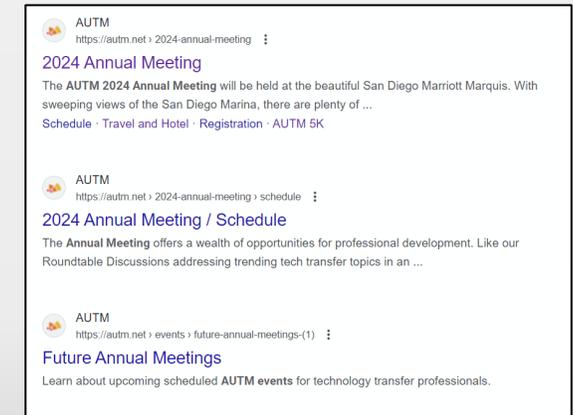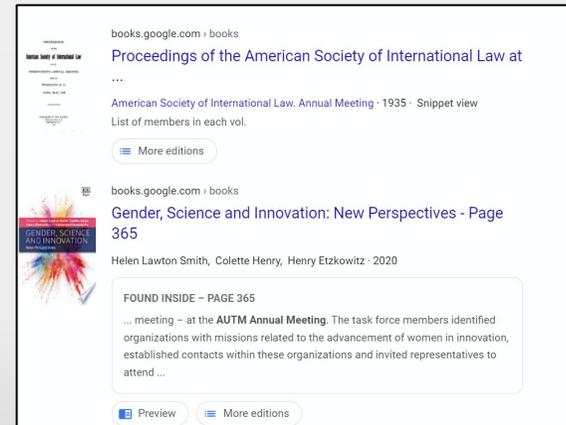
- *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003)
- *Perfect 10 v. Amazon.com,* 508 F.3d 1146 (9th Cir. 2007)
- *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009)
- *Authors Guild, Inc. v. HathiTrust.*, 755 F.3d 87 (2nd Cir. 2014)

Cases Relevant to Fair Use in the AI Context At the Supreme Court

Sun Java API Diagram
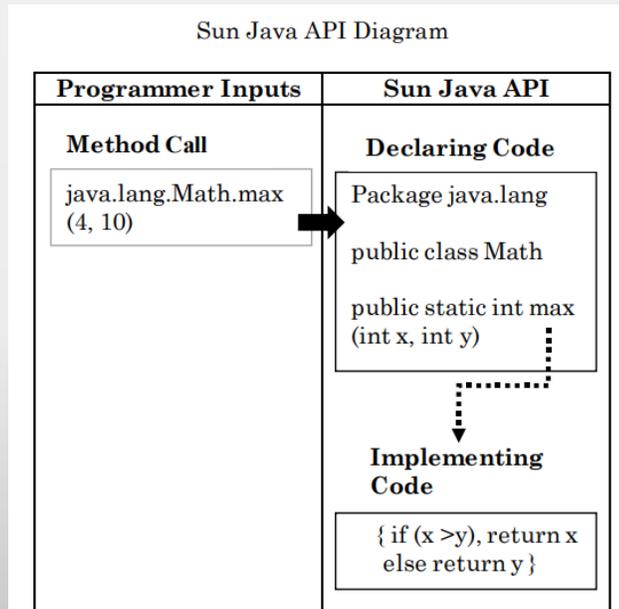
Google LLC v. Oracle Am., Inc., 141 S. Ct. 1183 (2021)

Figure 6. Warhol's orange silkscreen portrait of Prince superimposed on Goldsmith's portrait photograph.
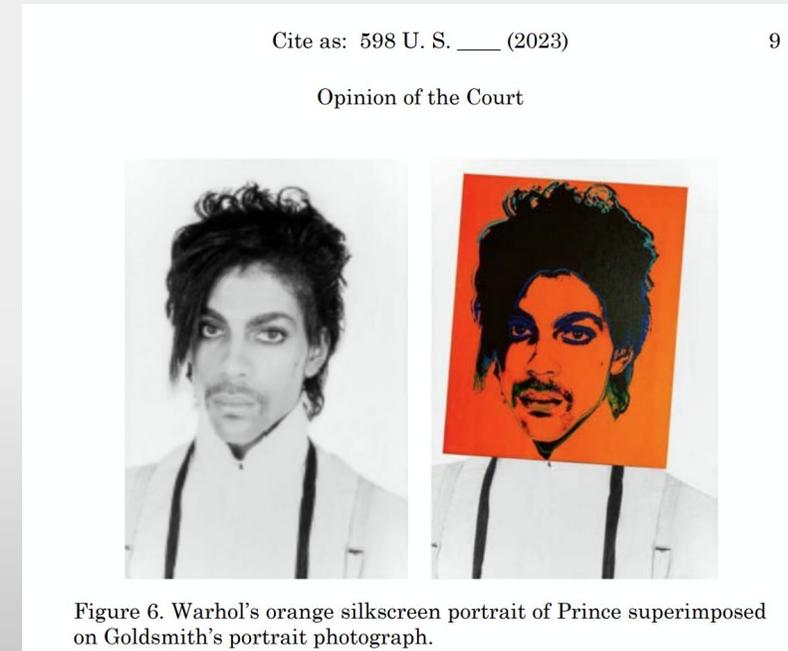
Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258 (2023)

At the Copyright Office

COPYRIGHT AND ARTIFICIAL INTELLIGENCE

Part 3: Generative AI Training PRE-PUBLICATION VERSION

A REPORT OF THE REGISTER OF COPYRIGHTS   MAY 2025
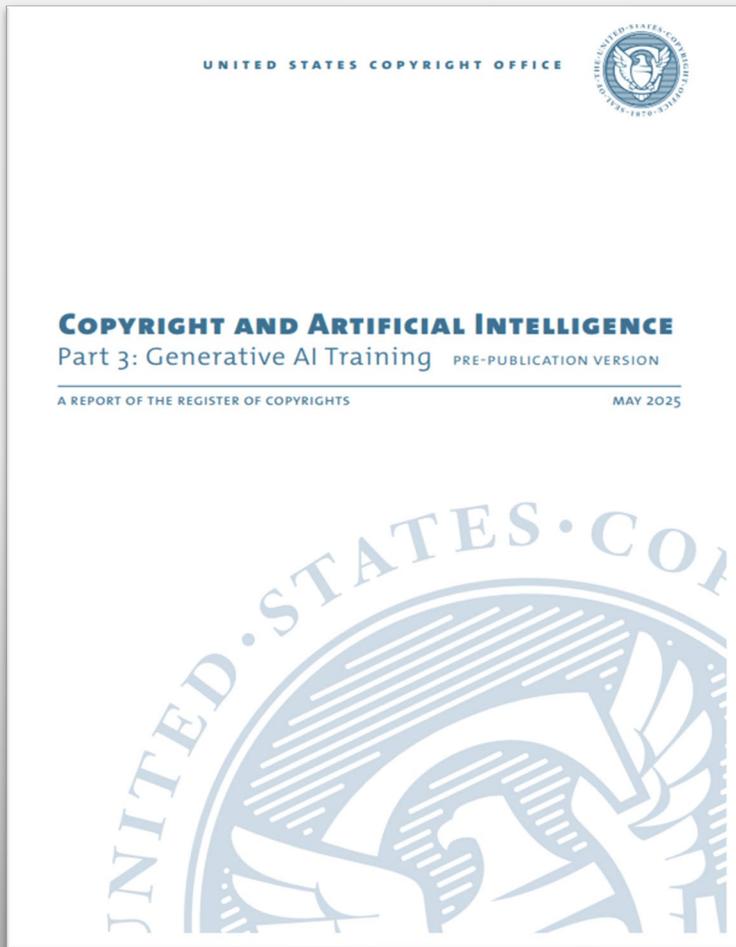
E. Weighing the Factors

It is for the courts to weigh the statutory factors together "in light of the purposes of copyright,"[415] with no mechanical computation or easy formula. How much each factor adds to the balance, and in which direction, will depend on the facts and circumstances of the particular case.

We observe, however, that the first and fourth factors can be expected to assume considerable weight in the analysis. Different uses of copyrighted works in AI training will be more transformative than others. And given the volume, speed and sophistication with which AI systems can generate outputs, and the vast number of works that may be used in training, the impact on the markets for copyrighted works could be of unprecedented scale.

As generative AI involves a spectrum of uses and impacts, it is not possible to prejudge litigation outcomes. The Office expects that some uses of copyrighted works for generative AI training will qualify as fair use, and some will not. On one end of the spectrum, uses for purposes of noncommercial research or analysis that do not enable portions of the works to be reproduced in the outputs are likely to be fair. On the other end, the copying of expressive works from pirate sources in order to generate unrestricted content that competes in the marketplace, when licensing is reasonably available, is unlikely to qualify as fair use. Many uses, however, will fall somewhere in between.

(Some) Cases Involving Generative AI

- *Thomson Reuters Enter. Centre GmbH et al v. ROSS Intelligence Inc.,* No. 1:20-cv-00613-SB (Del., May 6, 2020) (Partial Summary Judgment Granted on Feb. 11, 2025) (On Interlocutory Appeal)

- *Doe 1 v. Github, Inc.,* No: 4-22-cv-06823-JST (N.D. Cal., November 3, 2022) (On Interlocutory Appeal for DMCA Claims)

- *Andersen v. Stability AI Ltd,* No. 3:23-cv-00201-WHO (N.D. Cal., January 13, 2023) (Ongoing)

- *Getty Images (US) v. Stability AI,* No. 1:23-cv-00134-JLH (Del., February 2, 2023) (Venue Changing to ND Cal)

- *Getty Images (US) Inc. v. Stability AI Ltd,* HIGH COURT OF JUSTICE BUSINESS AND PROPERTY COURTS OF ENGLAND AND WALES INTELLECTUAL PROPERTY LIST (ChD),(2025)

Kadrey v. Meta Platforms, No. 3:23-cv-03417-VC (N.D. Cal. July 7, 2023) (Partial Summary Judgment Granted June 1, 2025)

- *In re Google Generative AI Copyright Litigation,* No. 5:23-cv-03440-EKL (N.D. Cal. July 1, 2023) (Ongoing)

- *Huckabee v. Bloomberg,* No. 1:23-cv-09152-LGS (S.D.N.Y., Oct. 17, 2023) (Motion to Dismiss Denied on November 24, 2025 based on questions of fact)

- *Concord Music Group, Inc. v. Anthropic PBC,* No. 5:24-cv-03811 (N.D. Cal., October 28, 2023) (Ongoing)

- *In re: OpenAI Inc. Copyright Infringement Litigation MDL,* No. 1:25-md-03143-SHS (S.D.N.Y, consolidated with *New York Times* and *Authors Guild* litigation) (ongoing)

- *Bartz v. Anthropic,* No. 3:24-cv-05417 (N.D. Cal.- SETTLED Sept. 5, 2025)

- *Advance Local Media v. Cohere,* No. 1:25-cv-01305 (S.D.N.Y. Feb. 13, 2025) (Motion to Dismiss Denied on November 13, 2025)

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

RICHARD KADREY, et al.,

Plaintiffs,

v.

META PLATFORMS, INC.,

Defendant.

Case No. 23-cv-03417-VC

QUESTIONS FOR HEARING

8. There must be some difference, from a fair use standpoint, between downloading pirated works and using them to train AI versus lawfully acquiring the works and then using them to train AI. The plaintiffs' argument that this is dispositive in their favor seems wrong, but Meta seems equally wrong to argue that it's entirely irrelevant. Assuming the Court rejects both sides' arguments on this point, how should the issue be treated?

1. … In other words, is the real question whether using the book to train the language model is sufficiently different from a human reading the book so as to make it transformative, without regard to what the language model ultimately enables people to produce after the copying?

2. … If the copying enables the model to produce both comparable and non-comparable works, how does that affect the analysis?

12. Imagine a case where the evidence showed that: (1) allowing use of protected books to train an AI model would, to a degree, diminish the market for the copied works; and (2) disallowing use of protected books to train the AI model would, to a degree, diminish the effectiveness of the AI model's ability to generate high-quality output. In that scenario, would the question of how easy or difficult it would be to obtain licenses to use the protected works for AI training be relevant to the fair use analysis?

Market vs Transformative Nature: Hearing on Motion for Summary Judgment, *Kadrey v. Meta Platforms*, No. 3:23-cv-03417-VC (N.D. Cal., May 1, 2025)

(4) Market Effects
- ""You are dramatically changing, **you might even say obliterating, the market for that person's work**, and you're saying you don't even have to pay a license to that person for using their work to create the product that's destroying the market for their work… **I just don't understand how that can be fair use.**"

- However, **must actually "show that the market for their actual copyrighted work is going to be <mark>dramatically affected</mark>"**

(1) Purpose and Character
- "I think that's probably where we come out, is that the use [of books to train Meta's large language models] is highly transformative"

"You could have a case where the use of the copyrighted works is highly transformative, but nonetheless it could have a massive effect on the market for the copyrighted works, such that it would not constitute fair use"

Fair Use Factors
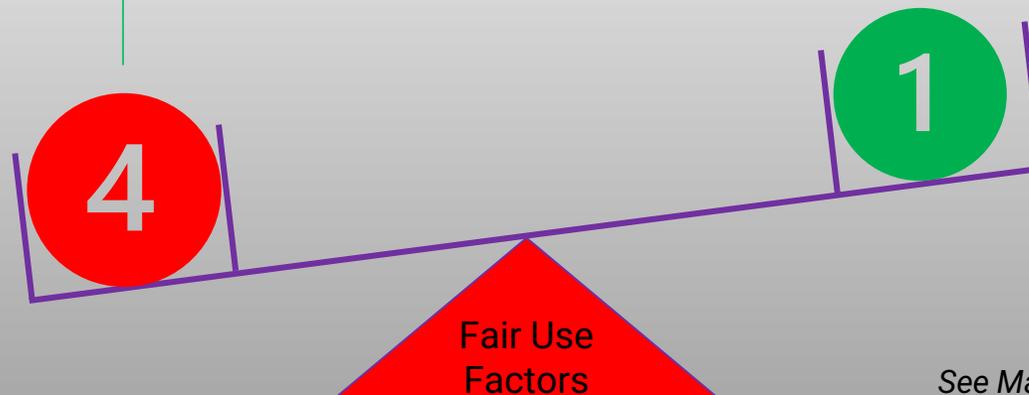
See Market Effect Key In Authors' IP Suit Against Meta, Judge Says, Law360 (May 1, 2025, 7:59 PM EDT), https://www.law360.com/ip/articles/2332741

March, 5, 2026

"In this case, because Meta's use of the works of these thirteen authors is highly transformative, the **plaintiffs needed to win decisively on the fourth factor** to win on fair use… . But the **plaintiffs presented no meaningful evidence on market dilution at all.** Therefore, on this record, Meta is entitled to summary judgment on its fair use defense to the claim that copying these plaintiffs' books for use as LLM training data was infringement"

(2) Nature of Original Work
- ❖ Books close to core of copyright
- ❖ Not a case of 'intermediate copying' since model relies on expression

(3) Amount Used
- ❖ Amount used reasonable given transformational purpose

(1) Purpose and Character
- ❖ Tool functions as more than a book
- ❖ Copying styles is not the same as copying expression
- ❖ Transformative nature outweighs commercial nature and fact of pirating and bad faith

(4) Market Effects
- ❖ Llama does not allow user to generate a meaningful portion of the works
- ❖ Licensing for transformative purpose not cognizable
- ❖ Not enough evidence to show indirect market substitution

2

1 3 4

Fair Use Factors

Splitting Out Uses: *Bartz v. Anthropic*, Decisions Partially Granting Summary Judgment, No. 3:24-cv-05417 (June 23, 2025)

	Training the Claude LLM	Creating Central Digital Library – Scanning Physical Copies	Creating Central Digital Library – Pirating Digital Copies
(1) Nature of Use	**Favors Fair Use:** ❖ Training is "quintessentially transformative" ❖ No outputs infringed ❖ Treats LLMs similar to comparable human activity	**Favors Fair Use:** ❖ Mere format change, since physical copies destroyed ❖ No distribution of scanned works ❖ *But*: LLM use distinguishable	**Points Against Fair Use:** ❖ Piracy not necessary for LLMs ❖ Digital library separate use than LLM, no justification ❖ Library had any use, not just LLM training
(2) Nature of Work	**Points Against Fair Use:** All works contained expressive elements, which was why they were chosen for incorporation in the LLM and Digital Library.		
(3) Amount Used	**Favors Fair Use:** ❖ Reasonably necessary to use whole books, given requirements to train LLMs	**Favors Fair Use:** ❖ Could already store hard copies ❖ Total copying needed to accurately digitize works	**Points Against Fair Use:** ❖ No entitlement to store pirated copies ❖ Any amount unreasonable
(4) Market Effect	**Favors Fair Use:** ❖ No evidence of knockoffs ❖ Competing works OK ❖ Licensing Market – no copyright act entitlement	**Neutral to Fair Use:** ❖ Deprived digital sale, ❖ Did not deprive physical sale ❖ *But:* Distributions actionable	**Points Against Fair Use:** ❖ Future intended fair use does not give right to pirated copy ❖ Pirated source displaced market – "copy for copy"
Fair Use?	**Yes, Fair Use.**	**Yes, Fair Use.**	**No, Not Fair Use.**

Anthropic / Author's Class Action Settlement

June 2025 Order

"The **downloaded pirated copies used to build a central library were not justified by a fair use**. **Every factor points against fair use.** Anthropic employees said copies of works (pirated ones, too) would be retained "forever" for "general purpose" even after Anthropic determined they would never be used for training LLMs. A separate justification was required for each use. None is even offered here except for Anthropic's pocketbook and convenience.

...

We will have a trial on the pirated copies used to create Anthropic's central library and the resulting damages, actual or **statutory** (**including for willfulness**). That Anthropic later bought a copy of a book it earlier stole off the internet will not absolve it of liability for the theft but it may affect the extent of **statutory damages....**"

September Settlement Agreement

Case 3:24-cv-05417-WHA   Document 363   Filed 09/05/25   Page 10 of 39

MEMORANDUM OF POINTS AND AUTHORITIES

I.   INTRODUCTION

If approved, this landmark settlement will be the largest publicly reported copyright recovery in history, larger than any other copyright class action settlement or any individual copyright case litigated to final judgment. The proposed settlement well surpasses other copyright recoveries, will provide meaningful compensation for each Class Work, and Class Counsel believes this will set a precedent of AI companies paying for their use of alleged pirated websites like Library Genesis and Pirate Library Mirror. Subject to court approval, the principal terms are:

- Anthropic will pay the Class at least *$1.5 billion dollars*, plus interest. With around 500,000 works in the Class, this amounts to an estimated gross recovery of *$3,000 per Class Work*.

- Anthropic will *destroy the LibGen and PiLiMi datasets* after the expiration of any litigation preservation or other court orders.

- In exchange, Anthropic will receive a *past release only* for conduct up to August 25, 2025. Claims arising out of conduct after August 25, 2025 won't be released by the Settlement, nor will any claims (past or future) arising out of allegedly infringing outputs from Anthropic's AI models.

This result is nothing short of remarkable.

March, 5, 2026

AI Generated Work (Ross)

Copyright-Protected Work (Thomson Reuters)

Public Domain

Question	West Headnote	Case Opinion
Does originality for copyright purposes mean that the work was independently created and has some minimal degree of creativity?	Originality, for copyright purposes, means that the work was independently created and has some minimal degree of creativity.	Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity.

"I hold that … Ross infringed 2,243 headnotes"

March, 5, 2026

(4) Market Effects
- ❖ Meant to Compete
- ❖ Affects Market for AI training data

(3) Amount Used
- ❖ Headnotes used for training only
- ❖ Output does not include work
- ❖ Irrelevant that not all headnotes used to train

(1) Purpose and Character
- ❖ Commercial
- ❖ Non-transformative
- ❖ Copying unnecessary

(2) Nature of Original Work
- ❖ Headnotes creative
- ❖ Less creative than others

Fair Use Factors

1 4 2 3
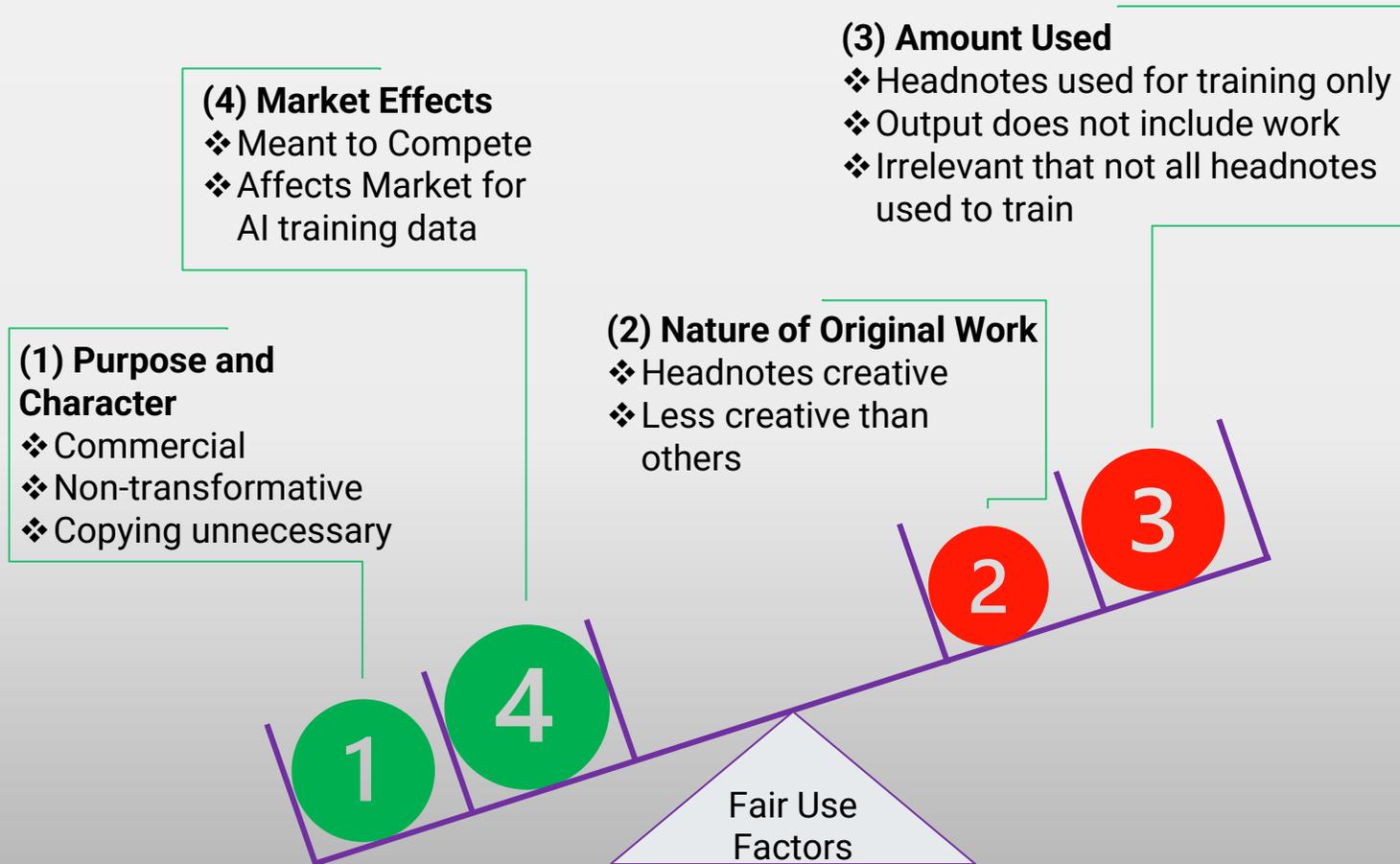
"E. Balancing the factors, I reject Ross' fair-use defense Factors one and four favor Thomson Reuters. Factors two and three favor Ross. Factor two matters less than the others, and factor four matters more. Weighing them all together, I grant summary judgment for Thomson Reuters on fair use."

"[C]ourts 'most frequently' address the defense at summary judgment, and a defendant's entitlement to a fair use defense must be' clearly established on the face' of a complaint 'to support dismissal.' … Bloomberg's claimed entitlement to a fair use defense does not meet this high standard, in large part because **the Court needs a robust factual record to conduct the fair use analysis in this case**."

(1) Purpose and Character
"[I]n examining the first factor, the Court must assess whether the allegedly infringing use was **commercial in nature**. … The Court lacks a sufficient factual record to make that determination. And, drawing all inferences in Plaintiffs' favor, the Amended Complaint suggests that Bloomberg's development of BloombergGPT was for a commercial purpose. Bloomberg is a finance news company. It allegedly built BloombergGPT as the world's first LLM AI program 'for finance.' FAC ¶ 44. **The plausible inference is that BloombergGPT will assist Bloomberg in conducting financial reporting, the core of Bloomberg's line of business, or developing new information products to sell to its financial industry customers.**"

(4) Market Effects
"In this case, analyzing this factor would conceivably require the Court to address at least the following questions:
Will widespread copying of literary works to train LLM AI programs result in a **deluge of AI-produced literary works, adversely affecting the market for the Works**? Because LLM AI programs have the potential to yield massive profits but require significant amounts of text to train, **is there a derivative market for literary works used to train AI**?
And, if so, **does unlicensed copying of literary works detrimentally harm an author's ability to participate in and profit from that derivative market**?"

1

4

Fair Use Factors

Amster Rothstein & Ebenstein LLP

March, 5, 2026

29

Getty brought a primary copy right claim known as the "Training & Development Claim".
Getty brought a secondary copy right claim under the Copyright, Designs and Patents Act 1988.

Getty abandoned the primary claim: "no evidence that the training and development of Stable Diffusion took place in the United Kingdom (such that what has been called 'the Training and Development Claim' has been abandoned)"

"the model **weights** are **not themselves an infringing** copy and they **do not store an infringing copy**."

"The fact that its development **involved the reproduction of Copyright Works** (through storing the images locally and in cloud computing resources and then exposing the model weights to those images) is of **no relevance**."

"In the circumstances, **an electronic copy stored in an intangible medium** (such as the AWS Cloud) is, in my judgment, capable of being an infringing copy and thus also capable of being "**an article**". (156) (AI models are subject to the same copyright infringement claims as tangible articles).

Key Take Aways

Training Set

L
L
M

Output

Lawfully obtained – Probably Fair Use

Unlawfully obtained – Probably Not Fair Use

Not substantially similar to source – Probably Not Infringement

Substantially similar to source – Probably Infringement

March, 5, 2026
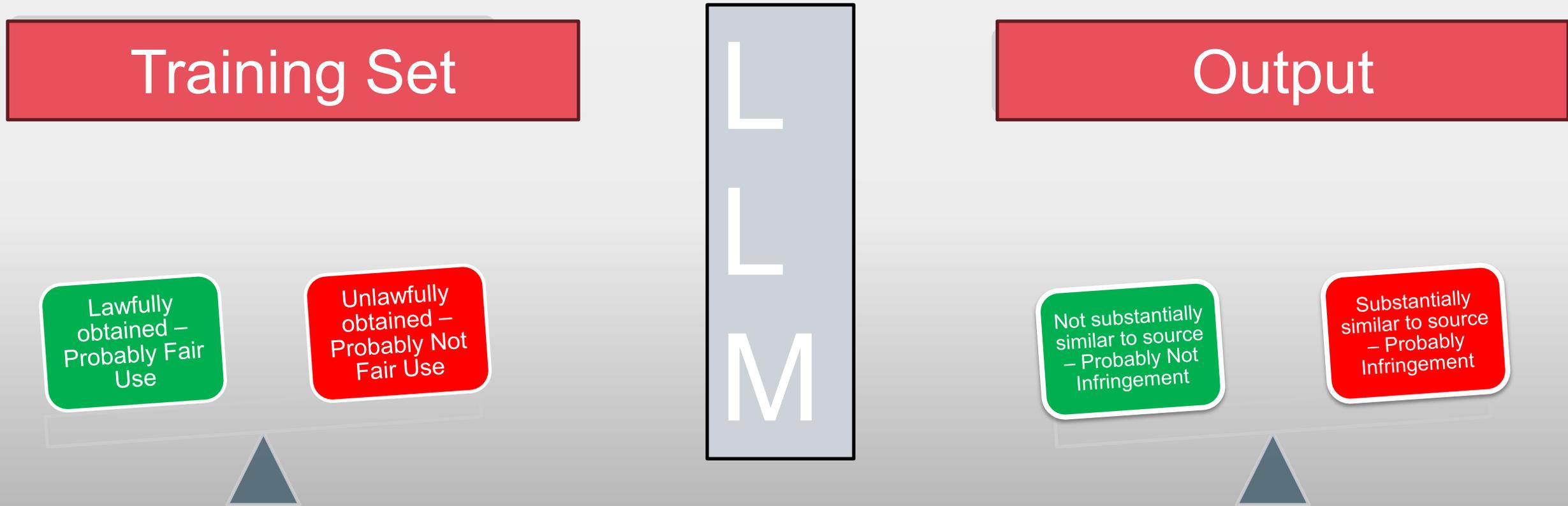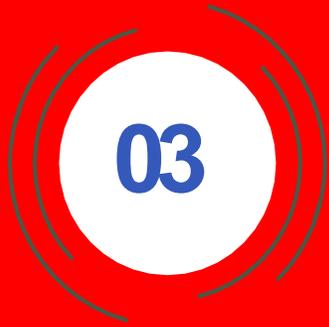
03 Policy considerations for organizational compliance
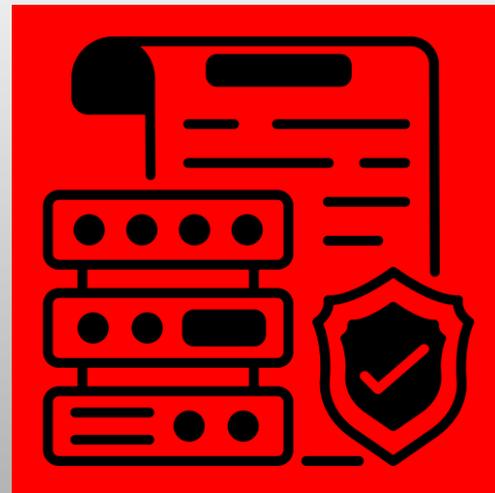
Legal & Regulatory Compliance

- **Privacy & Human Subjects Laws**
  - **HIPAA (health data, de-identification), FERPA (student records), GDPR/UK GDPR (international collaborations), state privacy laws (e.g., CCPA).**
  - **Institutional Review Board (IRB) policies and patient consents.**
- **Export Control / National Security**
  - **EAR/ITAR restrictions and sanctions lists may prohibit transfer of certain datasets (e.g., biomedical or dual-use research).**
- **Copyright & IP**
  - **Ensure the Organization has rights to the underlying works (e.g., texts, images, software outputs).**

Ethical & Mission Alignment

- ## Non-Discrimination & Responsible AI
  - **Policy guardrails to prevent data being used in biased, discriminatory, or harmful applications.**

- ## Alignment with Academic Mission
  - **Organization policies often restrict data licensing that conflicts with open research, freedom to publish, or educational access (e.g., permitted uses).**

- ## Transparency & Accountability
  - **Clarify how trained models may be disclosed or validated to avoid "black box" exploitation of organization resources (e.g., models provide reasoning).**
  - **Traceability if model or data are used in a regulated product.**

Governance & Institutional Risk

- ## Data Stewardship
  - **Many organizations adopt "data governance boards" or "research data offices" to evaluate requests for high-value datasets.**

- ## Reputation & Public Trust
  - **Licensing decisions must weigh potential negative publicity if data is misused (e.g., medical data being commercialized in controversial ways).**

- ## Conflicts of Interest
  - **Policies should ensure that administration, faculty, lab, etc. involved in licensing negotiations complies with conflict-of-interest and conflict-of-commitment rules.**

March, 5, 2026

Operational & Security Standards

- **Data Use & Security Requirements**
  - **Encryption, controlled access, audit trails, breach notification obligations.**

- **Data Lifecycle Policies**
  - **Clear rules on retention, deletion (e.g. GDPR deletion request), and return of data or destruction after license expiration.**

- **Cross-Border Transfers**
  - **Require compliance with institutional and governmental rules on storing or processing data outside the home jurisdiction (including non-U.S. citizens in the lab) (e.g. US Bulk Data Rule).**

March, 5, 2026

Financial & Value Capture

- ## Royalty & Revenue Sharing Policies
  - **Organizations may have internal rules on how revenue from data/IP licensing is distributed (collector, PI, department, institution).**

- ## Cost Recovery
  - **Policies may require cost recovery for data extraction, data curation, compilation, de-identification/anonymization, transformation, compliance, overhead etc.**

- ## Non-exclusive vs. Exclusive Use
  - **Default policy is usually non-exclusive licensing to preserve academic and future commercial opportunities.**
  - **Time-limited exclusive license or time-limited commercial embargos may be permitted on a strategic / tactical basis.**

Amster Rothstein & Ebenstein LLP

Social Responsibility

- ## Access Considerations
  - Some organizations set policies that data licensed to industry must also remain available for academic/nonprofit use.
- ## Global Equity
  - Policies may restrict licensing if the downstream use could disproportionately harm vulnerable populations.
- ## Open Science Compatibility
  - Balancing proprietary licensing with commitments to open data and public benefit.
  - Particularly Federally Funded Research requires open data.

Amster
Rothstein &
Ebenstein LLP

Policy Considerations in Licensing Data for AI Training

1. Legal & Regulatory Compliance

Privacy & Human Subjects Laws
- HIPAA (health data, deidentification), FERPA (student records), GDPR/UK GDPR (international collaborations), state privacy laws (e.g., CCPA).
- Institutional Review Board (IRB) policies and patient consents.

Export Control / National Security
- EAR/ITAR restrictions and sanctions lists may prohibit transfer of certain datasets (e.g., biomedical or dual-use research).

Copyright & IP
- Ensure the organization has rights to the underlying works (e.g., texts, images, software outputs).

2. Ethical & Mission Alignment

Non-Discrimination & Responsible AI
- Policy guardrails to prevent data being used in biased, discriminatory, or harmful applications.

Alignment with Academic Mission
- Organization policies often restrict data licensing that conflicts with open research, freedom to publish, or educational access (e.g. permitted uses).

Transparency & Accountability
- Clarify how trained models may be disclosed or validated to avoid "black box" exploitation of organization resources (e.g. models provide reasoning)
- Traceability if model or data are used in a regulated product.

3. Governance & Institutional Risk

Data Stewardship
- Many organizations adopt "data governance boards" or "research data offices" to evaluate requests for high-value datasets.

Reputation & Public Trust
- Licensing decisions must weigh potential negative publicity if data is misused (e.g., medical data being commercialized in controversial ways).

Conflicts of Interest
- Policies should ensure that administration, faculty, lab, etc. involvement in licensing negotiations complies with conflict-of-interest and conflict-of-commitment rules.

4. Operational & Security Standards

Data Use & Security Requirements
- Encryption, controlled access, audit trails, breach notification obligations.

Data Lifecycle Policies
- Clear rules on retention, deletion (e.g. GDPR deletion request), and return of data or destruction after license expiration.

Cross-Border Transfers
- Require compliance with institutional and governmental rules on storing or processing data outside the home jurisdiction (including non-U.S. citizens in the lab) (e.g. US Bulk Data Rule).

5. Financial & Value Capture

Royalty & Revenue Sharing Policies
- Organizations often have internal rules on how revenue from data/IP licensing is distributed (researcher vs. institution).

Cost Recovery
- Policies may require cost recovery for data curation, de-identification/anonymization, or compliance overhead.

Non-exclusive vs. Exclusive Use
- Default policy is usually non-exclusive licensing to preserve academic and future commercial opportunities.
- Time-limited exclusive license or time-limited commercial embargos may be permitted on a strategic / tactical basis.

6. Social Responsibility

Access Considerations
- Some organizations set policies that data licensed to industry must also remain available for academic/nonprofit use.
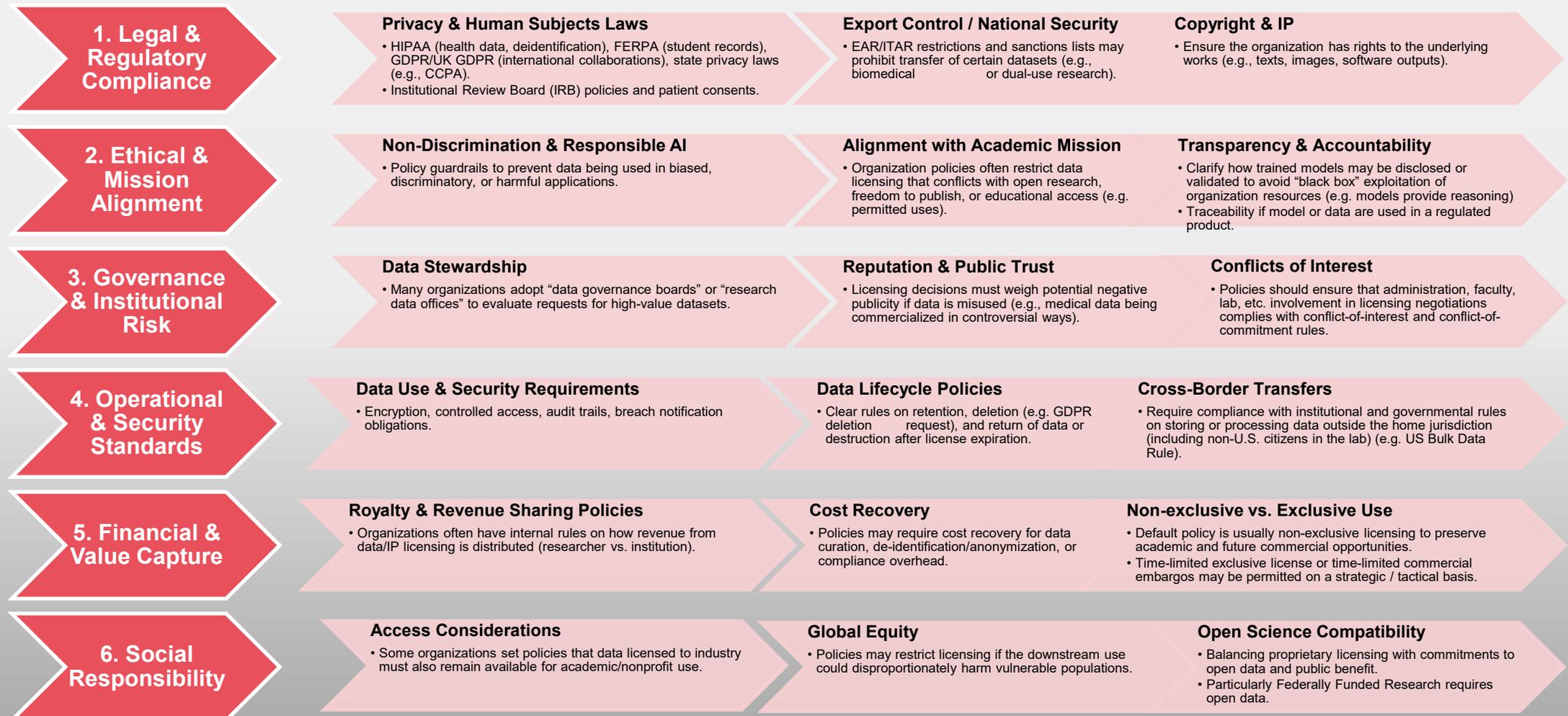
Global Equity
- Policies may restrict licensing if the downstream use could disproportionately harm vulnerable populations.

Open Science Compatibility
- Balancing proprietary licensing with commitments to open data and public benefit.
- Particularly Federally Funded Research requires open data.

March, 5, 2026

Bottom Line

Organization policy considerations in data licensing must ensure:

- **Legal compliance with**
  - privacy, IP, and export controls.
- **Mission alignment with**
  - academic and ethical values.
- **Institutional protection against**
  - reputational and liability risks.
- **Governance frameworks to ensure**
  - transparency, fairness, and oversight.

Questions and Discussion

Amster
Rothstein &
Ebenstein LLP

Connect with Our Presenters

ARE LLP

CHARLES R. MACEDO