

# Concept Note: Governance for Self-Managing Event Systems

This note outlines a potential PhD research direction focused on enabling large-scale event-driven systems to self-discover their operational structure, assess risk, and take safe, explainable actions. The work combines temporal modeling, machine learning, and governance principles, with applications in data infrastructure and AI safety.

## Problem Statement:

Modern data infrastructures (pipelines, schedulers, CDC systems) produce massive streams of events. Operators (SREs, data engineers) currently monitor, correlate, and intervene manually to handle failures or delays. The goal is to formalize this process: can a system learn from its own history to automatically surface what should happen, when, and what to do when things go wrong—without hand-maintained DAGs or crontabs?

## Proposed Approach:

- Ingest raw event data (e.g., MongoDB change streams → Kafka → ClickHouse).
- Learn job schedules, SLAs, and dependencies from history (temporal point processes, survival analysis).
- Detect late/missing runs and infer human interventions from anomalies.
- Use these patterns to train risk estimators and action policies.
- Constrain policies with governance guardrails (safety > throughput).

## Relevance:

- *DePaul & DICE*: Combines machine learning, temporal modeling, and governance research.
- *Industry (e.g., finance, data platforms)*: Safer, more autonomous infrastructure reduces operational risk and cost.
- *AI Safety*: Extends Geoffrey Hinton's "maternal instinct" concept, embedding protective bias and explainability into automated systems.

## First Milestones:

1. Construct benchmark datasets of event streams with labeled late/missing outcomes.
2. Evaluate baseline models (Poisson, Hawkes, survival) against learned DAG + SLA extraction.
3. Prototype a watcher that flags runs as MET/LATE/MISSING from historical patterns.
4. Publish initial findings in the context of safe automation for event-driven systems.

This research direction aims to bridge applied data science, infrastructure reliability, and AI governance. It offers both practical industry impact and a step toward principled frameworks for safe automation.