()

Why Defence AI needs synthetic data you can measure, not just generate

Artificial intelligence | Europe | United Kingdom

Type
Insight

Published
11 Mar 2026

Defence AI teams are turning to synthetic data because real operational data can be scarce, sensitive, or hard to move. But synthetic data only helps if you can show it represents the real conditions a model will face. A joint paper from Thales and Faculty

If you cannot evidence your training data, you cannot evidence your AI

Synthetic data is moving from a niche technique to a practical requirement in Defence AI. The reason is not convenience. It is constraint. Operational data can be sensitive by nature, platforms

sets out why data assurance must come before AI assurance, and what defendable evidence can look like.

may not be able to persist and return datasets, and ownership and storage rules can limit what can be shared and reused. Synthetic data can widen the training set without exposing what should stay protected, and it can help you represent conditions that are rare, denied, or difficult to capture at scale.

The risk is that synthetic data can also mislead you. Data that looks convincing to a person can still be statistically unlike the data a model will see in service. Models do not "see" like humans do. They learn patterns and draw a decision boundary, the point where the model flips from one class to another. If synthetic samples are too clean, too uniform, or too easy to separate, the model can learn the wrong boundary. It performs well on idealised data and then breaks when real-world overlap and noise show up.

From AI assurance to data assurance

If your goal is AI, you can deploy responsibly; assurance has to start earlier than most programmes expect. The joint paper's core point is simple. You cannot assure a model trained on synthetic data unless you first assure the synthetic data itself. Current approaches to synthetic data validation are inconsistent and often lack statistical rigour. That makes it hard to provide defendable evidence that synthetic examples are representative and equivalent to real examples for the intended use.

A useful way to think about assurance is fit for purpose. Data assurance methods that work for one dataset, sensor, or mission context may not transfer cleanly to another. The assurance question is not “is this data good?”. It is “is this data good enough for this use, under these conditions, with this model class?”.

What defendable evidence can look like

A practical assurance approach needs more than a visual check. The paper sets out a structured methodology that combines statistical measures and model-based tests to identify where synthetic and real distributions diverge, and what that divergence might mean downstream.

For example, statistical analysis of features such as frequency content and texture can reveal gaps between synthetic and real samples. In one study area, synthetic data showed lower spectral centroid than real data, suggesting the synthetic set carried less high-frequency content. This matters because many deep learning models are highly frequency sensitive. The point is not that a difference automatically makes the data unusable. It is those differences that help you target improvements and stop you making assumptions.

The second element is to test how a model actually separates classes when you mix real and synthetic. A common failure mode is that the model learns “real versus synthetic” as strongly as it learns “positive versus negative”. If that happens, you have learned a shortcut, not the task. The paper describes this effect using embedding analysis, where the data splits into distinct clusters by source. That is a warning sign for generalisation.

How to turn assurance into something repeatable

The end goal is not a one-off report. It is a repeatable pipeline that lets you benchmark synthetic data quality over time, track improvements, and build confidence that changes in generation methods do not introduce new blind spots. The paper recommends building automated assessment steps into the generation workflow so that quality and limitations are visible, comparable, and auditable across programmes.

There are also practical implications for how you generate data. Different generation methods have different failure modes, and those failure modes should be explicit assessment criteria. If a method tends to produce over-sharpened edges, tiling artefacts, or unrealistic uniformity, your assurance tests should be designed to detect those traits before they reach a training pipeline.

What happens next

Defence AI programmes need a shared way to discuss synthetic data quality that goes beyond “it looks realistic”. That means agreeing a small set of metrics and tests that can be explained to non-specialists, and that can be repeated across datasets and domains without exposing sensitive detail. It also means being honest about limitations, and using results to drive better generation, better augmentation, and better experimental design.

If you are working on AI assurance, test and evaluation, or data constraints, the question to ask is straightforward. What evidence would you accept to trust synthetic data as part of a training and validation pipeline, and how do we make that evidence portable between programmes without breaching classification boundaries?