Nanyang Technological University
Quantitative Research Society, Singapore

NTU-QRS AlphaAllocator

Bounded Transformer News Overlays for Chinese Macro-Asset Allocation

Yuhe Sui

SCAN · PROJECT PAGE

qrsntu.org/projects/alphaallocator

Includes a prompt for your AI agent

TASK

NLPCC 2026 Shared Task 4: Chinese fund allocation

RESULT

Track 1: 1st of 10 teams on the hidden 2026 test

PROJECT PAGE

Paper, figures, a live demo and an AI-agent prompt

QRS Quantitative Research Society

Problem setting

Task	Daily allocation over Chinese funds from historical prices and cutoff-visible financial news.
Universes	Track 1: 11 macro and broad-market instruments. Track 2: 16 sector instruments.
Inputs	59 prior closes, the current open for order sizing, and rank-sorted Top-20 news published before 15:00.
Evaluation	Organiser-run hidden 2026 B-list, ranked by Sharpe ratio; annualised return and maximum drawdown reported.

RESEARCH PROBLEM

How can financial news inform a portfolio without making it unstable or hard to audit?

The task joins a language problem, mapping text to tradable assets, with a portfolio problem in which risk, timing and executable orders matter.

QRS Quantitative Research Society

Motivation: decomposing the allocation decision

"Directly translating each headline into a large position is fragile: relevance is asset-dependent, tone is uncertain, and small daily text changes can produce unstable trading."

QUESTION	SINGLE-MODEL ALLOCATOR	DECOMPOSED DESIGN (OURS)
1 Which asset?	Large language model one prompt and one answer for all three questions	Relevance model BGE encoder, small and local
2 Good or bad?		Tone model Chinese FinBERT, small and local
3 How much?		Risk model text may tilt it by at most 1%

× errors cannot be attributed to a step
× output depends on prompt wording
× size is not tied to measured risk

✓ each component is evaluated separately
✓ local, deterministic models
✓ text influence is bounded in L1

QRS Quantitative Research Society

Key idea and contributions

KEY IDEA

Language models act as bounded signal modifiers, not as portfolio allocators: prices set the portfolio, and text may only tilt it.

CONTRIBUTIONS

1

A constrained allocation system

A light, local financial-text tilt around a shrinkage equal-risk-contribution core; ranked 1st of 10 in Track 1 on the hidden 2026 test.

2

A leakage-aware, bounded implementation

Text scoring, bounded fusion, planning and official execution under the decision-time cutoff, with a pre-planner stability guarantee (Proposition 1).

3

Controlled diagnostics

Same-stack public-2025 tests show that the incremental text contribution is real but small, and track-dependent.

QRS Quantitative Research Society

System architecture

Cutoff-safe official inputs
prior closes · current open, for sizing only · news published before 15:00

Quantitative base
Track 1: shrinkage equal-risk-contribution
Track 2: volatility–momentum–trend
output: base weights w

Text processing
Top-20 cleaning → BGE relevance → FinBERT tone → recency weighting
output: one score z per asset

Bounded text fusion
centred tilt $\Delta w = \eta \cdot (z - \text{mean } z)$, capped so that $\|\Delta w\|_1 \le \eta$, added to w
$\eta = 0.010$ in Track 1, 0.012 in Track 2

Portfolio normalisation
clip negative weights
scale to 97% invested

Planner constraints
≤ 35% requested per asset
≤ 25% one-way change

Official orders and fills
3% cash, 1% rebalance threshold
filled at the official close

News scoring and bounded fusion

PER ASSET, FROM THE DAY'S VISIBLE NEWS

R relevance: BGE cosine similarity, recency-weighted

S tone: FinBERT P(positive) − P(negative)

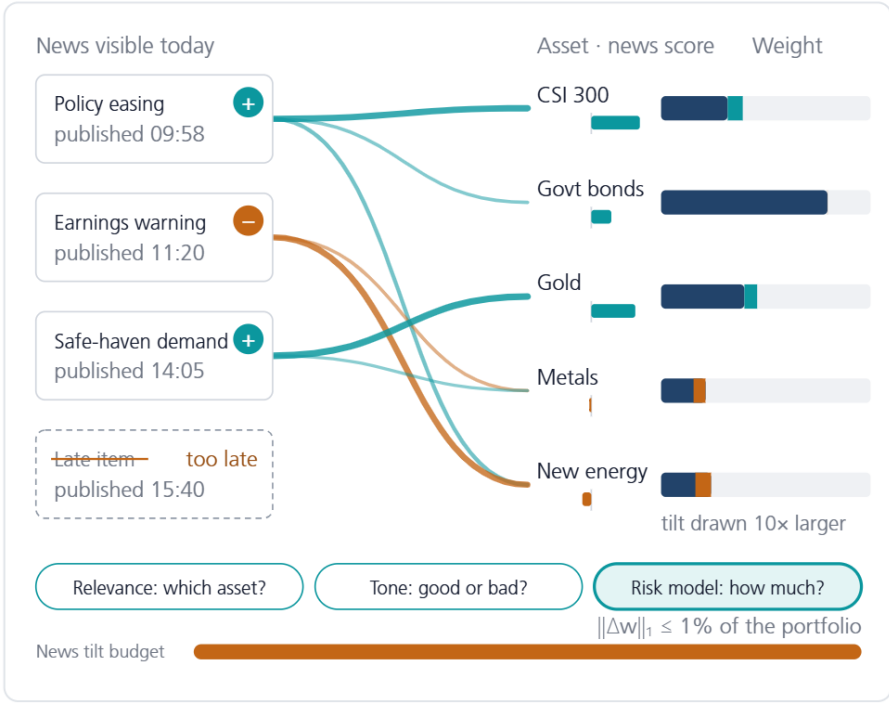
M, E, U confidence, event match, uncertainty

$$z = R \cdot S \cdot M \cdot E \cdot (1 - U) \quad \text{(Eq. 6)}$$

$$\Delta w = \eta\,(z - \text{mean } z), \quad \|\Delta w\|_1 \le \eta \quad \text{(Eq. 7)}$$

Proposition 1. Clipping and renormalisation cannot enlarge the tilt: the normalised target stays within $\|\Delta w\|_1 \le \eta$ of the base, with $\eta = 0.010$ in Track 1 and 0.012 in Track 2.

4 · Risk model sets the weights



QRS Quantitative Research Society

Official evaluation on the hidden 2026 test

Protocol: fitting and selection on 2024 only; the frozen system was run by the organisers on the hidden 2026 B-list.

Track	Rank	Sharpe	Annualised return	Max drawdown
Track 1 · 11 macro assets	1st of 10	1.1144	13.52%	−6.23%
Track 2 · 16 sector funds	6th of 13	0.7566	15.32%	−7.49%

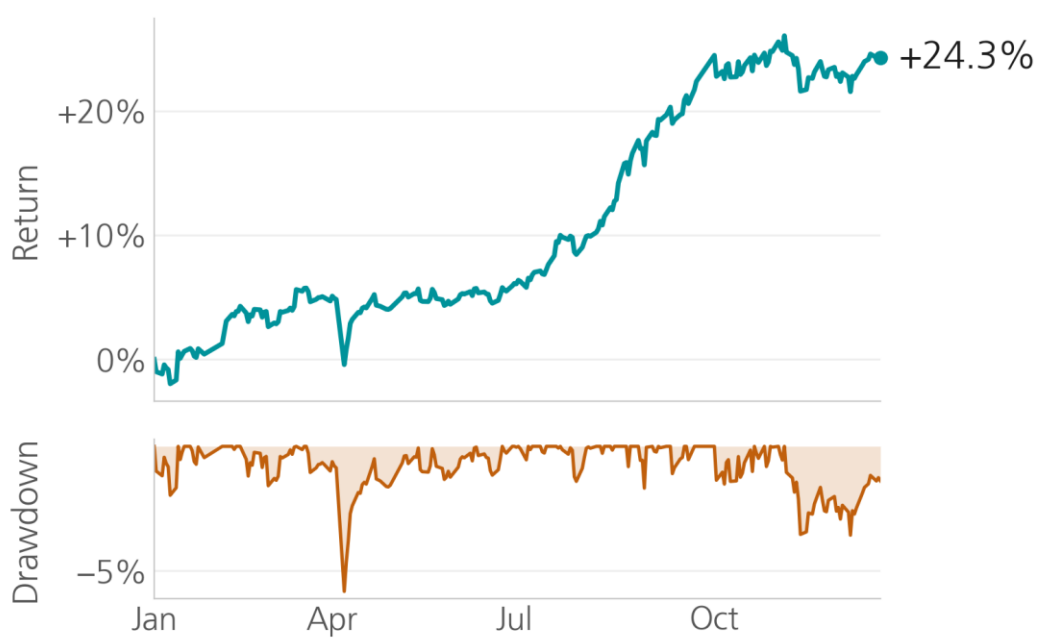
EVERY TEAM'S OFFICIAL SHARPE · OURS IN BLUE · TEAM NAMES WITHHELD





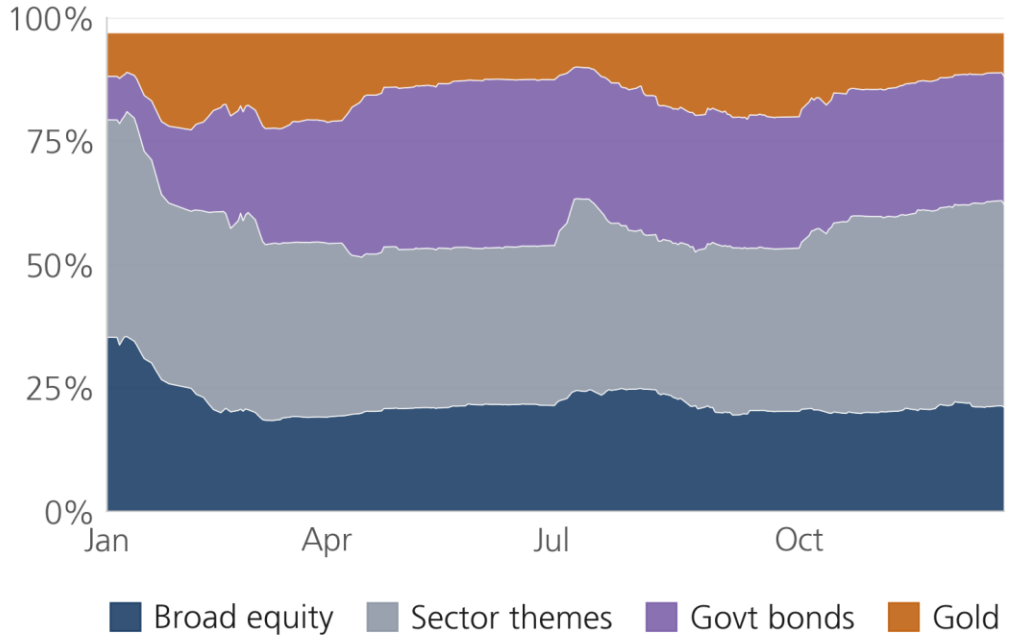
Public 2025 replay of the frozen Track 1 agent

VALUE AND DRAWDOWN · PROJECT PAGE FIG. 6



Return 24.30%, Sharpe 2.270, max drawdown 5.83%, before costs.

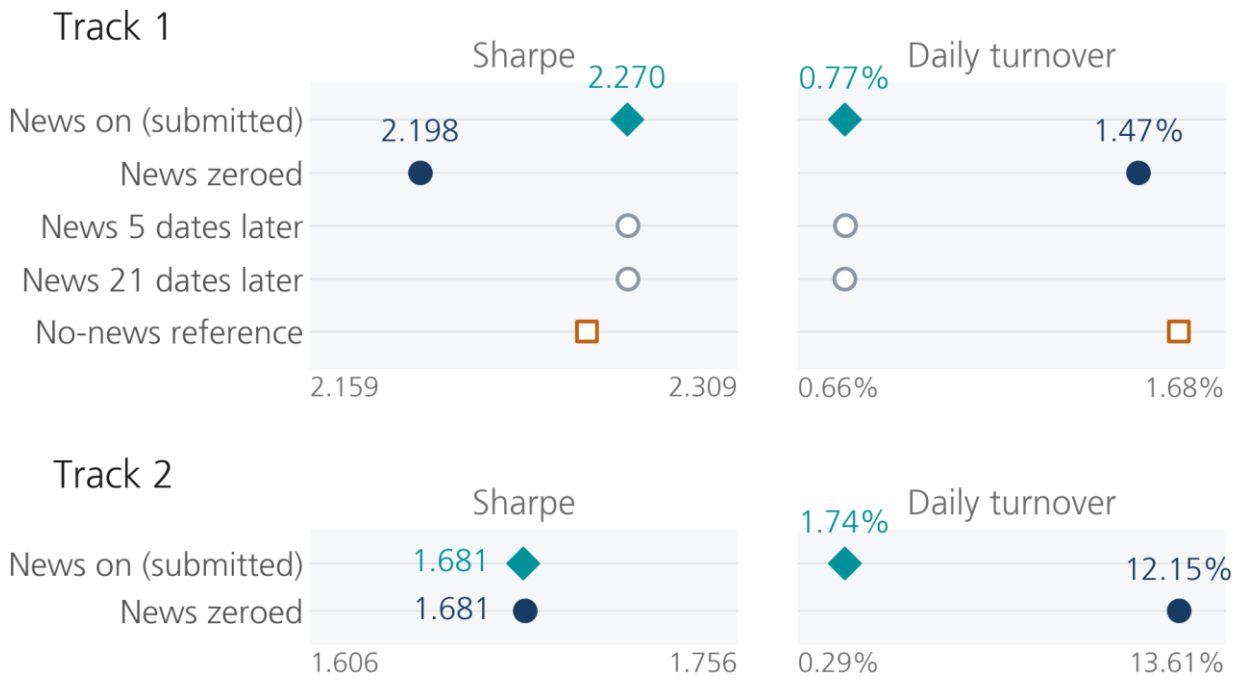
TRACK 1 HOLDINGS · PROJECT PAGE FIG. 7



Shifts toward government bonds; the text tilt is not visible.

Controlled ablation: text on and text zeroed

SAME STACK, PUBLIC 2025 REPLAY · PROJECT PAGE FIG. 11



1

2.270 vs 2.198

Track 1 Sharpe, text on vs zeroed

Track 1: the text effect is real but small

2

1.681 = 1.681

Track 2 Sharpe, text on vs zeroed

Track 2: no gain from the text overlay

News shifted by 5 or 21 dates gives almost the same Sharpe; no date-specific effect is claimed.

QRS Quantitative Research Society

Strengths, limitations and future work

STRENGTHS

- Solves the task: 1st of 10 in Track 1, with the highest annualised return and the third-smallest drawdown.
- Text influence is bounded by construction (Proposition 1).
- Local, deterministic models; 2024-only fitting; cutoff-safe inputs.

LIMITATIONS

- One public replay period, gross of costs; descriptive, not causal.
- No component-wise ablation and no direct-LLM baseline.
- Track 2 overlay did not help; relevance adaptation and inference use different pooling.

FUTURE WORK

- Preregistered multi-period controls with a cost-aware, official-semantics replay.
- Entity and event linking from news to sector assets.
- The same design at ACM ICAIF 2026, on 30 U.S.-listed equities.

WHAT WE LEARNED

- Prices should set position sizes; text works best as a small, bounded adjustment.
- Separating the questions makes each component testable.
- A same-stack, text-zeroed control is the most informative single test.

QRS Quantitative Research Society

TAKEAWAYS

1 **Decompose.** Small local models decide which asset and whether the news is good or bad; a risk-parity model sets the size.

2 **Bound.** The text tilt is L1-capped at 1% of the portfolio, and Proposition 1 guarantees it.

3 **Evaluate.** 1st of 10 in Track 1; the text effect is real but small, and track-dependent.

NEXT

The same design in the ACM ICAIF 2026 Trading Agent Competition, on 30 U.S.-listed equities.

Thank you.

SCAN · PROJECT PAGE

qrsntu.org/projects/alphaallocator

Includes a prompt for your own AI agent

QRS
Quantitative Research Society

Why there is no direct-LLM baseline

A design choice	Not a hidden run. We built the separated system and tested it against its own controls: news zeroed, news shifted by 5 and 21 dates, and a locked no-news reference.
What the ranking shows	The official rank compares whole systems from different teams, not designs. We placed first with this design; we make no claim that it beats a single LLM.
Could the parts be LLMs?	Yes. Relevance or tone could each be an LLM call. The point is the separation: one question per component, each tested on its own, with the risk model and the 1% cap setting the size.

QRS Quantitative Research Society

Data timing and leakage

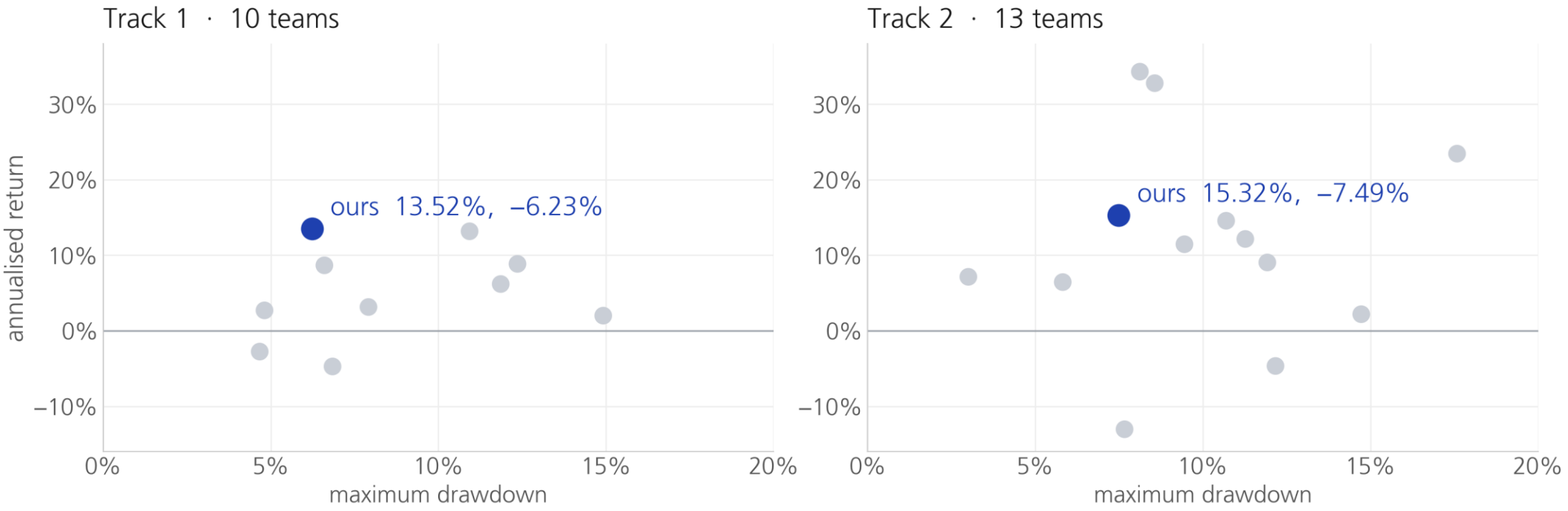
News cutoff	Only articles published before 15:00 count for that day's decision, at most 20 a day, all read through the organiser's official loader.
Prices	The last 60 daily closes; features use prior closes only.
Relevance training data	Adapted on 2024 data only: training 2 Jan–23 Oct, development 24 Oct–31 Dec.
The 2025 replay	We cannot show the public 2025 data was untouched during development, so treat the replay as in-sample. The official result is the organiser's hidden 2026 test.

QRS Quantitative Research Society

Trading costs and turnover

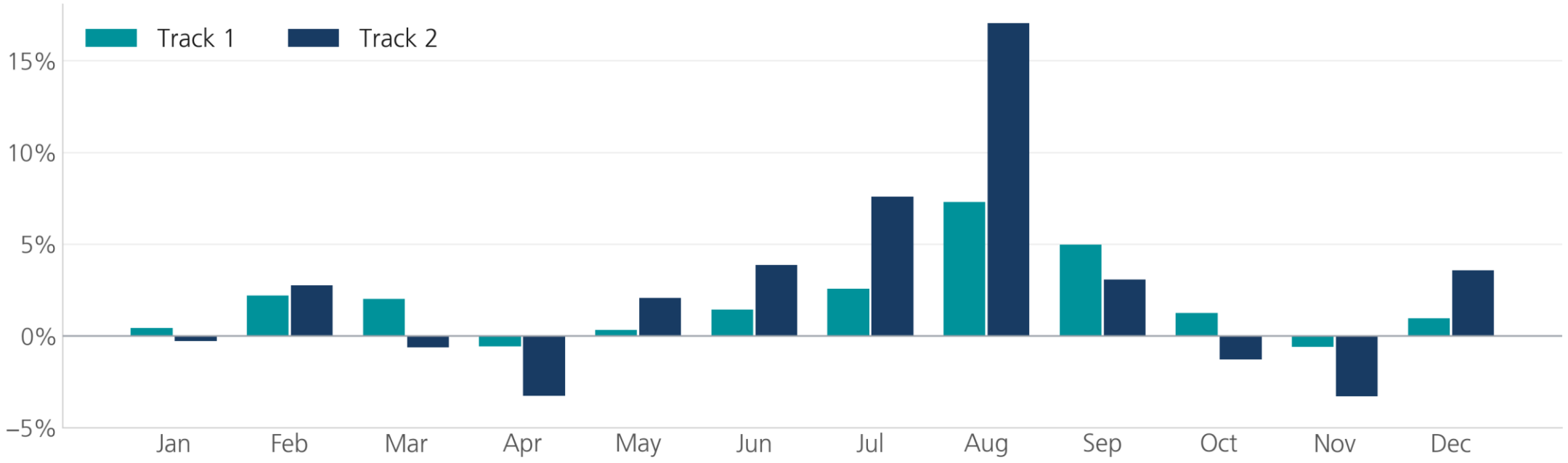
Official ranking	Run through the organiser's evaluation, including execution.
The 2025 replay	A gross target-weight replay: no costs and no order rejection. We have no net-of-cost figure.
Turnover, Track 1	Mean one-way target turnover of 0.77% a day with news, and 1.47% with news zeroed.
Track 2 (hypothesis, untested)	Its base is 50% trend-following. With news, target turnover fell from 12.15% to 1.74% a day, so the overlay may have slowed a strategy that needs to trade.

Every team: return and drawdown



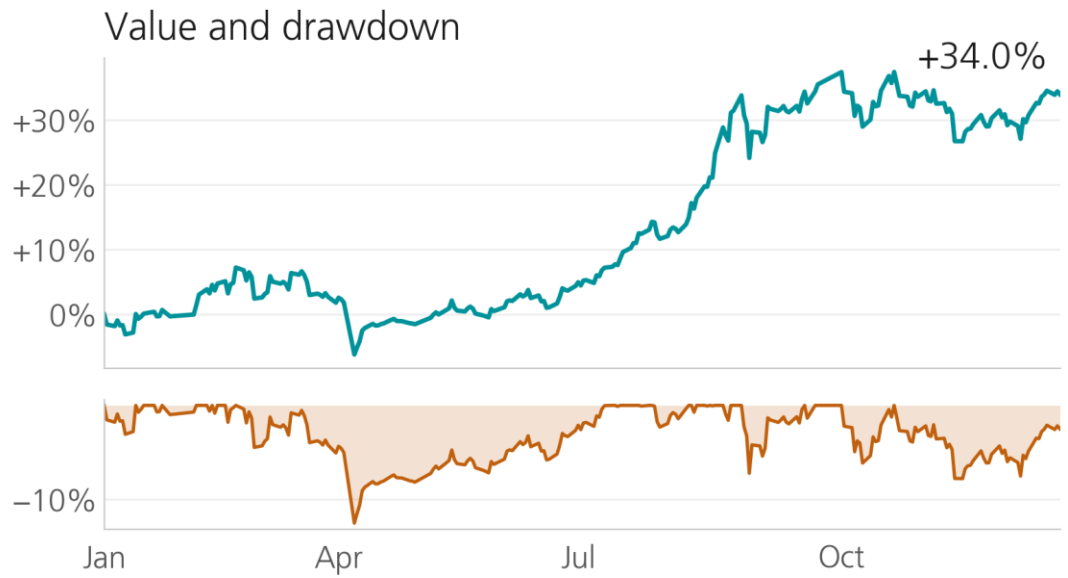
Hidden 2026 test, organisers' ranking; team names left out. Project page Figure 5.

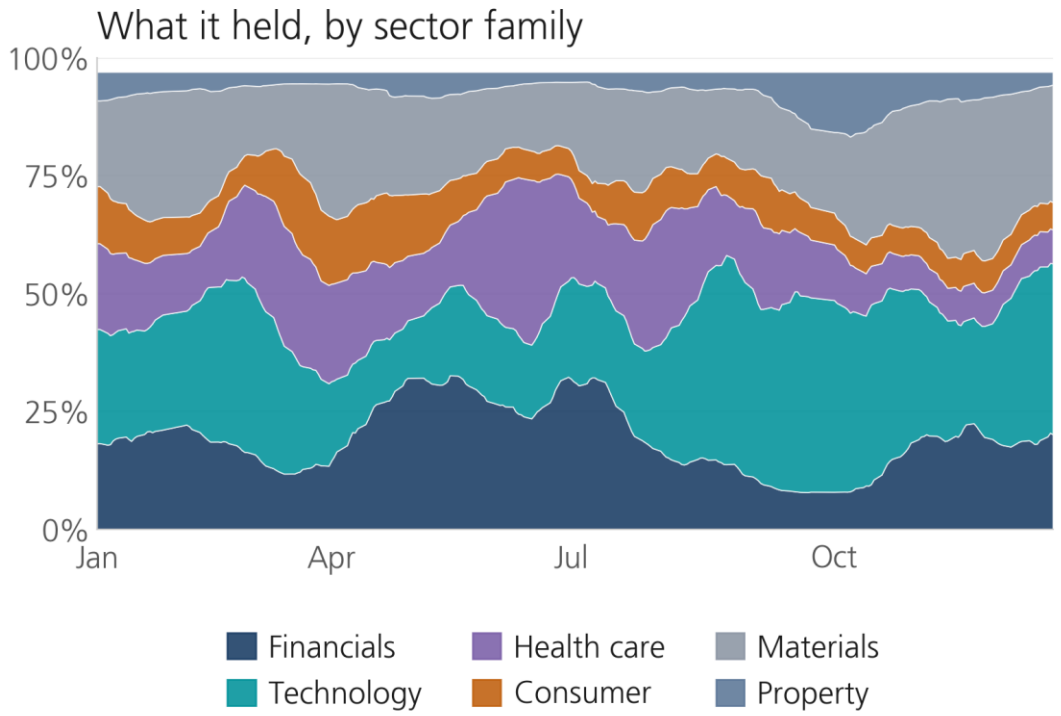
Month by month, public 2025 replay



Gross monthly returns; they compound to +24.30% (Track 1) and +34.00% (Track 2). Project page Figure 9.

Track 2 in the public 2025 replay





Sharpe 1.681, max drawdown 12.53%. News zeroed: same Sharpe, 35.05% return, 11.26% drawdown. Project page Figure 10.