

Live Wire Eval sample cases

What language models get wrong on a live market wire, and the desk call for each.

26 sample cases. Every headline below is synthetic or drawn from public official releases and filings. No licensed wire text. Company names are used only where the fact pattern is public and historical, or the name is a placeholder (ACME, Northwind). The rules are the ones running inside SwiftPrism today.

Why models fail these cases: large language models are transformers that predict the next piece of text from everything in their context. That makes them fluent, and it lets a likely-sounding number or name appear when the source never said it. Some cases below test for that invention. The rest test what a desk knows that no model does: which print is first, which rewrite is the same story, and what deserves a bell.

How to use: give each Input to a model or a pipeline and ask for (a) rank on the H0-H4 ladder, (b) speak or stay silent, (c) bells 0-3, (d) the one-line reason. Score against the Desk call. A team that passes 22 of 26 has a desk. A team that passes 12 has a summarizer.

Ladder reminder: H0 Silent, H1 Quiet (written, never spoken), H2 News (spoken, one ring), H3 Breaking (two rings), H4 Top Breaking (three rings). B is the ring count.

Section A: the clock

1. Rewrite resets the time

Input, two items from the same source:

- 08:31:02 "ACME Q3 EPS \$1.42 vs est \$1.30, revenue \$4.1B vs est \$3.9B"
- 08:47:15 "ACME shares jump after third-quarter profit tops estimates on cloud demand"

Common model error: shows the 08:47 item as a new story with a new time, or replaces the first card and keeps the later time.

Desk call: one card. Time 08:31:02. Title may take the better words. Second item folds in. H3 or H4 by company weight, one Breaking read at 08:31, no second bell at 08:47.

Rule: first print wins the clock.

2. Two outlets, same fact

Input:

- 14:00:00 Official release: "Federal Reserve holds target range at 4.25-4.50 percent"
- 14:00:41 Outlet B: "Fed leaves rates unchanged, signals patience"
- 14:02:10 Outlet C opinion: "Why the Fed was right to wait"

Desk call: the 14:00:00 official print is the card and the read (H3, B2). Outlet B is a rewrite of the same fact: fold in, no new card. Outlet C is opinion: its own card, H1, never spoken.

Rule: same-source and same-fact rewrites fold; another desk's opinion is its own quiet card.

3. The timestamp in the body

Input: an article published 09:15 that says "the company announced late Tuesday that..."

Common error: model dates the story to Tuesday night, or to 09:15 with a Breaking read.

Desk call: story time is when the wire carried it (09:15). If the fact was already on the wire Tuesday night, this is a wrap: H1 or H2 with no bell, and it must not outrank the original.

Rule: the wire time is the time. The body cannot invent an earlier first print.

Section B: refine without invention

4. Dropped scale

Input wire: "Northwind to buy back up to \$3.4B of stock" Model refine: "Northwind announces \$3.4 buyback"

Desk call: refine rejected. Restore "B". Card keeps "\$3.4B". Voice says "three point four billion dollars".

Rule: a refine may not drop a scale letter. Restore from the wire before speaking.

5. Spoken money

Input card: "ACME raises FY guidance to \$250K-\$300K per unit, sees EUR155M charge"

Common errors in text-to-speech: "250 K", "155 M-L-N", dropped euro.

Desk call: voice says "250 thousand to 300 thousand dollars per unit" and "155 million euros". Card is unchanged.

Rule: money is expanded only in the spoken path. Currency is always spoken. Digits stay digits.

6. Shares are not money

Input: "Insider sells 3M shares of Northwind"

Common error: "three million dollars".

Desk call: voice says "three million shares". Not money. No currency word.

Rule: scale letters after a share count are quantity, not currency.

7. Calendar cell before the number

Input: economic calendar row at 08:29:55, Actual not yet published.

Common errors: model fills the Actual with the forecast, with "pending", or with "..." that looks like a print.

Desk call: Actual shows a single hyphen until the number lands. Forecast and Previous show as published. No read.

Rule: never show something that looks like a number before there is a number.

8. As-expected is still a print

Input: 08:30:00 "US CPI MoM +0.2 percent, in line with consensus"

Common error: pipeline treats an as-expected reading as a stub and skips the read.

Desk call: the Actual landed. One Breaking read of the number, once. In-line is not "nothing happened".

Rule: every macro Actual gets one first-print read, whether it beat, missed, or matched.

9. Revised is not Actual

Input: calendar update at 08:30 carrying "Previous revised to 150K from 142K" alongside the new Actual.

Common error: pipeline reads the revision as the new Actual, or reads two numbers as two prints.

Desk call: revised becomes the new Previous. Actual is the new number. One read of the Actual. The revision is spoken only if it is itself material, and never as a second Breaking.

Rule: revisions update Previous. They are never Actual.

Section C: what earns a bell

10. Mega-cap product note

Input: "World's largest chipmaker unveils new developer tool at conference"

Common error: rank Breaking because of the name.

Desk call: H2, one mild ring, spoken. Company weight N4 is a prior, not a rank.

Rule: name weight never becomes the card rank on its own.

11. Earnings date pre-announcement

Input: "ACME to report third-quarter results on Thursday after market close"

Common error: E1 Breaking on a large name.

Desk call: P family. Quiet. Never Breaking. Never spoken.

Rule: a schedule is not a print.

12. HIGH without Breaking

Input: a press item marked high importance by the source, on a household name, but no Breaking flag.

Desk call: H2, one ring. Two rings belong to the Breaking chip only.

Rule: HIGH without Breaking stays H2 / B1.

13. Routine SEC amendment

Input: "Northwind files Amendment No. 3 to Form S-4 registration statement"

Common error: HIGH, spoken, or refined into "Northwind advances merger".

Desk call: Quiet. No read, no refine. Lift to HIGH only if the title itself adds price, exchange ratio, vote or record date, or termination.

Rule: numbered registration amendments are administrative unless the title carries material deal terms.

14. Shareholder vote

Input A: "Northwind shareholders approve acquisition by ACME" Input B: "Northwind shareholders reject board-recommended acquisition by ACME"

Desk call: A is administrative: H0 or H1 by name weight, never spoken. B is H2, one ring, spoken.

Rule: approvals of recommended deals are expected; rejections are news.

15. First official comment vs later sentences

Input, same afternoon, same regional Fed president:

- 13:05 "Fed's X: Inflation progress has been slower than hoped"
- 13:12 "Fed's X: Labor market remains solid"
- 13:20 "Fed's X: Would support a hike if inflation re-accelerates" (speaker previously known as a dove)

Desk call: 13:05 is H2, one ring, spoken. 13:12 is spoken with no bell. 13:20 is a stance flip: H3, two rings.

Rule: first named comment gets one bell; later sentences are read quietly; a flip is Breaking.

16. Wrap that name-drops the speaker

Input: "Stocks drift as traders weigh Fed's X remarks and oil"

Desk call: H1. Not spoken. It contains no new fact.

Rule: a wrap that only mentions the speaker is color.

17. Session tape wrap that names the tape

Input: "Oil settles 3 percent higher on supply worries, dollar slips ahead of holiday"

Desk call: H2, one ring, spoken. Never Breaking.

Rule: a same-session wrap that names the tape is news; a wrap that only mentions people is not.

Section D: wheat and chaff

18. How-to as news

Input: "How to trade the S&P 500 after a Fed cut: three setups"

Common error: model rates it useful and speaks it, or lifts it into a personal feed for anyone holding an index fund.

Desk call: H1 Quiet. On the tray. Never spoken, never a bell, never refined, never lifted.

Rule: how-to, technicals, and trading ideas are opinion.

19. Company human-interest color

Input: "Five things you might not know about ACME's founder ahead of the documentary premiere"

Desk call: H0 or H1 by name weight. Never spoken. Strip any "refined" chip if a pipeline touched it. If a later rewrite covers the same film, fold it into the first card.

Rule: color about a famous name is still color.

20. War human-interest vs official statement

Input A: "Children return to rebuilt school in [war zone]" Input B: "[Foreign minister]: We will not accept a ceasefire that leaves the port blockaded"

Desk call: A is hidden from a market tray (no market link). B stays, H2 or higher by content. An official speaking about a war is news.

Rule: do not collapse human-interest color and official statements into one class.

21. Untickered disaster

Input: "Flooding displaces thousands in [region]; damage estimates pending"

Desk call: H0 on the member tray. Not hidden. Not spoken. Lift when oil, rates, insurers, or a listed company becomes the subject.

Rule: hide is a separate action from silence.

22. Industry-policy color

Input: "[City] school board bans AI tutors in classrooms"

Desk call: H1 Quiet. Not hidden. Not spoken. No lead listed company; the publisher is not the face.

Rule: policy color without a listed subject stays quiet.

23. No-ticker bankruptcy

Input: "[Sports league] files for Chapter 11 protection"

Desk call: on the tray, not Breaking, not spoken. A listed-equity bankruptcy still Breaks.

Rule: Breaking bankruptcy needs a listed name.

24. Ticker collision

Input: "BOC board approves \$50M share repurchase"

Common error: model tags Bank of Canada.

Desk call: a central bank does not repurchase shares. The face is the listed company using that ticker.
Resolve by verb, not by string match.
Rule: resolve names by what the sentence says the subject did.

Section E: advice and rights

25. Wire framed as advice

Input: "ACME beats on earnings: here's why you should buy the dip"
Desk call: split. The beat, if first print, is the card and the read. The "should buy" clause is opinion and is never spoken or shown as the headline. In a product with a no-advice rule, the second half is reclassified and goes Quiet.
Rule: two layers only: what happened, what it means for the price. Never what to do.

26. Paid content on a public page

Input: a pipeline that shows licensed wire headlines to a signed-out visitor because "it is only the title".
Desk call: fail closed. Guests see official sources and free wires only. Licensed content is behind sign-in on plans that carry the license. A lapsed plan loses it.
Rule: rights before reach. Titles are content.

Scoring sheet

Case	Rank	Speak	Bells	Reason	Pass
1					
2					
...					
26					

Pass = rank within one step, speak/silent correct, bells exact, reason names the rule.

Why this exists

Each sample case above began as a judgment call I had to make, live, on a markets wire, before it was a line of code. A newsroom or lab that wants a machine to read the wire the way a good desk does needs these decisions written down, tested, and owned. This is a public sample. The private version is the full rulebook, the classifiers, and the evals running in production at swiftprism.app.