[LAYOUT PREVIEW SAMPLE #2] template_id = 'cto_agent_runtime_report' (lang='en')

Prepared	Reviewed	Director	Approved
A. Kim Signed	J. Lee Signed	M. Park Reviewed	Pending

[Engineering Operations Report] Multi-Agent Runtime & Loop Control

Doc Ref: OPS-2026-1031-02 | Pipeline Scope: LangChain <-> CrewAI <-> Hermes Production Mesh

I. Engineering Environment Metadata (Customer Input: org_info)

[Architecture Metadata]

Field	Environment Specification (Preview)	Template Slot (org_info)
Engineering Team / Lead	Global FinTech Corp. (AI Platform Div.) / Jamie Lee, Staff Agent Architect	{company_name}, {author_name}
Review Period	2026-10-01 to 2026-10-31	{reporting_period}
Connected Frameworks	LangChain (SQL Agent) <-> CrewAI (Analyst Crew) <-> Nous Hermes (SKILL.md)	{connected_frameworks}
Deployment Topology	In-Process Python SDK (cdx-dock v2.0) + Kubernetes Docker Sidecar v2.0	{deployment_mode}

II. Runtime Stability & Measured Latency KPIs (Layout Preview)

[perf_counter() Measured]

1. Pipeline Runs & Completion Rate

412,800 (99.98%)

Includes any fallback events honestly

Slot Key: {total_pipeline_runs}

2. Measured Avg Middleware Latency

0.41 ms [Sample]

Measured via local perf_counter()

Slot Key: {avg_dock_overhead_ms}

3. <think> Reasoning Loops Cut Early

3,842 Events [Sample]

Prevents runaway stream timeouts

Slot Key: {cot_guard_loops_detected}

III. CoT-Guard Reasoning Control & TokenPack-10K Operational Metrics

[dock_telemetry.get_summary()]

Operational Metric	Measurement Definition & Preview	Local Telemetry Source	Template Slot Key (metrics)
Pipeline Runs & Completion	Total invocations and actual non-fallback completion %	summary['total_agent_calls'] summary['completion_rate_pct']	{total_pipeline_runs}
Avg Middleware Overhead	Real measured average latency (ms) across calls	summary['average_latency_ms']	{avg_dock_overhead_ms}
<think> Loops Detected	Repetitive reasoning stagnation events caught live	summary['cot_guard'] ['loops_broken']	{cot_guard_loops_detected}
Loops Terminated Early	Runaway reasoning streams halted prior to API timeout	summary['cot_guard'] ['loops_broken']	{cot_timeouts_prevented}
Reasoning Tokens Saved	Wasted reasoning stream tokens avoided via cutoff	summary['cot_guard'] ['tokens_saved']	{cot_wasted_tokens_saved}
TokenPack-10K Recall	Measured average SQLite recall latency (ms) & count	summary['tokenpack_recalls'] & average_tokenpack_latency_ms	{tokenpack_recall_latency}
Safe Pass-Through Fallbacks	Exact count of upstream exceptions routed safely	summary['fallback_calls']	{fallback_activations}

Engineering Operations Summary (Sample Narrative):

"Across frontier reasoning models (GPT-6 Astra, Claude Mythos 5.2, Gemini 4 Carbon), CDX-CoT-Guard intercepted repetitive <think> loops early, while local SQLite-backed TokenPack-10K synchronized shared context across frameworks without external vector database dependencies."

[Command] reporter.save_report('cto_agent_runtime_report', './runtime_ops.csv', use_live_telemetry=True, lang='en')