

v3.4-SYN-002 – Schema-Guided Data Generation & Domain Simulation Engines

Document Title	Schema-Guided Data Generation & Domain Simulation Engines
Version	v3.4
Document ID	v3.4-SYN-002
Date	2025-03-22
Author	Take Back Your Data – Synthetic Intelligence Group
Document Type	Public / Certification / Internal

1. Purpose & Scope

This document defines the logic for schema-guided synthetic data generation and the design of domain simulation engines in MaxOneOpen. It enables high-fidelity training and evaluation without using real-world data, ensuring compliance, realism, and safety.

2. Schema-Guided Data Logic

- All synthetic data must be derived from deterministic, schema-bound generation rules
- Schema blueprints define structure, token class, field logic and boundary conditions
- Data must conform to type constraints, distribution rules and entropy targets
- Simulation domains are modular and fully composable across training scopes

3. Simulation Engine Architecture

Simulation Layer	Function Scope	Audit Path
Domain Modeler	Environment + agent logic	Schema link + scenario hash
Interaction Layer	Multi-agent causality traces	Event chain + entropy seal
Reward Engine	Learning signal simulation	Anchor tokens + logic map
Output Filter	Safety, bias, realism layer	ZK-sealed moderation proof

4. Certification Hooks

- Forks must generate synthetic data through schema-sealed simulation paths
- All data must be replayable, traceable, and policy-compliant
- Simulation engines must expose modular scope, entropy control, and token validation

5. Certification Triggers

- Use of freeform or non-schema data disqualifies fork
- Simulation results without trace log or safety filtering invalidate certification

6. Certification Relevance

MaxOneOpen-certified forks must rely on deterministic, auditable simulation engines and schema-guided data generation. Training workflows must avoid real-world leakage while maintaining policy-aligned realism.