

Framing Synthetic Validity – Understanding the SDG Logic beyond Data Mimicry

Version: 2.0

Status: Strategisches Kontextpapier

Projekt-ID: SDG-FRAME

1. Introduction: Beyond Defensive Narratives

This document is not a defense. It is a reframing. Synthetic data, as enabled by the SDG system, cannot be assessed through the lens of classical training datasets. The logic behind SDG represents a paradigmatic shift: from copying data to orchestrating generative contexts – fully documented, auditable, and controllable by the user.

2. From Static Data to Programmable Contexts

In the SDG framework, 'data' is not inherited from real-world sources. It is the procedural output of defined rules, plausibility standards, and adversarial robustness conditions. There is no direct data sourcing – only rule-based generation within tightly audited boundaries.

3. Validity is Constructed, Not Inherited

The SDG does not rely on prior truth to produce valid output. It builds validity from scratch – through a combination of validation cycles, bias resistance testing, diversity scoring, and continuous re-evaluation triggers. Every output can be traced back to its generating logic and tested against external standards.

4. The Role of the User: Authority Without Dependence

The SDG gives control to the user. The user defines the template space, the logic sets, and the operating parameters. The system merely enforces boundaries that prevent misuse – not content restrictions. This is not automation by default, but orchestration by design.

5. Auditability over Assumption

SDG outputs are not assumed to be trustworthy – they are provably auditable. The system logs, documents, and validates each generation process. Compliance is enforced at the

structural level, not as a post-hoc check. This model aligns with EU AI Act requirements and ISO 27001 audit expectations.

6. Addressing Public Misunderstanding

Skepticism toward synthetic data is valid – when synthetic means unregulated. But the SDG is different: every data point is traceable to its structural logic. There are no hidden sources, no copied records, no anonymized derivatives – only synthetic results derived from documented templates.

7. From Training Data to Rule Ecosystems

The SDG is not a static generator. It is the first modular, fully audit-ready system for generating dynamic training contexts at scale. It opens the door to future ecosystems of rule libraries, validated source models, and context-aware compliance engines.

8. Conclusion

Synthetic data in the SDG is not a shortcut – it is a new category of knowledge representation. The question is no longer whether data is synthetic or real – but whether it is auditable, intentional, and structurally valid. The SDG answers this with a system built from the ground up to prove it.