Making Synthetic Data Work for Transaction Monitoring: The Next Frontier in Financial Crime Prevention

By Oonagh van den Berg, CEO and Founder VRS[1]

Published: 05 December 2025[2]

Contents

---

[1] https://www.linkedin.com/in/oonaghvandenberg/

[2] Link: https://www.linkedin.com/pulse/making-synthetic-data-work-transaction-monitoring-oonagh-5btrf/

Introduction

Financial crime is evolving at a speed and scale that traditional monitoring frameworks were never built to handle. Criminal networks are now global, adaptive, and technologically sophisticated — and our controls, historically reliant on static rules and retrospective detection, have struggled to keep pace.

Across banks, fintechs, and payment institutions, one uncomfortable truth is emerging:

If we wait to "see it in the data", we are already too late.

To future-proof Transaction Monitoring (TM), organisations now need a different strategy — one built on simulation, experimentation, and proactive intelligence. This is where **synthetic data** steps in as one of the most transformative tools in modern compliance.

What Synthetic Data Actually Is — and Why It Matters

Synthetic data is *not* "fake" data. It is **artificially generated data that faithfully preserves the statistical properties, behaviours, and patterns of real datasets — without containing any actual customer information.**

Think of it as:

- Realistic, but not real.
- Behaviourally accurate, but privacy-safe.
- High-volume, high-variety, and controllable.

It gives compliance teams the ability to:

- **Train models** without breaching privacy rules.
- **Stress test controls** under rare or emerging typologies.
- **Simulate criminal behaviour** before it appears in production data.
- **Collaborate with vendors and regulators** without sharing sensitive information.

This shift — from *detect once it happens* to *simulate before it happens* — is the step change AML has been missing for two decades.

Why Synthetic Data Has Become Essential

The digital economy has intensified both the *speed* and *complexity* of financial crime. Scams morph weekly. Mule networks expand across borders. Sanctions evasion routes re-route in real time. Traditional TM systems simply cannot generate enough variety, volume, or edge-case coverage to train modern machine-learning detection tools.

Synthetic data fills this gap by:

1. Modelling Modern Criminal Behaviour

It can encode:

- Complete scam lifecycles (grooming → deposits → cash-out).
- Mule account patterns.
- Sanctions-evasion pathways.
- Corporate-laundering networks.
- Cross-channel fraud behaviour.

And it can do this at scale — thousands of variants, instantly.

2. Solving the Data Scarcity Problem

Financial crime is low-frequency but high-impact. With synthetic data, teams can generate:

- Balanced fraud vs. non-fraud datasets,
- Extremely rare typologies,
- Never-before-seen combinations of behaviours.

3. Protecting Privacy While Unlocking Innovation

Regulators are increasingly recognising synthetic data as a **privacy-enhancing technology**, not a loophole.

The FCA, ICO, European Commission, and central banks are explicitly exploring synthetic data for supervisory sandboxes, model testing, and secure analytics.

Synthetic data enables collaboration without compromising personal data — a game changer for regulated institutions.

The Hidden Problem: We Already Have the Data — We Just Don't Use It

Most institutions collect **~1,200 onboarding data points** during corporate or retail KYC. Yet, on average, only **80–120** fields are used in TM or screening models.

Why?

Because most banks were built for:

- *Regulatory evidence*, not intelligence.
- *Tick-box KYC*, not behavioural analytics.
- *Static rule engines*, not feature-rich ML models.

Legacy systems can't ingest high-dimensional data, and onboarding systems, CRMs, document repositories, and TM engines rarely talk to each other. As a result:

We are data-rich but insight-poor.

This isn't a problem of data availability. It's a problem of architecture, culture, and capability.

Would Using All Real Data Solve the Problem?

Not entirely.

Even if we used every onboarding field:

- Criminals will still innovate faster than historical data can capture.
- Rare typologies will remain rare.
- Real data reflects yesterday's crime, not tomorrow's.

Synthetic data is therefore not a substitute — **it is a force-multiplier.**

Real data = What has happened. Synthetic data = What *could* happen.

Together, they build systems capable of anticipating risk rather than reacting to it.

The Regulatory Stance: Cautious, But Increasingly Supportive

Regulators are no longer resisting synthetic data — they are shaping its future.

- **The FCA** created a Synthetic Data Expert Group and published extensive guidance.
- **The ICO** is recognising synthetic data as a privacy-enhancing technology.
- **The European Commission** sees it as key to digital finance innovation.
- **Supervisors** are exploring synthetic data for testing models without exposing real customer information.

The message is clear: **Synthetic data is acceptable — if governance, validation, and documentation are strong.**

This means institutions must demonstrate:

- Statistical fidelity
- Privacy protection
- Utility for model training and testing
- Behavioural consistency across real vs synthetic datasets
- Clear technical documentation and challenge

The same model-risk governance used for AI and ML applies here.

Culture, Capability, and the Real Blockers to Progress

Most organisations still underestimate how **cultural** the blockers are:

1. Compliance teams are trained to protect data, not exploit it.

Data is seen as a liability, not an intelligence asset.

2. Tech teams often lack typology expertise.

They don't always understand how criminals think.

3. Incentives are misaligned.

Teams are rewarded for “no breaches”, not for detection quality or false-positive reduction.

4. Skills gaps persist.

Data engineering, ML literacy, and model-governance capability are rare within classic compliance functions.

Overcoming these barriers requires more than a new tool. It requires a transformation in mindset, ownership, and talent.

A 3–5 Year Roadmap to a Data-First Monitoring Model

Below is a future-proofing roadmap you can use in teaching, strategy design, and internal leadership discussions.

Phase 1 — Fix the Foundations

- Map onboarding and monitoring data.
- Clean and standardise key entities.
- Establish golden sources.
- Build data ownership and stewardship.

Outcome: A stable, trusted data layer.

Phase 2 — Intelligence Activation

- Create model-ready features.
- Integrate onboarding data into TM and screening.
- Improve segmentation.
- Reduce noise and enhance prioritisation.

Outcome: Better detection using the data you already have.

Phase 3 — Synthetic Data Adoption

- Generate privacy-safe synthetic datasets.
- Use synthetic data for: • scenario design • model testing • typology simulation • AI/ML training
- Validate fidelity, privacy, and model-behaviour consistency.

Outcome: Proactive, simulated crime detection capabilities.

Phase 4 — Regulator-Ready AI Governance

- Establish AI governance frameworks.
- Document synthetic-data processes.
- Monitor model drift.
- Evidence responsible innovation for supervisors.

Outcome: A resilient, modern, regulator-aligned financial crime programme.

Discussion questions we should all be having within our organisations now!

1. What would be possible if TM systems could simulate emerging crime before it appeared in the data?
2. How do we balance privacy, innovation, and regulatory expectations?
3. What skills must future compliance teams develop to operate data-driven ML models?
4. How can we reframe KYC from "document collection" to "intelligence generation"?
5. What governance should be in place before implementing synthetic data?
6. How can institutions embed adversarial, criminal-thinking modelling ethically?
7. Which cultural behaviours within compliance functions must change first?
8. What prevents institutions from using the data they already have?

Closing Reflection

Synthetic data is not the future of financial crime prevention — **it is the present, and the institutions who adopt it early will define the next decade of AML innovation.**

The shift we are facing is not merely technological; it is behavioural, cultural, and strategic. It requires new skills, new governance structures, and new thinking about how criminals operate.

But the opportunity is immense:

Synthetic data enables a world where detection is no longer reactive but anticipatory. Where TM is not bound by history, but empowered by simulation. Where privacy and innovation reinforce, not oppose, each other.

The future of AML belongs to the institutions that recognise one simple truth:

Crime is evolving faster than our systems — but with the right data, we can evolve faster still.