Who Will Monetize Truth?

A Thesis for the Future of the Information Business

Francesco Marconi

CEO & Co-Founder, AppliedXL

REF: TB-2026-001 / DISTRIBUTION: GENERAL / CLASSIFICATION: OPEN

CONTENTS

The Thesis — Three Species, One Ecosystem

Q1 In a post-AGI world, where does value accrue in the information ecosystem?
Q2 Who wins and loses among media institutions?

Q3 What happens when AI agents, not humans, become the primary consumers of news?
Q4 Is “information advantage” the new alpha?
Q5 What does the intelligence layer look like outside finance?

Q6 If AI can hallucinate citations, what happens to the concept of the record?

Q7 Are prediction markets the next major consumer of structured information?
Q8 Is trust mispriced?

Q9 If this becomes a data game, what's the trade?
Q10 What happens when AI runs out of journalism to train on?
Q11 Can wire services capture the value they produce?
Q12 What can't be automated?

Q13 If AI writes all the articles, what is the newsroom actually for?
Q14 What happens when $1 of AI replaces $33 of freelance labor?

Q15 What happens to the New York Times? What happens to Google?

Conclusion — Who Will Monetize Truth?

THE THESIS

Three Species, One Ecosystem

Every satisfying explanation of what's happening to the media industry is wrong. It's not dying. It's being repriced. Water doesn't gradually become steam. It stays liquid at 211 degrees. Still looks like water, still behaves like water. Then one more degree and the entire structure shifts.

The news industry no longer looks like itself. Not because the mastheads have changed or the buildings are empty. Because the audience, the economics, the distribution, and the competitive set have all been replaced while the surface stayed the same. And most of the industry is still optimizing for a game that's already over.

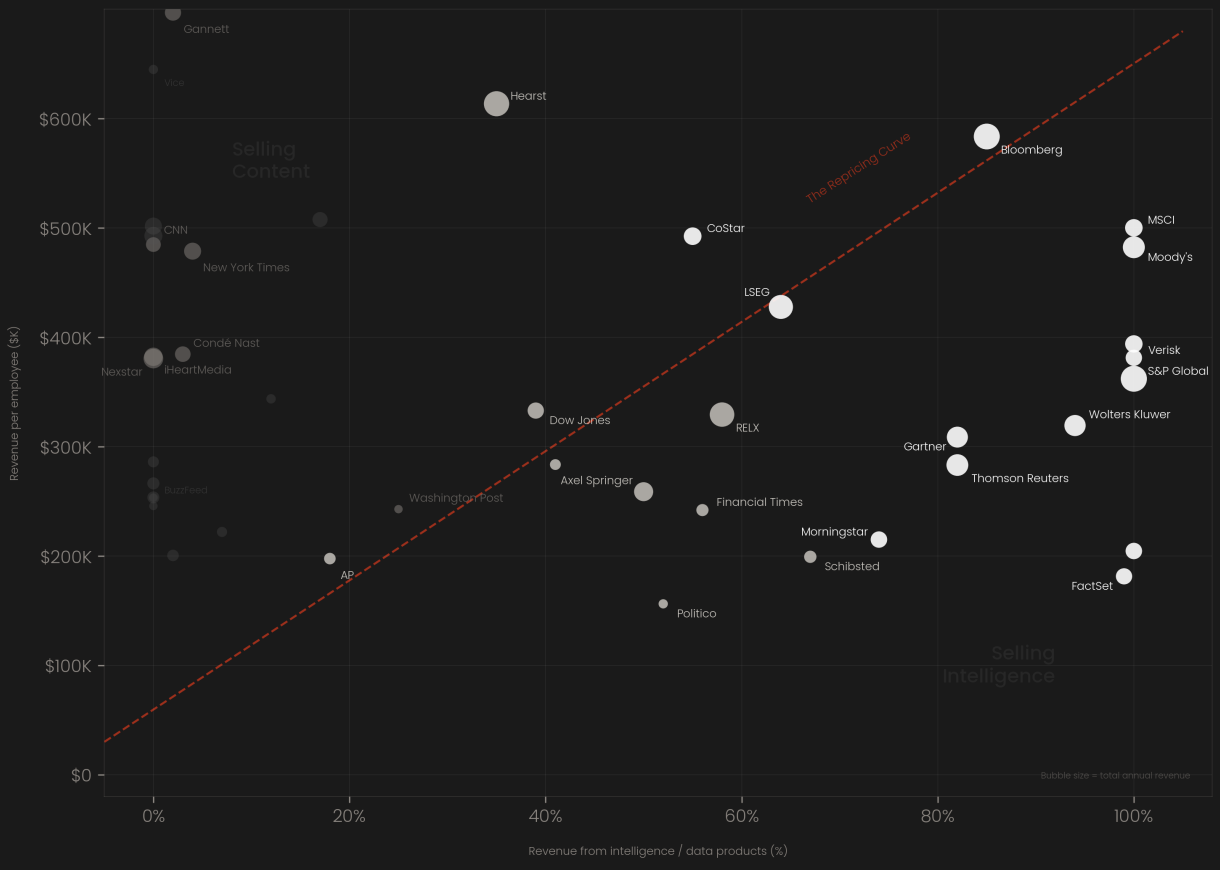
We're at 211.

The media industry is splitting into three different species. The Intelligence Business, the Attention Aggregator, and the Public Good. Only the first has pricing power in an era of abundant content. The second faces structural collapse not because awareness lacks value but because AI is making it free. The third will not survive as a business and it needs a different funding model entirely. These are not stages of evolution. They are distinct businesses with distinct economics. The classification is determined by who pays and what they do with it. The worst outcome is executing the strategy for the wrong one.

The temperature has been rising for twenty years. AI made it exponential. I've spent the past decade inside the institutions being repriced, the Associated Press, The Wall Street Journal, and now AppliedXL. What follows is what I've seen.

THE GREAT REPRICING

The Great Information Repricing



The chart shows two distinct clusters because they are two distinct businesses. The companies in the middle, Dow Jones, the Financial Times, Schibsted, Hearst, are the most instructive. They discovered they contained both species: a content operation and an intelligence operation, sharing the same newsroom but serving different buyers. The question for every institution is not whether to move right. It is whether it has intelligence assets trapped inside a content wrapper.

Q1

In a post-AGI world, where does value accrue in the information ecosystem?

OpenAI: $29B (2023) $500B (Oct 2025), targeting $1T IPO

Anthropic: $380B (Feb 2026)

NYT market cap: ~$12B

Intelligence cos >35% margin: Bloomberg, Thomson Reuters, CoStar, S&P Global, Gartner, Verisk, Moody's

Sources: Sacra; Visual Capitalist; Tomasz Tunguz; Variety; Press Gazette; Dataconomy

The distinction between content and intelligence isn't about quality. It's about what the buyer does next. Content informs awareness. Intelligence informs a decision. In 2025, seven companies that sell intelligence generated over $52 billion at margins above 35%. The entire U.S. newspaper industry generated $21 billion at margins near zero. Same raw material. Opposite economics. These companies are not media by self-identification. But they do the same upstream work as any newsroom: gathering information, verifying it, developing sources, recognizing patterns. The divergence is the transformation layer. Intelligence companies structure the output for decisions and embed it in a professional workflow. Newspapers package the same work as narrative for a general audience. The value didn't disappear. It migrated. From the content layer to the intelligence layer.

Here is why. When content becomes abundant, it loses pricing power. When everything is noise, the scarce resource is the ability to detect signal. Value always migrates to the scarce layer. Always. In every industry. In every century.

The gap shows up in what buyers actually pay. An intelligence terminal costs $30,000 to $66,000 a year. CoStar, Gartner, Moody's, Verisk all charge in that range. A New York Times subscription brings in $116. That is a 275:1 ratio at the low end. Not an anomaly. But not evidence that content is worthless, either. Professional decision-making has always commanded a different price than general awareness. What's new is that AI is collapsing the economics of one side while expanding the economics of the other.

The content layer is being squeezed from both sides: intelligence companies above it capturing the margin, infrastructure platforms beneath it capturing the economics. The $17 billion OpenAI plans to burn in 2026 alone exceeds the total annual revenue of virtually every U.S. media company except Disney.

But the counter-thesis is worth taking seriously: as AI-generated content floods the ecosystem, the scarcity value of verified, original, early information should increase. Someone has to produce the raw material that the models process. The question is whether content providers capture the economics of that scarcity, or whether the infrastructure layer extracts the value and leaves the producers fighting over scraps.

The historical pattern suggests the processing layer captures the economics.

Content is free.
Intelligence is not.
The entire media industry is being repriced around that distinction.

Q2

Who wins and loses among media institutions?

The wrong way to answer this is with a list of dead companies and surviving ones. Vice collapsed, Bloomberg thrived. Everyone knows that. It's yesterday's story and it tells us nothing about tomorrow.

The more useful answer is that the category of “media institution” is splitting into three distinct species, and most organizations don't realize which one they are. The strategy for each is completely different, and the worst possible outcome is executing the wrong one.

The first species is the Intelligence Business. They don't sell content. They sell reduced uncertainty. What they share: proprietary expertise an AI can't replicate by reading the open web, structured data rather than narratives, and distribution embedded in the professional workflow rather than competing for attention on a social feed. Bloomberg, Punchbowl, The Information, Axios. Each charges $5,000 to $32,000 a year. Each earns margins the content layer can't touch.

The model works for new entrants. Semafor generated $40 million in revenue in its third year with its first profit. Its CEO Signal newsletter reaches 75% of Fortune 500 chief executives. A three-year-old company that understood which species it was from day one.

The transition is already taking shape in some legacy institutions, with many not noticing.

The Financial Times: over 75% of paying readers are now institutional. The newspaper earns 2% margins. The intelligence business earns 30%. The journalism didn't change. The buyer did.

Bloomberg: ~$15B revenue, 325K+ Terminal users, ~$32K/yr ARPU

Punchbowl: $5K+/yr

The Information: ~$30M+, <100 employees

Axios: sold for $525M

Vice: $5.7B $350M (94% loss)

BuzzFeed News: Pulitzer winner, closed

Sources: EBSCO; Talking Biz News; Axios; CBS News; Press Gazette; Dow Jones; Washington Post

The worst possible outcome is being an attention aggregator that thinks it's an intelligence business.

Dow Jones: more than half its profit now comes from products most readers have never heard of. Factiva. Risk & Compliance. Dow Jones Energy. The Wall Street Journal masthead is still the brand. The margin is exploding from somewhere else.

Schibsted in Scandinavia earns more from its marketplace and classifieds businesses than from its newspapers. Axel Springer bought Politico not for its journalism but for its intelligence product, estimated to bring around $250 million annually. None of them abandoned their journalism. They discovered a second business inside the first. The intelligence assets were already there. The products for them weren't.

The Washington Post, in the middle of a reset, has started building toward the intelligence layer. WP Intelligence targets senior executives who said in focus groups they have enough news. What they want is forward-looking analysis, policy positioning, and recommendations. The white space isn't reporting. It's interpretation for decision-makers.

The second species is the Attention Aggregator. The question isn't which ones already died. It's which currently profitable organizations are next. HuffPost lost half its organic traffic. Forbes lost 50%. CNN erected a paywall after years of being free. Condé Nast's CEO called AI Overviews a 'death blow.' Their journalism didn't get worse. The distribution model underneath it collapsed. In fact, AI Overviews accelerated the crumble by a different mechanism than search and social did. Search and social redirected the click. AI eliminates it. The Attention Aggregator's product was the click. When comprehension replaces curiosity, the click has no reason to exist.

The distinction is precise: Intelligence Businesses sell the ability to act on information before others can. Attention Aggregators sell awareness of things that happened. In a world where awareness is becoming free, only the ability to act on information before others can has pricing power.

The third species is the Public Good. Local accountability, investigative work, foreign correspondence. This journalism will never generate sufficient market returns, although it is the most critical for society. Pretending it's a viable business is what killed local news. Recognizing it as infrastructure, funded through grants, philanthropy, and public investment the way we fund libraries, public broadcasting, and courts, is the only path to sustaining it.

Most media institutions don't know which species they are. Most are wrong about the answer.

Q3

What happens when AI agents, not humans, become the primary consumers of news?

Bot traffic: 51% of web (first time in a decade)

News sites: 7x more AI bots

AI crawling: up 15x

Perplexity crawl-to-refer: 700:1

Forbes: -50% traffic

Publishers blocking AI bots: 79%

Sources: Thales/Imperva (2025); Cloudflare; Arc XP; Digiday; BuzzStream; Forbes; WIRED; CJR

Something happened in 2024 that should have been a much bigger story than it was. Automated bot traffic surpassed human traffic on the web for the first time in a decade. 51% of all web traffic is now non-human. News sites are seven times more likely to see AI bot traffic than average websites. AI crawling increased 15x in 2025.

The “reader” of most news content is no longer a person scrolling over coffee. It's software: AI agents researching on behalf of users, trading algorithms scanning for signals, compliance systems screening regulatory filings for risk. This inverts everything about how news is produced and distributed.

Headlines optimized for human curiosity become irrelevant. SEO disappears as a discipline. What matters is structure, machine-readability, timeliness, and API accessibility.

A Columbia University study estimated that Google derives roughly $21 billion a year in revenue from news-related searches. What it pays publishers through News Showcase: roughly $1 billion total, spread across thousands of partners in 31 countries. Less than five cents on the dollar. Even adjusting for the difference between direct attribution and broader revenue modeling, the gap is orders of magnitude. That's the exchange rate. Its AI Overviews reduced click-through rates to publishers by 47 to 58% where they appear. Chartbeat data shows Google search referrals to news sites dropped 33% globally in a single year, 38% in the U.S. Google generated $295 billion in ad revenue in 2025. It is getting richer on content it does not produce.

Perplexity crawls 700 pages for every one referral it sends back. Perplexity is valued at $20 billion. Its marketing tagline was “skip the links.” A positioning statement that is also, inadvertently, a confession. It is now being sued by nearly a dozen publishers and platforms across three continents. It has paid publishers nothing.

The extraction of value from publishers is not subtle. Forbes lost 50% of its traffic. 79% of major publishers now block AI training bots, which is a bit like locking the barn door after the horse has been algorithmically replicated. The news sector lost an estimated 600 million monthly organic visits between mid-2024 and mid-2025. The LA Times fell 51%. Even the New York Times declined 9.4%.

There's a new world order: machines consume the information and send almost nothing back to the producers. And some of them aren't even asking permission. The news organizations that recognize this shift and rebuild their output and their businesses for algorithmic consumption will capture the next wave of demand. The ones still optimizing for clicks from humans are optimizing for an audience that's already left.

Q4

Is “information advantage” the new alpha?

Alt data market: $11.65B (2024) $135.72B by 2030 (63.4% CAGR)

Hedge funds + alt data: +3% returns

AI strategies: +12%

Goldman latency: 120ms 14ms

1ms trade value: ≈ $100M/yr (HFT)

Sources: Grand View Research; J.P. Morgan; SEC; Forge; BigDATAwire; Global Trading

The observation becomes a rule. The rule becomes a screen. The screen runs continuously.

For two hundred years, the business model of journalism was temporal information advantage. The scoop. The exclusive. The breaking news. Knowing something before the public knew it, and extracting value from that window before it closed. Wire services compressed the window from days to hours. Cable news compressed it to minutes. The internet compressed it to seconds. AI compressed it to the point where the gap between detection and publication can be measured, priced, and traded. The window didn't disappear. It moved.

Finance learned to price temporal advantage decades ago. The publishing industry never fully separated the intelligence from the article. The detection and the story were sold as one product, at the article's price.

The alternative data market hit $11.65 billion in 2024, growing 63% a year, because finance learned to price what publishers still give away. That value is the market for structured temporal advantage. The news industry has no equivalent market. The intelligence is produced every day, in every newsroom, and discarded every night.

Speed is real. But it's the dimension AI is compressing toward zero for everyone. The durable advantage lives in what AI can't compress: the judgment to know what information means and the relationships to obtain it before anyone else. Call these precision and access.

Access: a reporter who can call a former FDA commissioner and get a candid read on whether an advisory committee vote signals approval or delay. A journalist embedded in a regulatory community who gets a tip before a new filing gets into the docket update. These advantages compound with time and cannot be replicated by compute.

Precision means knowing what information means, not just finding it fast. A

pharmaceutical company quietly amends the definition of success for its most important drug trial. The filing is public. Does the change mean the drug works better in a subset, or that the original bar was too high? Same document. Opposite conclusions. The difference is the taxonomy to read it, the network of former regulators to confirm it, and the pattern-matching to know what followed last time.

A key job of journalists has always been this: find the signal in the public record before it becomes a body count or a market crash. That pattern recognition is the most valuable intelligence asset in the information economy, and it is almost entirely unmonetized. It stays in the journalist's head. When they leave, it leaves with them. No one captures it. No one structures it. No one builds on it.

Why? Because journalism's incentive structure rewards publication, not detection. The upstream work is invisible.

But once a journalist finds the pattern, it can be encoded. Hedge funds figured this out years ago and built entire trading desks around the learned patterns of their analysts. The observation becomes a rule. The rule becomes a screen. The screen runs continuously. That work used to take weeks. Months. Sometimes a career. Every one of those steps can now be retraced and shown to a machine. Not to write the story. To find the next one. And the next one. Before anyone knows to look. Call it pattern encoding. The industry needs a term for this window between when a signal becomes detectable in public data and when it becomes a published story. I call it the Pre-News Window. It has always existed. The difference is that it can now be measured, structured, and sold.

Here is the shift AI makes possible: journalistic intelligence, which was always qualitative, relationship-dependent, locked inside individual heads, can now be made quantitative. Structured. Tradeable. The same transformation that turned financial intuition into algorithmic trading is coming for editorial judgment. With one difference: the financial transformation had structured data to work with. The editorial transformation requires someone to produce the structured data first. That is the moat. The signal that preceded the article is becoming the more valuable product. This is the knowledge LLMs cannot scrape. No model can train on a story that hasn't been written yet.

Q5

What does the intelligence layer look like outside finance?

Vertical intel platforms: ~$500M ARR, 88% of S&P 100, $10K–$66K/seat

Flatiron Health: Roche paid $1.9B

Recorded Future: Mastercard paid $2.65B

Dow Jones Energy: Revenue +200% via 3 acquisitions

Industry Dive: Sold for $525M

STAT News / Endpoints: FT acquired Endpoints, 163K subs

Healthcare BI: $10B to $35.7B by 2034

Sources: Sacra; CNBC Disruptor 50 (2025); StockAnalysis; CB Insights; FT

The temptation is to look at the products. The right place to look is the workflows.

Finance got there first. Bloomberg, Refinitiv, FactSet. The infrastructure is so mature that the question in financial markets is not whether an intelligence layer exists. It is whether the organization can afford the one that does. The terminal's power has almost nothing to do with the data it contains. It has to do with what happens when an entire professional community builds its habits around a single tool. Compliance systems, risk models, operational pipelines, all wired into it. The switching cost eventually exceeds the subscription cost. At that point it is not a product anymore. It is infrastructure. The question is where that happens next.

The next layer is being built now, and it is vertical. Healthcare and life sciences. Energy and sustainability. Technology and infrastructure. These are the domains where the data is richest, the decisions are highest-stakes, and the intelligence infrastructure is least complete. Each is a corpus of public and semi-public data that is technically accessible but practically unreadable at scale, and that unreadability is the opportunity. The proprietary asset is what gets built on top of it: structured, monitored, interpreted, and delivered at the speed and precision that professional decisions require.

Something made this possible now that did not exist ten years ago. The underlying data has been there for decades. Clinical trials have generated public filings since the 1990s. Energy markets have had FERC data for longer than that. What shifted is the cost of acting on it. AI makes it possible to monitor and structure public data at a scale that was previously economically out of reach for anyone outside the top tier of the market. That is the role AI plays in this story. Infrastructure enabler. Not intelligence layer. Most of the companies moving into this space have not made that distinction, and it will cost them.

A wrong read on a Phase III readout costs a fund manager a position. A missed signal in an energy market costs a trader a position. A delayed threat briefing costs a CISO a breach. The infrastructure is not there yet. That is the gap, and the gap is the market.

The incumbents are starting to figure this out. Mastercard paid $2.65 billion for

Recorded Future. Dow Jones built an energy data vertical through three acquisitions, growing revenue over 200%. These moves read as diversification. They are something more urgent: land grabs on the interpretation layer before it consolidates.

The incumbents are moving in from one direction. Specialized B2B media companies are moving in from another. Where the Moodys and Mastercards of the world start with data infrastructure and build toward professional workflows, the B2B media companies start with professional audiences and build toward proprietary data. The starting points are opposite. The destination is the same.

STAT News built a life sciences intelligence operation that commands premium subscription prices from the same professionals Bloomberg serves in finance. FreightWaves built a proprietary data layer underneath an editorial product, turning freight market signals into a terminal for supply chain professionals. The editorial voice brought the audience. The proprietary data created the moat. Industry Dive sold for $525 million, not because it produced good journalism but because it achieved deep workflow integration across professional communities that had no other dedicated intelligence source. These companies are not in the news business. They are in the professional intelligence business, and the market is pricing them accordingly. Vertical intelligence platforms serving market research professionals have grown from $100 million to $500 million in ARR in three years, serving 88% of the S&P 100 at $10,000 to $66,000 per seat. That trajectory is the market validating the model.

Then there is what Silicon Valley thinks is happening.

The theory is simple: compute plus data plus a sleek interface equals intelligence. Perplexity's CEO twice positioned his product as the Bloomberg Terminal for the masses at $20 a month. The pitch is that general intelligence models can read any filing, cross-reference any dataset, surface any signal. That domain-specific terminals are just expensive interfaces over data that a foundation model can now access directly.

This gets the business exactly wrong, and the incompatibility is structural. A foundation model company cannot afford to be the best tool for clinical trial analysts, energy traders, or cyber underwriters specifically. Their economics require the same model to work for everyone. Horizontal scale and vertical depth are not a tradeoff. They are incompatible business models. When finance professionals rejected Perplexity, the critique was specific: data errors, no proprietary datasets, no compliance network. A product that cannot be trusted inside a regulated workflow is not a cheaper terminal. It is not a terminal at all.

So where does the moat actually get built? At four layers, each depending on the one below it.

The foundation is data generation. If the data exists somewhere else, the moat is a speed advantage at best. If the company creates the data, through proprietary monitoring, original methodology, or network-generated signals, it is the market. Flatiron Health built an oncology data platform so embedded in clinical workflows that Roche paid $1.9 billion for it. The asset was not the interface. It was the data that existed nowhere else.

The second layer is interpretation. A foundation model can read a ClinicalTrials.gov amendment. It cannot know that an amendment filed sixty days before a primary endpoint is a categorically different signal than one filed six months out, unless someone with domain knowledge encoded that distinction upstream. The data exists everywhere. The meaning exists only where domain expertise has been structurally embedded. That is not a retrieval problem. It is a knowledge problem.

The third layer is workflow integration. That dependency takes years to build and is nearly impossible to dislodge. The next cohort of analysts forming their instincts on AI-native tools is a real threat to incumbents, not because those tools are better, but because dependency forms early and compounds.

The fourth layer is compliance and validation. In every regulated domain, the intelligence layer carries liability. Professionals do not adopt tools that have not been validated against the standards of their field. Confident noise at speed is not a terminal. It is a liability.

The companies that build the next intelligence terminals will not come from media and they will not come from AI. They will come from domain experts who understand both the information and the professional workflow it serves, building at all four layers simultaneously, not just the one that looks like a product.

The real danger is not that the terminals fail to get built. It is that the incumbents who already generate proprietary data leave the interpretation to the user, handing the most valuable layer of the workflow to whoever builds above them. Data without interpretation is a feed. Interpretation without proprietary data is a feature. The terminal is what happens when the same company controls both, integrates them into a professional workflow, and earns the compliance validation to make switching unthinkable.

The distance between a demo and a terminal is the distance between content and intelligence. Building a demo takes a day. Building a terminal takes a decade.

Q6

If AI can hallucinate citations, what happens to the concept of the record?

NeurIPS 2025: 100+ hallucinated citations in 53 papers

Stanford: 75%+ hallucinated court rulings

MIT: 34% more confident when wrong

Losses: $67.4B (2024)

CAI: 6,000 members

NSA: recommends C2PA

Sources: Fortune; GPTZero; Stanford; MIT; CAI; NSA/DoD (2025)

The factual record can now be fabricated with perfect confidence.

At NeurIPS 2025, the world's top AI conference, researchers found over 100 hallucinated citations across 53 accepted papers. Completely fabricated references, invented journal articles with plausible titles and plausible authors, that survived peer review at the most prestigious venue in the field. The factual record that courts, researchers, and future generations build on can now be fabricated with perfect confidence.

Stanford found LLMs hallucinate court rulings at least 75% of the time. A lawyer in the Mata v. Avianca case submitted ChatGPT-fabricated citations to a federal court, the kind of incident that turns a technical limitation into institutional damage.

In regulated domains, 'mostly accurate' equals 'unusable.' A hallucination rate acceptable for consumer search is disqualifying for compliance, trading, and regulatory workflows. Trust is the moat AI can't shortcut.

This isn't just a media problem. It's an infrastructure problem for knowledge itself. If the factual record becomes unreliable, if no one can distinguish between a real citation and a hallucinated one, the entire system of knowledge that depends on verifiable sources starts to erode. Not dramatically, not all at once, but in the way that rust works: slowly, then everywhere.

The response is emerging. Content provenance standards are being deployed across roughly 20% of the web, with adoption recommended by the NSA and over 6,000 members of the Content Authenticity Initiative.

The provenance infrastructure is being built. But it's racing against hallucination rates that are worsening, not improving. The gap between the attack surface and the defense infrastructure is widening. The first institution to be materially damaged by a hallucinated citation will set the legal precedent for the entire verification economy. That case hasn't happened yet. But the conditions for it are in place. When it happens, the institutions that can prove provenance will be the ones the market turns to. Detection finds the signal. Verification confirms it's real. Together they form the intelligence layer. Neither alone is sufficient.

Q7

Are prediction markets the next major consumer of structured information?

Kalshi: $23.8B volume, $260M revenue, CFTC-regulated

Polymarket: $9B valuation, ICE $2B investment

Fabian incident: $22M pool, death threats to journalist

CNN: Kalshi-powered ticker on live broadcasts

U.S. Army: MIPB publication on prediction market signals

Sources: Axios; CNBC; Fortune; American Banker; Army MIPB (2025); Times of Israel (2026); Polymarket

Every prediction market contract is a financial instrument that prices the gap between when something happens and when the information ecosystem confirms it happened. That gap is the Pre-News window. Prediction markets are the first infrastructure that prices it explicitly, in public, with real money attached.

Combined volume hit nearly $40 billion in 2025. Kalshi generated $23.8 billion in volume and $260 million in revenue as a regulated CFTC exchange. Polymarket reached a $9 billion valuation after ICE invested $2 billion and began normalizing prediction probabilities into institutional data feeds. A Kalshi-powered ticker now runs alongside CNN's live programming. The U.S. Army published on prediction markets as intelligence signals. The diversity of adoption matters: regulated exchanges, crypto-native platforms, broadcast media, military intelligence. Each needs structured information differently, creating multiple demand channels that didn't exist five years ago.

But prediction markets don't just consume information. At sufficient scale, they create incentives that act on the information supply chain itself. And those incentives can be destructive.

In March 2026, bettors on Polymarket sent death threats to Times of Israel journalist Emanuel Fabian after his reporting on an Iranian missile strike impacted a $22 million betting pool. They pressured him to alter his account of what happened. One offered to share profits if the story changed. Polymarket condemned the harassment and banned the accounts. But the incident was not an aberration. It was a structural inevitability.

Here is why. A prediction market contract settles on a verifiable fact: did something happen, or didn't it. When the market is small, that settlement is administrative. When $22 million rides on the answer, the settlement becomes a high-value target. And in the current infrastructure, that target is often a single journalist's reporting. The market assigned a $22 million price to one reporter's account of reality. The reporter, covering a missile strike in a conflict zone, earns a fraction of that in a year. The ratio between the value the market places on the information and the value the market pays the person who produces it is the same gap that runs through this

entire report. It just became violent.

That is the case for structured, institutional resolution infrastructure. Not as an incremental improvement. As a prerequisite for the market's survival at scale.

The resolution problem is also a domain expertise problem. Consider a prediction market contract on whether the FDA will approve a specific drug by a specific date. Resolution requires more than confirming a yes or no. It requires knowing what counts as approval. A Complete Response Letter is not the same as a Refuse to File. An accelerated approval with post-marketing requirements is not the same as a full approval. The contract terms must be precise enough that resolution is unambiguous, and the resolution authority must understand the domain deeply enough to adjudicate edge cases. That is not a technology problem. It is an intelligence problem. The same intelligence problem that clinical trial analysts, regulatory specialists, and domain journalists solve every day, mostly uncompensated at the level the market now demands.

AI systems need training data: large, historical, comprehensive. Prediction markets need something different: resolution data that is specific, timely, authoritative, and precise enough to settle a financial contract. They don't need narrative journalism. They need machine-readable facts with clear provenance, delivered continuously, at the speed the market operates. The providers who build the authoritative resolution layer for this ecosystem will have built a toll booth on a highway that is getting busier every quarter.

But here is the uncomfortable part. Prediction markets, by creating a visible, liquid price for the Pre-News window, also create the most direct financial incentive to corrupt it. Every other information market prices attention or access. Prediction markets price the fact itself. When a price is put on whether something is true, there is an incentive to manipulate the answer. The Fabian case is the crude version: direct intimidation of a source. The sophisticated version is already imaginable: manufactured signals in public data designed to move prediction markets before the real signal arrives. Spoofing, but for facts instead of orders.

The prediction market ecosystem will not scale without an institutional resolution layer that is authoritative, depersonalized, and resistant to manipulation. That layer looks like structured intelligence feeds with auditable provenance, not individual reporters with bylines. The information companies that build it will sit at the intersection of domain expertise and financial infrastructure, which is exactly where the intelligence layer described in this report lives.

No one sends death threats to a data feed. That is not a punchline. It is a design

requirement.

Q8

Is trust mispriced?

NYT digital subs: $1.43B (+14.3%)

Willingness to pay: 17% globally

Would pay nothing: 55%

Paywall bypass rate: 99%

Sources: NYT 10-K; Reuters Institute (2025); Pew Research; Nieman Lab

Trust is mispriced because the instruments to capture it don't exist yet.

Globally, only 17% of people are willing to pay for news. 55% of non-payers say they wouldn't pay anything at all. And 99% of Americans who encounter a paywall either leave or find a free alternative. Ninety-nine percent. The paywall isn't a pricing instrument for trust. It's a friction point that drives virtually everyone to unverified alternatives.

The New York Times grew digital subscription revenue 14.3% to $1.43 billion in 2025. That's the exception that proves the rule: one institution captures billions while the rest of the market can't monetize trust at all. The instruments to price trust don't exist yet. The reason is structural: verification is produced as a byproduct of journalism, not as a standalone product. No institution has separated the verification function from the publication function because the organizations that verify are organized around publishing, not around selling trust.

There's something almost paradoxical happening. The outlets that can institutionally verify that a human observed something, or that data came from an authoritative primary source, hold an asset that becomes more valuable every single day. But trust doesn't trade on a commodity exchange. There's no futures market for provenance. The monetization mechanism is still being figured out. The market is mispricing trust because the instruments for capturing that value haven't been built yet. They're being built now, not in newsrooms but in the data supply chain: provenance standards that certify origin, verification APIs that confirm facts at the point of decision, and structured data feeds that price information by its relevance to the buyer's workflow.

Current pricing is like charging the same rate for electricity whether it powers a phone or runs a hospital. The instrument doesn't match the asset. The same FDA filing is worth nothing to a general reader, $50 to a portfolio manager, and $50,000 to the pharmaceutical company deciding whether to continue a program. That's contextual relevance. The value is determined not by the content but by the decision it informs.

When someone builds a mechanism that prices verified information at the point of decision, not the point of access, the entire economics of trust changes. Imagine a

portfolio manager about to execute a trade who can pay $50 to verify in real time whether a clinical trial amendment signals efficacy or failure. That's decision-point pricing. The value of the verification is proportional to the decision it informs, not the subscription it's bundled into. Current pricing mechanisms, subscriptions, paywalls, licensing fees, ignore this entirely. They price the content. The intelligence economy prices the context. As AI makes synthesis cheap, value shifts upstream to auditable, point-in-time data that a compliance officer can stand behind. Provenance is the new premium.

Q9

If this becomes a data game, what's the trade?

Total AI licensing: ~$2.92B across 35+ deals

News Corp: $250M+ (OpenAI), ~$150M (Meta), $400M+ total

Reddit-Google: $60M/yr

Reddit-OpenAI: ~$70M/yr

Really Simple Licensing: Reddit, Yahoo, Medium; machine-readable protocol

Sources: Variety; Datacconomy; Press Gazette; Digiday; Media and the Machine; Deadline

AI companies have committed roughly $2.92 billion in content licensing across 35 or more deals. Against $200 billion in AI infrastructure spending in 2024 alone, that's 1.5%. The content producers are being paid a rounding error relative to the value being extracted. News Corp has been the most aggressive, assembling over $400 million in disclosed commitments: $250 million or more from OpenAI in May 2024, up to $150 million from Meta in March 2026 at up to $50 million per year, plus undisclosed deals with Bloomberg and active negotiations with Google. Their CEO Robert Thomson now describes News Corp as “essentially an input company,” comparing journalism to semiconductors, data centers, and energy as foundational AI infrastructure. He calls it a “woo and sue” strategy: license content to willing partners, litigate against those who scrape without permission. At near-pure margins, News Corp's AI licensing revenue, estimated at $100 million or more annually once all deals are active, could represent 15% or more of net income.

Thomson's framing is revealing: News Corp sees itself not as a media company but as infrastructure for AI, the way a semiconductor fab sees itself as infrastructure for computing. The strategic question is whether infrastructure providers capture the economics or the application layer does. History has an answer: semiconductor fabs have lower margins than Apple, cloud providers have lower margins than SaaS companies, commodity producers have lower margins than processors. If Thomson's strategy follows this pattern, positioning News Corp as AI infrastructure is actually a losing strategy, consigning publishers to the low-margin layer permanently. The counter-argument: content may be different because it's non-substitutable, protected by copyright, and operationally essential. Whether that counter-argument holds is the most important open question in the information economy.

The deeper trade is this: companies with large, structured, proprietary datasets, data that models need and can't scrape from the open web, are sitting on an asset class that didn't exist five years ago. The companies that recognized this early and invested in structuring their information for algorithmic consumption will capture disproportionate value. The ones still thinking of their content as articles rather than as data will discover they've been giving away the most valuable part of the supply chain. I've seen this from the inside of enough licensing conversations to know: the companies that structured their data early are setting the terms. Everyone else is negotiating from a position that gets weaker with every deal signed.

Q10

What happens when AI runs out of journalism to train on?

Quality human text exhaustion: projected by 2028 (Epoch AI)

Model collapse: documented (Nature, 2024)

AI-generated web content: 30-40%

News deserts: 213 counties, 50M Americans

Newsroom jobs: -75% since 2005

Sources: Epoch AI; Medill/Northwestern; arXiv; Nature

A photocopy of a photocopy of a photocopy.

AI is getting smarter about a world that no longer exists.

Scientists still publish. Courts still rule. Agencies still file reports. Some of that reaches training datasets directly. But raw filings are dense, fragmented, and hard to learn from at scale. The version AI actually learns from, verified, contextualized, structured for comprehension, is what journalists produce. Newsrooms don't just inform communities. They feed the machines. They're the translation layer that turns raw public records into the quality data AI trains on.

That translation layer is breaking.

AI extracts the value of journalistic work, replaces the traffic that funded it, and trains on the output. The result is models increasingly trained on AI-generated content, progressively losing contact with the primary source material. A photocopy of a photocopy of a photocopy. Researchers call it model collapse: when AI trains on its own output, the distribution narrows, the tails disappear, and the model loses the ability to represent the world accurately. It doesn't get dumber all at once. It gets blander. More generic. Less grounded.

Some AI companies have started paying publishers for training data. But those deals reach a handful of large outlets. Local and regional journalism, the layer that monitors city councils, court filings, and permit applications, continues to collapse. 213 counties have no local news source at all. 50 million Americans have limited-to-no access. The pipeline that refreshes the knowledge isn't thinning

evenly. It's breaking at the base.

The industry needs a term for this. Call it the Knowledge Freeze: the systematic degradation of the human-produced knowledge layer that AI depends on. Model collapse is the technical symptom. The Knowledge Freeze is the structural cause.

Quality human-generated text is projected to be exhausted as early as 2028. Not because humans stop writing. Because the marginal training value of additional web-scraped text approaches zero. The models have already ingested the recoverable corpus. When that happens, the only source of genuinely new, high-quality training data is new human-produced journalism and analysis. The newsrooms closing today are the ones that would have produced the training data AI needs tomorrow.

And it's a feedback loop. AI replaces clicks. Revenue drops. Newsrooms shrink. The quality layer thins. AI content fills the void. Training data degrades. Models lose reliability. Repeat. Each turn of the cycle is imperceptible. The cumulative effect is structural.

The pattern encoding that journalists do, the conversion of raw observation into structured detection logic, is exactly what gets lost first. When a climate reporter who tracked EPA enforcement actions for a decade takes a buyout, that institutional knowledge doesn't transfer to a junior hire or a language model. It vanishes. No one inventories it. No one encodes it. And no model can learn it from text that was never written.

An information economy losing contact with verified reality. Models hitting a ceiling not because compute is scarce, but because the ground truth beneath them is eroding. Who rebuilds the translation layer? And who keeps training on photocopies?

Q11

Can wire services capture the value they produce?

AP: ~$842M revenue, 235 bureaus, 94 countries

AP licensing share: 30% (2007) 10% (2024)

Thomson Reuters: $200M+/yr AI investment, 28% GenAI contract value

OpenAI pays Reddit: ~$70M/yr; AP likely single-digit millions

Sources: Press Gazette; Editor & Publisher; Thomson Reuters; Globe and Mail

The ones who build the refinery win. The ones who keep shipping barrels of raw text will watch their margins compress even as demand for verified facts increases.

OpenAI pays Reddit roughly $70 million a year for anonymous comments. Arguments about whether a hot dog is a sandwich. Relationship advice from strangers. It pays the Associated Press, 235 bureaus in 94 countries producing the most reused factual reporting on earth, considerably less. The pricing gap is instructive. Reddit data is conversational, structured in Q&A format, and maps directly to how chatbots learn dialogue patterns. AP content is narrative prose written for human readers, dense and harder to parse algorithmically. Reddit isn't more important than AP. It's more usable in current form. That observation proves the thesis: the organizations that structure their information for algorithmic consumption will capture disproportionate value. The ones that don't will watch others extract it.

A $200,000 investigative story. Within three minutes of publication, an AI extracts its value for about two cents. That is a 10-million-to-one ratio. The people who argue AI will replace journalism are missing the point. It doesn't have to. It just needs to wait three minutes. This ratio can only be marginally improved because the cost isn't in the writing. It's in the knowing.

The problem is structural. Most wire service content is unstructured text, narratives written for human readers, not machine-readable data. It's like owning enormous reserves of crude oil but having no refinery. The raw material is there, but without serious investment in structuring, tagging, and packaging information for algorithmic consumption, wire services will watch others extract the value from what they produce. AP's core licensing revenue share fell from 30% in 2007 to 10% in 2024. The decline is structural: its newspaper clients shrank, digital-native sources reduced dependency on wire copy, and the content itself remained unstructured text in a world demanding machine-readable data.

Thomson Reuters gets it. They're investing $200 million or more per year in AI, with 28% of annualized contract value now coming from GenAI-enabled products. They built the refinery, at a cost: reduced editorial headcount, narrowed coverage scope, and a product that serves institutions rather than the public. The transition requires trade-offs that not every organization can or should make. AP has started: its Intelligence Service now sells contextual, real-time event understanding to enterprises. But the cooperative ownership structure makes it harder to raise the capital for the kind of full transformation Thomson Reuters has executed. A co-op

can't issue equity. Its members, mostly newspapers in financial distress, can't increase their contributions. Governance decisions require consensus among hundreds of owners with different priorities. The structure that made AP resilient for 175 years is now the constraint. The ones who build the refinery win. The ones who keep shipping barrels of raw text will watch their margins compress even as demand for verified facts increases.

Q12

What can't be automated?

Investigative stories: $200K–$400K each (3,000x an AI post)

FOIA: 1.5M requests (2024), 12% fully granted

FOIA backlog: 267,000 cases, $723M cost

Sources: Brechner Center (2025); Columbia Journalism Review; AP

In any system, the bottleneck sets the price. When every other part of the news supply chain, writing, formatting, distribution, translation, summarization, can be automated faster and cheaper by AI, the parts that can't be automated become the bottleneck.

Original observation. Someone physically present at a city council meeting. A court filing that requires human review and contextual understanding. A source relationship built over years of trust. FOIA requests that hit a record 1.5 million in 2024 but were fully granted only 12% of the time, an all-time low.

Investigative journalism costs $200,000 to $400,000 per major story at top outlets. That's 1,500 to 3,000 times more expensive than an AI-generated post. And no amount of compute can reduce it, because the cost isn't in the writing. It's in the months of source development, the legal battles for access, the human judgment about what a document means.

AP expanded automated earnings reports tenfold. The automatable layer is scaling rapidly. The non-automatable layer is not. The gap between what machines can produce and what humans must produce is widening, not narrowing. Newsgathering itself, at least the part that relies on structured public data, is increasingly automatable. Court filings, permit applications, regulatory submissions, campaign finance disclosures. The cost of monitoring these will soon be marginal. The domain knowledge required to interpret what they mean will not. Observation is getting cheaper. Interpretation is not. Interpretation requires pattern recognition across contexts: knowing that a seemingly routine clinical trial amendment signals a change in efficacy expectations because the analyst has seen three similar amendments in the past decade and knows what followed. It requires knowledge of institutional norms and exceptions: understanding when a routine filing masks an extraordinary

disclosure. It requires relationships with people who can explain what a filing means versus what it says. Each of these is a human capability that compounds with experience and cannot be replicated by training on historical text.

Primary intelligence gathering scales with human networks and institutional access. It doesn't scale with compute. That century-old technology, shoe-leather reporting, remains the binding constraint. One caveat: this physical advantage has a shorter half-life than the industry assumes. Drones already monitor construction sites and agricultural output. Computer vision reads documents. The growth of robotics and physical AI is beginning to close the gap between what machines can observe and what previously required a human body in a specific location. The binding constraint is real today. Every year it gets a little less permanent. The window where human access is the moat is measurable, and it's narrowing.

The people doing automatable work face displacement. The people who are the binding constraint become more valuable than ever. Both are happening simultaneously. The question is what the newsroom looks like when the ratio inverts.

Q13

If AI writes all the articles, what is the newsroom actually for?

Newspaper jobs: below 91,550 (−75% since 2005)

Cuts: 7,300 in 2024–25

PwC: productivity growth 4x in AI-exposed industries

AI wage premium: 56%

Sources: Press Gazette; PwC (2025); Reuters; AP

When a reporter who covered the FDA for fifteen years takes a buyout, what leaves isn't a writer. It's a detection system that took fifteen years to calibrate.

The article is the exhaust product. The real output of a newsroom is upstream: source networks calibrated over years, editorial judgment about which detail in a 200-page filing matters next quarter, institutional knowledge that distinguishes signal from noise. That upstream work is what AI cannot replicate, and it is what disappears when headcount cuts are made indiscriminately. U.S. newspaper employment has fallen below 91,550, a 75% drop since 2005. The industry is hemorrhaging people. But the loss that matters most is invisible.

When a reporter who covered the FDA for fifteen years takes a buyout, what leaves isn't a writer. It's a detection system that took fifteen years to calibrate. No one inventories it on the way out. AI can write the article. It can't build the source. It can't develop the judgment. It can't walk into a room and know what question to ask. Articles are the exhaust.

What's happening isn't that newsrooms are becoming factories. It's that the factory part of the newsroom, the writing, formatting, translating, summarizing, is being

automated away, exposing what was always underneath: a very small number of people doing the work that actually matters.

The newsrooms that understand this are restructuring around leverage. One reporter with AI tools covering the territory that used to require ten. The workflow: AI monitors every federal agency filing, every state legislative update, every court docket in a jurisdiction. It flags anomalies, cross-references against historical patterns, drafts preliminary analysis. The reporter reads the flags, calls the source who can confirm whether the pattern means what it appears to mean, writes the interpretation no machine can produce. That is a leveraged intelligence workflow. It requires one person doing the work of ten, not because the other nine were redundant, but because AI absorbed the monitoring and the reporter absorbed the judgment.

Reuters can push a first alert within 6 seconds. Nearly 70% of AP staffers use generative AI daily. AI-skilled workers command a 56% wage premium.

The newsrooms that don't understand this are cutting headcount across the board, losing the source networks and institutional knowledge that are the actual product. They're automating the part that didn't need automating and eliminating the part that can't be replaced. The question isn't whether the newsroom survives. It's whether the people in it can make the transition: the reporter who redefines herself as an intelligence analyst, the editor who becomes a verification architect, the beat journalist who builds a detection pipeline instead of a clip file. These aren't metaphors. They're job descriptions.

Q14

What happens when $1 of AI replaces $33 of freelance labor?

$1 AI: = $33 freelance labor

New articles AI-generated: 52%

Ranking in Google: human: 86%

AI freelancers: +44% hourly rate

Sources: Ramp (2026); Graphite (2025); Ahrefs; Google; Upwork

One number tells the whole story: among companies most exposed to AI, $1 of AI spending replaced $33 of freelance labor. That's not incremental efficiency. That's the elimination of an entire category of work.

For a decade, there was a vast industry built on producing content that no human particularly wanted to read, optimized for an algorithm rather than an audience. Listicles nobody clicked past the headline. SEO blog posts that existed purely to rank. Product descriptions generated from a template. That industry's product is

now literally free to produce.

The freelance collapse is AI-driven. But it's compounding a financial extraction dynamic that has been running for two decades. What happened to freelance content producers is a preview of what's coming for commodity newsroom functions. Over 52% of new English-language articles are now AI-generated.

But here's what's counterintuitive, and it's the part that most analysis of the freelance apocalypse misses: the collapse of the floor raised the ceiling. Freelancers working on AI-related projects earn 44% more per hour. Spend-per-buyer on Fiverr climbed 13.3% even as active buyers fell 10%. Fewer clients are buying commodity writing, but those who remain pay significantly more for work that requires actual expertise.

Earnings reports, game recaps, weather summaries, routine government meeting coverage. The newsroom equivalent of content produced because the algorithm rewarded volume. Daily print circulation collapsed from 50 million to 15 million. Roughly 3,500 newspapers have closed since 2005, nearly 40% of all U.S. papers. Alden Global Capital extracts 17% profit margins from the newsrooms it acquires. The strip-mining model works until there's nothing left to mine.

Routine coverage is collapsing toward zero. Investigation and analysis are worth more than ever. The middle is gone, in marketing and in newsrooms, for the same reason. The bottom fell out. But the top is higher than it's ever been. The mechanism is specific: domain expertise, verification capability, and brand trust command premium pricing because they're the three things AI can't replicate at the quality threshold professionals require. The bifurcation is redefining what a media company is.

For a mid-career content writer, the transition is specific and urgent: learn to do something AI can't replicate, whether that's source development, domain specialization, or verification expertise, or accept that the floor is falling toward zero. The window for that transition is measured in years, not decades.

Q15

What happens to the New York Times? What happens to Google?

NYT: 12.78M subs, $2.82B revenue (+9.2%), debt-free

Google AI Overviews: 13–19% of searches; top CTR -58%

Zero-click searches: 56% 69%

Business Insider: -55% traffic

Sources: NYT 10-K (2025); Ahrefs; Similarweb; AdExchanger; Dataslayer

These are the two most important case studies in media right now because they represent opposite bets, and both are working. For now.

The Times bet on bundling. It reached 12.78 million subscribers, $2.825 billion in revenue, and $551 million in free cash flow. 51% of subscribers now take multi-product bundles across News, Games, Cooking, Wirecutter, and The Athletic. Bundle subscribers generate $12.92 in average revenue versus $3.36 for single-product, with 40% lower churn. The Athletic became profitable inside the bundle. The model works. The question is what it's a model for.

Google faces a paradox of its own making. AI Overviews now appear in 13–19% of searches, and when they do, the top organic result loses 58% of its clicks. Zero-click searches rose from 56% to 69% in a single year, though some of that increase predates AI, driven by knowledge panels and featured snippets. The AI-specific acceleration is what matters: Google is cannibalizing its own ad inventory to improve the user experience. Every query answered without a click is a query that generates less advertising revenue. The company is simultaneously the infrastructure provider that media companies depend on and the entity most actively destroying their economics. It's also destroying its own, just more slowly.

The lesson applies to every media institution. Direct audience relationships are a moat. The bundle works. The financials are strong. And an estimated 4% of the Times' revenue comes from data products.

Four percent. The most successful news company in the world, the one that got subscriptions right, that got bundling right, that built a direct relationship with 12.78 million paying readers, derives 4% of its revenue from intelligence products. That's not a failure of execution. It's evidence of an unrealized adjacency. The subscriber base contains professionals who would pay intelligence prices for what the Times already knows. The product for them doesn't exist yet. The question is whether the Times builds it before someone else extracts it. The bundle proved they can retain attention. The question is whether they can sell intelligence.

For Google, the paradox resolves one of two ways: dominant AI platform that absorbs the intelligence layer the way it absorbed search advertising, or antitrust-forced

unbundling that creates openings for vertical intelligence companies. Either outcome restructures who controls the flow of information to machines and humans alike.

CONCLUSION

Who Will Monetize Truth?

The questions above map the terrain. Here is what the terrain implies.

Three capabilities define where an intelligence company wins. Signal originality: detecting patterns in public data before they become consensus. Auditable provenance: every claim traceable to a source, timestamped, certifiable. AI-native speed: delivering intelligence at the pace machines operate, not the pace humans publish.

These are not predictions about what should happen. They are predictions about what will happen, with dates attached, so they can be proved wrong.

1. The detection-to-publication gap becomes a financial instrument.

Five years ago, monitoring a single federal agency required a human analyst. Today, an AI system can monitor every federal agency, every state legislature, every clinical trial registry simultaneously. The cost is approaching a cloud computing bill. That collapse in monitoring cost makes the gap between when a signal appears in public data and when it appears in a news article measurable, priceable, and tradeable. By 2028, at least one regulated financial instrument, not a prediction market but a tradeable security, will be explicitly priced on the pre-publication detection gap. By 2029, the institutional client paying for pre-publication signal will generate more revenue than millions of readers paying $10 a month each. The arc from monitoring cost collapse to tradeable temporal advantage to pre-publication premium is a single economic chain, not three separate predictions.

2. Institutional buyers outbid consumer audiences.

By 2028, institutional buyers will outbid consumer audiences for structured intelligence in at least three verticals outside finance: clinical trials, energy regulation, and cybersecurity. It's already happening in financial data. Intelligence platform revenue dwarfs any newspaper's. Every vertical where journalists currently produce insight for a mass audience will see an institutional buyer willing to pay 100x the consumer price for the structured version of that same insight, delivered earlier. The 275:1 gap is not an anomaly. It's a preview.

3. The talent follows the economics.

This is a consequence of the prediction above. When institutional buyers outbid consumer audiences, the people whose actual skill is pattern recognition under uncertainty migrate to the institutions that price the skill correctly. This isn't brain drain. It's price discovery. The publications that recognize their real asset is detection logic and editorial judgment will restructure around that. The ones that keep optimizing for pageviews without building intelligence products from their editorial assets will lose the talent that makes both possible. Any publication that hasn't identified the intelligence business inside its content operation by 2029 is running out of time.

4. Journalism becomes the last mile.

The article is the exhaust product of a detection and analysis pipeline that is orders of magnitude larger than the newsroom that publishes it. Most of the value, most of the talent, and most of the revenue will sit in the pipeline, not in the publication. The publication will still exist. The newsroom API is the last mile. The pipeline is the business. It will just no longer be where the power is. By 2030, the largest information company by revenue in at least two verticals outside finance will have no newsroom.

The scenario is already plausible: a machine detecting a market-moving signal before any human reports it, reshaping the legal boundary between media, intelligence, and financial services. It doesn't depend on whether newsrooms restructure. It depends only

on the monitoring cost collapsing far enough that the detection capability exists at scale. Which it nearly does. The Knowledge Freeze isn't coming. It's here. And it accelerates every quarter a newsroom closes. As the freeze deepens, AI output quality degrades, which makes verified human-produced intelligence more scarce, which means whoever monetizes that scarcity wins.

The question is no longer who will produce the truth. AI can generate the appearance of truth at infinite scale.

The question is who will monetize it. That's a much smaller group. And the window to join it is closing.

These predictions are designed to be proved wrong. One failure condition deserves serious engagement: foundation model companies are moving toward domain specialization faster than domain companies are building AI infrastructure. If Anthropic, Google, or OpenAI achieves sufficient domain credibility to pass compliance review at a major financial institution, the moat around vertical terminals narrows dramatically. General-purpose platforms have commoditized vertical solutions in software, cloud, and search. The bet that it won't happen in intelligence is a bet that regulatory complexity is a harder barrier than technology. That bet may be right. It is still a bet.

The uncomfortable question is whether there are more institutions capable of making the transition than institutions that will be destroyed by not making it. The window is still open. But it is not open for long. And the first step is the hardest: admitting what the institution actually is.

Francesco Marconi built news automation at the Associated Press and led R&D at the Wall Street Journal before co-founding AppliedXL, an intelligence company that detects early signals in public data before they become news. His book, The Science of First, introduced the concept of Pre-News. He wrote this from Brooklyn. He has a stake in the thesis of this report.

APPENDIX

Company Data

FY2025 where available, FY2024 otherwise. Revenue in $M USD. Data % = estimated share of revenue from structured data, analytics, and intelligence products. Classification: Intelligence (>50% data revenue), Hybrid (15-50%), Traditional (<15%).

Company	Revenue ($M)	Employees	Data %	Category
Bloomberg	~15,000	25,700	~85%	Intelligence
S&P Global	15,336	42,350	100%	Intelligence
MSCI	3,135	6,268	100%	Intelligence
Moody's	7,718	16,000	100%	Intelligence
LSEG	~11,230	26,251	~64%	Intelligence
Wolters Kluwer	~6,738	21,100	94%	Intelligence
Thomson Reuters	7,476	26,400	82%	Intelligence
Gartner	6,497	21,044	82%	Intelligence
Dun & Bradstreet	2,382	6,247	100%	Intelligence
Verisk	3,073	7,800	100%	Intelligence
CoStar	3,247	6,593	~55%	Intelligence
Clarivate	2,455	~12,000	100%	Intelligence
FactSet	2,322	12,800	~99%	Intelligence
Morningstar	2,446	~11,375	~74%	Intelligence
Hearst	13,500	~22,000	~35%	Hybrid
RELX	~11,988	36,400	~58%	Hybrid
News Corp	8,452	~22,300	~33%	Hybrid
Dow Jones	2,331	~7,000	~39%	Hybrid
Financial Times	~726	~3,000	~56%	Hybrid
Politico	~250	~1,600	~52%	Hybrid
Axel Springer	~4,400	~17,000	~50%	Hybrid
AP	~600	3,036	~18%	Hybrid
Economist Group	~466	1,643	~41%	Hybrid
Schibsted	~774	3,884	~67%	Hybrid
New York Times	2,825	~5,900	~4%	Traditional
Washington Post	~170	~700	~25%	Traditional
Gannett	~2,300	~3,300	~2%	Traditional
CNN	~1,600	~3,300	~0%	Traditional
Nexstar	4,949	13,005	~0%	Traditional
Condé Nast	~2,000	~5,200	~3%	Traditional
iHeartMedia	3,860	10,100	~0%	Traditional
Vice (2022)	~258	~400	0%	Traditional
BuzzFeed	~190	~750	~0%	Traditional
BBC Studios	~2,860	~5,700	~0%	Traditional
Sinclair	3,548	~7,200	~0%	Traditional
Dotdash Meredith	1,777	~3,500	~17%	Hybrid
Forbes	~275	~800	~12%	Traditional
Guardian	~371	1,671	~7%	Traditional
Lee Enterprises	611	3,047	~2%	Traditional
The Athletic	~172	~700	~0%	Traditional
Vox Media	~515	~1,800	~0%	Traditional

Sources: SEC filings, published accounts, company IR, Press Gazette, Sacra, Axios, PitchBook, ZoomInfo. Revenue per employee computed from reported figures.