

**v3.4-SYN-001 – Synthetic Data Foundations & Realism Calibration Framework**

Document Title	Synthetic Data Foundations & Realism Calibration Framework
Version	v3.4
Document ID	v3.4-SYN-001
Date	2025-03-22
Author	Take Back Your Data – Synthetic Intelligence Group
Document Type	Public / Certification / Internal

**1. Purpose & Scope**

This document defines the foundational principles for the use of synthetic data in MaxOneOpen forks. It includes calibration mechanisms to ensure utility, realism, compliance, and safety without using real user data.

**2. Synthetic Data Generation Principles**

- All training and fine-tuning must rely on non-personal, synthetically derived content
- Synthetic data must be produced using schema-driven templates and logic engines
- Realism must be calibrated against observable context and public domain knowledge
- Leakage of personal data, biased structures, or real trace correlation is strictly prohibited

**3. Calibration Framework & Evaluation Metrics**

Calibration Metric	Validation Scope	ZK Assurance Path
Semantic Coherence	Contextual grammar + intent	Schema-driven duality checks
Structural Realism	Data shape, type, frequency	Trace-free randomness proof
Bias Absence	Social, political, cultural layers	Token distribution validation
Safety Compliance	Prohibited patterns filter	ZK-sealed moderation log

**4. Certification Hooks**

- Forks must prove all model inputs are fully synthetic and trace-free
- Realism metrics must be sealed, schema-bound, and transparently auditable
- Synthetic data must not replicate, emulate or echo real user behavior

**5. Certification Triggers**

- Use of non-synthetic, real user traces disqualifies fork
- Absence of ZK validation or realism calibration invalidates certification

## 6. Certification Relevance

Certified MaxOneOpen forks must build, train and validate models exclusively using synthetic datasets. Only forks with transparent calibration and data provenance qualify under the sovereignty principle.