

IFC Workshop on Data Security and Privacy

Producing Shared Digital Asset Indicators Without Exposing Microdata

Privacy enhancing technology (PET) for shared regulatory and fraud modelling

Vincent Lee Wai Seng • Nur Aisyah Mohd Nasir • Chun Onn Pam

Bank Negara Malaysia

14–15 September 2026
Lisbon, Portugal

The views expressed herein are those of the authors and do not necessarily reflect those of the Central Bank of Malaysia.

Research Question and Presentation Outline

Research question

Can privacy-enhancing technologies support timely indicators for Malaysia's regulators and collaborative fraud modelling without disclosing the raw data of digital asset exchangers (DAX)?

Scope

- Regulatory indicators
- Federated learning
- POC via synthetic data

Structure

01

Context

Global crypto flows and data needs from Malaysia's DAX

Slides 3–4

02

Proposed approach

Two use cases leveraging on trusted execution environment (TEE)

Slides 5–7

03

Proof of concept

Methodology and evaluation framework

Slides 8–10

04

Conclusion

Limitations and path to pilot

Slide 11

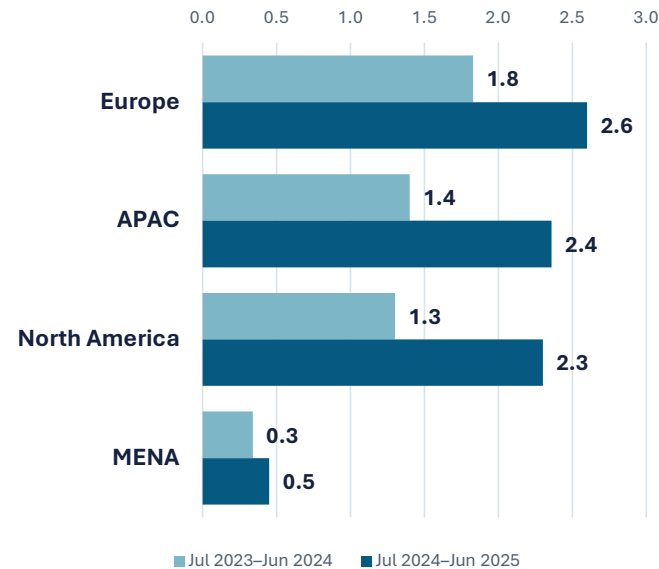
Global digital asset growth sharpens Malaysia’s visibility challenge

Europe remains largest and APAC grows fastest, raising Malaysia’s need for consistent digital asset (DA) data across authorities

REGIONAL SCALE AND GROWTH

Estimated on-chain value received (US\$tn; approximate)

Jul 2023–Jun 2024 vs Jul 2024–Jun 2025



MALAYSIA DA INFLOW PROFILE

(Source: Chainalysis)

Top source countries for DA inflows (Est.)

United States	28.9%
Japan	6.6%
India	5.7%
South Korea	5.6%

CENTRALISED DIGITAL ASSET EXCHANGES (DAX)

Est. **50.6%** of Malaysia inflows are facilitated by DAX (Jan 2024–May 2026)

Token share of Malaysia DA inflows (Est.)

49.1%	15%	35.9%
Stablecoin	BTC, ETH	Others

STRUCTURAL SHIFTS AT GLOBAL LEVEL

Stablecoins	USDT and USDC increasingly support remittances, payments, and settlement.
Cross-border use	Flows include household remittances, business payments, capital transfers and exchange-to-wallet movements.
Participation	Retail investment remains strong in emerging markets; institutional activity has risen with ETFs and regulatory clarity.

Limitation of digital asset data providers: Data by geographical area is estimated from clustering, attribution and web traffic sources.

PROBLEM

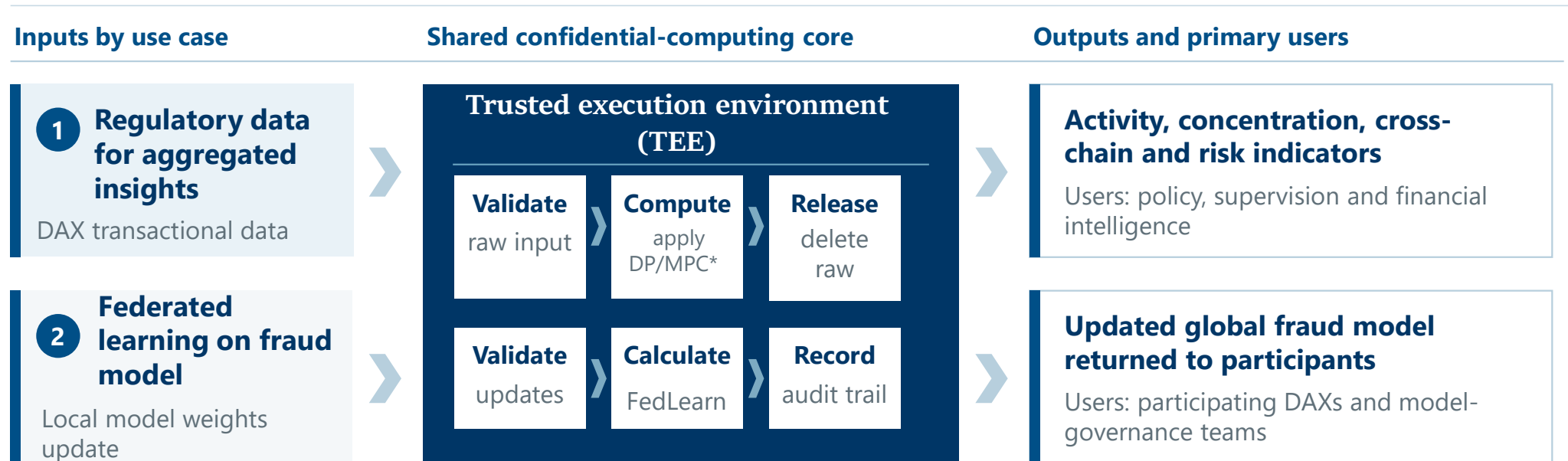
Malaysia's regulators need tailored views of DAX activity

Securities Commission Malaysia (SC) licenses DAXs; Bank Negara Malaysia (BNM) needs policy, supervisory and financial intelligence views.

Shared DAX data	Mandate-specific regulatory perspective		If data is requested separately
DAX Activity Transactions Cross-border links Accounts & wallets	➤ SC Malaysia	Licensing under capital market framework regulates and supervises DAXs, digital asset market conduct, investor protection, compliance	1 Fragmented visibility Each authority sees only part of the activity
	➤ BNM- Policy	Cross-border DAX surveillance assess the potential implications of digital asset activities on monetary and financial stability.	2 Duplicated reporting DAXs prepare overlapping extracts
	➤ BNM – Financial Intelligence	Financial crime & digital-asset fraud account-level investigations of financial crime and illicit digital asset activities.	3 Sensitive-data exposure Similar account data moves repeatedly
	➤ BNM – Supervision (potential future role)	Emerging risk patterns more to monitor stablecoin activity, since DA is not yet recognised as payment instrument in Malaysia	4 Siloed detection Linked cross-platform patterns may be missed
Proposed solution	A common data foundation can support distinct regulatory and supervisory views without repeating the data transfer.		

Proposed confidential-computing architecture

Both use cases rely on the same attestation, access and output controls.



* DP= Differential privacy (e.g. Laplace/Gaussian noise), MPC = secure multi-party computation

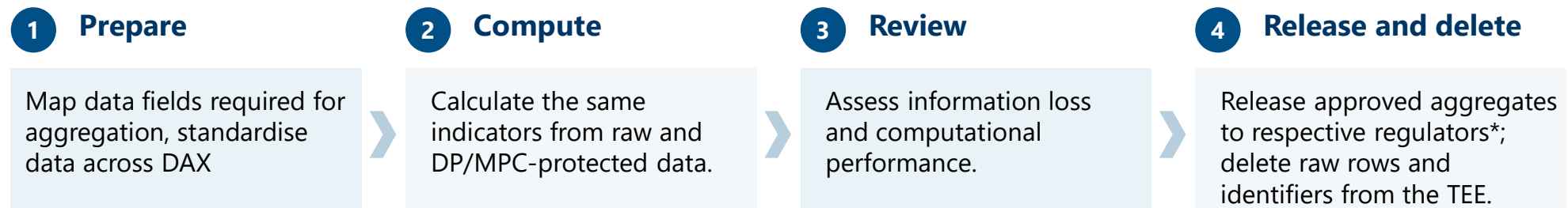
What is TEE?

A trusted execution environment (TEE) is like a locked room inside a computer: sensitive data and code can be processed there without being exposed to the rest of the system. It also facilitates remote attestation, control access and audit trail.

Use case #1- Producing regulatory indicators

Approved queries are run on DAX records; only reviewed aggregates are released.

From approved query to controlled release



*Output for regulators

Policy and statistical analysis

- Total aggregated transaction value and volume (time series)
- Breakdown by DAX, by digital asset/token type, transaction type
- Top corridors for digital asset transfers
- Total digital asset holding value/balance by digital asset/token type

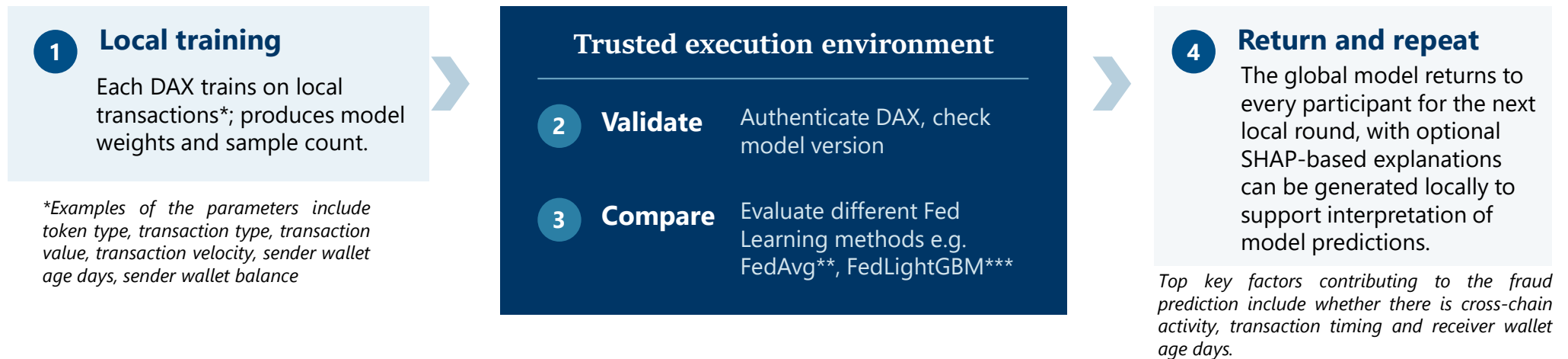
Potential supervisory and financial-intelligence analysis

- Average anomaly scores per transaction
- Transaction velocity trends
- High risk corridors for digital asset transfers
- Total number of accounts

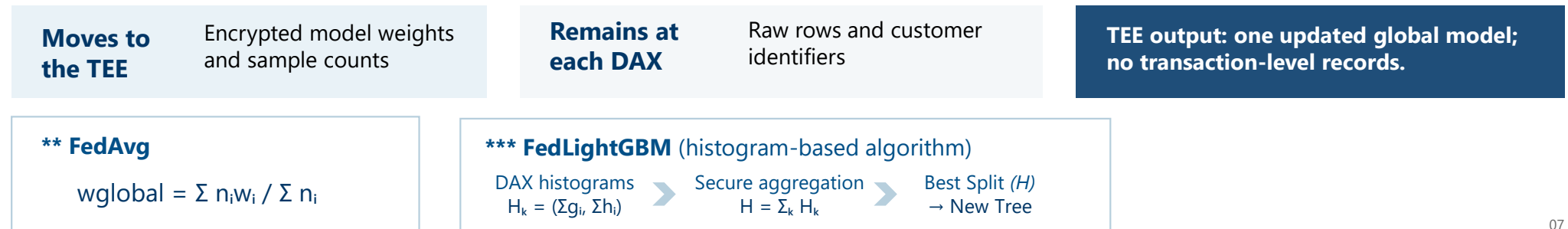
Use case #2- Federated model aggregation for fraud detection

Each DAX trains its fraud detection model locally; the TEE validates and aggregates model updates via federated methods.

A repeating federated-learning round



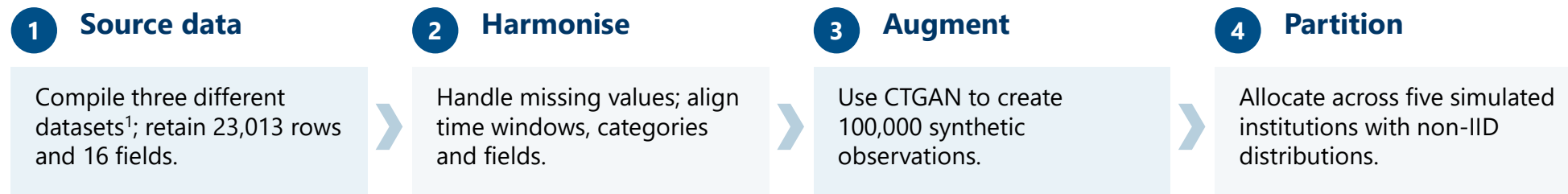
Data boundary



PoC methodology and evaluation

Synthetic digital-asset transaction data are used to test both use cases.

Build the synthetic test environment



One prepared dataset feeds two evaluation tracks

Part 1 · Regulatory indicators

Before aggregation: apply DP with Laplace noise or multi-party computation (MPC)² → produce identical regulatory indicators → compare both with the unprotected reference

Evaluation: information loss and performance in applying DP vs MPC

Part 2 · Federated learning

Train identical local models using the data items → aggregate validated updates with federated approaches (FedAvg, FedLightGBM) → compare federated, siloed and centralised baselines

Evaluation: model prediction accuracy (silo vs federated)

Sources:

1. Kaggle Crypto Dataset; GitHub Crypto Fraud Dataset; BNM Cash Balance of Payments Data;
2. Schoenmakers (2020);

Choosing the appropriate TEE infrastructure for the PoC

The platform choice prioritises enterprise fit and implementation effort, not only hardware capability.

TEE options considered

Cloud provider	How data is protected	Hardware options	Changes needed for the PoC	Fit with existing environment
Azure	Protects the entire virtual machine	AMD SEV or Intel TDX processors	Works with the current setup; little or no redesign of existing codes	Best fit; works with the existing enterprise environment
Google Cloud	Protects the entire virtual machine	AMD SEV or Intel TDX; confidential GPU	Simple setup with little redesign	Would add another cloud environment to manage
AWS	Creates a separate locked area inside a server	AWS Nitro	Application must be split and redesigned	More complex to build and operate
Alibaba Cloud	Protects the virtual machine or a separate secure area	Intel TDX, separate enclave	Simple setup with little redesign	Less aligned with the current environment

PoC evaluates trade-offs across utility, performance and model accuracy

Comparing information loss, computing requirements and prediction accuracy scores across privacy-enhancing approaches.

Evaluation Metrics (on the synthetic dataset)

Use-case 1

1 Indicator utility

Laplace DP incurs information loss while MPC preserves accuracy

Method	Info loss (% mean diff)
No PET	Nil
Gaussian*	Median % diff: ~1.07%
Laplace*	Median % diff: ~0.46%
MPC**	Nil

2 Computing performance

Computation power required doesn't differ much across PETs

Method	Memory usage
No PET	Median ~137.27 MB
Gaussian*	Median ~427 MB
Laplace*	Median ~397 MB
MPC	Median ~431 MB

Use-case 2

3 Fraud model accuracy

Federated approach improves the fraud detection scores, as compared to silo local model of DAX

Model approach	AUC	Recall
Silo local DAX (average)	0.7433	75.3%
FedAvg	0.7526	76.9%
FedLightGBM	0.7596	80.1%

* Evaluated based on epsilon ranging from 0.1 to 20 [the lower the epsilon, the higher the noise is introduced (higher privacy)]

** MPC preserves the original aggregate value (no noise added) while providing stronger confidentiality during computation.

CONCLUSION

Constraints, future considerations and next step

A scoped pilot can test which PETs are practical for the use cases based on actual data.

Solution constraints		Future considerations	
DP/MPC: utility and performance	Finding the right balance between privacy, accuracy and computing resources [1]		Hybrid DP + MPC Combine shuffle-model DP with MPC to reduce central trust and improve utility for distributed analytics [7].
TEE: residual attack surface	Side-channel and rollback risks remain; patching and protected state are essential [2–3]		Advanced confidential computing Ensure TEE is embedded with rollback-resistant storage [8].
FL: data and communication	Non-IID data can affect convergence; communication is a key constraint [4]		Adaptive FL optimisation Explore federated learning optimisation techniques (e.g. FedProx) to handle non-IID data with fewer communication rounds [9].
FL: leakage and poisoning	Gradients can leak training data; model poisoning by malicious participants [5–6]		Verifiable robust FL Extend secure aggregation with zero-knowledge update validation and norm bounds to limit leakage and poisoning [10].
What's next: Potential pilot	Beyond the PoC, BNM, SC and participating DAXs could test the approach in a potential pilot using actual data.		

References: [1] Kairouz et al. (2017); [2] Fei et al. (2021); [3] Matetic et al. (2017); [4] McMahan et al. (2017); [5] Zhu et al. (2019); [6] Bagdasaryan et al. (2020); [7] Brandao & Peralta (2021); [8] Chu et al. (2025); [9] Li et al. (2020); [10] Lycklama et al. (2023).