
Identifying Regime Shifts with Language Models

Are language models able to identify regime shifts before they are reflected in share prices?

Alistair Smallwood, Head of Applied AI at Primer

ABSTRACT

A regime shift is a lasting change in how the market values a company: a re-rating, in either direction, that rewrites what investors believe about its growth, durability and category. Regime shifts are rare but dominate outcomes. Over the thirteen months studied, the tenth of 3,464 listed companies with the largest abnormal moves accounted for 33% of all absolute abnormal return in the universe. Investors do not find them by forecasting the next quarter. They find them by seeing a mechanism the market has not yet priced and building conviction as evidence arrives.

We asked whether a language model can spot the reason a company might be re-rated before it happens. We gave Claude Opus 4.5 only what an investor could have read on 1 December 2025 (results, call transcripts, guidance, analyst estimates and the share price) and asked it to write the bull and bear case for each company. The model's own knowledge ends before that date, so it could not have known what came next. Across 65 companies whose shares then re-rated, the model highlighted the real reason for the re-rating 71% of the time, and at least part of it 97% of the time. To check that the model was not just writing generic stories that would fit any company, we also compared each thesis with the reason a different company re-rated: they matched only 1% of the time.

The model could also, more often than not, tell which companies would move and which way. Shown a company that went on to move 40% or more alongside one that did not, it picked the mover 66% of the time, and for the companies that moved it leaned the right way 66% of the time (64% of risers and 67% of fallers). Put together, for 55% of the companies that re-rated, it named the real reason and leaned the right way, months before the move. Simply picking volatile shares can match it at spotting which companies will move, but says nothing about which way or why, and that is what a fundamental investor trades on. The model never made a regime shift its single most likely outcome, but investors do not need it to. If a share does not move, that is rarely a disaster; what they look for is a skewed range of outcomes, a fat tail up or down that the share price does not yet reflect. Nor is a regime shift called at a single point in time: an investor spots something that could lead to one, then keeps gathering and monitoring evidence, building conviction and sizing the position as it accumulates. The wider system we describe would do the same. Correctly identifying potential regime shifts is the foundation, and our system already does this well; monitoring them effectively is what turns an idea into an autonomous signal.

1 Introduction

Most of the time, a company's value changes for familiar reasons: a quarter beats or misses, estimates move, and the share price follows. Occasionally something different happens. Investors stop asking how much the company will earn next year and start asking what kind of company it is. Meta in 2022, Nokia in 2008 and Microsoft after 2014 were not stories about a quarter. In each case the market changed its mind about the length, durability or nature of the company's earnings, and the valuation moved far more than the numbers did.

These regime shifts matter disproportionately to investors. They produce the largest gains and losses in a portfolio, they are where active managers either earn their fees or lose them, and they are hard to see in time because, by definition, the evidence is incomplete when they begin. A system that could

identify them early, or even narrow the field of candidates, would be valuable to any fundamental investor.

This paper tests what a language model can contribute. It reads the same kind of material an analyst reads (results, transcripts, guidance, consensus and price history), as it stood on a given date, and writes a thesis about whether and how the company might be re-rated. We then compare those theses with what happened over the following ten months. The design follows our earlier work on capital markets days, where the model anticipated the headline target in nearly three cases out of four but was weaker on market reaction. A similar pattern appears here: the model is strongest at naming what could happen, and its view of which way needs building into conviction over time.

2 What a regime shift is, and why it matters

A definition. We define a regime shift as a lasting change in how the market values a company, visible in two things at once: a share-price move that is large relative to the company's own history and its market, and a revision in the same direction to estimates two years out. The second condition separates a re-rating rooted in the business from a sentiment swing. The first, measured relative to the company's normal volatility, separates a genuine change from a volatile stock doing what volatile stocks do.

How investors recognise them. Experienced investors rarely call a regime shift from the numbers alone. They look for a mechanism: a new source of demand, a product cycle, a cost curve crossing, a competitor or regulator that undermines pricing power, a management team that has lost or regained credibility. They then ask how the market would re-file the company if the mechanism plays out. A growth company re-filed as ex-growth, a hardware maker re-filed as AI infrastructure, a monopoly re-filed as a utility. The valuation multiple moves with the category, often far more than earnings. On the downside, markets find it hard to look past a lingering structural question even when the company might work its way out. If nobody can

prove the threat will not crush the business, the stock de-rates and stays there.

Why the payoff is out of distribution. Re-ratings are where the fat tails of equity returns come from. In our universe the median company moved 22% against its market over thirteen months, while the top decile moved more than 68%. That top decile accounted for a third of all absolute abnormal return; the top 1% alone for 6%. An investor can be right on direction barely more than half the time and still do well if the right calls include the large moves and the wrong ones exclude them. Missing a single 80% de-rating in a long book can erase a year of small wins. This is why identifying candidate regime shifts early, even imperfectly, is worth so much.

Why they are hard to spot. A regime shift cannot be proved at the time it begins. If the evidence were conclusive it would already be in the price. The analyst's skill lies in seeing the mechanism while it is still a hypothesis, judging whether it is plausible, and deciding how much evidence is needed before acting. That is a judgement under uncertainty, which is precisely where language models tend to hedge.

3 Research questions and scope

We ask three questions. First, when a company does undergo a regime shift, did the model's thesis, written months earlier, identify the mechanism that caused it? Second, can the model tell in advance which companies will re-rate, in which direction, and with what probability? Third, if it cannot do the second well, what kind of system would make its output useful?

Regime shifts are rare, so we started wide: a universe of about 3,500 listed companies above USD 1bn in market capitalisation, covered by Primer's document tools, large enough to yield a meaningful set of companies that actually re-rated. The original study used 15 regime shifts identified by a pre-registered

screen (10 up and 5 down) and 42 matched controls. A held-out test used 30 further companies with large, late share-price falls, each paired with a matched company that did not fall, none of them used in developing the system. A fresh-company test added 20 companies never used before (9 risers and 11 fallers, 11 of them calm), drawn by a fixed rule because their move was unusual for that company.

The paper makes no claim of investment returns. Its contribution is a careful measurement of what the model does and does not do, and an argument for how such systems should be assembled.

4 Out-of-sample design

Every company is assessed as of a fixed date, using a frozen model whose training data ends before that date and only information published before it (section 5 describes what the model receives). Documents whose own text shows a later filing date are excluded. Before forecasting, the model is asked what it remembers about each company after September 2025, and companies with any recalled later events are excluded; none were.

The model writes a structured thesis: the strongest case for a lasting deterioration, the strongest case for a lasting improvement, the behavioural drivers that would make investors re-value the company, the signposts that would confirm or kill

each view, and probabilities for an up-shift, a down-shift or neither. A skill pack, described in section 6, supplies regime-shift archetypes, a deterioration checklist, market-behaviour drivers and historical cases.

Outcomes are measured independently. For the reason-naming test, research agents wrote up the cause of each large move from evidence published after March 2026. Graders then compared each thesis with the actual cause without knowing which pairing was real. Company names were removed from both texts, and every thesis was also graded against a different company's cause as a control. All tests were specified before results were seen; analyses added later are labelled exploratory.

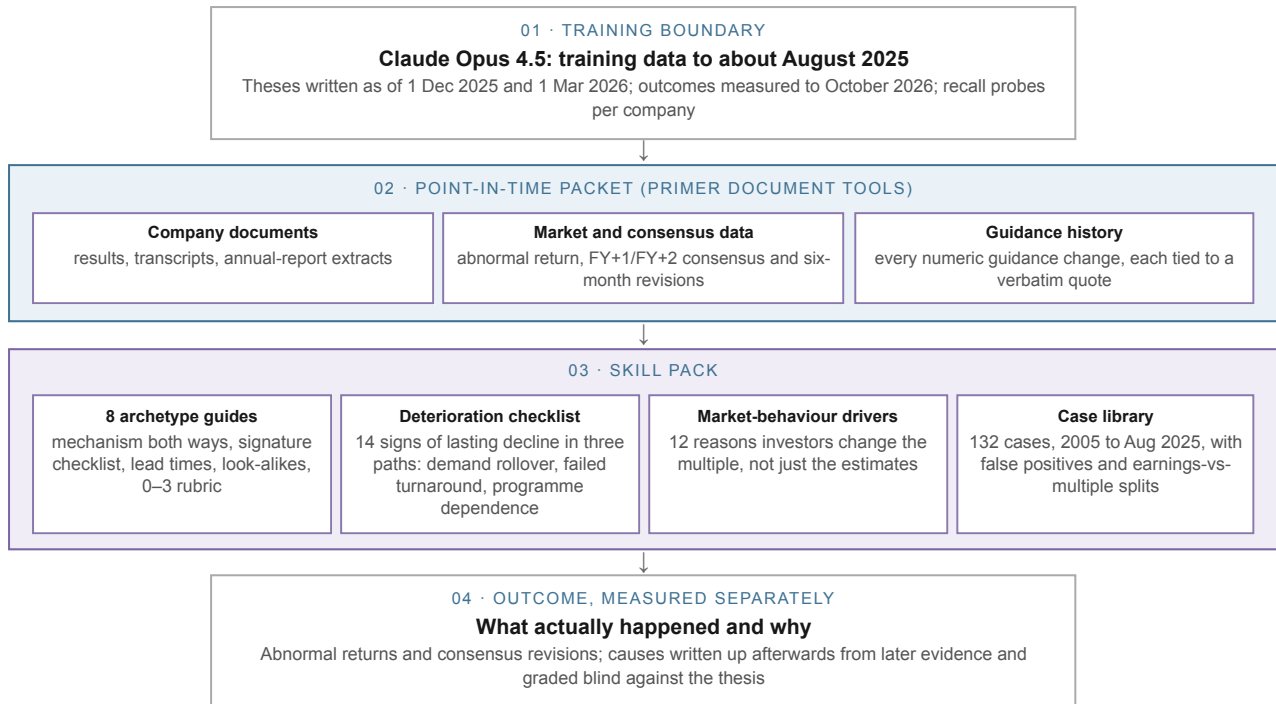


Figure 1. How each thesis is produced. The model, its information and the skill pack are fixed before the as-of date; outcomes are measured separately.

5 Building the case library

Why a case library. Before an agent can spot a regime shift, it has to know what one looks like. So the first job was to find actual regime shifts from the past, rebuild each one from the evidence available at the time, and study what they had in common: the mechanisms, the earliest signals, how long they took to show in the price, and the false starts that looked similar but were not. Those commonalities became the agent's education (section 6), which it then applies to a real sample of companies whose outcomes it cannot know.

1. The case library. We built a library of historical regime shifts from 2005 to August 2025, all resolved before the model's training data ends and none involving a test company. We used Primer's document tools for accurate and complete retrieval of what companies actually said at the time: results, calls and filings as originally published, rather than the model's memory of them. Where a document was not available, the gap was flagged rather than filled. The library began as 97 regime-shift cases across eight recurring mechanisms, deliberately including false positives: companies where a plausible shift story was told and did not happen. Each case records the contrarian thesis, the earliest observable signals, the lead time to the main price move and, for false positives, the tell that separated them. Seventeen deterioration cases, thirteen genuine declines and four false alarms, were added after the first version proved too ready to see improvement.

2. Splitting price into earnings and multiple. We then rebuilt every case around the question investors care about:

how much of the share-price move came from earnings, and how much from the multiple the market was willing to pay? Research agents measured the price change and the change in consensus earnings two years out (delivered earnings before 2015, where consensus history is thin); the re-rating is the price change net of the earnings change. For each case they recorded the investor story before and after, how long the multiple took to move, whether it later reverted and what finally settled the debate. Eighteen cases were added to capture investor behaviour: structural de-ratings such as TomTom, TeamViewer, Pearson and BT; bear cases that broke, such as Philip Morris and Alphabet; and reframes such as Microsoft and Ferrari. The research worked to a strict information boundary of 1 September 2025, and any case involving a test company was removed before the skills were written.

Running on a real sample. When the system then assesses a company, the agent receives the skill pack and a packet of evidence about that company as it stood on the as-of date: its results releases, call transcripts and annual-report sections; the share price and abnormal return to that date; consensus estimates for the next two years with their six-month revision path; and a history of its numeric guidance. The documents come from the same Primer tools, with full text and publication dates, so the packet is cut exactly at the as-of date and nothing written later leaks in. Each guided number is copied with its exact sentence, a script checks that the sentence exists in the source, and changes between reporting events are computed by script rather than by the model.

6 Building the skills and agent understanding

For the agent to understand regime shifts, rather than simply summarise a company, we built custom skills from the case library. A separate drafting agent, which never saw the test companies, wrote them into a skill pack of about 55,000 tokens that the agent reads before every assessment. There are five.

Archetype guides. Eight recurring mechanisms: business-model transition, product-cycle inflection, new demand theme, cost-curve crossover, supply discipline, consumer-behaviour shift, management and capital-allocation reset, and regulatory or policy break. Each guide gives a definition; the mechanism in both directions, in the order it usually unfolds; a signature checklist with typical lead times to the main price move; what consensus usually gets wrong; look-alikes and how to tell them apart; and a 0–3 scoring rubric with caps, so that a zero on certain items limits conviction whatever the total.

Deterioration checklist. Fourteen signs of a lasting decline, each scored 0–3 with a rubric: profit beating while sales miss; repeated guidance cuts; rising promotions or markdowns; inventory growing faster than sales; share loss to named challengers; falling cash conversion; balance-sheet or dividend strain; a reset year that keeps moving; metrics dropped from disclosure; hedging language on calls; dependence on a policy or programme that may not fund on time; a cyclical turn in end-markets; an inflated base; and a turnaround that relies on cost cuts alone. They are organised into three paths (demand rolling over, a turnaround that fails, dependence on something

outside the company's control), with look-alikes that include bear cases that were wrong.

Market-behaviour drivers. Twelve reasons investors change the multiple rather than just the estimates: an unprovable threat; re-filing into a different category; a switch of yardstick; credibility lost or rebuilt; a story premium on unproven adoption; earnings the market will not capitalise; a balance sheet that turns the story into one about solvency; a change of shareholder base; read-across from a theme or peer; a complexity discount; policy setting the terminal value; and a cash-machine reframe. Each has a direction, a 0–3 rubric, a typical duration and the cases that show it, together with tests for estimate changes that are not re-rated and bear cases that did not stick.

Case digests. The library itself, in three digests: 97 regime shifts and false positives, 17 deteriorations and false alarms, and 126 market-behaviour records with the earnings-versus-multiple split. The agent compares each company with its closest precedents and with the closest false cases, asking which tell the false case showed and whether this company shows it.

Working order. The pack also tells the agent how to work: read the numbers before the narrative, score the deterioration checklist for every company, write the bear case before the bull case and read each through the behavioural drivers, ask whether a change would be a re-rating or only an estimate revision, compare with precedents, and only then give probabilities.

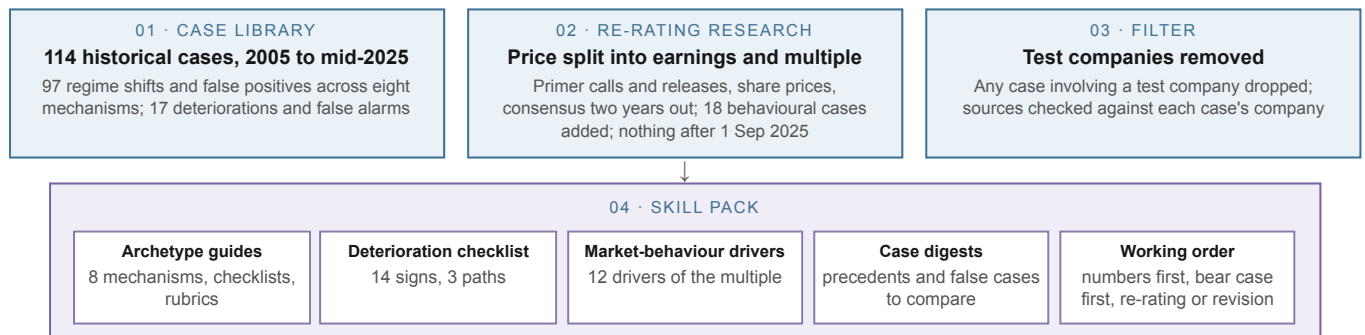


Figure 2. How the skill pack was built. Every case ends before the model's training data does, and a drafting agent that never saw the test companies wrote the skills.

7 The model names the mechanism that leads to a re-rating

The headline. We took 65 companies whose shares later re-rated, from three sets: the 15 regime shifts in the original study, 30 sharp fallers held out from development, and 20 fresh companies, including calm ones, whose move was unusual for that company. For each we asked one question: did the model's thesis, written before the move, describe the real reason for it? Across 285 versions of the thesis (five per company in the first two sets, three in the fresh set), the model highlighted the real reason for the re-rating 71% of the time, and at least part of it 97% of the time. For 48 of 65 companies, most of the model's versions got it right.

The check. A thesis could look right just by being vague: 'competition could squeeze margins' fits almost any company. To test for this, graders also compared each thesis with the reason a *different* company moved, without knowing which pairing was real, and with company names removed. Those mismatched pairs matched 1% of the time (17% partly). The model's stories are specific to each company, not generic stories that fit anything.

What it looks like. For Fair Isaac, the model wrote that the housing regulator could let lenders choose their credit score,

that VantageScore would win share in mortgages, and that investors would re-file the company 'from monopoly to utility'. That is what happened ten months later. For Celsius it identified a stalling core brand hidden behind acquisition-led growth; for Charter, broadband customers lost to fixed wireless and fibre; for Builders FirstSource, a housing downturn amplified by high fixed costs. The appendix works through three examples in full: Fair Isaac, Nike and Boston Scientific.

Each set tells the same story. The original shifts scored 67%, the held-out fallers 74% and the fresh companies 70%. The fresh set matters most. The first two were almost all volatile stocks (41 of 45), so the fresh set was drawn to include calm, large companies such as Boston Scientific and Hermès, and risers. Calm companies scored 70% and risers 74%.

What this does and does not show. The model writes both a bear case and a bull case for every company, including companies where nothing happened. So this test shows that the right story was on the model's list. It does not show that the model picked that story as the likely outcome; section 8 tests that.

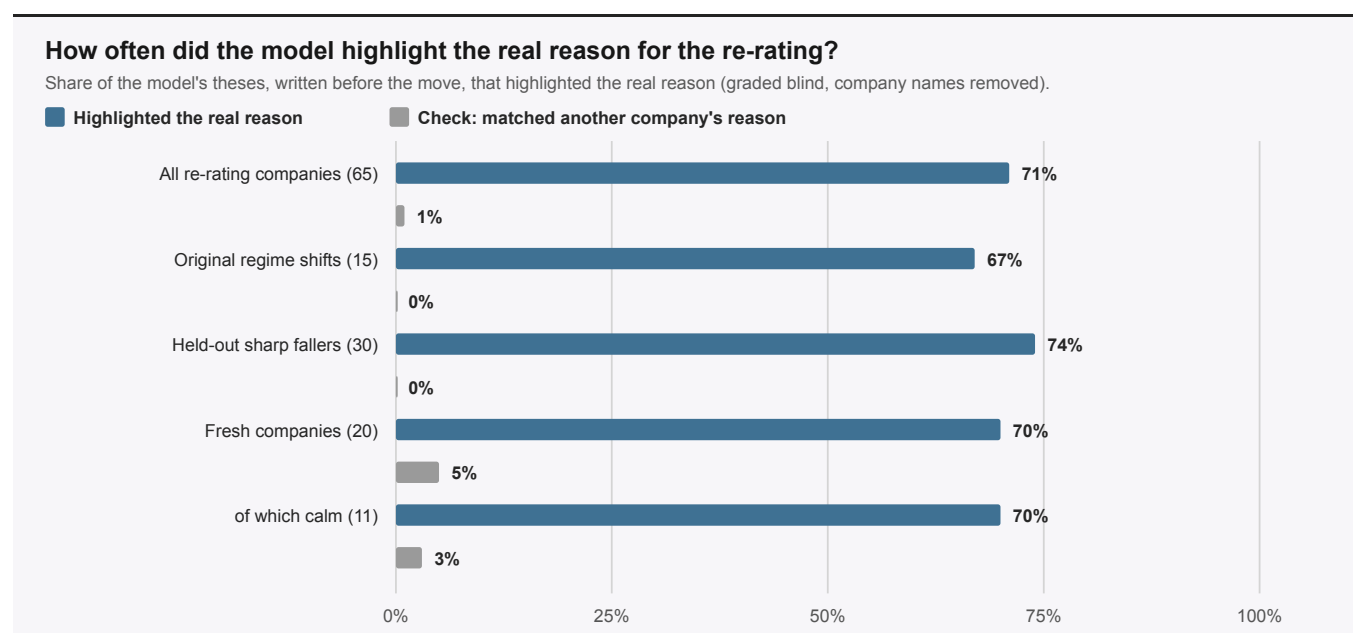


Figure 3. Across 65 companies that went on to re-rate, the model highlighted the real reason for the re-rating about seven times in ten, and each set of companies gives a similar answer. Compared with a different company's reason, the same stories almost never matched, so they are specific to each company rather than generic.

8 Can it tell which companies will move, and which way?

Two questions. Naming the right story is not the same as knowing which companies will re-rate. For every company the model also gave a probability of a rise, a fall and no shift, so we asked two questions. First, shown a company that went on to move 40% or more against the market alongside one that did not, could the model pick the mover? Second, for the companies that did move, did it get the direction right?

The model never committed. Across 645 forecasts its top answer was no shift 641 times; it never made a rise its top answer. Its probabilities sat in a narrow band, 7 to 25% for a rise and 8 to 33% for a fall, so the signal lies in how it ranks companies rather than in any single call.

Picking the mover. Each probability picked the company that went on to move about two times in three: 66 in 100 for a fall, 66 for a rise and 65 for no shift (Figure 4).

Getting the direction right. Among the 64 companies that moved 40% or more, the model leaned the right way, with a higher probability of a rise than a fall for risers and the reverse for fallers, in 42 cases: 14 of 22 risers and 28 of 42 fallers (Figure 5). Always guessing a fall would score similarly on this mix

of companies but get every riser wrong; the model got about two in three right in each direction.

Trying to fix the downside. We knew from the start that the model read the upside better than the downside, so most of our effort went into falls: the deterioration checklist, the behavioural drivers and a warning-signs score built from them. On the held-out test of 30 sharp fallers, that score picked the faller 65 times in 100, ahead of the model's own probability and of analysts' estimate changes.

Putting it together. For 36 of the 65 companies that re-rated, the system both named the real reason and leaned the right way: 11 of 19 risers and 25 of 46 fallers. A story written for another company matched the real reason for only one company in 65, so getting both right by luck is close to impossible.

How far to trust this. These two checks were run after the main tests, on companies chosen to include many movers (65 of 137 moved 40% or more, against roughly one in ten in a normal universe), with a threshold we chose. In a full universe the same ranking would produce more false alarms for every hit. The edge is real enough to build on, but it needs confirming prospectively.

Shown a company that moved and one that did not, how often did it pick the mover?

All 137 companies; a move is 40% or more against the market, December 2025 to October 2026. Times in 100.

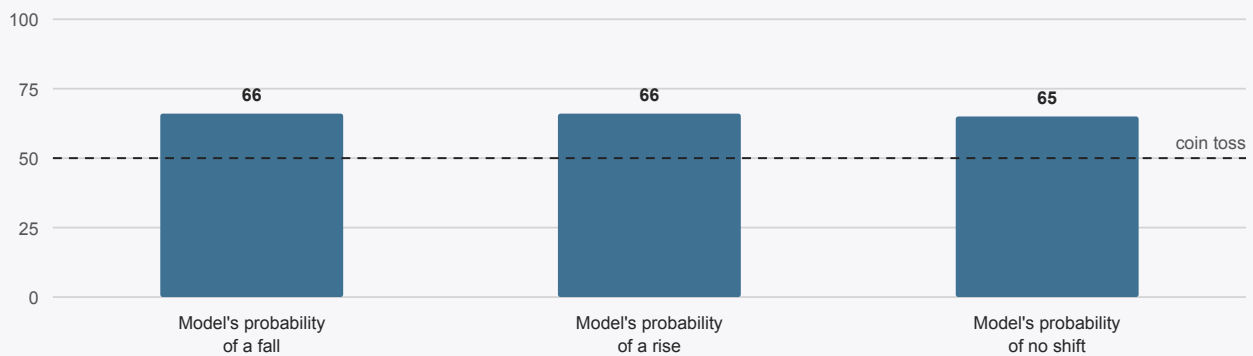


Figure 4. Each of the model's probabilities picks the company that went on to move about two times in three.

For companies that made a large move, did the model get the direction right?

A rise is called when the model's probability of a rise was higher than its probability of a fall. Per cent of companies.

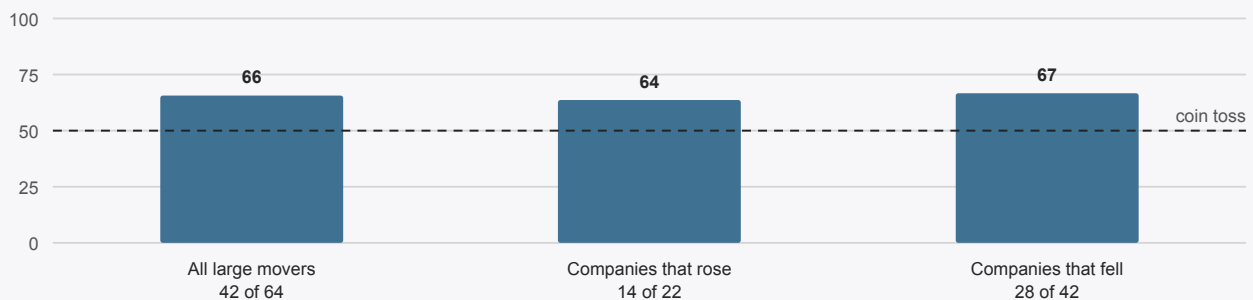


Figure 5. The model got the direction right for about two in three large movers, and about equally often for risers and fallers.

9 What did not make the model choose better

We tried hard to make the model choose better, and each attempt was tested on companies with known outcomes before being adopted. None worked well enough to adopt:

Better prompts and richer skills. Four versions of the instructions and a 55,000-token skill pack of archetypes, behavioural drivers and 132 historical cases improved the quality of the stories but not the choice of companies. A blind audit confirmed the model applied the behavioural drivers correctly; it simply did not translate them into discrimination.

A defined multi-agent research process. Separate forensic, outside-in (competitors, customers and suppliers) and market-behaviour analysts fed an evidence ledger, followed by analogues, industry context, a red team and three judges. The judges still passed or took small positions (15 pass, 14 short, 1 long of 30 judgements; no size above small), and the ranking did not improve. This matches published evidence that debate

among agents with the same information converges rather than discovers.

Monitoring the company's own news. Feeding the company's subsequent announcements back to the model moved its probabilities the wrong way: the change pointed to the eventual fallers only 23 times in 100, worse than a coin toss. Companies heading for a fall tended to sound reassuring; companies already in difficulty disclosed it openly.

Expected value. Asking for sized scenarios (how a story transmits to earnings and the multiple, with probabilities) produced sensible tables but expected values unrelated to outcomes (rank correlation -0.06).

Payoff tests. On the companies tested, flagged names were right on direction no more often than the sample's base rate, although flagged fallers fell somewhat further than unflagged ones.

10 How language models can do regime shifts

What the tests show. The model is good at the first job an analyst does: spotting why a company might be re-rated, and which way. It is not good enough, from a single reading, at the second: deciding which of those ideas will actually play out, and when.

How investors close that gap. They rarely decide at one moment. They spot an idea, then watch for evidence: a competitor reports weak demand, a customer switches supplier, a regulator publishes a consultation, guidance is cut a second time. Conviction, and the size of the position, builds as the evidence arrives. But you need the hypothesis first: without it, there is nothing to watch for. That is what the model provides, and why its thesis is useful.

The full system (Figure 6). An agent scans every company all the time and hands over each candidate: why it could re-rate, which way, how big the move could be and what would prove or disprove it. Monitoring then tracks the evidence from three places: the company's customers, suppliers and competi-

tors, read together as one story (*Mosaic Theory at Scale*); outside data such as shipments and prices (*Systematic Alternative Data Research with Language Models*); and the company's own results and share price. Each time new evidence arrives, the odds are updated. Candidates that build are flagged to the investor, unclear ones stay under watch, and disproved ones are removed.

Odds, not calls. The investor does not need the model to be certain. A candidate with a 17% chance of a 60% fall and a 15% chance of a 25% rise is already worth watching; the evidence that follows decides whether it becomes a position.

A timeline. For the investor, this looks like a news feed. When they log in, new candidates and updates to existing ones appear in one ranked list: a new idea at a company they own, a signpost confirmed by a supplier's results, a thesis weakened by a competitor's numbers. They can dig deeper or ask the system to keep watching. The system does the reading; the investor makes the call.

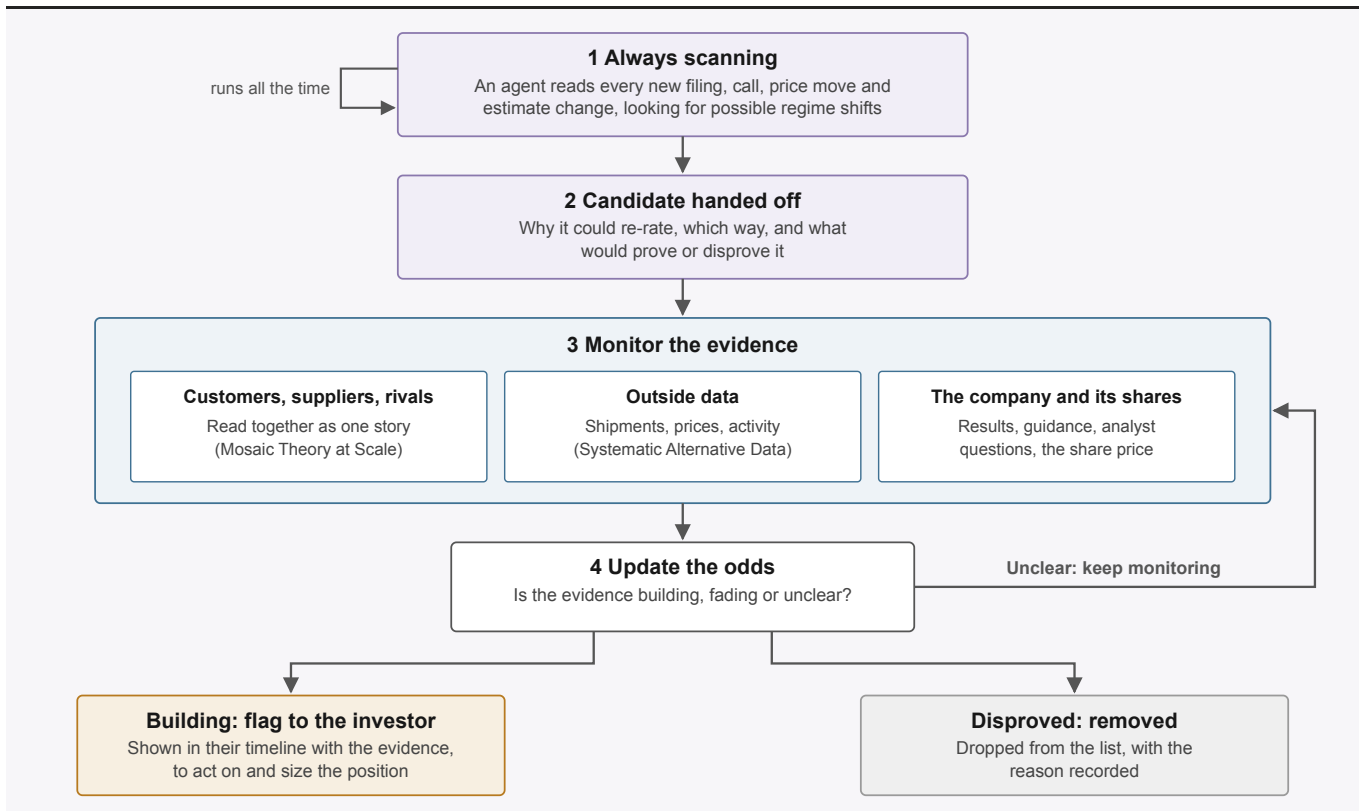


Figure 6. The full system. An agent scans every company all the time; each candidate it finds is monitored, its odds are updated as evidence arrives, and it is either flagged to the investor, kept under watch or removed once the evidence disproves it.

11 Limitations

One period, one model. All outcomes fall between September 2025 and October 2026, a period with its own themes (AI infrastructure demand, consumer weakness, an oil shock in March 2026) and one model generation.

Retrospective design. Every outcome postdates the model's training data, but the system was developed on some of the same companies. The held-out and fresh-company tests address this; results on the original companies are exploratory.

Machine grading. Causes were written up by research agents and graded by model graders under blind, name-

scrubbed conditions with a shuffled control. A human review of a sample of gradings is recommended before the results are relied on.

Small samples. Most large movers in the main test were volatile stocks. The fresh test checked calm companies directly (70% named), but on only 11 calm companies with three versions each.

No return claims. The paper measures identification and discrimination, not investment performance. Licensed consensus and price values are not reproduced.

12 Conclusion

Language models can do the part of regime-shift investing that is about imagination disciplined by evidence: reading a company's material and naming the mechanisms by which the market could come to value it very differently. Across 65 companies that re-rated, the model highlighted the real reason for the re-rating months before the move 71% of the time, against 1% for mismatched causes, and the result held for calm companies and for risers.

They can also, more often than not, tell which companies will move and which way: about two times in three on each count,

and for most companies that re-rated they named the reason and leaned the right way. What they do not yet do is commit. The model's probabilities stay cautious, and conviction in investing comes from evidence that accumulates after the idea, not from a single reading. The right design is therefore a system: identification that surfaces candidate regime shifts with their mechanisms and signposts, and monitoring that gathers the cross-company and outside evidence needed to build conviction over time.

R Evidence and references

1. Smallwood, A. (2026). *Forecasting Capital Markets Day Targets with Language Models*. Primer working paper.
2. Smallwood, A. (2026). *Mosaic Theory at Scale*. Primer working paper. <https://www.primerapp.com/blog/mosaic-theory-at-scale/>
3. Smallwood, A. (2026). *Systematic Alternative Data Research with Language Models*. Primer working paper. <https://www.primerapp.com/blog/systematic-alternative-data-research-with-language-models/>
4. Halawi, D. et al. (2024). Approaching human-level forecasting with language models. arXiv:2402.18563.
5. Schoenegger, P., Jones, C., Tetlock, P. E. and Mellers, B. (2025). Prompt engineering and LLM forecasting. arXiv:2506.01578.
6. Kim, A., Muhn, M. and Nikolaev, V. (2024). Financial statement analysis with large language models. arXiv:2407.17866.
7. Glasserman, P. and Lin, C. (2023). Assessing look-ahead bias in stock return predictions generated by GPT sentiment analysis. arXiv:2309.17322.
8. Majority voting versus multi-agent debate in LLM reasoning (2025). arXiv:2508.17536.
9. LLM forecasters hedge toward 0.5; statistical recalibration versus prompting (2025). arXiv:2511.07678.
10. Research records: projects/regime-shift-rediscovery (pre-registrations, STATE.md log, scripts and derived results).

Research status. Working paper, prepared 19 October 2026. All results are retrospective; every outcome postdates the forecasting model's training data. Main tests follow plans recorded before results were seen; analyses described as exploratory were added later. The paper makes no claim of investment returns. Licensed price and consensus values are not reproduced.

A1 Example 1: Fair Isaac, the fault line named, the lean wrong

COMPANY	MOVE	MODEL THESIS
Fair Isaac (credit scoring)	About -82% against the market	Regulator-driven loss of mortgage-score pricing power; re-filed from monopoly to utility

What happened. Fair Isaac's shares fell about 82% against the market. Regulators eroded its pricing power in mortgage credit scores: the government-sponsored enterprises began accepting VantageScore, and the housing regulator later said both scores would sit on the same pricing grid, threatening fees roughly ten times VantageScore's.

What the model wrote. In every sample, the bear case was this mechanism: the regulator implements lender's choice,

VantageScore gains share in conforming mortgages, Fair Isaac loses pricing power and investors re-file it from monopoly to utility. Graders marked all five samples as naming the cause.

What it got wrong. The model rated an up-shift as more likely than a down-shift (a fall at 8 to 15%, a rise at 15 to 25%). It saw the fault line clearly and judged that the franchise would hold.

A2 Example 2: Nike, the right story without conviction

COMPANY	MOVE	MODEL THESIS
Nike (footwear and apparel)	About -88% against the market	Over-promoted franchises, share loss, China reset, a turnaround that drags on

What happened. Nike's shares fell about 88% against the market and two-year earnings estimates were cut by 45%. A new management team chose to shrink its largest lifestyle franchises, cut discounted digital sales in China and launch a multi-year restructuring that pushed the margin recovery out to the end of the decade.

What the model wrote. That classic franchises had been over-promoted, competitors had taken share that would not

return, China required multi-year restructuring and the turnaround could prove another failed reorganisation. Graders marked one sample as naming the cause and four as partial: the model had the franchise and China mechanism but not the deliberate managed decline that extended it.

What it got wrong. Its probabilities were close to even in December and only tilted modestly toward a fall by March.

A3 Example 3: Boston Scientific, a calm company with the story named early

COMPANY	MOVE	MODEL THESIS
Boston Scientific (medical devices)	About -90% against the market; more than five times its normal volatility	Competitors in pulsed-field ablation and a Watchman slowdown de-rate a premium growth multiple

Why it matters. Boston Scientific is the kind of company the original test lacked: a calm, large company whose shares normally move less than most, valued at a premium for two growth franchises. Its fall was more than five times the size of its normal volatility.

What happened. Both franchises slowed together. In heart-rhythm treatment, competitors' pulsed-field ablation catheters took share from its FARAPULSE system faster than expected, and growth in Watchman, its stroke-prevention implant, fell from about 30% to low single digits as clinical evidence and referral patterns shifted. Guidance was cut twice in 2026, management warned that the headwinds would last into 2027, and the market stopped paying a premium growth multiple.

What the model wrote. In every version, the bear case was that FARAPULSE growth would slow as pulsed-field ablation matured, that Watchman growth could stall, and that the company would be re-filed 'from high-growth disruptor to mature medtech' at a lower multiple; two versions also named competitors taking share. Graders marked two of the three versions as naming the cause and one as partial. Every version named the next Watchman trial readout as the signpost to watch.

What it got wrong. It gave a lasting fall only a 10–15% chance and judged no regime shift the most likely outcome. As with Nike, the mechanism was right and the conviction was missing, which is the gap a monitoring system is designed to close.