White paper · Public edition v.2 · APOCALYPSE™ Research Foundation

A 17-Agent Mixture-of-Refinement Early-Warning System for Global Instability

Author: Prof. Dr. Stelios Bekiros

Project: APOCALYPSE / PROMETHEUS
Date: May 2026

"The Index does not foretell the future. It listens to the world."

Abstract

PROMETHEUS is a neural early-warning system designed to detect the precursors of global instability before they crystallise into crises. It ingests 119 heterogeneous live data sources distributed across four published categories — Macro & Geopolitical, Planetary & Seismic, Solar & Heliospheric, and Social & Prediction — and produces a single daily calibrated number, the Global Apocalypse Index (GAI), alongside a seven-day probabilistic forecast band.

The architecture is organised around 17 agents: 16 internal specialist agents, each embodying a distinct published precursor theory, and one web-facing publishing agent. The 16 internal agents are not independent models. They are doctrinal kernels within a unified Mixture-of-Refinement neural network, routed by a Bayesian mixture allocator and fused via Bayesian posterior pooling. The routing is balanced by a Sinkhorn algorithm operating under a number-theoretically derived connectivity constraint; no single precursor theory may dominate the output.

Training proceeds walk-forward over a timeline stretching from 1900 to the present, with strong regularisation and early stopping governed by a validation metric derived from temporal hold-out. A genetic algorithm outer loop, implementing multi-objective Pareto optimisation, tunes the mixture parameters after gradient training completes. Newly admitted sources with short observation histories are buffered through a structured donor / extrapolation / synthesis completion scheme so that flat or zero-valued sparklines never appear in the public dashboard. The canonical checkpoint achieves an AUC of 0.906 on stratified hold-out, a walk-forward backtest AUC of 0.833, a false-alarm rate of 3.2% at 80% detection coverage, and an expected calibration error of 0.096 — clearing the

three-meter sign-off required for public deployment.

This paper describes the system's design rationale, data architecture, agent structure, training methodology, and evaluation results at a level suitable for a scientifically literate non-specialist audience. Proprietary kernel internals, hyper-parameters, and warmstart mechanisms are withheld; the academic publication and the Technical / Math Reference provide the full methodological record.

1. Introduction

Why Early Warning Matters

Every major global disruption — financial collapse, geophysical catastrophe, geopolitical escalation, pandemic — is preceded by observable precursors. These precursors are rarely found in a single signal. They are conjunctions: a financial stress indicator rising, a geomagnetic index becoming anomalous, social media conflict tone darkening, and a seismic cluster building — all simultaneously, in a pattern that no single-domain expert monitors end to end. Human attention is domain-local. Crises are multi-domain.

The practical consequence is that most early-warning infrastructure today operates as a collection of independent single-domain dashboards. A financial analyst watches implied volatility. A seismologist watches earthquake catalogues. A space-weather forecaster watches solar-wind data. None of these specialists has an integrated view; no existing system weighs these channels against each other in a principled, continuously updated probability estimate.

PROMETHEUS is built to address that gap directly.

Why Single-Source Signals Fail

There is a deeper theoretical reason why single-source monitoring is inadequate for catastrophic-event precursion, and it goes beyond the practical difficulty of cross-domain surveillance. Catastrophic regime changes — the class of events that constitute genuine global instability — are characterised by critical slowing down: the system's internal dynamics slow, its state space contracts, and precursor signals become increasingly coherent across previously uncorrelated channels. The precursor is, in a precise mathematical sense, a cross-domain signature. A single-channel index, however well-calibrated for its own domain, cannot detect it.

Moreover, different precursor theories — Thom's catastrophe geometry, Varotsos's natural-time formalism, Hawkes's self-exciting process model, extreme value theory — make empirically distinct predictions. Their signals are not always consistent. A principled early-warning system must be able to hold all of them active simultaneously, weighting each according to its current predictive contribution, rather than committing in advance to any single theoretical framework.

Why a Mixture of Doctrinal Precursor Theories

The core design principle of PROMETHEUS is doctrinal pluralism: every established precursor theory that has empirical support in the literature is assigned a dedicated computational kernel within the network. These kernels are trained jointly, and their relative contribution to the final output is determined by the data, not by the designer. When a regime shift is approaching via a catastrophe-geometry route, the catastrophe kernel will naturally receive higher weight. When a social cascade is the primary driver, the Hawkes kernel will dominate. When multiple theories converge — which is the empirically strongest precursor signature — the mixture allocator reflects that convergence in the output probability.

This is the Mixture-of-Refinement (MoR) principle, and it is both the technical heart and the philosophical foundation of the system.

2. Data Architecture

119 Live Adapters Across 4 Categories

PROMETHEUS draws on 119 production data sources, all wired through live adapters, distributed across four published categories so that no single category and no single source can determine the output. The four categories are:

1. Macro & Geopolitical — implied volatility indices, bond market stress measures, nowcast GDP estimates, economic policy uncertainty, geopolitical risk indices, multilateral macroeconomic feeds (World Bank, IMF, OECD, BIS), conflict event catalogues, and climatic and disaster series.
2. Planetary & Seismic — global earthquake catalogues (USGS FDSN, EMSC, IRIS), early earthquake warning feeds, planetary ephemeris and geomagnetic storm indices, ocean and cryosphere panels, and the Schumann-resonance Earth-ionosphere channels.
3. Solar & Heliospheric — daily sunspot counts, planetary K-index, solar X-ray flux, coronal mass ejection catalogues, solar-wind speed and magnetic field orientation, and geomagnetic baselines.
4. Social & Prediction — conflict event counts and tone (GDELT), prediction-market sentiment, social-search velocity, public-health surveillance, and democracy indices.

All 119 sources are publicly available or obtainable through open institutional feeds. None is proprietary financial data; none requires a commercial licence for replication. This is a deliberate methodological commitment to reproducibility.

The source manifest is provisioned at 200 slots as a Hetzner storage-precaution headroom for future admissions. The 119 active sources are the only ones contributing to the live index; of these, 88 are inherited from the previous panel and 31 were newly admitted on 17 May 2026. The remaining 81 slots are reserved for future sources that pass the polarity-validation and admission gates.

The Polarity-Validated Panel

A critical pre-training verification step is the polarity diagnostic: every source in the panel must correlate positively with observed crisis severity in the historical record. After a direction-tagging pass — which inverts sources where raw values decrease during crises, such as the Dst geomagnetic storm index and the GDELT global tone score — every one of the 119 active production sources passes the polarity check before being admitted to the panel. This ensures that every channel points in the same semantic direction: higher value means higher risk.

Sources that fail network access during any given inference run are masked rather than imputed; the network is trained to operate on partially available panels, and masked channels do not propagate noise into the output. Newly admitted sources with very short observation histories are buffered through a donor / extrapolation / synthesis completion scheme during their warmup window, so that flat or zero-valued sparklines never appear in the public dashboard.

3. Agent Architecture

PROMETHEUS comprises 17 agents: 16 internal computational kernels within a single unified model and one outward-facing publishing agent. Each internal kernel applies a distinct mathematical lens to the same time-series panel — catastrophe and chaos geometry, recurrence quantification, helix-cyclic structure, extreme value tails, multi-resolution wavelets, natural-time criticality, heliospheric and Schumann electromagnetic precursors, planetary seismic and orbital signals, Hawkes self-excitation of social and conflict streams, a broad statistical-indicator bank, a number-theoretic activation, an evolutionary outer-loop optimiser, and an I/O substrate — and contributes to a shared mixture output whose contribution weights are routed dynamically by the data. The Orchestrator (Agent 01) is the fusion stage, holding the Bayesian posterior layer that pools the expert outputs into a single calibrated probability under a Dirichlet prior that prevents any single expert from dominating; the Website agent (Agent 17) is the public interface, the sole consumer of the model's output. Detailed per-agent mathematical specifications — kernel formulations, hyperparameters, routing logic, and ablation results — are documented in the Technical / Math Reference, distributed under separate sensitivity classification.

4. The Mixture-of-Refinement Core

The 16 doctrinal kernels are the experts in a Mixture-of-Refinement architecture. The key design challenge for any mixture model is preventing expert collapse: the tendency for routing mechanisms to converge on one or two dominant experts and ignore the rest. PROMETHEUS addresses this through two structural innovations.

The first is the routing mechanism itself. Rather than a simple top-k hard assignment — the standard approach in mixture-of-experts architectures — the system uses a Bayesian Latent

Mixture Allocator that performs soft assignment across all 16 experts, balanced by a Sinkhorn iteration that enforces approximate load balancing. No expert is ever completely silenced. No expert can completely monopolise the output: a dominance guard imposes a ceiling on the fraction of the posterior any single kernel may contribute.

The second is the attention connectivity structure between experts. The inter-agent information exchange is governed by a connectivity pattern derived from Ramanujan expander graph theory, which provides a provably near-optimal balance between information propagation speed (short path lengths between any two experts) and propagation balance (no expert disproportionately amplified by the structure of the communication graph). This prevents the routing mechanism from developing structural preferences for particular experts independent of the data.

The fusion stage — Agent 01's Bayesian posterior pooling — combines the routed expert outputs in log-odds space with a learned Dirichlet prior, producing a single calibrated probability that is directly interpretable as a statement about global instability risk.

5. Training and Evaluation Methodology

Walk–Forward Training

PROMETHEUS is trained in walk-forward mode over a timeline from 1900 to the present. Walk-forward training means that at each training step, the model sees only data from the past relative to the forecast point — never information from the future. This is the only valid evaluation protocol for a time-series forecasting system: train-test leakage on temporal data is not a statistical curiosity but a fundamental evaluation error that produces models whose apparent performance evaporates on deployment.

The validation set is the most recent period of the timeline, held out entirely from gradient training. The primary evaluation metric — AUC, the area under the receiver operating characteristic curve — is computed on this temporal hold-out and tracked with an exponential moving average to smooth epoch-to-epoch fluctuation.

Regularisation and Early Stopping

The canonical training configuration employs strong weight decay to resist memorisation of historical crisis patterns. Early stopping is triggered by the validation AUC metric: when the held-out performance fails to improve for three consecutive epochs, training terminates. This is the mechanism that prevents the model from fitting the particular distribution of its training initialisation rather than the underlying generative process.

A composite loss function combines a Brier probability score, a focal loss component that increases the relative weight of correctly anticipating crisis periods, a lead-time reward that encourages early detection, and a Bayesian KL-regularisation term that keeps the expert mixture from collapsing to a degenerate posterior.

Genetic Algorithm Tuning

After gradient training, a genetic algorithm outer loop performs multi-objective Pareto optimisation over the mixture allocation parameters and zone classification thresholds. The three objectives being simultaneously optimised are: maximising the area under the ROC curve, maximising the detection lead time, and minimising the false-alarm rate. No human-set weights govern this trade-off; the Pareto front represents the full achievable trade-off surface, and the operating point is selected from it.

Three-Meter Sign-Off

The system is not deployed until it passes all three of the following metrics on the canonical evaluation set:

1. AUC on temporal hold-out — measures the system's ability to rank crisis periods above non-crisis periods, over the most recent data it has not been trained on.
2. False-alarm rate at 80% detection coverage (FAR@POD=0.80) — measures how many non-crisis periods are incorrectly flagged as high risk when the system is operating at a sensitivity level that captures 80% of genuine crises.
3. Expected calibration error (ECE) — measures how closely the system's stated probability estimates correspond to empirical frequencies: if the system says 70% probability, do crises actually occur about 70% of the time in those conditions?

Author Biography

Prof. Dr. Stelios Bekiros is Chair Professor of Finance, Economics and Artificial Intelligence, with research affiliations across European and North American institutions, including positions in econometrics, nonlinear dynamics, and machine learning for complex financial and physical systems. His work spans catastrophe theory, multi-agent computational intelligence, extreme value analysis, and the application of deep learning to non-stationary multi-source time series. APOCALYPSE / PROMETHEUS is the synthesis of his research programme in these areas.