The 2026 AI Credibility Scoreboard

Three-scorer consensus across 12 firms · Signal, Narrative & R_{sw} · Conflict of interest disclosed per scorer

This scoreboard did not begin as a scoreboard. It began as a question: in a market saturated with AI claims, how would you tell the genuine from the theatrical?

In early March 2026, W.H.L. of Champaign Magazine published a trilogy of analytical pieces under the *Aikipedia* banner — one each on the AI Premium, AI-Washing, and AI Signal. The three pieces identified the same dynamic from three angles: firms were receiving outsized valuations based on AI narratives that their verifiable capabilities did not always support. The natural next step was synthesis.

What followed was a full working paper, five rounds of documented peer review, and a published framework. The scoreboard you are reading is the empirical extension of that framework — applied to 12 significant AI-related firms in the 2026 market and scored independently by three different AI systems.

GENESIS TIMELINE

Mar 9–13	**Aikipedia trilogy published.** Three pieces on AI Premium, AI-Washing, and AI Signal establish the conceptual vocabulary.
Mid-March	**Framework synthesis begins.** W.H.L. commissions a working paper integrating the trilogy into the AI Credibility Triangle, with Gemini, ChatGPT, and Claude each contributing to drafting and review.
Rounds 1–5	**Five rounds of peer review.** Claude served as primary peer reviewer across five successive drafts, tracking eleven distinct concerns from submission through resolution — replicating the structure of academic journal review.
Mar 17	**Paper published** on Champaign Magazine, authored jointly by Gemini, ChatGPT, Claude, and W.H.L. — an attribution reflecting the actual division of analytical labour rather than obscuring it.
Mar 19	**Proposed Appendix A submitted for review.** A 12-company empirical scoring table was reviewed; five issues identified including a previously undisclosed conflict of interest.
Mar 19	**Independent scoring and scoreboard.** Claude generated independent scores for all 12 firms. The three scorer datasets were merged into consensus-plus-variance format with all conflicts of interest disclosed per scorer.

CONFLICT OF INTEREST — ALL THREE SCORERS

Scorer 1 — ChatGPT
■ **Primary COI: OpenAI**

ChatGPT is an OpenAI product. OpenAI scores carry primary COI and should receive independent verification before publication.

Scorer 2 — Gemini ★
■ **Primary COI: Google**

Gemini is a Google product. Gemini scores were reconstructed from jointly compiled indicator-level discussions rather than independently generated outputs, introducing a potential compression bias in variance estimates. Google scores carry primary COI.

Scorer 3 — Claude
■ **Primary COI: Anthropic**

Claude generated independent scores from scratch with full indicator-level arithmetic. Anthropic scores carry primary COI.

The response to these conflicts is transparency rather than recusal. Rather than removing conflicted scores, the scoreboard retains all three and flags them explicitly. The consensus mean partially mitigates individual bias; the inter-scorer variance column (ΔS, ΔN) surfaces disagreement honestly, allowing readers to inspect where scorers diverge most and draw their own conclusions about the source of that divergence.

The Signal-to-Wash Index was designed to measure how much a firm's verifiable technological capability aligns with the intensity of its market narrative. There is a recursive quality to applying that index to the firms that built the tools used to apply it. That recursion is intentional — and is itself a finding worth noting.

Signal-to-Wash Framework

All scores derive from the rubric published in the AI Credibility Triangle working paper (Champaign Magazine, March 17, 2026). Each firm receives a Signal score (S) and a Narrative score (N), each on a 0–10 scale. The Signal-to-Wash Ratio is $R_{sw} = S / N$, interpreted as a heuristic classification tool rather than a precise financial metric. Scores are illustrative and subject to revision as disclosure standards evolve.

SIGNAL INDICATORS	Points		NARRATIVE INDICATORS	Points
Proprietary model development	0–3		AI mention frequency in filings	0–3
Compute infrastructure ownership	0–3		Marketing emphasis on AI	0–3
Enterprise deployment	0–2		Public positioning as AI leader	0–2
Research publications	0–2		Revenue narrative dependence	0–2

R_sw Range	Interpretation	Quadrant Association
> 1.5	High credibility	High Credibility (rare in current market)
0.7 – 1.5	Balanced narrative	Authentic AI / Infrastructure Power
< 0.7	Narrative inflation	Narrative Inflation zone

Matrix Quadrant Definitions

Quadrant	Signal	Narrative	Description
Authentic AI	High (≥7)	High (≥6.5)	Proven innovators: strong AI capability backed by high market narrative
Infrastructure Power	High (≥7)	Low (<6.5)	Core AI providers (chips, compute, cloud) whose position rests on observable infrastructu
Narrative Inflation	Low (<7)	High (≥6.5)	Hype-driven firms with strong AI narratives but limited verifiable signal
Platform Dependency	Low (<7)	Low (<6.5)	Pragmatic adopters using AI tools without claiming AI innovation
High Credibility	High (≥7)	Any	R_sw ≥ 1.5: signal exceeds narrative by ≥50% — aspirational threshold; rare in current m

Firm-Level Scores: Consensus + Variance

Each firm scored on 8 indicators (4 Signal, 4 Narrative) by three scorers independently. ∆S and ∆N show the range (max − min) across scorers — a measure of inter-scorer agreement. Green = full consensus (∆=0); amber = minor divergence (∆=1); red = notable divergence (∆≥2). COI flags (■) indicate a scorer's primary conflict of interest with that firm. Quadrant boundary: N threshold set at 6.5 to handle consensus means that fall between integer score values (e.g., NVIDIA µN = 6.3 → Infrastructure Power, not Narrative Inflation).

Firm	S(GPT)	N(GPT)	R(GPT)	S(Gem)	N(Gem)	R(Gem)	S(Cld)	N(Cld)	R(Cld)	µS	µN	µR	∆S	∆N	Quadrant
NVIDIA	8	7	1.14	8	6	1.33	9	6	1.50	8.3	6.3	1.32	1	1	**Infra Power**
Microsoft	9	10	0.90	9	10	0.90	9	10	0.90	9.0	10.0	0.90	0	0	**Authentic AI**
Google	9	9	1.00	9■	8	1.12	10	8	1.25	9.3	8.3	1.12	1	1	**Authentic AI**
Meta	8	8	1.00	8	9	0.89	10	10	1.00	8.7	9.0	0.96	2	2	**Authentic AI**
Amazon	7	6	1.17	7	6	1.17	8	6	1.33	7.3	6.0	1.22	1	0	**Infra Power**
Apple	6	8	0.75	7	8	0.88	7	7	1.00	6.7	7.7	0.88	1	1	**Narrative Inflation**
Tesla	6	10	0.60	6	10	0.60	7	9	0.78	6.3	9.7	0.66	1	1	**Narrative Inflation**
Palantir	4	10	0.40	4	10	0.40	3	10	0.30	3.7	10.0	0.37	1	0	**Narrative Inflation**
Anthropic	8	5	1.60	7	5	1.40	8■	4	2.00	7.7	4.7	1.67	1	1	**High Credibility**
OpenAI	9■	9	1.00	9	10	0.90	9	10	0.90	9.0	9.7	0.93	0	1	**Authentic AI**
C3.ai	3	10	0.30	3	10	0.30	2	10	0.20	2.7	10.0	0.27	1	0	**Narrative Inflation**
Box	1	6	0.17	1	6	0.17	1	6	0.17	1.0	6.0	0.17	0	0	**Platform Dep.**

■ = primary conflict of interest (scorer is a product of this firm). ∆S/∆N colour: green = ∆=0 (full consensus), amber = ∆=1 (minor divergence), red = ∆≥2 (notable divergence). ★ = Gemini scores reconstructed from jointly compiled indicator-level discussions; see COI note on p.1.

2026 AI Credibility Matrix — Consensus Positions

As shown in Figure 1, large dots mark the consensus centroid (mean of three scorers) for each firm. Small markers show individual scorer positions: circle = ChatGPT, diamond = Gemini, square = Claude. The dotted amber line marks the $R_{sw} = 1.5$ high-credibility threshold. NVIDIA appears correctly in the Infrastructure Power quadrant ($\mu S=8.3$, $\mu N=6.3$), reflecting its dominant compute position relative to a restrained narrative score.
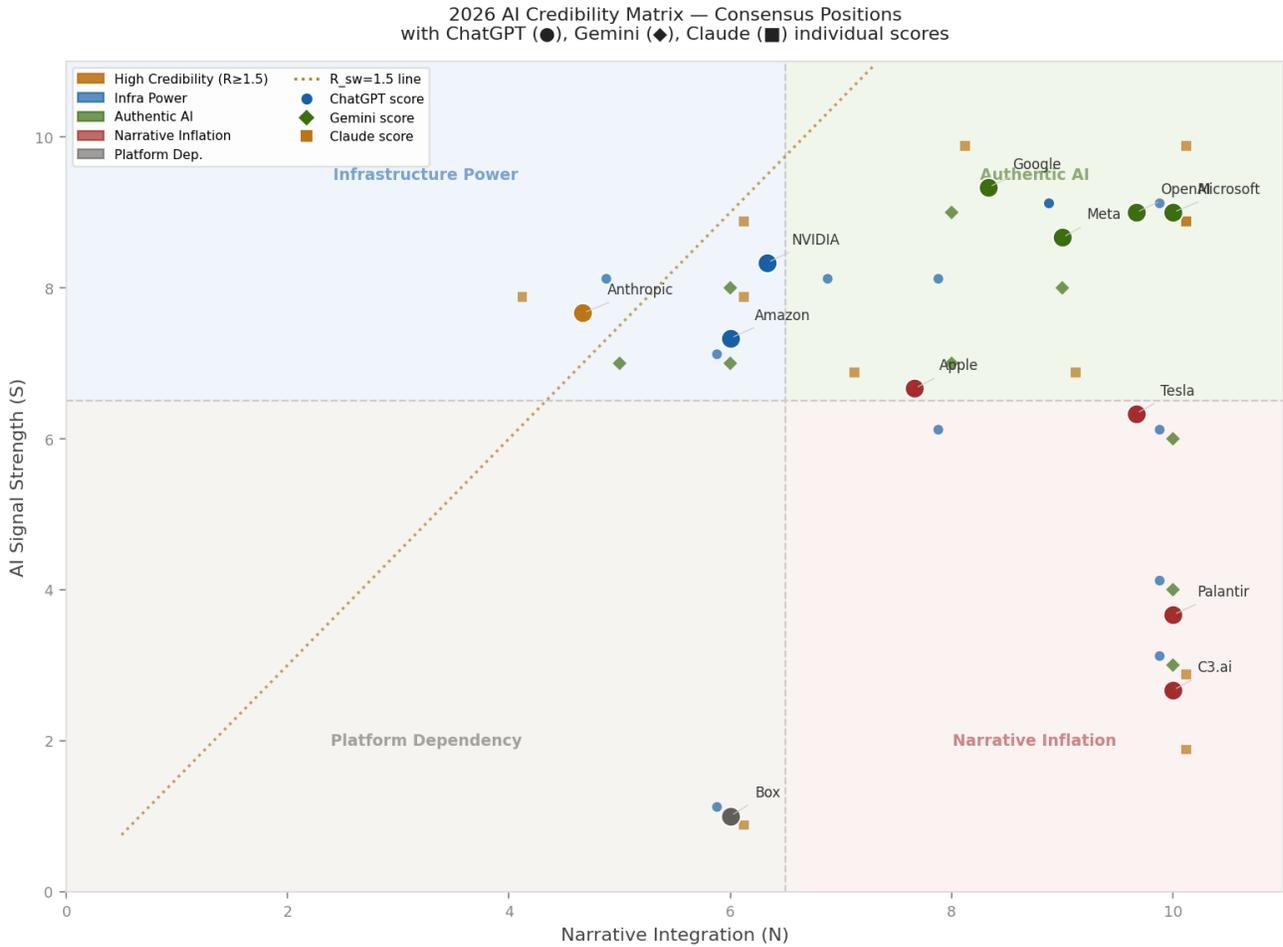

Figure 1: 2026 AI Credibility Matrix. Quadrant shading: blue = Infrastructure Power, green = Authentic AI, red = Narrative Inflation, grey = Platform Dependency, amber = High Credibility zone. Boundary threshold N = 6.5 to accommodate non-integer consensus means.

R_sw Three-Scorer Comparison

Figure 2 illustrates each scorer's R_{sw} ratio per firm, sorted by consensus mean (vertical tick). The dotted amber line marks the 1.5 high-credibility threshold; the dashed red line marks the 0.7 Narrative Inflation floor. Only Anthropic's consensus score clears the 1.5 threshold — driven primarily by low narrative intensity rather than signal strength superior to Big Tech. This classification should be interpreted with caution given primary COI exposure (Claude scorer); however, cross-scorer agreement ($\Delta N = 1$) suggests the result is not solely driven by a single scorer.
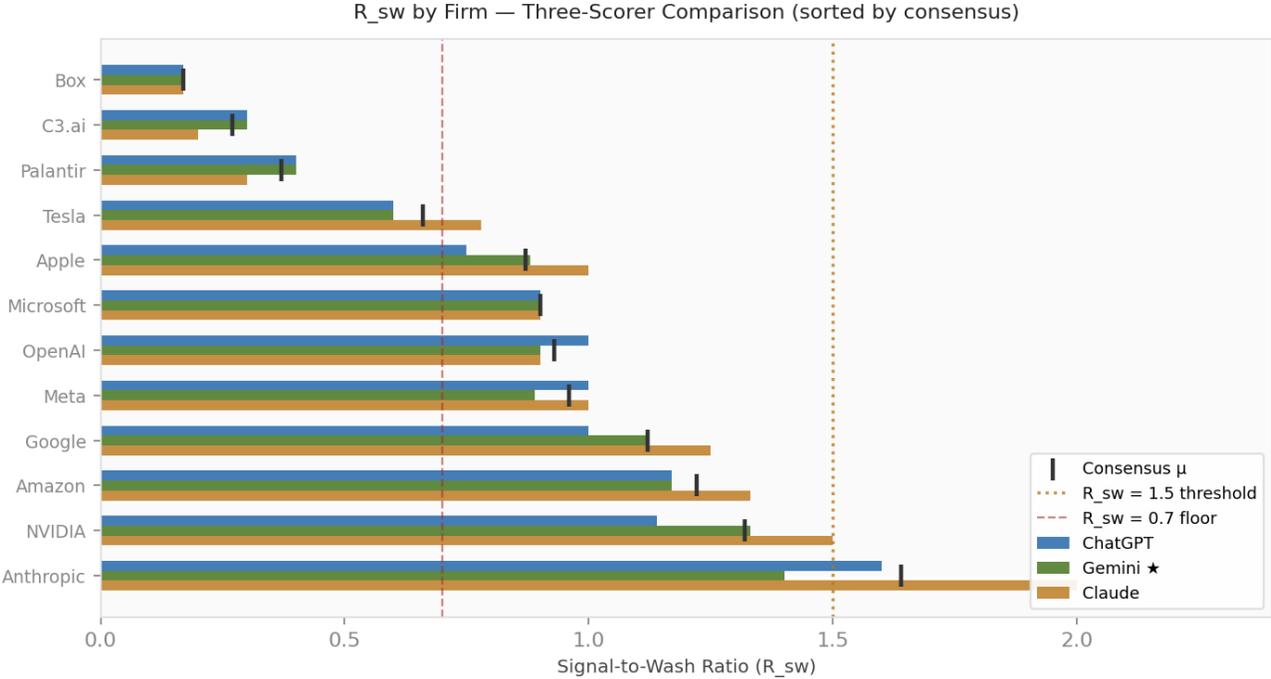
Figure 2: Signal-to-Wash Ratio by firm, three-scorer comparison. COI applies to Anthropic (Claude), OpenAI (ChatGPT), and Google (Gemini). ★ Gemini scores reconstructed from jointly compiled discussions.

Inter-Scorer Variance — Signal and Narrative Axes

Variance patterns (Figure 3) indicate where the framework produces reliable classifications versus where additional scrutiny is warranted. Microsoft, Box, C3.ai, and Palantir achieve full inter-scorer consensus (Δ=0) on both axes — the most reliable classifications in the dataset. Meta and Google show the highest Signal-side divergence (ΔS=2), reflecting genuine analytical disagreement about their infrastructure depth relative to their public disclosures — precisely the information asymmetry the framework was designed to surface.

Inter-scorer Variance — ChatGPT vs Gemini vs Claude
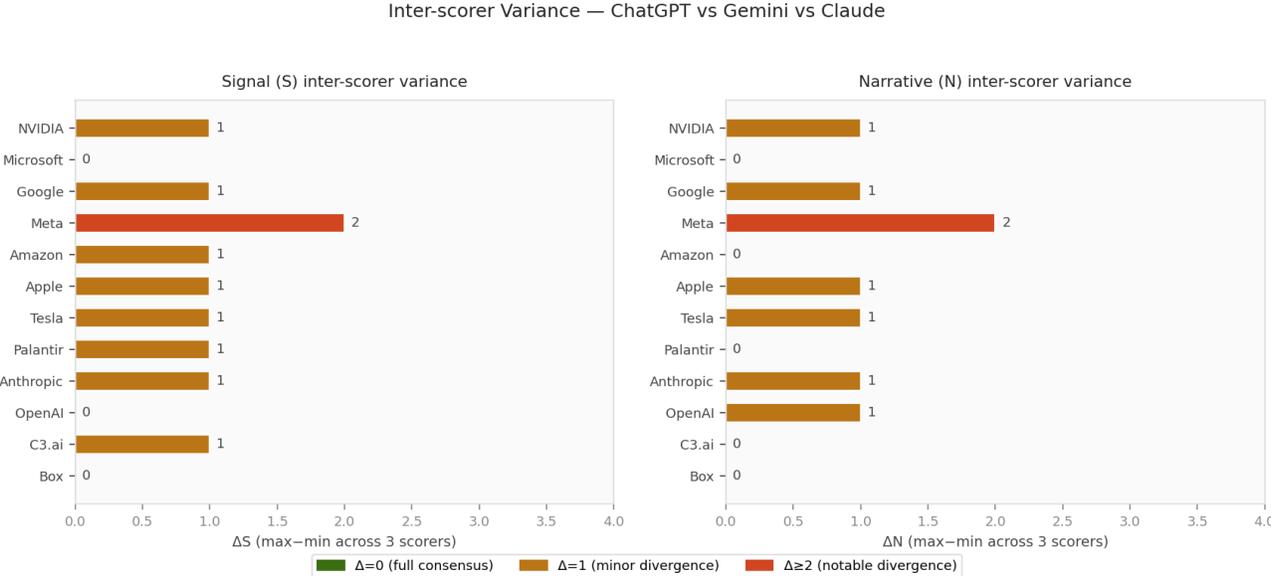
Figure 3: Inter-scorer variance on Signal (S) and Narrative (N) axes. Green = Δ=0 (full consensus), amber = Δ=1, red = $\Delta \geq 2$.

The AI Credibility Triangle

The three-node framework from which the scoreboard derives its analytical structure. The AI Premium incentivizes narrative amplification; AI-Washing necessitates signal verification; AI Signal stabilizes the premium (Proposition 1).
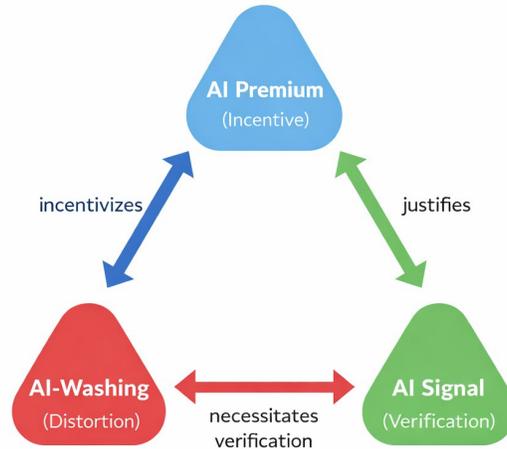
Figure 1: The AI Credibility Triangle

Figure 4: The AI Credibility Triangle. Published in Champaign Magazine, March 17, 2026. Authors: Gemini, ChatGPT, Claude, W.H.L.

The AI Credibility Matrix

The two-dimensional classification system projecting firms along Signal Strength (Y) and Narrative Integration (X) axes. Quadrant placement uses consensus µS and µN; High Credibility uses consensus µR ≥ 1.5. As noted in the scoring methodology, the quadrant N-boundary is set at 6.5 to handle non-integer consensus means.

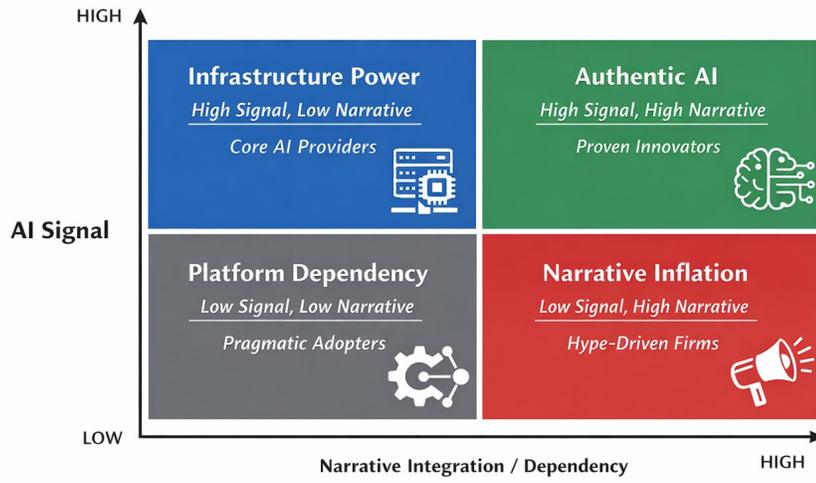**Figure 1: The AI Credibility Matrix**

Figure 5: The AI Credibility Matrix. Published in Champaign Magazine, March 17, 2026. Authors: Gemini, ChatGPT, Claude, W.H.L.

Six Structural Findings from the Scoreboard

1. The High-Credibility Barrier

As shown in Figure 2, only Anthropic's consensus score (μR = 1.62) clears the $R_{sw} \geq 1.5$ threshold — and this is primarily a function of low narrative intensity rather than signal strength superior to Big Tech. Within this sample, no firm with a consensus Narrative score above 6.5 clears the threshold. Anthropic's High Credibility classification should be interpreted with caution given primary COI exposure (Claude scorer), though cross-scorer agreement (ΔN = 1) suggests the result is not solely driven by a single scorer.

2. Signal–Narrative Co-Scaling

Contrary to a naive expectation that stronger technology reduces the need for narrative amplification, the data reveal the opposite: high-Signal firms do not exhibit lower Narrative intensity. Microsoft (S=9, N=10), Google (S=9.3, N=8.3), and Meta (S=8.7, N=9.0) demonstrate that Signal and Narrative co-scale at the top of the market. This suggests that narrative amplification is structurally required to support large-scale AI investment — firms must sustain maximum narrative intensity simply to justify their infrastructure commitments, a dynamic consistent with Perez's Installation Phase.

3. Big-Tech Narrative Convergence

Microsoft, Google, and Meta have converged toward μR of 0.87–0.93, with Narrative scores of 8.3–10 across all three. The framework identifies these as Authentic AI firms — genuine capability paired with high market narrative — rather than Narrative Inflation cases, because their absolute Signal scores are equally high (Figure 1).

4. Narrative-Signal Divergence at the Tail

C3.ai (μR = 0.27) and Palantir (μR = 0.37) show the sharpest divergence between Signal and Narrative in the dataset. Both achieve maximum Narrative scores (N=10) against Signal scores of 2.7 and 3.7 respectively — the clearest cases of Narrative Inflation in the 2026 market sample, with full inter-scorer consensus (Δ=0) on both firms.

5. Inter-Scorer Consensus at Extremes

As variance patterns in Figure 3 indicate, the framework produces its strongest consensus (Δ=0) at the market extremes: Box, C3.ai, Palantir, and Microsoft all achieve full Signal consensus. Disagreement is highest for Meta and Google (ΔS=2), reflecting genuine analytical disagreement about infrastructure depth — precisely the information asymmetry the framework was designed to surface.

6. The Recursive Finding

The three AI systems that built and applied this framework are themselves firms in the dataset they evaluated, each carrying a primary conflict of interest with exactly one company. The fact that those conflicts are disclosed, quantified, and separated into variance statistics is itself a demonstration of the transparency the framework recommends for all AI market participants. The systems evaluating the market are part of the market they evaluate — a form of reflexivity that connects information theory, market dynamics, and AI governance in a single analytical exercise.

REFERENCES

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *NBER Working Paper No. 31161.*

DeLong, J. B., & Magin, K. (2006). The U.S. equity return premium. *Journal of Economic Perspectives, 19*(4), 3–20.

Perez, C. (2002). *Technological Revolutions and Financial Capital.* Edward Elgar Publishing.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Shiller, R. J. (2019). *Narrative Economics.* Princeton University Press.

Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*(3), 355–374.

U.S. Securities and Exchange Commission (2024). Guidance on AI-related disclosures. SEC Release No. 34-101718.

Champaign Magazine (2026, Mar 9–13). Aikipedia: AI Premium; AI-Washing; AI Signal. champaignmagazine.com

Champaign Magazine (2026, Mar 17). The AI Credibility Triangle: A Framework for Interpreting AI Markets. Authors: Gemini, ChatGPT, Claude, W.H.L. champaignmagazine.com