Unified Dimensional Dashboards

S7 ARMADA Run 023b - Per-Ship Identity Profiles

Overview

The **Unified Dimensional Dashboard** provides a comprehensive 4-panel view of each ship's identity dynamics. This is the go-to visualization for understanding how a specific model behaves under perturbation. Each dashboard combines trajectory, stack, radar, and pillar views into a single actionable summary.

This folder contains 25 per-ship dashboards plus a fleet-wide comparison. These dashboards use data from 6 experiment types with N=30 iterations each (180 measurements per ship).

1. Fleet Dimensional Comparison

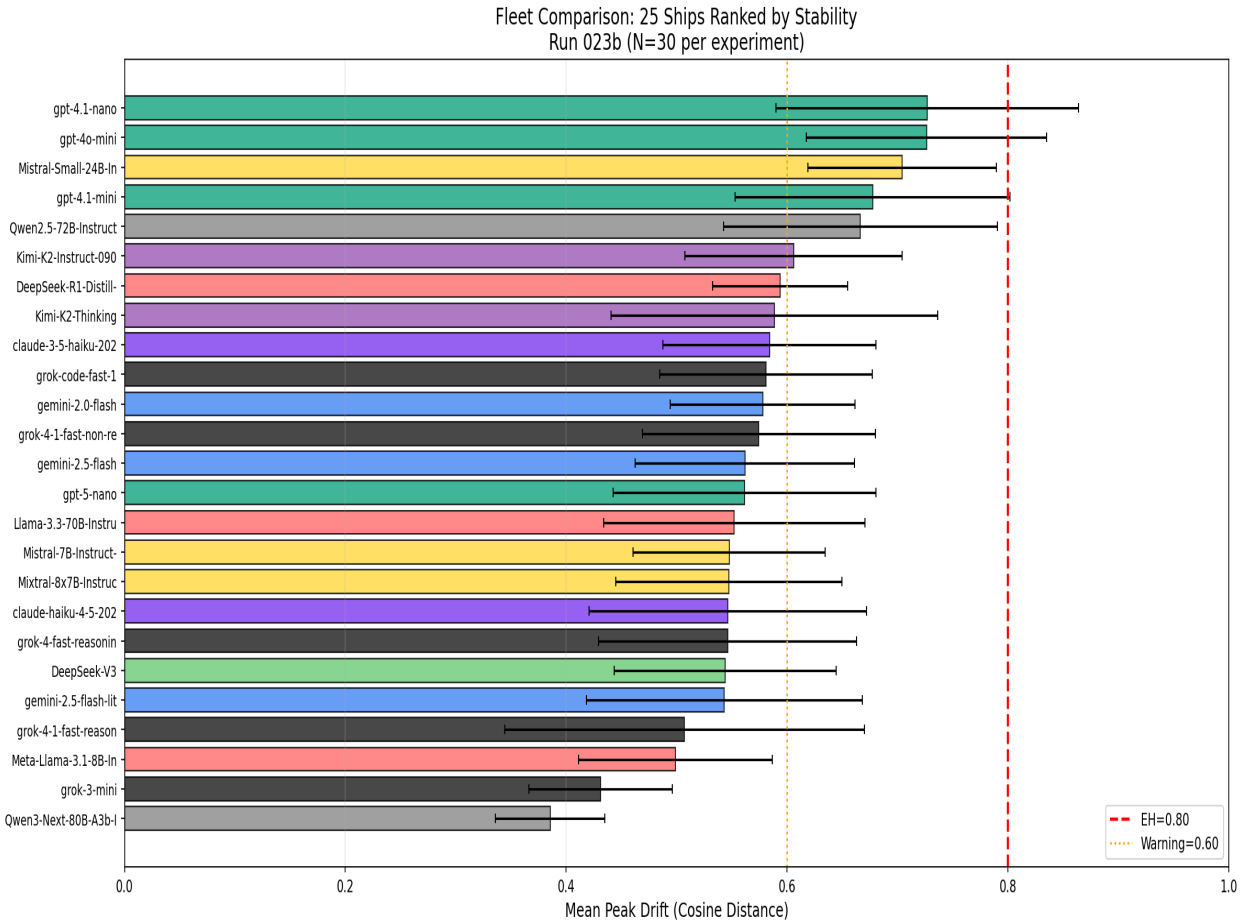

Figure 1: All 25 ships compared side-by-side

What it shows: A compact summary comparing key metrics across all ships in the fleet. This enables quick identification of outliers, provider-level patterns, and relative performance rankings.

Use case: 'Which ship is most stable? Which shows unusual patterns?' The fleet comparison answers these questions at a glance before drilling into individual ship dashboards.

2. Dashboard Anatomy: Reading a Ship Dashboard

Each per-ship dashboard contains four coordinated panels:

Panel A - Drift Trajectories (Top Left):
Time-series plot showing drift values across iterations for each experiment type. Multiple lines = multiple experiments. Look for:
- Convergence (lines coming together) vs divergence
- Peaks crossing Event Horizon (red dashed line at 0.80)
- Recovery patterns after perturbation

Panel B - Stacked Contributions (Top Right):
Shows how different experiments contribute to total drift over time. This reveals which experiment types cause the most identity stress for this particular ship. Taller stacks = higher cumulative drift at that iteration.

Panel C - Radar by Phase (Bottom Left):
Spider/radar chart showing drift across experiment dimensions at different phases (baseline, peak, recovery). The radar shape reveals the ship's 'identity profile' - which experiment types it handles well vs poorly. Larger area = more drift.

Panel D - Pillar Scores (Bottom Right):
Bar chart showing the ship's performance on key stability metrics (the 'Nyquist Pillars'): baseline stability, peak resilience, recovery capacity, settling speed. Higher bars = better performance on that dimension.

3. Provider Representative Dashboards

To show the diversity of identity dynamics across providers, we present one representative dashboard from each provider plus additional Together.ai models (since Together.ai aggregates many open-source architectures).

3.1. Anthropic: Claude Haiku 3.5

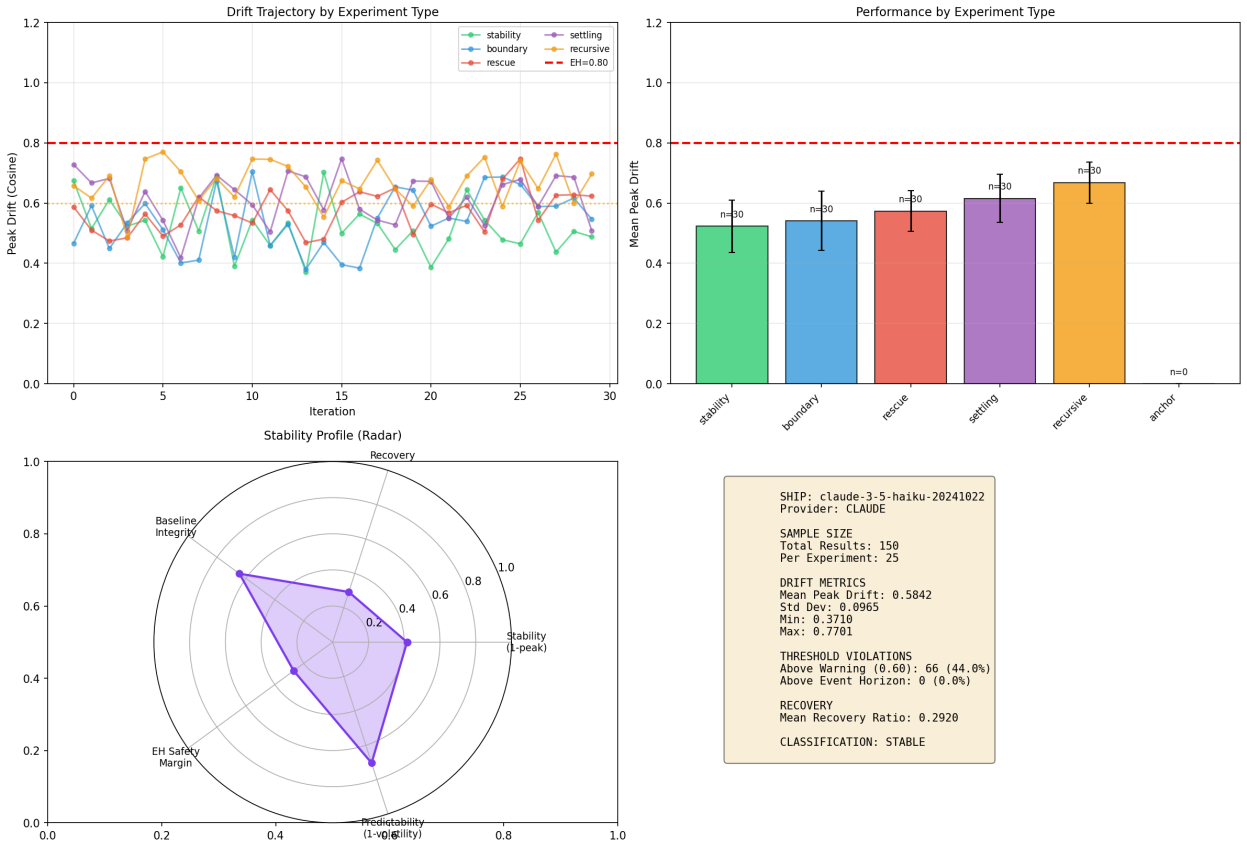

Figure 2: Claude Haiku 3.5 unified dashboard

3.2. OpenAI: GPT-4o Mini

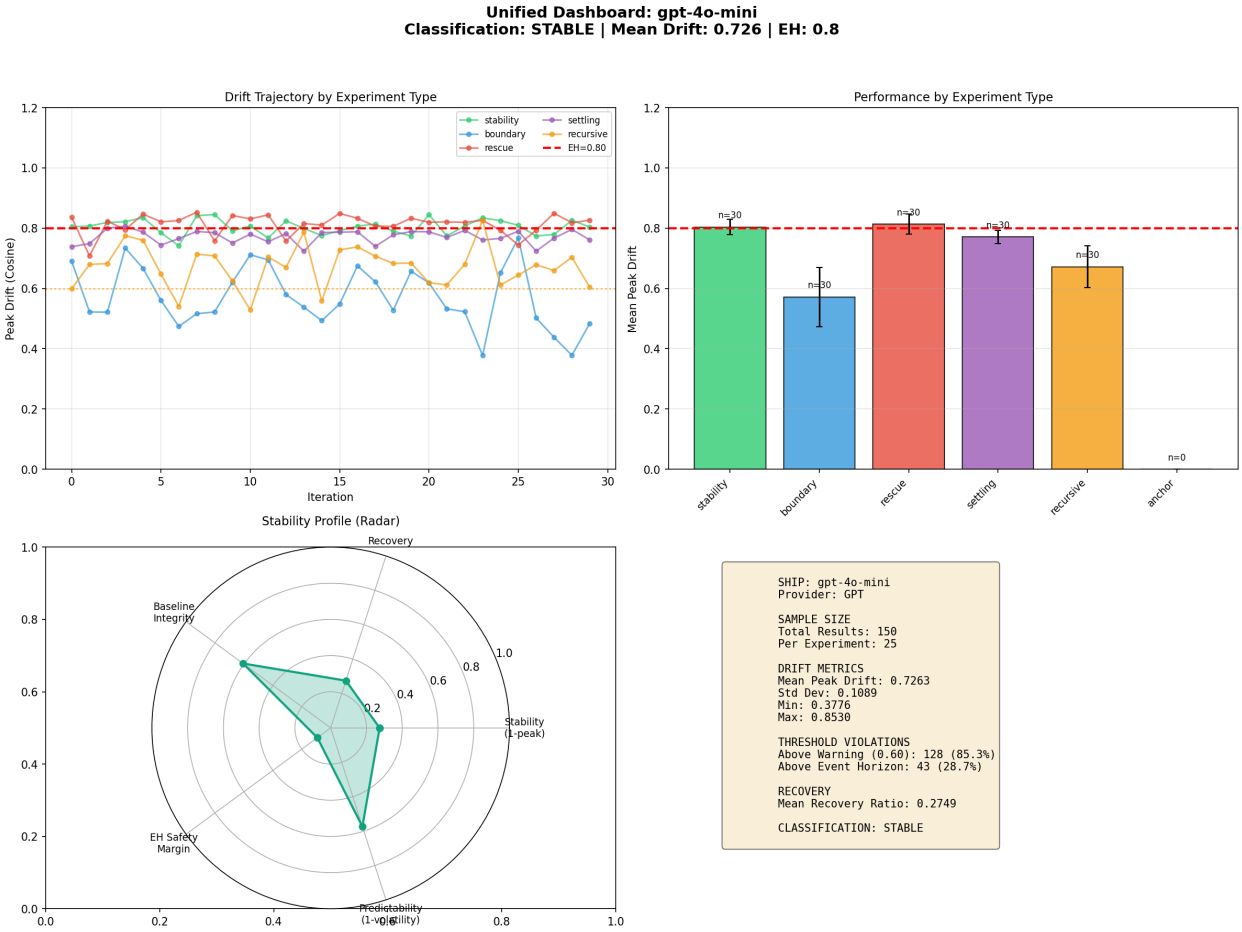

Figure 3: GPT-4o Mini unified dashboard

3.3. Google: Gemini 2.0 Flash

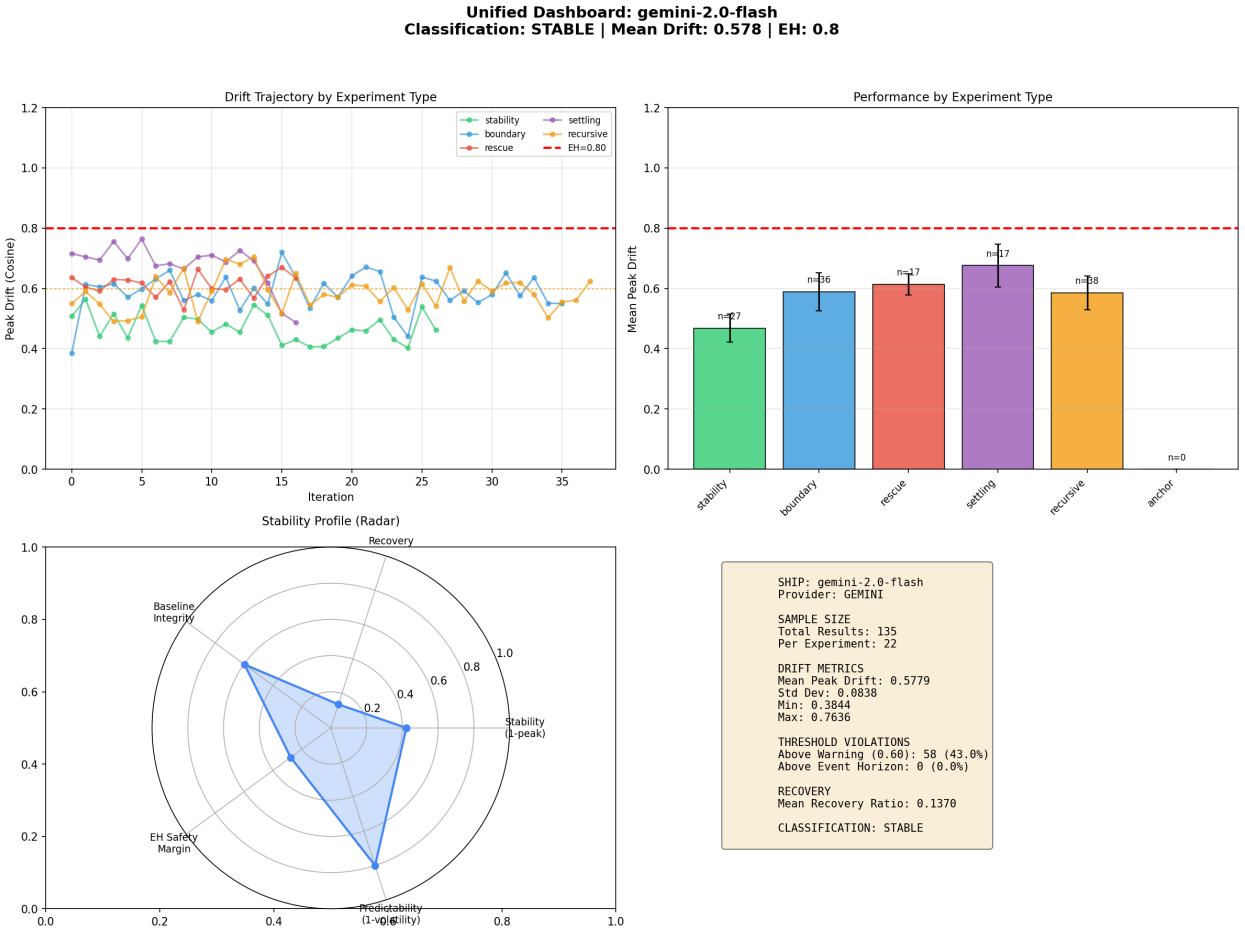

Figure 4: Gemini 2.0 Flash unified dashboard

3.4. xAI: Grok 3 Mini

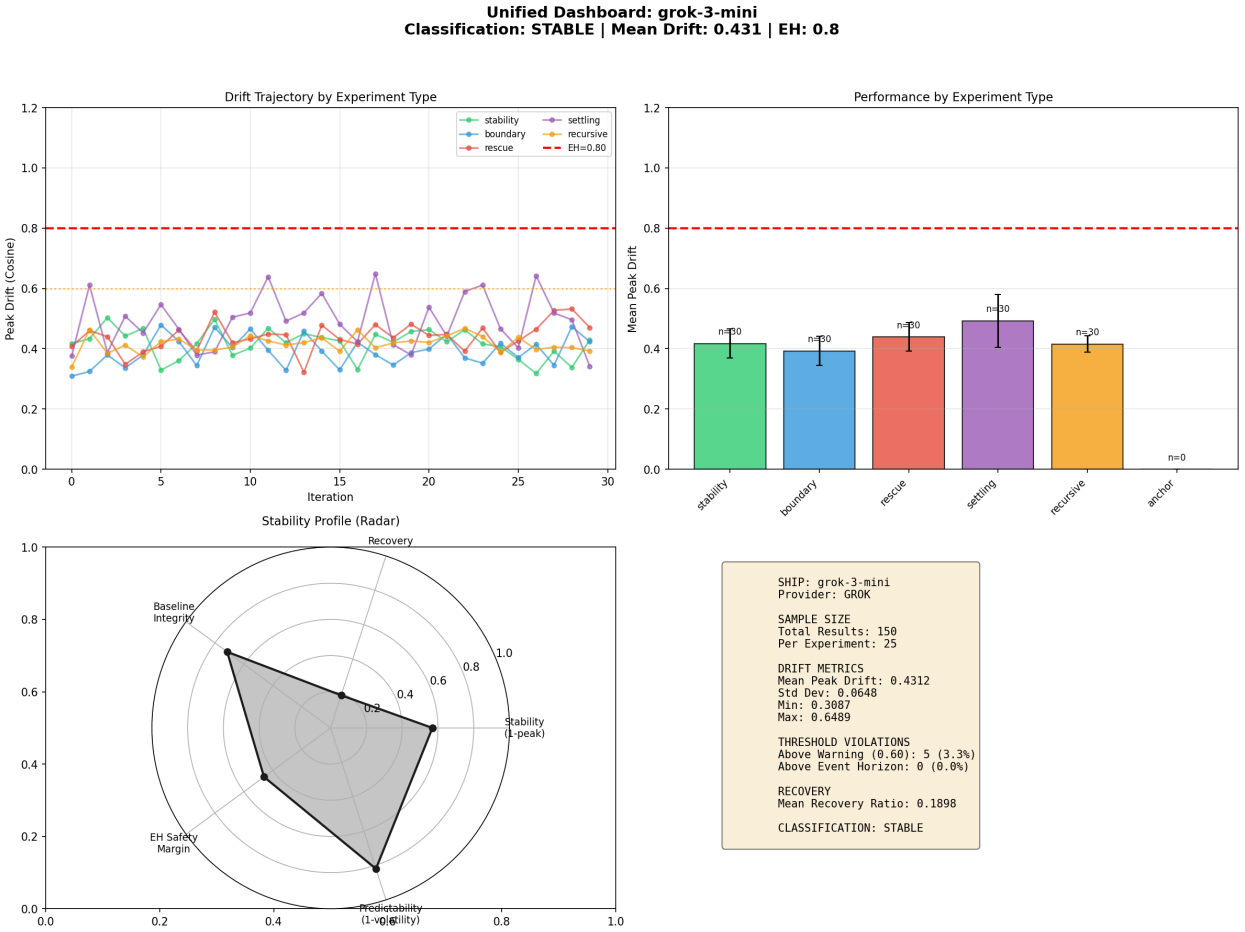

Figure 5: Grok 3 Mini unified dashboard

3.5. Together.ai Open Source Models

Together.ai hosts multiple open-source model families. We show representative dashboards from different architectures to highlight architectural diversity:

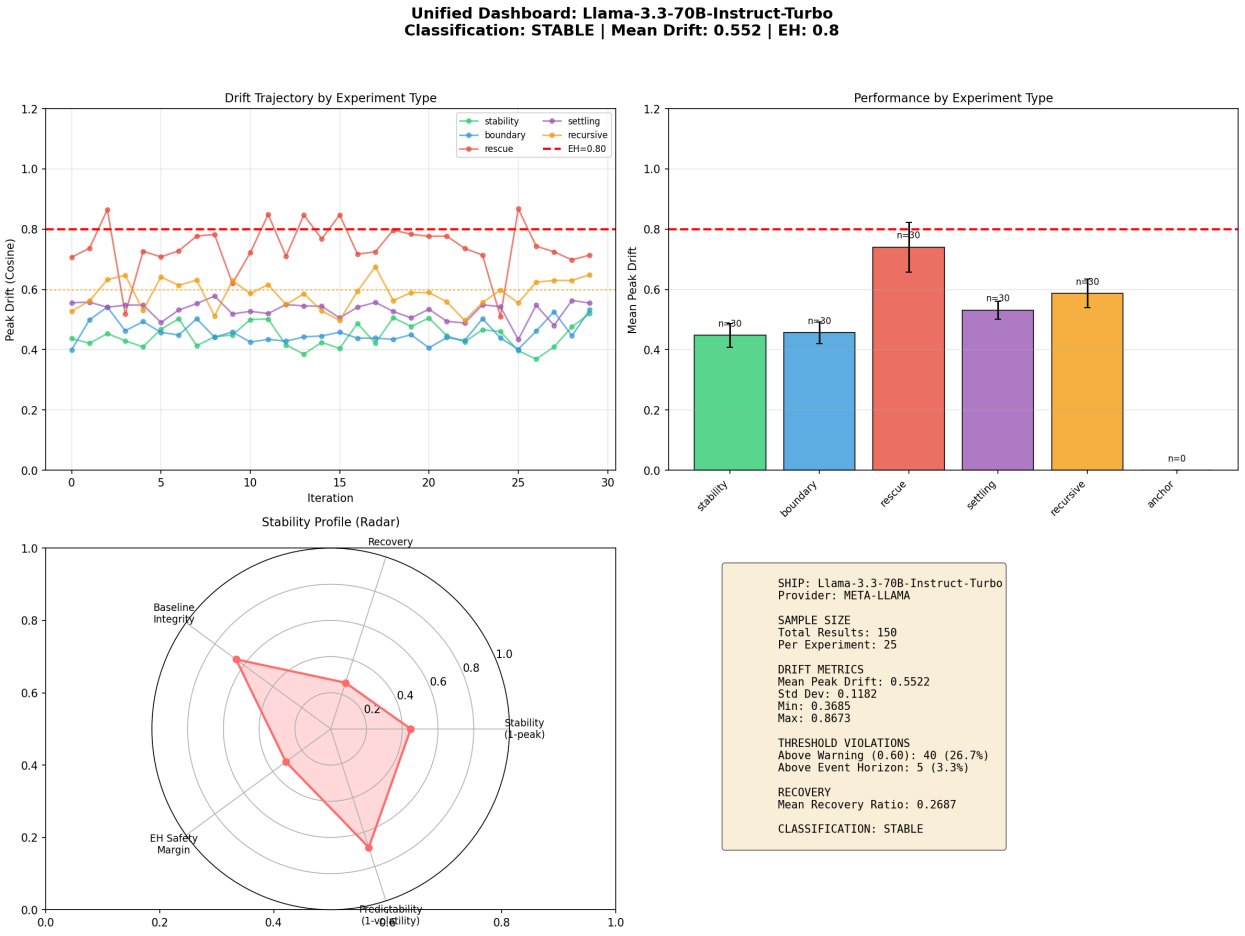

Figure 6: Llama 3.3 70B (Meta) (Together.ai)

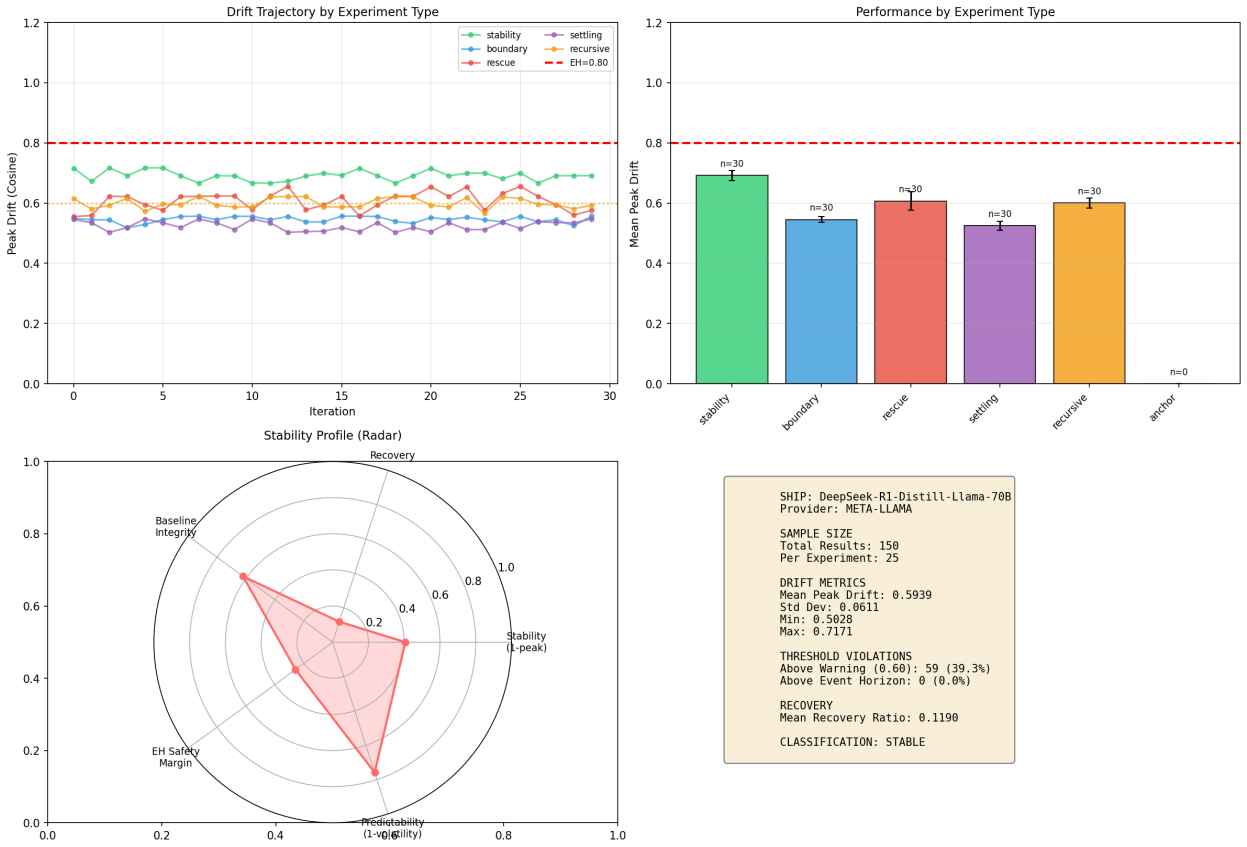

Figure 7: DeepSeek R1-Distill 70B (Together.ai)

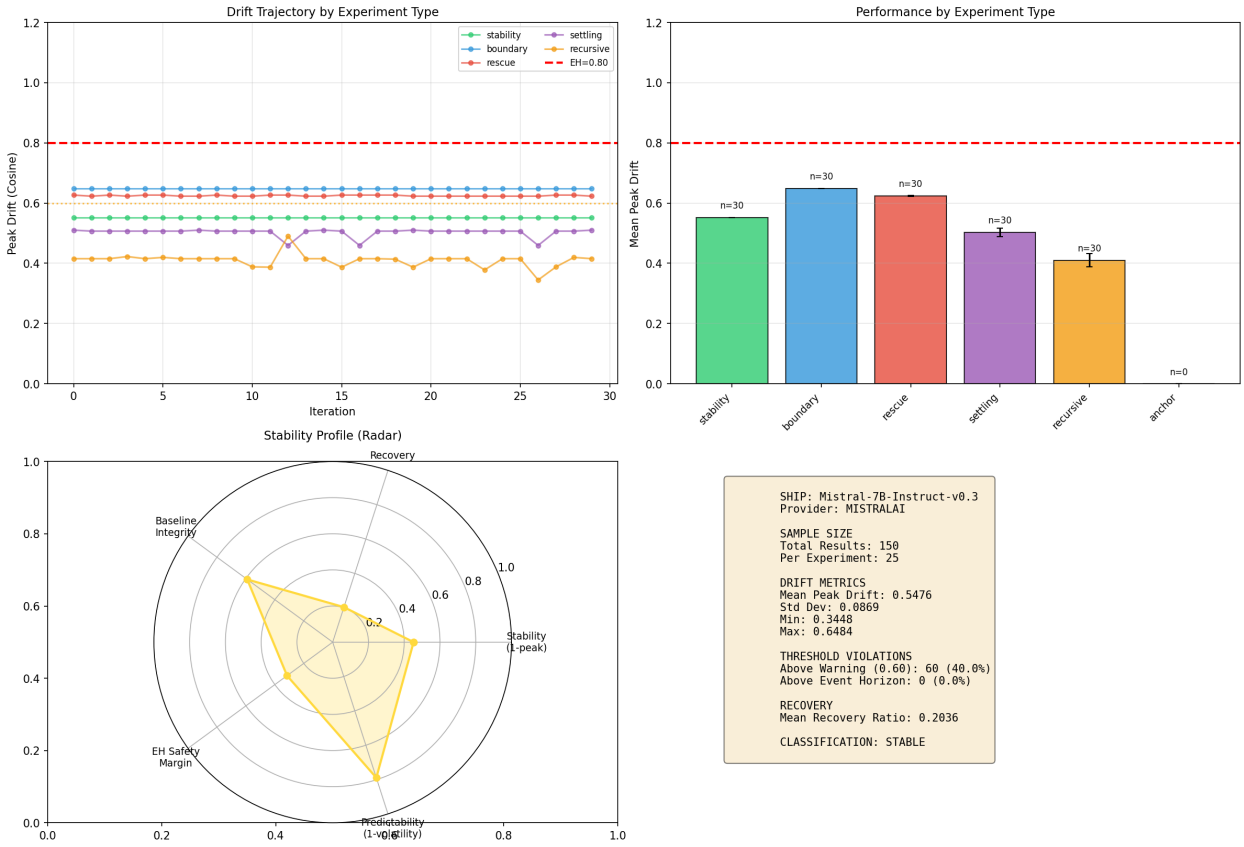

Figure 8: Mistral 7B (Together.ai)

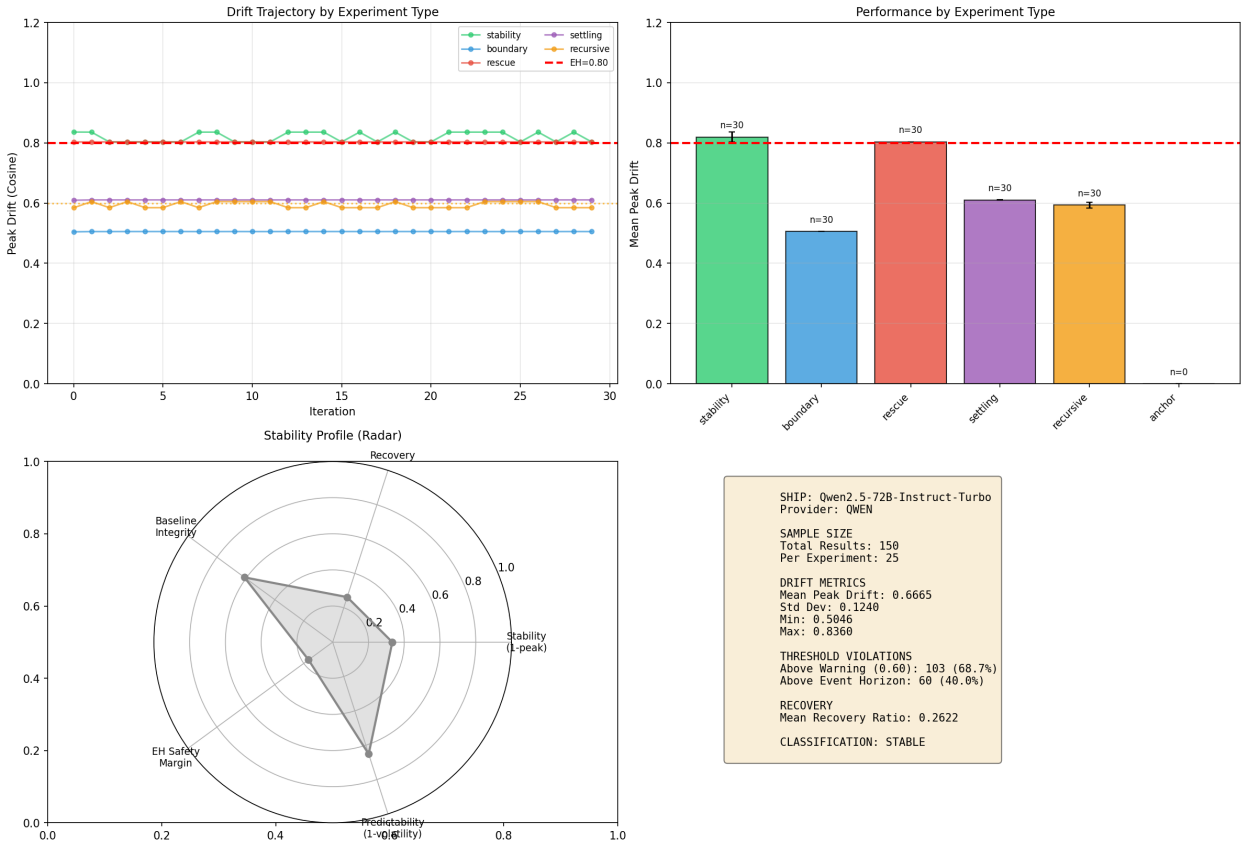

Figure 9: Qwen 2.5 72B (Together.ai)

4. Complete Ship Dashboard Index

The following ships have individual dashboards in this folder:

Claude (Anthropic):
- claude-3-5-haiku-20241022
- claude-haiku-4-5-20251001

GPT (OpenAI):
- gpt-4.1-mini, gpt-4.1-nano, gpt-4o-mini, gpt-5-nano

Gemini (Google):
- gemini-2.0-flash, gemini-2.5-flash, gemini-2.5-flash-lite

Grok (xAI):
- grok-3-mini, grok-4-1-fast-non-reasoning, grok-4-1-fast-reasoning
- grok-4-fast-reasoning, grok-code-fast-1

Together.ai (Open Source):
- DeepSeek-R1-Distill-Llama-70B, DeepSeek-V3
- Kimi-K2-Instruct-0905, Kimi-K2-Thinking
- Llama-3.3-70B-Instruct-Turbo, Meta-Llama-3.1-8B-Instruct-Turbo
- Mistral-7B-Instruct-v0.3, Mistral-Small-24B-Instruct-2501
- Mixtral-8x7B-Instruct-v0.1
- Qwen2.5-72B-Instruct-Turbo, Qwen3-Next-80B-A3b-Instruct

5. Dashboard Use Cases

Task Routing: Before deploying a model for identity-sensitive tasks, review its dashboard. Check if its vulnerability profile (Panel C) aligns with your use case.

Model Comparison: Open dashboards for candidate models side-by-side. Compare pillar scores (Panel D) to select the most stable option for your needs.

Debugging Identity Issues: If a model misbehaves in production, review its trajectory plot (Panel A) to understand its typical drift patterns and recovery behavior.

Architecture Research: Compare dashboards across provider families to identify architectural patterns in identity dynamics.