SYZ::ACE Sovereign AI Gateway — FOSS-First White Paper

Version 1.0 — October 2025

Executive Intent

Demonstrate that SYZ::ACE can operate a Sovereign AI Gateway — a unified entrypoint for AI/LLM traffic that runs entirely under SYZ control, exposes a standardized API, orchestrates multi-provider routing, and enforces policy, compliance, and observability natively via ACE subsystems.

Architectural Domains to Cover

Each domain must be independently verifiable using FOSS-only components. Core domains include ingress/API surface, routing engine, policy layer, observability, caching, cost ledgering, configuration, deployment, testing, and documentation.

Core Proving Points

We must demonstrate OpenAI-compatible proxying, dynamic multi-provider routing, policy enforcement (JWT/OIDC, rate limits, redaction), metrics collection, caching, cost ledgering, and deployability under full SYZ control.

Architecture Highlights

• FastAPI-based ingress implementing OpenAI-compatible endpoints. • YAML-configured routing engine with pluggable providers. • Policy layer for auth, redaction, and rate limiting. • Redis-backed caching, Prometheus metrics, and ledger hooks. • Docker Compose and Helm deployments. • All dependencies under OSI-approved licenses.

Proving Steps

1. Implement API proxy and OpenAI conformance tests. 2. Add routing and provider adapters. 3. Enforce JWT/OIDC and rate limits. 4. Integrate Redis caching and metrics. 5. Push usage to ACE::LEDGER. 6. Validate entirely offline control-plane operation.

Differentiation vs Proprietary Approaches

Unlike ngrok or other closed ingress systems, SYZ::AI_GATEWAY is: • Fully self-hosted and sovereign. • Integrated with ACE observability and ledger. • OpenAI-compatible and

provider-neutral. • Licensed under FOSS terms with reproducible builds.

Risk and Mitigation

Potential risks include token miscounting, provider exhaustion, latency inflation, and PII exposure. Mitigations include accurate token metering, key rotation, local caching, and redaction middleware.

Success Criteria

■ Works with OpenAI SDK unmodified. ■ Multi-provider routing verified. ■ Metrics live under /metrics. ■ Ledger entries populated. ■ 100% FOSS stack. ■ Deployable with single compose up.

Strategic Significance

This project positions SYZ as a sovereign ingress provider for AI traffic, showcasing the AIaC principle (AI-as-Code). It demonstrates autonomy, compliance, and sovereignty — establishing ACE as the backbone of verifiable, agentic AI infrastructure.