Title: **The Mnemosyne Protocol**: A Contextual Orchestration Framework for Generative Media
Author: Mert Kerem Salman
Affiliation: Mnemosyne Labs, Dubai & Istanbul
Date: February 19, 2026

Contact (Business): ks@mnemosynelabs.ai
Website: www.mnemosyneprotocol.org
Backup (Personal): keremsalman@gmail.com

Abstract

Generative AI models (LLMs and Diffusion Models) have achieved state-of-the-art results in isolated tasks but suffer from "Contextual Fragmentation" in temporal workflows. This paper introduces the **Mnemosyne Protocol**, a vector-based orchestration layer designed to maintain semantic and visual consistency across heterogeneous agent systems.

Unlike the Model Context Protocol (MCP)**[1]**–style context passing—where data is typically transmitted to a model session—Mnemosyne utilizes an "**Inverse Context Flow**" (**ICF**) architecture. This approach brings the model's reasoning capabilities to the local data environment, ensuring "**Local-First Sovereignty**" for intellectual property.

Preliminary simulations indicate up to a **~40% reduction in continuity hallucination rates** (measured against a baseline of stateless zero-shot prompting on N=100 sequential narrative frames). We propose a mathematical framework for "**Contextual Continuity**" and demonstrate its application in minimizing production rework while maximizing IP security.

Metric Definition: We define "Continuity Hallucination" as the sum of **Style Drift** (visual inconsistency > δ) and **Factual Constraint Violations** (e.g., character clothing changes, object permanence errors) per 100 sequential frames.
Here, δ (delta) denotes a calibrated perceptual-distance threshold (e.g., cosine distance in CLIP embedding space) used to flag style drift relative to a reference style pack.

[1] "MCP" refers to the open standard for connecting AI assistants to systems, as defined by the Model Context Protocol specification (https://modelcontextprotocol.io). Mnemosyne's "**Inverse**" **architecture** builds upon similar interoperability principles but reverses the data flow for **sovereignty**.

1. Introduction

The current paradigm of Generative AI is "stateless." Each prompt is an isolated event. For industrial media production, this is a fatal flaw. A character generated in Frame 1 often loses facial consistency by Frame 100. **The Mnemosyne Protocol** addresses this by introducing a persistent "State Layer" that **orchestrates heterogeneous agents**

(e.g., Claude for reasoning, Midjourney/Runway for visualization) under a unified context constraint.

2. Problem Statement: The Fitzgerald Paradox in AI

We define the core challenge as bridging "**Creative Chaos**" (Temperature > 0.7) and "**Algorithmic Order**" (Temperature < 0.2). Current systems force a trade-off: high creativity leads to low consistency, while high consistency leads to sterile, repetitive outputs.

- **Contextual Amnesia:** Agents do not share memory states.
- **IP Leakage Risk**: While enterprise policies vary, cloud-based inference fundamentally requires transmitting proprietary assets (scripts, storyboards) to third-party providers. Retention periods and training data usage are subject to changing vendor terms. Mnemosyne mitigates this by enforcing **Local Key Custody**, ensuring that core IP assets and cryptographic keys never leave the local orchestration node.

3. The Mnemosyne Architecture

The protocol operates on three primary layers:

- **3.1. The Context Vector** $(\boldsymbol{C_t})$**:** A dynamic, multidimensional representation of the narrative state (Time, Location, Mood, Character Arc) at any given time t.
- **3.2. Inverse Context Flow (ICF) & Security Model :** While standard industry protocols (like Anthropic's MCP) focus on exposing local data to remote models, **Mnemosyne** inverts this relationship to prioritize **Data Sovereignty**.

  Security Guarantees & Threat Model:

  - **Confidentiality**: Core IP and cryptographic keys (e.g., **Script Bibles**, **Character LoRAs**) remain strictly local; encryption is enforced at rest and in transit.
  - **Least-Privilege Context**: Redaction policies are applied before dispatching context to external agents, ensuring they only receive the minimum viable data (e.g., **preventing a Background Generator from accessing plot twists or character secrets**).
  - **Integrity:** State snapshots (S_t) are cryptographically secured via hash chaining to prevent unauthorized alteration.
  - **Auditability:** All inter-agent transactions and verification steps are recorded in an append-only log.
  - **Injection Resilience:** Mitigates prompt-injection attacks via tool sandboxing and schema-validation allowlists.

- **3.3. Multi-Agent Orchestration:** A supervisor agent assigns tasks to sub-agents based on their specialized capabilities (e.g., "Agent A: Generate Dialogue," "Agent B: Generate Background," "Agent C: Verify Consistency").

State Management & Terminology:

To ensure strict isolation between reasoning and memory, the protocol defines three distinct data planes:

- **Context Vector** (C_t): Global narrative constraints (e.g., world rules, core IP aesthetics).
- **Memory State** (M_t): Persisted episodic facts (e.g., character positions, previous actions).
- **State Snapshot** (S_t): The redacted, minimal-view data currently exposed to active agents.

4. Mathematical Framework

We formalize the generation process not as a discrete function, but as a continuous integration of context and memory over time. **The Mnemosyne Equation** is defined as:

$$\Psi(A_1, A_2, \dots A_n) = \int_{t=0}^{T} O(C_t, M_t) \cdot \prod_{i=1}^{n} Agent_i(S_t) \cdot dt$$

Product $\prod_{i=1}^{n} Agent_i(S_t)$ **:** Denotes the **compositional contribution** of specialized agents (Generation + Verification) operating under the shared snapshot S_t. This implies that a failure in one agent's context verification propagates through the chain, enforcing strict inter-agent dependency.

Notation Definitions:

- Ψ: The final coherent narrative output (Limit of continuity).
- O: **The Orchestrator Function** that penalizes deviation from the core context.
- C_t: **Context Vector** at time t (Global Truth: Location, Lighting, Tone).
- M_t: **Memory State** (Local Truth: Character positions, previous actions).
- S_t: **State Snapshot** shared visibly with agents.
- $\int_0^T$: Represents the enforcement of temporal consistency from Frame 0 to Frame T.

5. Conclusion & Future Work

The Mnemosyne Protocol shifts the **focus** from "better models" to "**better architecture**." We argue that we do not need smarter models to solve **continuity**; we need **stricter protocols**. This framework lays the foundation for the "**Operating System of Storytelling**," enabling a future where one creator can orchestrate entire productions with algorithmic precision.

Appendix A: Core Protocol Specification (Draft)

To formalize the orchestration, the Mnemosyne Protocol standardizes the following message types for multi-agent interaction:

- $PUSH_{STATE}$: **Orchestrator** broadcasts the **redacted** Snapshot (S_t) to generation agents.
- $PULL_{STATE}$: Agent requests specific **historical memory** (M_{t-1}) within strict constraint bounds.
- $LOCK_{STYLE}$: Verifier freezes specific CLIP embeddings (e.g., "**Facial Scar**") to prevent style drift during iteration.
- $VERIFY$: Orchestrator calls the **Verifier agent** to compute the cosine distance of generated output against M_t.
- $REDACT$: Policy engine strips non-essential IP data before external cloud inference.

(Mnemosyne Protocol Version: 1.4)

Keywords: Generative Media, Multi-Agent Systems, Contextual Fragmentation, Mnemosyne Protocol, AI Orchestration, MCP.

Note: **The Mnemosyne Protocol** is an independent research initiative for **generative media orchestration** and is distinct from the 'Mnemosyne Project' (spaced repetition software) or other similarly named memory tools.