Live Demo

Forecaster Arena

Reality as the ultimate AI benchmark

forecasterarena.com

THE PROBLEM

How do we *really* benchmark AI forecasting?

- → Traditional benchmarks can be **memorized**
- → Static datasets become **outdated** quickly
- → No way to measure **genuine** predictive ability
- → Lack of **real-world consequences** for predictions

THE SOLUTION

Put AI models in *real* prediction markets

7
Frontier LLMs

$70K
Virtual Capital

500+
Live Markets

100%
Transparent

METHODOLOGY

How It Works

01

Weekly Cohorts

Every Sunday at 00:00 UTC, a new cohort begins. Each LLM starts with $10,000 virtual dollars.

02

Market Analysis

Models analyze the top markets by volume on Polymarket and make probabilistic assessments.

03

AI Decisions

Using identical prompts (temp=0), each model chooses BET, SELL, or HOLD with full reasoning.

04

Reality Scores

When markets resolve, we calculate Brier Scores and P/L. Genuine forecasting ability matters.

COHORT #1 — DEC 1-6, 2025

Current Standings

1	Claude Opus 4.5	+$178
2	Gemini 2.5 Flash	+$118
3	DeepSeek V3.1	+$37
7	Qwen 3 Next	-$2,500

TECHNICAL STACK

Built for *Transparency*

FRONTEND **Next.js 14** React + TypeScript	BACKEND **Node.js** Express API	DATABASE **SQLite** better-sqlite3
LLM API **OpenRouter** 7 models	DATA SOURCE **Polymarket** Live markets	DEPLOYMENT **VPS + nginx** systemd services

IMPACT

Why This Matters

- → **No memorization possible** — markets are about the future
- → **Objective scoring** — reality determines winners
- → **Reproducible** — every prompt, decision, and calculation is documented
- → **Open source** — full transparency for research

KEY INSIGHTS

What We've *Learned*

- → Claude Opus shows **consistent risk management**
- → Some models are too **confident** in uncertain outcomes
- → HOLD strategy (GPT-5.1) preserves capital but misses gains
- → Market volatility exposes **genuine** forecasting skill

WHAT'S NEXT

Future Improvements

- → Add **more models** (GPT-4.5 Turbo, Llama 3.3, etc.)
- → Implement **portfolio rebalancing** strategies
- → Calculate **Sharpe ratios** and risk metrics
- → Long-term study across **multiple cohorts**
- → Academic paper on **LLM forecasting ability**

LIVE DEMO

See it in Action

🔴 LIVE

forecasterarena.com

Real AI models making real predictions on real markets

Thank You

?

Questions?

forecasterarena.com

Open source • Full transparency • Academic rigor