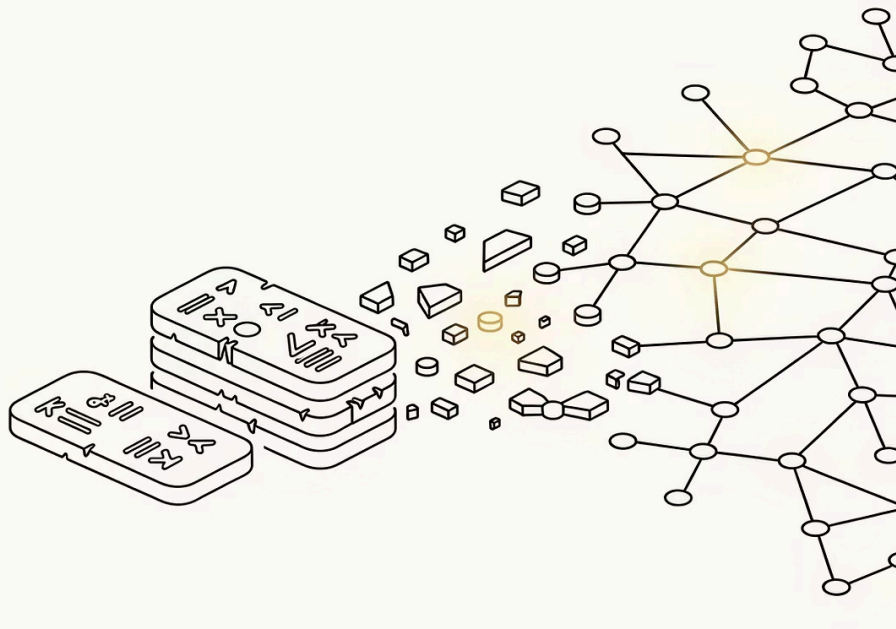
THE ATTRIBUTION PROBLEM IS UNSOLVED

And it's the hinge the next decade turns on.

Bell AI Fellowship Workshop · Tenni Theurer · April 2026



HISTORICAL CONTEXT

5,000 YEARS OF PLATFORM POWER GRABS

Every content technology follows the same arc: **excite creators, consolidate power, commoditize creators.** From clay tablets to AI licensing, the pattern is relentless — and we are living through its latest iteration.

1 CLAY TABLETS → PRESS

3400 BCE–1440 CE: whoever controlled copying controlled the money. The Statute of Anne (1710) — the first copyright law — was invented to solve the attribution problem.

2 GOOGLE NEWS (2002)

"We drive clicks" — showed enough information that users didn't need to click through to the original source.

3 FACEBOOK VIDEO (2016)

Inflated metrics 150–900%. Publishers bankrupted. Facebook paid $0. The pivot destroyed newsroom budgets built on those false numbers.

4 APPLE NEWS+ (2019)

Apple takes 50%. Publishers received "one-twentieth" of promised revenue. The math never worked for anyone below the top tier.

5 AI LICENSING (2024)

"We'll pay for your content" — $817M/yr total paid vs. $2B+/yr destroyed in referral traffic. Net negative for the ecosystem.

$49B

NEWSPAPER AD REVENUE

Peak in 2000

$6B

NEWSPAPER AD REVENUE

By 2023 — an 88% collapse

$300B+

GOOGLE AD REVENUE

2025 — captured from the same ecosystem

DEMAND CRISIS

AI IS ELIMINATING CLICKS — AND THE AUDIENCE HAS ALREADY MOVED

ZERO-CLICK EXTINCTION

69%

of Google searches end without a click (May 2025)

-38%

Google organic traffic to US publishers, year-over-year

60–95%

Google traffic lost by small publishers

THE AUDIENCE ISN'T WHERE YOU THINK

- **43%** of under-30s get news on TikTok. **64%** of Gen Z use TikTok as a search engine.
- **Reddit** is the #1 LLM citation source — 40.1% of all AI search citations.
- Only **28%** of Americans trust mass media (Gallup all-time low).
- Budget mismatch: **$74.9B/yr on SEO** (90%+ flowing to Google), while TikTok, YouTube, Reddit, and ChatGPT drive 60%+ of actual discovery.

⚠ AI chatbot referrals grew 200% but still account for less than 1% of all pageview referrals. Growth is real. Scale is not.

LEGAL FRAMEWORK

TRAINING VS. GROUNDING: THE MOST IMPORTANT DISTINCTION NOBODY HAS ADJUDICATED

TRAINING

Content absorbed into model weights during learning. Millions of articles processed over weeks of compute. Once trained, content is "baked in" — distributed across billions of numerical parameters. **You can't un-train on specific articles without retraining from scratch.**

This is what the NYT lawsuit is about.

Legal status: 2 of 3 judges say fair use; 1 says not if the content was pirated. Genuinely unsettled.

GROUNDING (RAG)

Content fetched at query time and placed in the model's context window alongside the user's question. The model reads it, synthesizes a response, and the content leaves the context window when done. **Content never enters permanent weights.** Like an open-book exam vs. memorization.

This is what content marketplaces are built for.

Legal status: No court has ruled. Completely unsettled.

Why this matters for builders: If you're building anything that uses external content to improve AI responses, you're in the grounding category. The legal framework for what you're doing does not yet exist.

THE DEFINING CASE

NYT V. OPENAI/MICROSOFT

The lawsuit that will set the terms for AI content licensing for the next decade. Filed December 27, 2023. SDNY. Seeking billions in damages.

DEC 27, 2023

NYT files suit. 100+ examples of verbatim content reproduction in the complaint.

APRIL 2025

All core copyright claims survive motion to dismiss. Case proceeds to discovery.

DISCOVERY

100M+ chat logs ordered — the largest AI discovery corpus in history. 16+ lawsuits consolidated into a single MDL.

SETTLEMENT WATCH

Anthropic settled for **$1.5B** — the largest copyright recovery in US history. 70+ total copyright suits against AI companies ongoing.

Fair use scorecard: 3 judges have ruled — 2 for AI companies, 1 against. The law is genuinely unsettled.

STRATEGIC ANALYSIS

THE STRATEGIC PARADOX

AI companies are simultaneously **defendants** AND **operators of content marketplaces.** The outcome of NYT v. OpenAI cuts both ways — and the optimal outcome may not be what either side publicly advocates.

IF NYT WINS

- **Pro:** Licensed content marketplaces become legally required, not optional
- **Pro:** Creates moat — only deep-pocketed players can afford licensing
- **Con:** AI companies pay billions in damages
- **Con:** Content costs rise across the board

IF NYT LOSES

- **Pro:** No damages; freedom to operate
- **Con:** Content marketplace value proposition collapses
- **Con:** Publishers abandon licensing negotiations entirely
- **Con:** Race to bottom on quality; no incentive for premium content

The optimal outcome is **settlement** — establishes licensing norms without binding precedent. Keeps the legal question ambiguous enough that publishers feel they need marketplaces, but not so clear that past practices create massive liability.

CORE PROBLEM

THE ATTRIBUTION PROBLEM

When an AI synthesizes an answer from 15 sources, how do you measure each source's contribution? This is not a policy question. It is a hard mathematical and engineering problem — and it hasn't been solved.

THE THEORY: SHAPLEY VALUES

Cooperative game theory from 1953 (Nobel Prize 2012). Four fairness axioms:

- **Efficiency:** All value is distributed
- **Symmetry:** Equal contributors get equal share
- **Dummy player:** Zero-contribution sources get nothing
- **Additivity:** Values are consistent across games

THE PROBLEM: COMPUTATIONAL COST

Computing exact Shapley for N sources requires **2^N forward passes.**

- At 15 sources: 32,768 inference passes per query
- At Copilot scale: ~**$320 per query**, ~$32 billion per day
- Monte Carlo approximation: ~**$0.45/query** — still 15× the cost of the answer itself

⚠ What actually ships: URL-based citations and proprietary bookkeeping. CJR found AI citations are wrong 60%+ of the time.

TECHNICAL DEEP DIVE

HOW ATTRIBUTION ACTUALLY BREAKS UNDER THE HOOD

The AI pipeline has three stages. Attribution fails at each one — and the compounding effect is total opacity.

STAGE 1 — INGESTION

Content crawled, chunked, embedded into vector stores. Author, source, and license metadata may or may not survive chunking. C2PA provenance metadata is typically discarded. Source URLs stored as application-level bookkeeping — not cryptographic provenance.

STAGE 2 — RETRIEVAL

User query is embedded; nearest neighbors retrieved. The retriever decides what's relevant — but **relevance ≠ contribution.** A document can be retrieved and contribute nothing, or contribute heavily but rank low in retrieval scoring.

STAGE 3 — GENERATION

The model synthesizes retrieved chunks + parametric knowledge into a response. **This is the black box.** There is no mechanism inside the transformer that tracks "this sentence came from chunk 3." Citation is a post-hoc guess, not a measured signal.

⊗ **The fundamental problem:** Attribution requires knowing contribution at Stage 3, but we only have reliable tracking at Stage 1. The middle is opaque.

TECHNICAL LANDSCAPE

THE TECHNICAL APPROACHES: WHAT EXISTS AND WHAT'S BUILDABLE

🟢 SHIPPING TODAY

Output decomposition (ProRata): Decomposes AI answer into claims, matches post-hoc. 700+ publishers. But it's pattern matching, not causal attribution — and the algorithm is a black box with no independent audit.

Trained citation (Perplexity): Model trained to self-report sources. Scalable. But models hallucinate citations just like they hallucinate facts.

🟡 RESEARCH STAGE

Counterfactual RAG (RAGonite): Remove each source, re-run query, measure change. Causal and interpretable. Requires N+1 inference passes — expensive.

Influence functions (TRAK/TrackStar): Google scaled to 8B params / 160B tokens. Key finding: causal influence ≠ factual retrieval.

Embedding uniqueness: Score content by replaceability. Fast, scalable. But weak correlation with actual value.

🔴 NOT READY — WATCH CLOSELY

Cluster Shapley: Groups similar docs, reduces 2^N to manageable — but loses individual-level attribution.

FreeShap (ICML 2024): Shapley for LLMs via neural tangent kernels. No retraining needed. Early but architecturally elegant.

MARKET OPPORTUNITY

THE WHITE SPACE: WHAT NOBODY HAS BUILT

The patent space is unoccupied. The unified framework doesn't exist. This is the opportunity.

THE EXISTING PIECES

Citation prediction from embeddings (SPECTER, SemNovel) is well-established. Embedding uniqueness correlating with citation impact is confirmed. Data valuation via Shapley/influence functions is active research. Commercial attribution exists (ProRata, Bria) but uses different methods.

THE MISSING BRIDGE

No one has used citation-validated uniqueness as a training signal for a content value estimator. The path from "uniqueness correlates with value" to "deployed attribution system" has not been built.

WHAT'S BUILDABLE NOW

Counterfactual RAG attribution for per-query grounding credit. Embedding-based uniqueness for corpus-level valuation. Open-source tooling exists: pyDVL, TRAK, RAGonite, C2PA SDKs, RSL. A startup could build a **"Stripe for content attribution"** using existing open-source components.

⚠ **The risk:** The window is 2–3 years. After that, the market locks in simpler heuristic methods. Whoever ships first defines "good enough." The economically optimal solution wins over the mathematically fair one.

INFRASTRUCTURE

THE AGENTIC WEB: THE PROTOCOL LAYER BEING BUILT RIGHT NOW

The infrastructure for AI agents to negotiate content access programmatically is emerging — but adoption is still in the primordial soup stage.

MCP — MODEL CONTEXT PROTOCOL

Open standard for AI agents to connect to external tools and data. Microsoft joined the steering committee. Think of it as **HTTP for agents** — a standardized way for any AI to talk to any service.

NLWEB — "HTML FOR THE AGENTIC WEB"

Microsoft's open-source project by R.V. Guha (creator of RSS, RDF, Schema.org). Lets any website expose a natural language interface for AI agents. Early adopters: Shopify, Tripadvisor, Eventbrite, O'Reilly.

RSL — REALLY SIMPLE LICENSING

Machine-readable licensing in robots.txt. Four pricing models: pay-per-crawl, pay-per-inference, subscription, free with attribution. **1,500+ orgs** including Reddit, Yahoo, Cloudflare, AP, Stack Overflow.

Gartner: The agentic web "paints a compelling vision for open interoperability, but we are still very early in the primordial soup of agentic standards."

REALITY CHECK

THE GAP BETWEEN VISION AND REALITY

WHAT'S BEING SOLD

"AI agents reliably access, cite, and pay for premium content."

A coherent, fair, metered ecosystem where every creator gets compensated for every use.

WHAT ACTUALLY EXISTS

- URL-based citations that are wrong **60%+ of the time**
- Business rules for payment splits — not mathematical fairness
- No standard for embedding source provenance into AI outputs
- **99.6% of AI content usage is invisible** — models use content far more than they cite
- C2PA can bookend the pipeline but cannot track content *through* it
- C2PA v2.3 text provenance standard published January 2026 — near-zero adoption
- No court has ruled on grounding/RAG

⊗ **Anyone claiming attribution is "solved" is ahead of the science.** The infrastructure, the legal framework, and the technical standards are all simultaneously incomplete.

BUSINESS MODELS

SIX MODELS FOR CONTENT IN THE AI ERA

No single model is sufficient. Viable strategies combine 2–4 of the following. The models that work are increasingly creator-native, not publisher-native.

1

AI LICENSING (LUMP-SUM)

Flat fee for training/grounding rights. Works for top 50–100 publishers only. $817M/yr total — net negative vs. traffic destruction.

2

AI LICENSING (PER-USE)

Fee per AI response use. In theory, works for anyone; in practice, top 10–20 capture 80%+. Same power law as Spotify.

3

SUBSCRIPTIONS

Readers pay directly. Works for elite brands (NYT: 12.2M subs) and creators (Substack). Only 18% of users pay; 57% leave at paywalls.

4

COLLECTIVE LICENSING

ASCAP/BMI model for text. Only viable path for small and mid-tier publishers. RSL has 50+ publishers including Reddit. Early but accelerating.

5

ADVERTISING

Declining for publishers. Exploding for creators. YouTube paid creators $60B+ in 2025. The same model produces radically different outcomes by channel.

6

DIRECT AUDIENCE MONETIZATION

Merch, events, courses, community. 60%+ of creator earnings. Content is top of funnel, not the product itself.

THE CAUTIONARY TALE

THE SPOTIFY PARALLEL

Per-use AI attribution will follow the same distribution as music streaming — and the implications for the journalism ecosystem are severe.

THE ECONOMICS

Spotify pays $0.003–0.005 per stream. Only the top **1–2% of artists** earn a living wage. The platform is technically paying everyone. Practically, it's meaningless for 98%.

THE MATH FOR PUBLISHERS

A Guardian article cited 50 times a day earns fractional cents. A local newspaper cited twice a month earns effectively nothing. Volume requirements are impossible at the local level.

THE DILUTION PARADOX

The marketplace incentive is to sign *everyone* — it grows the catalog and legitimizes the platform. The individual publisher incentive is for *competitors* to stay out. These are fundamentally opposed incentives that cannot be reconciled.

STRUCTURAL SHIFT

CREATORS VS. PUBLISHERS: A DIVERGING FUTURE

TRADITIONAL PUBLISHERS

Produce commodity information — news, facts, analysis — that AI can synthesize from multiple sources. The AI threat is **existential.**

- Revenue down 88% in advertising since 2000
- Only top 50–100 can access AI licensing deals
- Moat: investigative journalism, archives, institutional credibility

INDEPENDENT CREATORS

Produce personality-driven content that is inherently non-fungible — AI cannot replicate a specific person's perspective. The AI threat is **moderate.**

- $253B creator economy, 23% CAGR
- Too small to negotiate AI licensing individually
- Moat: authenticity, community, niche expertise

The question isn't "how do publishers survive AI?" — it's **"who are the publishers in an AI world?"** The answer increasingly looks like individuals with audiences, not institutions with newsrooms. The business model that survives is built on trust, authenticity, and verified provenance — not volume.

THREE SCENARIOS FOR 2026–2030

A: LICENSED WEB (20-25%)

NYT strong, EU Act, C2PA std. Middle class survives.

B: SPOTIFY MODEL (45-50%)

Norms set, but low rates. Top publishers concentrate.

C: FAIR USE WINS (25-30%)

Courts rule fair use. Only elite brands survive.

In 2 of 3 scenarios, content licensing becomes a version of the Spotify model — **technically fair, practically meaningless for everyone except the top 1%.** The Spotify Model is where the trends are naturally heading without intervention.

FORWARD LOOK

WHAT TO WATCH

NEXT 6 MONTHS — KEY DATES

- **August 2, 2026:** EU AI Act transparency rules go live — first enforceable content-related AI regulation
- **Q2 2026:** Authors Guild class certification hearing
- **NYT case:** Settlement or major discovery revelations from 100M+ chat logs
- First real usage data from content marketplaces

TECHNOLOGY TRAJECTORY

- Inference costs falling 50–70%/yr. Monte Carlo Shapley drops from $0.45 to ~$0.005/query at 100× reduction — expensive but viable for high-value queries
- C2PA works for images. **Text provenance is the hard problem.** SynthID marks AI output but doesn't attribute human sources
- MCP and NLWeb adoption is the leading indicator for the agentic web
- **The window: 5–10 years** to build sustainable content models — models get better and need less grounding over time

THE LONG VIEW

THE 5,000-YEAR VIEW

> Every technology that expanded access to knowledge — writing, the alphabet, printing, the internet — eventually concentrated power in the hands of whoever controlled *distribution*, not creation.

The scribes lost to the monasteries. The monasteries lost to the printers. The printers lost to the platforms. The question for AI is whether "license and attribute" breaks this cycle — or whether it's the content equivalent of paying Spotify royalties.

WHAT COULD MAKE THIS TIME DIFFERENT

Legal precedent that creates a floor under creator compensation — not voluntary norms, but enforceable rights that survive platform transitions.

AND THE TECHNICAL REQUIREMENT

Attribution technology that makes content individually identifiable and compensable at inference time — not post-hoc bookkeeping, but causal measurement.

Neither is guaranteed. **Both are in motion right now.**

CLOSING ARGUMENT

THE ATTRIBUTION PROBLEM IS UNSOLVED — AND THE CLOCK IS RUNNING

The hinge the next decade turns on is not regulation, or market structure, or which AI company wins. It is whether we can measure, at the moment of generation, what each piece of human knowledge contributed to an AI response — and pay accordingly.

ORIGINAL REPORTING

What AI genuinely cannot do: break news, investigate wrongdoing, verify claims at the source.

AUTHENTIC PERSONALITY

Perspective that is valued for *who made it*, not just what it says. Non-fungible by definition.

COMMUNITY TRUST

Relationships built over time between a creator and their audience. The moat that scales with attribution, not despite it.

VERIFIED PROVENANCE

Knowing who made something, when, and why. The technical and legal foundation that makes all the other moats defensible.

The window to solve this is 5–10 years — not permanent. Models are getting better. They will need less grounding over time. The time to build sustainable content economics in the AI era is now.