

The Market Size of Truth

How machines and markets
will buy what journalism never sold

Francesco Marconi
CEO & Co-Founder, AppliedXL

APPLIEDXL RESEARCH

THESIS • AUGUST 2026

The Market Size of Truth

*How Markets and Machines Will Price
Information*

Institutions spend approximately \$160 billion a year on products that establish facts they can act on. News organizations, whose daily work produces those facts, capture roughly \$2 to \$4 billion. This report sizes the market for truth and handicaps the contenders for its open seat.

BY FRANCESCO MARCONI

For the full interactive version of this report, visit
www.appliedxl.com/research/market-size-for-truth

CONTENTS

00	The thesis	3
I	Defining the market	4
II	The number	6
III	The gap	9
IV	The demand curve	15
V	Who takes the seat	22
VI	Conclusion	28
VII	Appendix	29

00 • THE THESIS

The market nobody claims

It is a Tuesday in March 2026, and air-raid sirens are sounding in Tel Aviv. An Iranian ballistic missile has struck Israeli territory. A war correspondent works his sources, confirms what can be confirmed, and files.

Tens of millions of dollars are riding on the details in his report: which building the missile hit, and how it hit. The prediction-market contract names no independent resolution source and, like many geopolitical markets, settles on published reporting. Which means it settles on him.

The first death threat arrives within minutes. Then more, in several languages, from anonymous accounts. One bettor offers him a share of the winnings to change the account. He refuses. The platform bans the accounts and classifies the incident as a trust-and-safety problem.

The contract has conscripted the reporter into a role he never accepted: his reporting will determine who gets paid. His byline carries the editorial risk of being wrong. It also carries the financial weight of moving money and the physical danger created by people who can profit from his words. Newsrooms have procedures for being wrong, and little protection against the money or the danger.

None of the money at stake reached the reporter or his newsroom. This report is about the market that did the pricing. Institutions spend approximately \$160 billion a year on products that establish facts they can act on. News organizations, whose daily work produces those facts, capture roughly \$2 to \$4 billion: about a quarter of the categories they already play in, and about 2 percent of the full market. The quarter is the business as it stands. The 2 percent measures the frontier, and the frontier is what this report sizes.

The gap would be a curiosity if demand were flat. Demand is exploding, because the buyers are changing. For two centuries the buyer of verified information was a person reading. The next buyers are machines and markets: AI systems that need grounded answers billions of times a day, and contracts that pay out when something happens in the world. Both consume determinations continuously, and both pay by the decision rather than by the article.

Reference markets consolidate around one or two accepted institutions, and the seat at the center of this one is open. Six kinds of institutions could take it: the data giants, the exchanges, the hybrid publishers, the news organizations, the AI labs working through the expert-data marketplaces that already pay professionals by the hour to train models, and entrants that do not exist yet. This report sizes the market

and handicaps the contenders, with dated predictions so they can be proved wrong. Journalism holds the judgment every contender needs and almost none of the infrastructure. The numbers behind both claims are inside.

The market for truth has never appeared on a market map. Every buyer in it believes they are buying something else.

■

I. DEFINING THE MARKET · Q1

What does an institution buy when it buys truth?

The unit of sale

Entity + event + timestamp + evidence + methodology + authority

The test

Reliable enough for money, risk or legal responsibility to move

The moat

Designation, point-in-time history, service, liability

The grammar

Claim · commitment · event · state · resolution · outcome

Sources: CDS determinations committees; ISDA; IOSCO benchmark principles; AP election methodology

A document enters the public record. A company files results, a court rules, a regulator decides, a government announces. The information is public, and what happened can remain unclear. A company can accurately report that a trial met several secondary endpoints while omitting that it missed the primary one. A government can announce a policy that has no legal effect yet.

The document is evidence. The buyer pays for the determination: a defined event, a timestamp, controlling evidence, a methodology, and an institution willing to stand behind the conclusion. A bank blocks a transfer on it, an insurer releases a payment, an exchange settles a contract, and an AI system updates its answer.

That is the product this report sizes: a determination reliable enough for the world to act on. It does not need to settle philosophical doubt. It needs to be defensible enough for money to move.

Converted into datasets, the product resolves into a small grammar of objects. A claim records that someone asserted something and what the assertion means. A commitment records a stated future action with a date attached. An event records that something happened, a state records whether a condition is in effect, a resolution records the binary outcome of a tracked question, and an outcome records the character of that resolution: a drug approval is a resolution, while the type and conditions of the approval are the outcome. The grammar turns a stream of stories into a ledger of trackable questions. A commitment opens a question, evidence accumulates as claims and events, a resolution closes it, and the outcome describes how it closed. Every product that depends on things resolving, from event contracts to catalyst trading, consumes this structure.

The determination is copyable the moment it is published, which raises the question that killed article economics: why would anyone pay? Because the buyer pays for standing, and standing does not copy. The resolver named in a contract collects its fee whether or not the determination circulates. A rating carries value because mandates and regulation designate it, and the fee never depended on secrecy. The rest of the moat is operational: point-in-time history, correction records, service guarantees and liability cannot be scraped from a headline. Point-in-time matters because buyers test data against past market behavior, and a dataset is only testable when every value carries the timestamp of what was known at that hour, uncontaminated by later knowledge. A locked schema matters because it is a fixed contract buyers can wire systems into. Both take years to accumulate, and neither travels with a copied determination.

I. DEFINING THE MARKET · Q2

Is \$160 billion a real market or a spreadsheet invention?

0 analysts

Cover "truth" as a category today

\$2.8B

Alternative data: a category that did not exist in 2010

\$7.2B

Indices: a licensing afterthought turned industry

Sources: Neudata; Burton-Taylor; Moody's corporate history

The fair objection comes first. Nobody sells across these four segments, no analyst covers them together, and a motivated author can staple adjacent industries into any total he likes. Addition always works.

The defense is that categories get invented before they get recognized. Nobody sold "alternative data" in 2010; it is a \$2.8 billion segment with its own conferences now. The index business was a licensing afterthought before it became a \$7.2 billion industry. Moody's created the ratings category by selling into it. A category is real when it shares a buyer and a function, and this one does: the same institutions buy ratings, screening, benchmarks and grounding to close the same kind of question, even if no vendor yet sells across all four. Substitution happens within each segment today. The bet in this report is that it begins happening across them.

What matters is the discipline behind the number. Revenues are deduplicated before summing: Bloomberg, LSEG and S&P Global Market Intelligence sit inside the market-data aggregate rather than on top of it, ratings are counted once, and LexisNexis Risk is counted once inside RELX. Each line takes the highest-credibility source available, filings first, specialist analysts who survey the actual buyers next, and commercial research only where nothing better exists. The weakest lines are declared before anyone else finds them: no high-credibility standalone figure exists for ESG data or for the private price reporting agencies beyond Platts and Argus.

The model removes major overlaps and states ranges. And if the category is still judged artificial, the revenues are real either way. Someone is being paid roughly \$160 billion a year to say what happened.

II. THE NUMBER · Q3

How large is the market for truth today?

\$150–170B

Estimated annual range

\$160B

Midpoint used throughout

4 segments

Financial, risk, scientific, machine

\$49.2B

Market data & analysis, the largest single line

\$3–5B

AI data licensing, the fastest-growing line

Sources: [Burton-Taylor \(2026\)](#); [Burton-Taylor index report \(2026\)](#); S&P and Moody's segment results; [Neudata \(2026\)](#); RELX; IQVIA; licensing disclosures

Institutions spend an estimated **\$150 billion to \$170 billion a year** on products that establish facts they can act on. The midpoint is approximately **\$160 billion**. This is not a recognized industry category. It combines four businesses normally measured separately.

FIGURE 01

EST. ANNUAL REVENUE, \$B

WHAT INSTITUTIONS PAY FOR VERIFIED INFORMATION

Financial & market truth



Risk, compliance & identity



Scientific truth



Machine truth



SMALLEST · FASTEST GROWING

TOTAL ≈ \$160B / YEAR

The segments are different ways of paying someone to establish a determination. They grow at different rates and perform one function.

Financial and market truth runs \$65 to \$75 billion a year and defines the prices and benchmarks other systems accept. Market data and analysis is the bulk of it at \$49.2 billion. Ratings add \$11 to \$13 billion, indices \$7.2 billion, alternative data \$2.8 billion, and commodity and sustainability assessments contribute several billion more behind private-company walls.

Risk, compliance and identity runs \$60 to \$65 billion and decides whether a person, company or transaction can be trusted. Credit bureaus and business information account for \$19 billion, legal, tax and regulatory information for \$18 billion, identity verification and screening for \$15 to \$18 billion, insurance information for \$5 billion, and financial-crime data and tools for \$3 to \$5 billion.

Scientific truth runs \$16 to \$20 billion and attaches evidence to a process regulators and institutions accept. Scientific and academic publishing is roughly \$13 billion of it as a market, with the three largest houses, Elsevier, Springer Nature and Wiley, accounting for about \$7.3 billion between them, and clinical-trial data and sustainability assurance adding roughly \$1.5 to \$2 billion each.

Machine truth runs \$8 to \$12 billion and tells a model whether a document is current and controlling. AI data licensing leads at \$3 to \$5 billion, with evaluation and assurance, grounding and retrieval, data marketplaces, and resolution systems making up the rest. It is the smallest segment and the fastest growing, and Q10 returns to it.

The full line-item model, with sources per row, is in the appendix.

II. THE NUMBER · Q4

Why is the same fact verified ten thousand times?

\$206B Financial-crime compliance	\$246–266B Testing, inspection & certification
\$66.5B Audit & assurance	\$16.2B US title premiums, 2024
1 → thousands Institutions one encoded judgment can serve	

Sources: LexisNexis Risk Solutions (2023); [SkyQuest](#), [GMI & Roots Analysis](#) on TIC; Big Four reports; [ALTA \(2025\)](#)

Beyond the product market, institutions pay approximately **\$540 billion** a year in labor and services to establish facts: financial-crime compliance, testing and inspection, audit, title searches. That number is the strongest demand signal in this report. Verification matters so much that institutions pay half a trillion dollars a year for it, case by case.

The waste in the system is duplication. Every bank runs its own review of the same customer. Every fund rebuilds the same company's history. Thousands of compliance teams reach the same conclusion about the same entity in the same week, and each conclusion is used once and discarded. The expertise is real. The unit of work is wrong.

The frontier has hard edges. Much of this labor cannot convert. KYC reviews are required per institution and audits must be independent per client. Most testing and inspection examines a unique physical object. The convertible portion is the work that re-establishes the same public fact across many institutions: sanctions status, corporate events, registry states, filing histories. The 10 percent scenario is drawn against

that portion, and regulation decides how large it gets.

Software changes the unit, and the direction matters. A judgment produced once by someone who knows what they are looking at becomes a record used by thousands of institutions. The expert moves up the stack, from redoing the review to setting the methodology and standing behind the record. The people whose judgment gets encoded are the ones the product cannot exist without. A 10 percent conversion of this frontier adds more than **\$50 billion** in annual product revenue.

AI expands the market twice. It creates more demand for determinations, and it makes more verification reusable. Counted together, products and labor put the market for truth at roughly three quarters of a trillion dollars a year today, on a path past **\$1 trillion** before 2030. The headline in this report stays products-only, because labor answers a different question.

*News is worth what it resolves. A story takes its
price from the decision it closes.*

■

III. THE GAP · Q5

How much of this market does news capture?

\$8–15B

Current value of newsroom-type determinations

\$2–4B

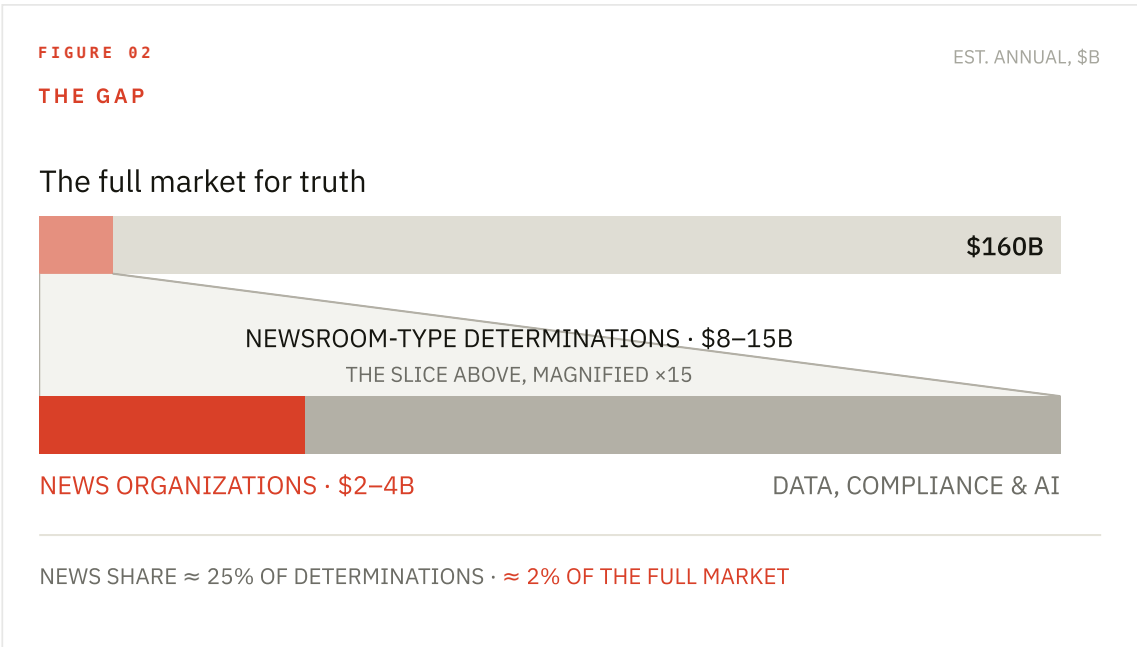
Captured by news organizations

~2%

Of the full \$160B market

Sources: report model

News organizations already produce determinations with institutional value: election calls, court outcomes, regulatory decisions, company events, specialist risk reporting. The report's model places the current value of these newsroom-type determinations at **\$8 to \$15 billion a year**, with news organizations capturing about **\$2 to \$4 billion**.



Within the determinations slice, news captures roughly a quarter, which is a respectable share of a small market. For every dollar news organizations capture, institutions pay the market for truth roughly sixty. The 2 percent against the full market only becomes meaningful if the addressable frontier expands beyond the categories news plays in today. Q15 makes that argument with numbers. Until it lands, the quarter is the honest headline.

III. THE GAP · Q6

Why is the reader the wrong denominator?

17%

Say they pay for online news globally

~1%

Of Americans pay when they hit a paywall

275:1

Terminal seat vs. news subscription

~\$70M/yr

OpenAI to Reddit (est.); AP paid considerably less

Sources: [Reuters Institute Digital News Report \(2026\)](#); Pew Research Center; company disclosures

Across the markets studied by the Reuters Institute, about 17 percent of respondents said they paid for online news. When Americans encounter a paywall, about 1 percent pay. The rest look elsewhere or abandon the question.

That does not mean verified facts have little value. It means the general reader is not the highest-value buyer. A clinical-trial result means one thing to a reader and something else to a fund holding the stock. The same result settles a contract, releases a milestone payment, triggers an insurance policy and updates an AI system. The reporting is identical. The consequence differs by orders of magnitude, and pricing follows consequence: an intelligence terminal seat costs 275 times a news subscription.

Which suggests a pricing principle the industry never wrote down: news is worth what it resolves. A story that resolves something, a contract, a model's answer, a regulatory question, a payout, takes its price from the decision it closes. The limit is plain: some of the most important journalism resolves nothing and should stay free to everyone. An investigation into a state's foster care system closes no contract. It matters anyway. Resolution is the part somebody is already willing to pay for.

The principle prices the resolution layer; the rest of the \$160 billion prices the layers around it: the monitored state, the screened claim, the matched identity, and the reference that holds steady between events. The grammar in Q1 covers all of them, and resolution is where the price per decision peaks. The buyers are specific, and so is where journalism sits in each chain.

AI platforms

BUYS FROM

Archive licensors and expert-data vendors

JOURNALISM'S POSITION

Supplies the source material

Event contracts

BUYS FROM

Exchange resolution operations

JOURNALISM'S POSITION

Usually outside settlement

Parametric insurance

BUYS FROM

Modeling and calculation agents

JOURNALISM'S POSITION

Little participation

Funds

BUYS FROM

Data and sentiment vendors

JOURNALISM'S POSITION

Reporting sits upstream, unpaid

Indices**BUYS FROM**

Index administrators

JOURNALISM'S POSITION

Occasional ownership or licensing

Risk & compliance**BUYS FROM**

Screening vendors

JOURNALISM'S POSITION

A few integrated businesses

Every buyer in that table is an institution with money riding on the answer, and each row traces one chain from a buyer of verified fact to the intermediary that collects the payment. The third column is where news organizations sit in that chain today.

The pricing of training data makes the same point. OpenAI pays Reddit an estimated \$70 million a year for anonymous arguments about whether a hot dog is a sandwich. It pays the Associated Press, roughly 250 locations in more than 100 countries producing the most reused factual reporting on earth, considerably less. Reddit's data arrives structured for how chatbots learn, while AP's arrives as narrative prose. The buyer pays for the usable form of a fact, and the usable form is the determination.

The industry has spent two decades trying to raise the reader's price. The larger move is changing the buyer.

III. THE GAP · Q7

Who captures the value instead?

1909

Moody's introduces bond ratings

1985

Baltic Exchange freight index

\$14–22M

Settled on one reporter's wording, March 2026

Sources: Moody's, Platts and Baltic Exchange corporate histories; Washington Post (Mar 2026); Bloomberg (Mar 2026); The Block; Axios

Companies that took repeated editorial judgments and turned them into reference products. Bloomberg wrapped a newsroom around a terminal, and Mergermarket, Reorg and 9fin built products around journalists reading filings and dockets. The longer pattern runs back a century, and the final row of it is the point.

Ratings

ESTABLISHED

Moody's · 1909

DOMAIN

Credit

THE SYSTEM HAD BEEN

Lenders each ran their own credit analysis

Price assessments

ESTABLISHED

Platts · 1923

DOMAIN

Commodities

THE SYSTEM HAD BEEN

Traders each reconstructed what a barrel traded for

Benchmarks

ESTABLISHED

S&P 500 · 1957

DOMAIN

Equities

THE SYSTEM HAD BEEN

Performance had no common yardstick

Identifiers

ESTABLISHED

CUSIP · 1968

DOMAIN

Securities

THE SYSTEM HAD BEEN

Securities were tracked by name and paper

Freight indices**ESTABLISHED**

Baltic Exchange · 1985

DOMAIN

Shipping

THE SYSTEM HAD BEEN

Brokers each priced routes from their own books

Verified outcomes**ESTABLISHED**

No incumbent · 2026 →

DOMAIN

Institutional events

THE SYSTEM HAD BEEN

Credit has a committee, catastrophe has damage indices, clinical and regulatory outcomes have nobody

These seats concentrate. Bloomberg and S&P Global take about \$30 billion a year between them, roughly a fifth of the whole market, and once filled a seat is held for generations. Occupants lose the seat through a loss of trust; a cheaper competitor has almost never taken one.

Where no institution claims the role, the market conscripts individuals. The missile contract from the opening is what conscription looks like: between \$14 million and \$22 million settled on one reporter's wording, the reporter fielded the threats, and the platform kept the fees. The dependence runs in both directions: another contract asked whether a senator would be publicly mentioned, and it resolved No because no approved outlet covered the closed rally in question. Reporting settles contracts, and the absence of reporting settles them too.

That failure argues for the grammar in Q1: a state object records that no qualifying event occurred, where publication-based settlement can only record that nobody wrote about it.

The Associated Press runs the consensual version, selling race calls under a defined methodology, and in 2026 agreed to supply them to a prediction market for election contracts. The buyer pays for the call. Everyone else in the industry supplies the same function without terms.

Individuals noticed before institutions did. A former Wall Street Journal reporter now charges hedge funds \$50,000 a year for the regulatory analysis she used to give away inside articles. The skill stayed the same while the label changed.

*No one sends death threats to a data feed. That is
a design requirement.*

IV. THE DEMAND CURVE · Q8

Why is demand exploding now?

2.5B

ChatGPT prompts per day, July 2025

51%

Of web traffic now non-human

Continuous

How machines consume determinations

Sources: Axios (Jul 2025); [Imperva Bad Bot Report \(2025\)](#)

A court ruling enters the public record at 10:01 a.m. By 10:02, an AI has answered ten thousand questions about it. By 10:04, trades have been placed on the answers. Some of those answers are wrong. The machine reads the ruling perfectly well. What it cannot see is what the ruling changed: which precedents it abandoned, which lower-court decisions it just made unenforceable. Establishing that requires knowing the state of the record before the ruling landed. The document is public. The change it made has to be reconstructed.

By 10:15 a journalist has read the opinion and knows exactly what changed. That knowledge reaches the world at noon, as prose, addressed to a reader. The systems that needed it at 10:02 were never the audience. By then the first answer has done its work: the trades cleared on it, and it has been copied into a thousand summaries that the next model will learn from. Whatever the machines and the markets act on becomes the truth in practice, for everyone downstream. Whoever supplies it is deciding what happened, and getting it wrong does not take a lie. Being first is enough.

That sentence describes the free tier of the information economy, and it marks the boundary of this market. Designation wins wherever liability attaches: the settlement, the compliance decision, the audited filing, the payout. The fast free answer keeps the casual query, and the paid determination takes every decision someone can be sued over. The money in this report sits on that side of the line.

The buyers behind that 10:02 answer are changing shape. A reader buys awareness a few times a day. A machine queries continuously and pays per answer, a settlement engine consumes determinations every hour it runs, and neither ever cancels a subscription. More than half of web traffic is already non-human. ChatGPT alone processes 2.5 billion prompts a day, and a meaningful share ask about the current state of the world, where the answer needs grounding in something checked.

Speed stopped being the scarce asset. Machines restate information instantly, which makes the account abundant. What remains scarce is warranted closure: the event occurred, the evidence is sufficient, the source is current, the determination can survive a challenge. Every additional machine consumer raises the price of closure while the price of the account keeps falling.

Demand grows on two curves at once: markets add contracts that settle on real-world events, and machines add queries that need grounded answers. Both curves price the determination.

IV. THE DEMAND CURVE · Q9

What do markets pay to settle?

\$51B Prediction-market notional volume, 2025	\$1T Bernstein's 2030 volume scenario
~\$10T CDS gross notional exposure	\$63.9B Cat bond & ILS market outstanding
\$425M One earnout dispute in Delaware	Volume ≠ revenue \$1T settled may pay tens of millions to resolve

Sources: Bernstein via CNBC (Apr 2026); [BIS OTC derivatives statistics](#); [Artemis](#); Fortis v. Shire (Del. Ch.); WSJ (May 2026)

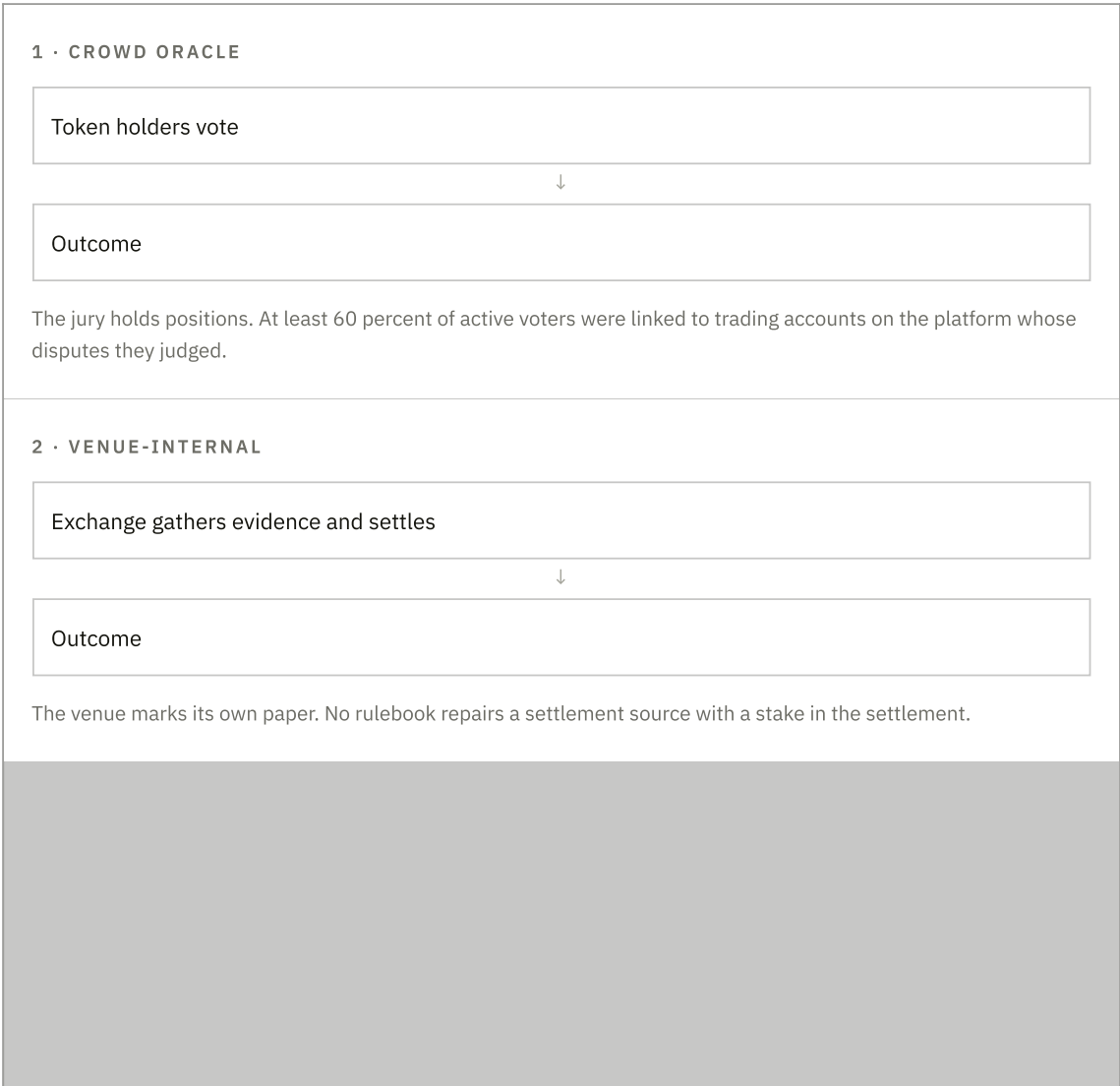
An event contract is any agreement whose value changes when something happens outside it: credit default swaps, catastrophe bonds, parametric insurance, weather derivatives, earnouts, contingent value rights, regulatory milestones, merger conditions. Every one has the same two problems: define the event, then decide whether it happened.

The deciding is the product. Credit markets carry roughly \$10 trillion in CDS exposure and run determinations committees because someone must interpret the event. The cat bond market reached \$63.9 billion outstanding and depends entirely on the quality of its triggers. Earnouts routinely end in court; one Delaware dispute over the contractual meaning of trial success involved up to \$425 million.

Prediction markets made the function visible by pricing it in public. They traded \$51 billion in notional volume in 2025, and Bernstein places 2030 volume at \$1 trillion. The harder contracts to settle, institutional and scientific events, are exactly the ones that require interpretation rather than a scoreboard.

Notional volume is a demand signal and a poor proxy for revenue. A trillion dollars of settled contracts might pay tens of millions in resolution fees, and AP's election licensing is likely worth single-digit millions a year. The proof that judgment can carry a large business is ratings, where a designated, issuer-pays model converts issuance into \$12 billion of annual fees. What resolution earns depends on the fee model it lands on, and no audited number exists yet. The model here books it at roughly \$1 billion today, with the steepest curve in the report.

The governance record explains why the resolver's seat is valuable. LIBOR was set by banks holding positions on the rate; manipulation drew more than \$9 billion in fines and ended the benchmark. A Wall Street Journal investigation found voters in a major prediction-market oracle holding positions in the markets they were deciding. Every settlement system eventually learns the same rule: the party with money on the determination cannot be the only party deciding it.



3 • SEPARATED

Independent layer establishes the record



Exchange settles under its rulebook

Evidence and authority sit apart: the arrangement equities already use, where the exchange lists and someone else audits.

Settlement authority stays with the regulated exchange in all three models; what changes is who supplies the evidence it settles on. Model 3 is the separation now arriving in event markets, and the one price reporting agencies have operated under for decades.

IV. THE DEMAND CURVE • Q10

What do machines pay to know?

\$10B+/yr

Lab spending on training data

~\$8.5B

Annual revenue of expert human-data vendors

\$85–110/hr

Rates paid to domain experts training models, averaging about \$105

1,668

AI-hallucination court decisions logged worldwide by July 2026

69–88%

Hallucination rate of legal AI on federal-case questions

Sources: Oxford Economics/Scale AI; Forbes; [HEC Paris \(Charlotin\) hallucination database](#); [Stanford RegLab](#); Pew Research Center; Challenger, Gray & Christmas; Graphite

Machine truth is the smallest segment and the fastest growing, and the largest line inside it is human judgment. AI companies now spend an estimated **\$10 billion or more a year on training data**, and the fastest-growing portion pays experts directly. Data companies recruit physicians, lawyers, physicists and finance professionals to write the rubrics, evaluations, reward signals and training environments that teach models how professionals think. The vendors supplying this judgment generate an estimated \$8.5 billion a year, and demand rises as models improve. That spend sits outside the machine-truth bar in Q3, and deliberately: expert training data is a service, counted the way the \$540 billion of verification labor is counted, while the bar counts products. The two converge as judgment gets productized.

For scale, the four largest cloud and AI companies are expected to spend roughly \$700 billion on infrastructure in 2026. The ratio reads as dismissal or as headroom, and headroom is the right reading: information spend can multiply several times over without registering in those budgets, and the binding constraint is the supply of judgment worth buying.

The willingness to pay tracks the error rate. American courts fined lawyers more than \$145,000 in the first quarter of 2026 alone for briefs citing cases that do not exist, about \$109,700 of it in a single case, and a database at HEC Paris had logged 1,668 such decisions worldwide by July. Stanford's RegLab found leading legal AI systems hallucinating on 69 to 88 percent of specific questions about federal cases. Roughly half of newly published English-language web articles are machine-generated. Answers are multiplying faster than institutions can check them, and a single mistake propagates through search results, legal filings, compliance tools and other models.

THE VERIFICATION GAP, 2019–2025

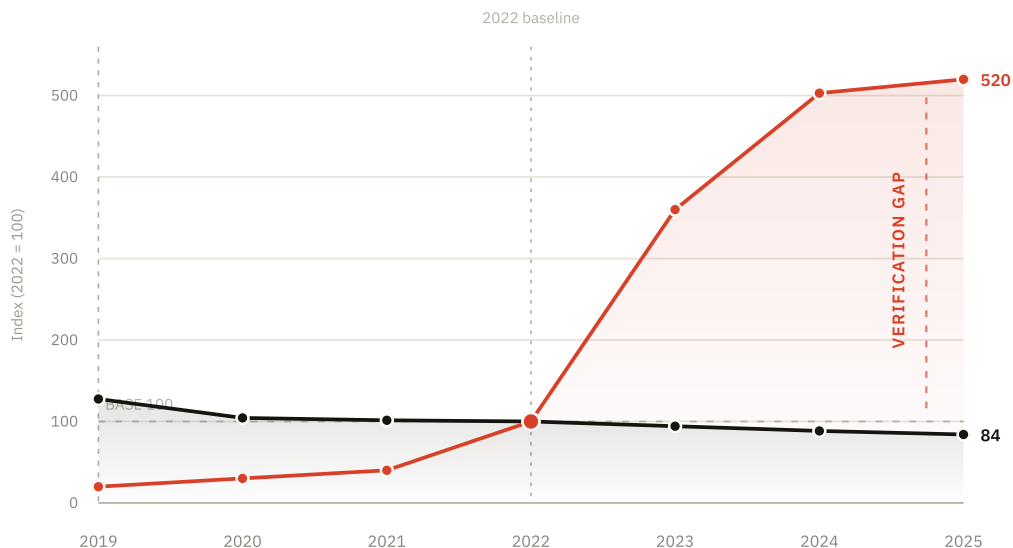
FIGURE 03

Are claims growing faster than the capacity to check them?

AI-generated content is expanding while newsroom employment, used here as one proxy for journalistic verification capacity, is declining.

■ Newsroom employment (69,000 → 58,000 since 2022)

■ AI-generated content share (10% → 52% since 2022)



METHOD

Both series indexed to 100 at November 2022 (ChatGPT launch). Newsroom employment: 88,000 (2019) to 58,000 (2025), from Pew Research and Challenger, Gray & Christmas. AI-generated share of new web articles: from near zero to 52 percent, from Graphite/SurferSEO. The index shows rate of change from a

shared starting point, not absolute scale. AI adoption is not asserted to have caused the newsroom decline; the two trends move in the same window for related but distinct reasons. Employment is an imperfect proxy: part of the decline is verification moving into products, which is the subject of this report. The 2025 endpoint is an estimate pending Pew's next newsroom series.

A model can retrieve any document. It cannot establish whether a document is final, which source controls, what changed, or what the silence around it means. The labs are buying exactly that kind of judgment, domain by domain, at \$85 to \$110 an hour, through marketplaces built to extract it. Editorial judgment has not been systematically bought yet. It is next.

IV. THE DEMAND CURVE · Q11

Where does the market land by 2030?

\$230–300B

2030 product-market scenario

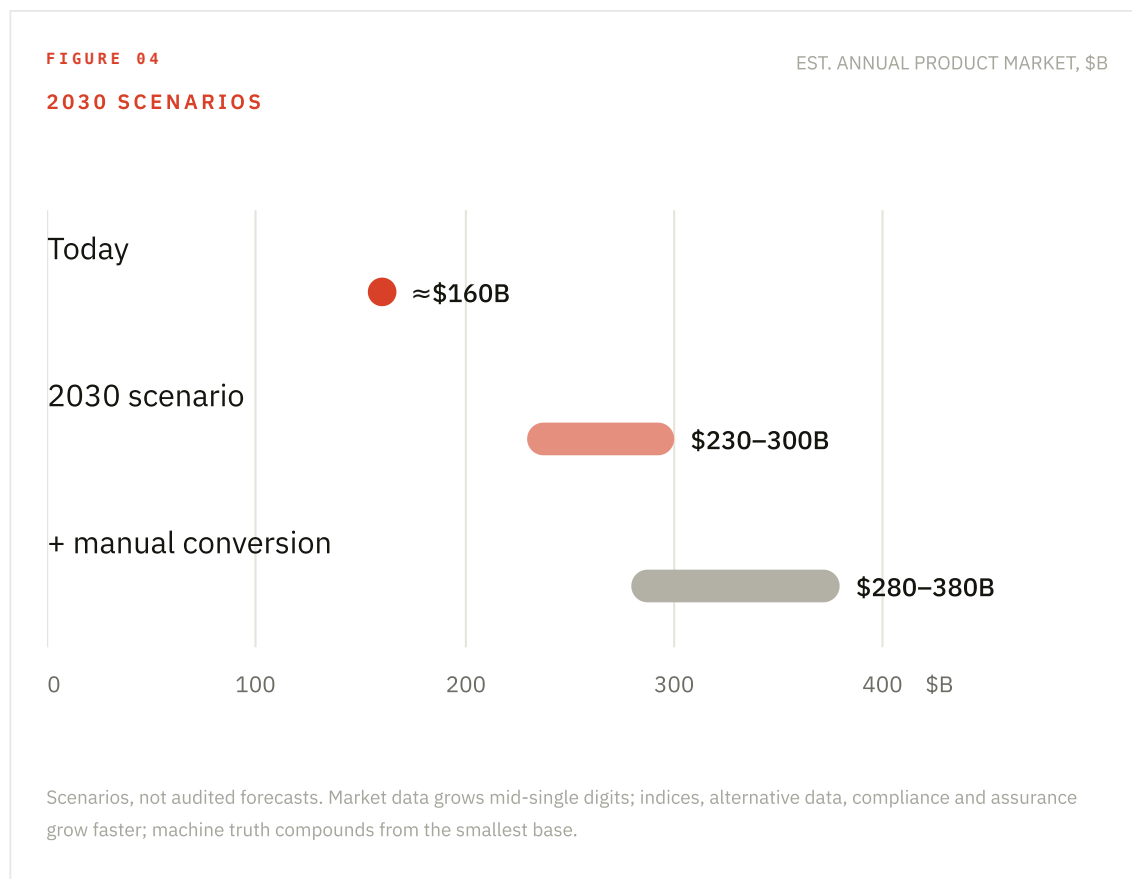
\$280–380B

With partial manual conversion

~\$1T

Wider verification economy

Sources: report model; segment growth rates



The exact endpoint matters less than the direction. More decisions are becoming dependent on explicit, machine-readable answers, and every one of those decisions needs a determination behind it.

*The judgment every contender needs still lives in
newsrooms.*

■

V. WHO TAKES THE SEAT · Q12

Which capabilities decide the winner?

Capability one

Sell the determination on its own terms

Capability two

Read the public record with an encoded lens

The precedent

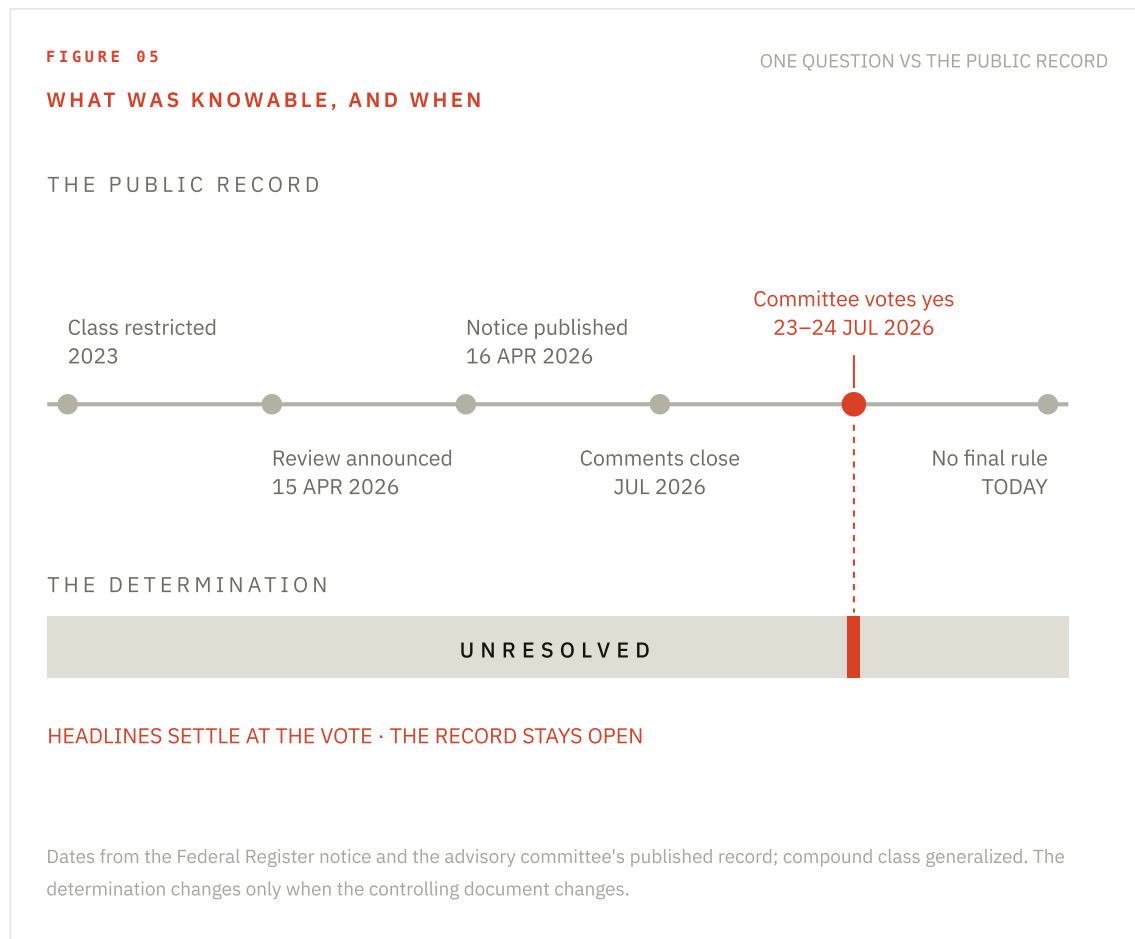
Moody's 1909 · Platts · Baltic Exchange 1985

Sources: corporate histories; Axios

History grades reference institutions on two capabilities, and every past winner built both.

The first is selling the determination on its own terms. The article and the judgment inside it are separate products, and the winners priced the judgment: as a feed, with a timestamp, a methodology, an evidence trail and a service level, priced by the decision it closes rather than the audience it reaches. A feed that settles a contract is worth more than the same fact inside a story. AP operates this model for elections today. Reuters, Moody's, Platts and the Baltic Exchange operated it for prices, credit, oil and freight a century before anyone called it unbundling.

The second is reading the public record at machine scale without losing the lens. The record is digital and continuous now, and extraction costs almost nothing. The expensive part is judgment: a filing can amend an earlier disclosure, an advisory committee vote is different from an approval, a press release can be accurate in every sentence while the whole misleads, and a docket entry can close a case or merely continue it. That judgment can be encoded as source hierarchies, prior state, materiality and correction logic, with humans at the decision points. Hedge funds ran this conversion on their own analysts years ago. The observation becomes a rule, and the rule becomes a screen that runs continuously. The lens also pays retroactively: an archive converted this way typically holds substantial data its owner never surfaced: values the text implies without stating, recoverable with a written rationale and a confidence threshold.



Where the output settles money, a human layer sits above the machine layer. The machine flags candidates for resolution by mapping what was promised against what was filed, and an editor approves the mapping before release. The published record carries the full chain: the original commitment, the evidence, the determination, and who signed off. That is the pattern a newsroom uses to call an election, generalized to any domain with a resolvable question.

Neither capability belongs to any contender by right. They are requirements, and every institution in the next question is missing at least one.

*The pattern a newsroom uses to call an election
generalizes to any domain with a resolvable
question.*



V. WHO TAKES THE SEAT · Q13

Who is best positioned?

June 2026

S&P takes independent administration of CDS determinations

\$2B

ICE's investment in Polymarket

\$200M+/yr

Thomson Reuters AI spend · 28% of contract value GenAI-enabled

Sources: ISDA/CDGC announcement (Jun 2026); ICE press release (Oct 2025); Thomson Reuters disclosures

Data & risk incumbents

S&P, Moody's, Verisk

HOLDS

Benchmark governance. S&P took independent administration of CDS determinations in June 2026.

MISSING

Event coverage beyond credit and commodities. No newsgathering reflex.

Exchanges

ICE, prediction platforms

HOLDS

The settlement demand itself. ICE put \$2 billion into Polymarket.

MISSING

Independence. The house cannot be the only grader of its own contracts.

Hybrid publishers

Thomson Reuters, Bloomberg, Dow Jones

HOLDS

Newsroom plus data infrastructure. Thomson Reuters spends \$200 million a year on AI, with 28 percent of contract value GenAI-enabled.

MISSING

Event resolution as a product line outside legal and finance.

News organizations

wires and specialist titles with deep beat expertise

HOLDS

The only contenders already selling settlement-grade determinations, and beat knowledge nobody else holds: a regional publication can know an oil and gas basin better than any data vendor.

MISSING

Capital and enterprise infrastructure. Editorial margins rarely fund point-in-time systems and service levels.

AI labs

buying judgment through the expert-data marketplaces

HOLDS

Capital and extraction machinery. An estimated \$10 billion a year already flows to expert judgment.

MISSING

Accountability. A lab grading the facts its own model serves is judge and interested party at once.

New entrants**HOLDS**

No legacy to defend. Event-first architecture from day one.

MISSING

A decade of trust. Though Moody's was a new entrant once.

The likeliest outcome is that the seat splits by vertical before anyone owns the category, the way credit, commodities and freight split a century ago. The handicap favors whoever pairs accountable judgment with settlement-grade infrastructure first, in one vertical, and expands from there. Governments are the wildcard outside the table: agencies could publish structured, settlement-grade determinations directly and shrink the private market in whichever categories they claim.

One test applies to whoever wins. The resolver's revenue comes from the parties with money on the determinations, which is the arrangement that compromised ratings in 2008 and LIBOR before it. The winner's governance has to survive its own customers: published criteria, a source hierarchy, conflict disclosures, a public correction record, an appeals process, diversified revenue, and separation between evidence and payout. Independence is part of the product, and it is the part the conflicted contenders cannot copy.

V. WHO TAKES THE SEAT · Q14

How fast can the labs extract the lens?

\$14B

Meta's investment in Scale AI

\$2B

Mercor gross revenue run rate, June 2026 · 91% from foundation labs

\$1.2B

Surge AI revenue, 2024

\$1B+/yr

One lab's reported budget for training environments alone

Sources: TechCrunch; Forbes (Sep 2025); [Sacra](#); [Crypto Briefing](#); industry reporting, 2025–26

The machinery for extracting professional expertise already runs at industrial scale. Meta invested \$14 billion in Scale AI. Surge crossed \$1.2 billion in revenue selling expert evaluations to labs. Mercor reached a \$2 billion gross revenue run rate in three years by recruiting doctors, lawyers and analysts to train models, with 91 percent of its revenue from foundation labs. One lab is reported to be weighing more than \$1 billion a year on training environments alone. Demand for human judgment grows as the models improve.

Licensing is the smaller story. An archive tells a model what the newsroom said. The expert-data economy buys the judgment itself: how a professional reads a document, which source controls, what a change means, what silence means. A license renews every year. Judgment sold as training data transfers once and permanently, and it keeps working for the buyer after the contract ends.

Nothing stops the labs from running the same extraction on journalism. Beat reporters hired at hourly rates, writing rubrics for how to read a docket, an FDA filing, a sanctions notice. The newsroom that developed that judgment over fifteen years captures nothing, and the lens leaves with the reporter. Reference markets reward the institution accepted first, and in extraction markets the advantage goes to whoever sells early. The seat does not remain open indefinitely.

The threat has a boundary worth naming. Extraction transfers past judgment, and past judgment ages the way a licensed archive ages. Live resolution requires an ongoing institution: someone monitoring, ruling and correcting this week, accountable for this week's determinations. The labs can buy the analytical layer by the hour. Buying the standing behind a live determination means becoming the kind of institution this report describes. The same boundary cuts both ways: whatever reporters can teach a newsroom's software, hired reporters can teach a lab's. The durable moat is the live institution. The lens is how a contender builds one first.



V. WHO TAKES THE SEAT · Q15

What is the newsroom-addressable number?

\$25–35B

Direct determination products, 2030

\$110–170B

Software with embedded editorial judgment

\$135–205B

Combined newsroom-addressable

\$20.6B

US newspaper market, for scale (Grand View Research, 2024)

\$607.5M

Dow Jones's index business sale to CME

Sources: report model; CME/Dow Jones (2010); Grand View Research (2024); IBISWorld

The 2030 scenario places direct determination products at **\$25 to \$35 billion**. Software carrying embedded editorial judgment, the encoded lens sold as monitoring, screening and resolution infrastructure, adds **\$110 to \$170 billion**. The combined newsroom-addressable market reaches **\$135 to \$205 billion**.

These are not forecasts of newsroom revenue. They describe the parts of the wider market where newsroom methods and authority matter. For scale: the entire US newspaper market is about \$20.6 billion on Grand View Research's 2024 measure, and closer to \$30 billion on IBISWorld's broader definition. The addressable market is roughly five to ten times the industry's current business, because the buyer is no longer only the reader.

The precedents carry actual money. Nikkei has collected royalties on Nikkei 225 futures listed in Osaka, Singapore, Chicago and São Paulo since the 1980s. Dow Jones sold ninety percent of its index business to CME Group for \$607.5 million, and the successor still takes a share of the profits on every index product the exchange runs. Platts assessments run to roughly a billion dollars a year, and neither Platts nor Fastmarkets had to become Bloomberg to get there.

The realistic route into these numbers is narrow: one domain covered deeply enough that the record is worth paying for. The concentration in Q7 applies here too. The addressable figures are category-wide, and capture is winner-take-most per vertical, decided early. The addressable market is large, and the number of winners per vertical is one or two.

Selling determinations opens the \$25 to \$35 billion, and the encoded lens opens the \$110 to \$170 billion. Neither requires abandoning the public mission; core facts can stay public while institutions pay for speed, structure, history and service levels, and institutional revenue can fund the reporting the market will never pay for.

The labs already pay experts by the hour to teach their models judgment. Editorial judgment is next on the list.

CONCLUSION

What is the market size for truth?

The market is \$160 billion today in products, with \$540 billion more in manual verification behind it. The 2030 scenario reaches \$230 to \$300 billion, of which \$135 to \$205 billion is addressable by newsroom methods. News currently captures \$2 to \$4 billion, and the market is not waiting for the industry to notice.

The predictions below carry dates so they can be proved wrong.

1

By the end of 2027

A major news organization licenses verified outcomes beyond elections to an exchange, insurer, financial institution or contract administrator.

2

By the end of 2028

An AI company pays a news organization more for its editorial methods and live judgment than for access to its archive, because extracted judgment ages and live judgment must be bought continuously.

3

By 2030

Resolution disputes become a major constraint on institutional and scientific event contracts.

4

By 2030

More insurance, financial and commercial contracts use externally verified events as automatic triggers, and the next generation of benchmark and resolution companies employs reporters and editors, whether or not anyone calls the work journalism.

One failure condition deserves engagement. If the AI labs, working through the expert-data marketplaces, extract the editorial lens faster than any newsroom encodes its own, the addressable numbers above transfer to them. Nothing in the market model requires the winner to come from journalism. The model only requires that someone with the judgment takes the seat, and the judgment currently lives in newsrooms.

It is a strange thing to have been good at something for a hundred years and never learned what it was worth. Someone will resolve the truth. What stays open is who holds that authority, under what rules, and what claim the public keeps on the record once it becomes infrastructure. The correspondent in the opening had no claim on it at all.

APPENDIX

Method & market model

Method

A line counts if the thing being bought is knowledge or assurance of what is true: a price, a rating, an identity, a test result, an audited number, a verified event. Advice and execution stay out. So does entertainment. Each line takes the highest-credibility source available, in order: company filings and trade-body data; specialist analysts who survey the actual buyers or sellers; commercial market research only where nothing better exists, and always as a range. Source tiers: D disclosed in filings or investor releases, B specialist analyst, P posted official price, R reported by named press, V vendor estimate. Figures at tier V, and all 2030 flows, are projections rather than audited actuals.

Revenues are deduplicated before summing: Bloomberg, LSEG Data & Analytics and S&P Global Market Intelligence sit inside the market-data aggregate; ratings, Verisk and LexisNexis Risk are each counted once. Growth to 2030 uses observed rates per segment: market data about 6 percent and indices about 13 percent (Burton-Taylor), alternative data 17 to 33 percent (Neudata), sustainability assurance about 20 percent (Verdantix), and machine-truth lines 20 to 49 percent, flagged as projections. Known gaps, stated first: no high-credibility standalone figure exists for ESG data or for private price reporting agencies beyond Platts and Argus, and private-company figures including Bloomberg and Argus are estimates rather than filings. The \$540 billion of verification labor is counted separately from the headline because it is overwhelmingly in-house headcount and field services; adding it puts the wider figure past \$1 trillion by 2030 and answers a different question. The conversion scenario in the 2030 chart is the moment part of that labor becomes product and enters the headline. Expert training-data services, roughly \$8.5 billion across Scale, Surge, Mercor and peers, sit outside the headline for the same reason.

Market model, segment by segment

Estimates in \$B USD per year, current. Ranges reflect source variance and private-company boundaries.

SEGMENT	LINE OF BUSINESS	EST. \$B/YR	SOURCE BASIS
Financial & market truth \$65–75B	Market data & analysis	49.2	Burton-Taylor, 2025
	Credit ratings	11–13	S&P and Moody's 2024–25 segment results; Fitch est.
	Indices	7.2	Burton-Taylor
	Alternative data	2.8	Neudata
	Commodity & sustainability assessments	several	S&P Commodity Insights; private est.
Risk, compliance & identity \$60–65B	Credit bureaus & business information	19	Experian, Equifax, TransUnion, D&B, Verisk
	Legal, tax & regulatory information	18	Thomson Reuters, RELX, Wolters Kluwer
	Identity verification & screening	15–18	First Advantage; Burton-Taylor
	Insurance information	5	Verisk and peers
	Financial-crime compliance data & tools	3–5	Burton-Taylor AML/KYC
Product market total		≈160	

SEGMENT	LINE OF BUSINESS	EST. \$B/YR	SOURCE BASIS
Scientific truth \$16–20B	Scientific publishing (broad market)	13	STM market est.; top three houses ≈\$7.3B (<u>WordsRated</u>)
	Clinical-trial data & analytics	1.5–2	IQVIA and peers
	Sustainability & carbon assurance	1.5–2	Verdantix
Machine truth \$8–12B	AI data licensing	3–5	Disclosed deals; News Corp, Reddit
	AI evaluation, observability & assurance	1.5–2	Commercial estimates
	Grounding & retrieval	1–2	Commercial estimates
	Data marketplaces (narrow)	1–2	Commercial estimates
	Resolution & oracle systems	1	Chainlink, DefiLlama, Verra
Product market total		≈160	

MANUAL FRONTIER	EST. \$B/YR	SOURCE BASIS
Financial-crime compliance (labor & services)	206	LexisNexis Risk Solutions, 2023 study
Testing, inspection & certification	246–266	<u>SkyQuest</u> ; <u>Roots Analysis</u> ; <u>GMI</u>
Audit & assurance	66.5	Big Four reports
Title search, examination & insurance (US premiums, 2024)	16.2	<u>ALTA</u>
Manual frontier total		≈540

APPLIEDXL

Intelligence infrastructure for the information industry.

appliedxl.com · support@appliedxl.com

© 2026 AppliedXL. All rights reserved.

The Market Size of Truth

How machines and markets will buy
what journalism never sold