

Projekt-ID. SDG-M3

Synthetic Data Generator (SDG) – Data Generation Engine Specification

Version: 2.0

Status: 100/100 Validiert

Executive Summary

The Data Generation Engine (DGE) forms the core of the Synthetic Data Generator (SDG) framework. It is responsible for creating high-quality, fully synthetic datasets covering textual, tabular, and structured formats. The DGE ensures that all generated data meets strict quality, diversity, and plausibility standards and operates fully in compliance with TBYD and MaxOne architectural principles.

Scope and Objective

This document specifies the architecture, functions, and validation criteria for the SDG Data Generation Engine. It defines the operational boundaries, interfaces, and quality control measures required for decentralized, privacy-first synthetic data generation.

Technical Background

The DGE is designed to operate natively on edge infrastructures, leveraging:

- Modular generation models (text, tables, structured data)
- Edge-local execution with no external dependency
- Zero-Knowledge compliance
- Adversarial noise simulation capabilities

Core Components

- Text Data Generator: Creation of natural language datasets
- Tabular Data Generator: Generation of structured datasets for tabular systems
- Structured Data Engine: Specialized for JSON, XML, and hierarchical data formats
- Diversity Injection Module: Enhances variability and representation
- Noise Simulation Module: Introduces adversarial and stress-test scenarios

Interfaces and Integration Points

The DGE interfaces include:

- MaxTune: Policy-controlled generation initiation
- Validation Engine: Immediate post-generation quality control
- MaxAudit: Event logging and validation result reporting
- MaxReg: Compliance checking during generation if applicable

Validation and Testing Criteria

Every generated data batch is tested against:

- Diversity Index Thresholds
- Plausibility Checks (Naturalness, Context Integrity)
- Bias Detection (Cultural, Linguistic, Topical)
- Robustness against Adversarial Perturbations

Compliance and Auditability

The Data Generation Engine fully adheres to:

- GDPR / DSGVO requirements
- Zero-Trust and Zero-Knowledge operational principles
- MaxOne security and modular audit standards
- TBYD 100/100 validation protocol for functional verification