

ANALYZING EARNINGS CALLS

Kerry Back

Thu Sep 17, 2026

Today

- 1 Where earnings call transcripts come from, and what they cost
- 2 A question: can a model read a call and tell you to buy?
- 3 How fast you would have to act on the answer
- 4 Directions we could take it

WHERE THE DATA IS

Transcripts, Free

Motley Fool

The archive, scraped. 2021 to now, roughly 45,000 calls. Prepared remarks and Q&A in full.

EarningsCall

Clean API, good SDK. The free tier is two companies: Apple and Microsoft.

SEC EDGAR

8-K Item 2.02 is the press release, not the call. Public domain, timestamped to the second.

The Fool archive paginates at 50 pages, about three weeks. The monthly sitemap has no cap and is how you reach 2021.

Transcripts, Paid

EarningsCall

9,000+ companies, speaker-segmented, MIT-licensed SDK

Financial Modeling Prep

transcripts on a paid tier; you may already hold a key

API Ninjas

one endpoint, history to 2000, premium below 2026

Capital IQ via WRDS

the research standard, 2004 on, ~5,000 US firms

Rice does not subscribe to Capital IQ Transcripts. `ciq_transcripts` returns permission denied; the sample is 20 calls from 2005.

FREE AND CURRENT

Yahoo for returns, Finnhub for consensus. Current to yesterday. Runs in your container.

RESEARCH GRADE, BUT STALE

CRSP daily ends 2025-12-31. The market index file ends 2024-12-31. IBES ends 2026-05-14.

WRDS cannot label a call from last month. That is not a small detail; it decides which sample you can study.

The Other Half: Labels

THE QUESTION

Can a Model Read a Call and Tell You to Buy?

A transcript is 43,000 characters of management telling you how the quarter went. Somewhere in there is information about what the stock is worth.

The question is not whether a model can summarize it. It obviously can. The question is whether anything it produces survives contact with a price.

How Fast Would We Have to Act?

The call is public the moment it ends. Our transcripts are not.

4 days

MEDIAN LAG, CALL TO FOOL
TRANSCRIPT

53%

POSTED 3+ DAYS AFTER THE
CALL

5 min

WHAT THE PROFESSIONALS
ALLOW THEMSELVES

Measured on 449 of our own transcripts. Anything that trades on the announcement itself is gone before we can read it.

Where the Move Actually Is

So the announcement day is out of reach. Is that where the money was?

	MEAN ABSOLUTE MOVE	IS IT PREDICTABLE?
Announcement day	8.1%	yes, partly. R-squared 0.05 on the surprise alone
Next 60 days	17.3%	not from the surprise. R-squared 0.001

Two thirds of the movement happens after the day we cannot trade. None of it that we know how to forecast. Size and predictability are different things.

POSSIBLE DIRECTIONS

Four Architectures

Fine-tune

Train a model on transcripts paired with the subsequent price move. End to end, no features.

Extract, then model

The LLM reads the call and emits numbers. A separate ML model turns those into a recommendation.

Embed

Turn the transcript into a vector. Use the vector as a feature list.

Just ask

One prompt, one answer. The thing all three of the above have to beat.

Fine-Tuning

Show a model thousands of transcripts and what the stock did next, and let it find its own features.

FOR

Nobody has to guess which features matter. The model is optimized for the actual target.

AGAINST

Expensive, hard to interpret, and easy to overfit. What did it learn? You cannot read it off.

Extract, Then Model

The model scores the call on dimensions you define. A regression turns those scores into a position.

Guidance revision, margin durability, how much management hedges, whether the quarter leans on one-off items.

The appeal is that the middle layer is inspectable. When it works you can say which dimension carried it; when it fails you can see where.

Skip the judgment. Turn the whole transcript into a few hundred numbers and regress on those.

Cheap

Orders of magnitude less than generation.
You can embed the whole corpus.

Opaque

Dimension 137 is not “margin pressure.”
There is nothing to put on a slide.

Embeddings

Just Ask It

Give the model the transcript and ask for a number. No features, no fitting, no training.

This is the control, and it is the one experiment nobody wants to run. If a one-line prompt does as well as a fine-tuned model, that is the result, and it is worth knowing before you spend a week on the alternative.

Three More Angles

Remarks vs Q&A

The prepared remarks are scripted. The Q&A is not. Split the document and see which half carries the signal.

Change, not level

Some CEOs always sound confident. Compare a call to the same firm's last call and score what moved.

Model as analyst

Ask it what to look for rather than what to conclude. Then test its hypotheses yourself.

Preliminary, one month, 181 to 198 events. Enough to steer you, not enough to believe.

Announcement-day reaction

ten extracted features added R-squared 0.096 over the surprise, $p = 0.007$

60-day drift

nothing. Would need roughly 80x the sample to see a 2% effect

Volatility, level

64% explained by its own past. No room

Volatility, change

benchmark only 0.058, and the obvious predictors are exhausted

What I Already Know

Your Turn

- 1 Pick a direction. Argue for it before you start building.
- 2 Decide what your prediction would have to beat, and write that down first.
- 3 Build the smallest version that could produce a number.
- 4 Tuesday: what you tried, what it gave you, and whether you believe it.

The infrastructure is in the repo: transcripts, extraction, labels, scoring. Take what helps and ignore the rest.