BIS Innovation Hub

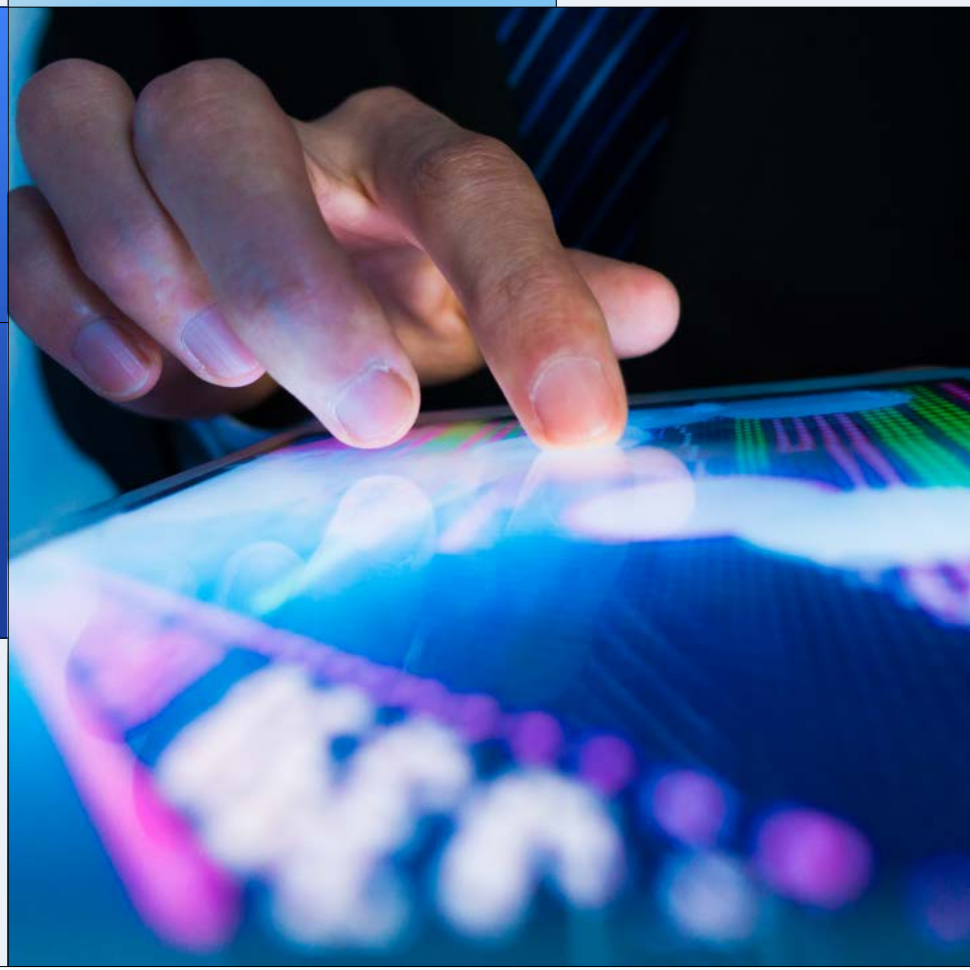
Bank of England

Project Hertha

Identifying financial crime patterns in real-time retail payment systems

Contents

- **Executive summary**
- **1. Motivation and hypotheses**
- **2. Results**
- **3. Key insights**
- **4. Areas for further research**



Executive summary

Project Hertha is a joint project between the BIS Innovation Hub's London Centre and the Bank of England. The project explored how transaction analytics could help identify financial crime patterns in real-time retail payment systems, while using the minimum set of data points.

Motivation

Combatting financial crime is essential to maintaining trust in the financial system. It has been estimated[1] that $3 trillion of money laundering and terrorist financing flow through the global financial system every year. Addressing this is increasingly urgent as **new technologies are also enabling new financial crime threats.**

To evade detection, **criminals operate in complex networks** which include many accounts across multiple financial institutions. Earlier initiatives, including the BIS Innovation Hub's Project Aurora, demonstrated the potential of network analytics to identify this activity in network-wide data.

Electronic payment systems process transactions across many participants, which gives them a network-wide view. **Project Hertha tested the application of modern artificial intelligence (AI) techniques to help spot complex and coordinated criminal activity in payment system data.** It measured the added value of such transaction analytics relative to a modelled benchmark of banks and payment service providers (PSPs) monitoring accounts in isolation.

Synthetic data set

The experiments were conducted using a **complex simulated synthetic transaction data set,** developed as part of the project. It includes data on 1.8 million bank accounts and 308 million transactions. The data set was built using an AI model trained to simulate realistic transaction patterns. While no real customer data was used in the exercise, the data set was designed to be representative of an ecosystem of retail payments in a single jurisdiction.

1 **Nasdaq Verafin**, Global financial crime report, 2024.

Executive summary

Findings

Project Hertha found that payment system analytics could be a valuable supplementary tool to help banks and PSPs spot suspicious activity. Key findings from the project include:

- Working in isolation, **payment system operators identified fewer illicit accounts** relative to banks and PSPs (39% vs 44%).
- Using findings from payment system analytics **helped banks and PSPs find 12% more illicit accounts** than they would otherwise have found.
- Payment system analytics was **particularly valuable for spotting novel financial crime patterns.** When trying to spot previously unseen behaviours, it helped achieve a 26% improvement.

The results have been achieved while using a minimal number of data points, demonstrating that **advanced models can effectively draw on network patterns rather than personal data.** They also assume that no private data are shared with the payment system operator.

The results demonstrate promise, but also show there are limits to the application and effectiveness of system analytics. It is just **one piece of the puzzle.** The introduction of a similar solution would also raise complex practical, legal and regulatory issues. Analysing these was beyond the scope of Project Hertha. The concept explored in the project does not assume any changes in the responsibilities of individual institutions.

Key insights

Results have also pointed at a few helpful practical insights:

- Payment system analytics proved **most effective when targeted at identifying more complex schemes** involving many accounts across different banks and PSPs. For some schemes, it doubled detection accuracy.
- To achieve the best results, **algorithms need to be trained on confirmed past cases.** Unsupervised algorithms were found to be far less effective.
- Likewise, the ongoing effective operation of payment system analytics requires **banks and PSPs to continuously provide feedback on outcomes** for accounts flagged by the model.
- **Explainable AI approaches could provide additional valuable information** to aid banks and PSPs in investigations and reporting, such as reasons why an account was flagged.

Further experiments could test similar approaches for cross-border and large-value payment systems as well as cryptoasset networks. These were out of scope for Project Hertha.

// The results have been achieved while using a minimal number of data points, demonstrating that advanced models can draw on network patterns rather than personal data. //

Section 1

Motivation and hypotheses

Collaboration between financial institutions is essential to combat the rise of financial crime. Project Hertha focused on the role of electronic payment systems.

Motivation and hypotheses: Background

Combatting financial crime is essential to maintaining trust in the financial system. Financial institutions have a critical responsibility to identify and prevent financial crime. However, they are often unable to address financial crime threats working in isolation.

Criminals often use complex chains of transactions across different financial institutions and payment methods to conceal illicit activity. This fragmentation can make it challenging for any individual institution to identify and prevent these activities. While there are notable examples of collaboration,[2] most financial crime prevention efforts continue to happen in silos.

Payment system operators have visibility on the flow of funds across all financial institutions within their network. These network-wide data have the potential to be used to help banks and payment service providers (PSPs) spot patterns that they might not be able to detect individually.

New technologies provide improved opportunities to identify and prevent financial crime, while balancing the competing objectives of protecting privacy and managing operational costs. Advanced artificial intelligence (AI)-enabled models can help to identify complex patterns in the data. Meanwhile, synthetic data generation can enable training models more effectively, particularly where there are legal or practical barriers to obtaining real data.

It is estimated that $3 trillion of illicit funds are laundered through the financial system every year globally.[3] Consumers and businesses have faced $485 billion in losses from payments, cheque and credit card fraud.[4]

It is also likely that the vast majority of illicit activity remains unreported and undetected. Europol has estimated that countries currently intercept and recover less than 2% of all illicit fund flows.[5]

New technologies are enabling new threats to the integrity of the financial system. For instance, generative AI can facilitate impersonation scams and forge electronic documents. Novel forms of payment (such as cryptoassets) are used for money laundering through privacy coins and mixers.

2 **The Future of Financial Intelligence Sharing,** Case studies of the use of privacy preserving analysis to tackle financial crime, January 2021.

3 **Nasdaq Verafin**, Global financial crime report, 2024.

4 **Nasdaq Verafin**, Global financial crime report, 2024.

5 **Europol,** The other side of the coin: an analysis of financial and economic crime, 2024.

Motivation and hypotheses: Research question

Project Hertha explored how transaction analytics could help identify financial crime patterns in real-time retail payment systems, while using the minimum set of data points.

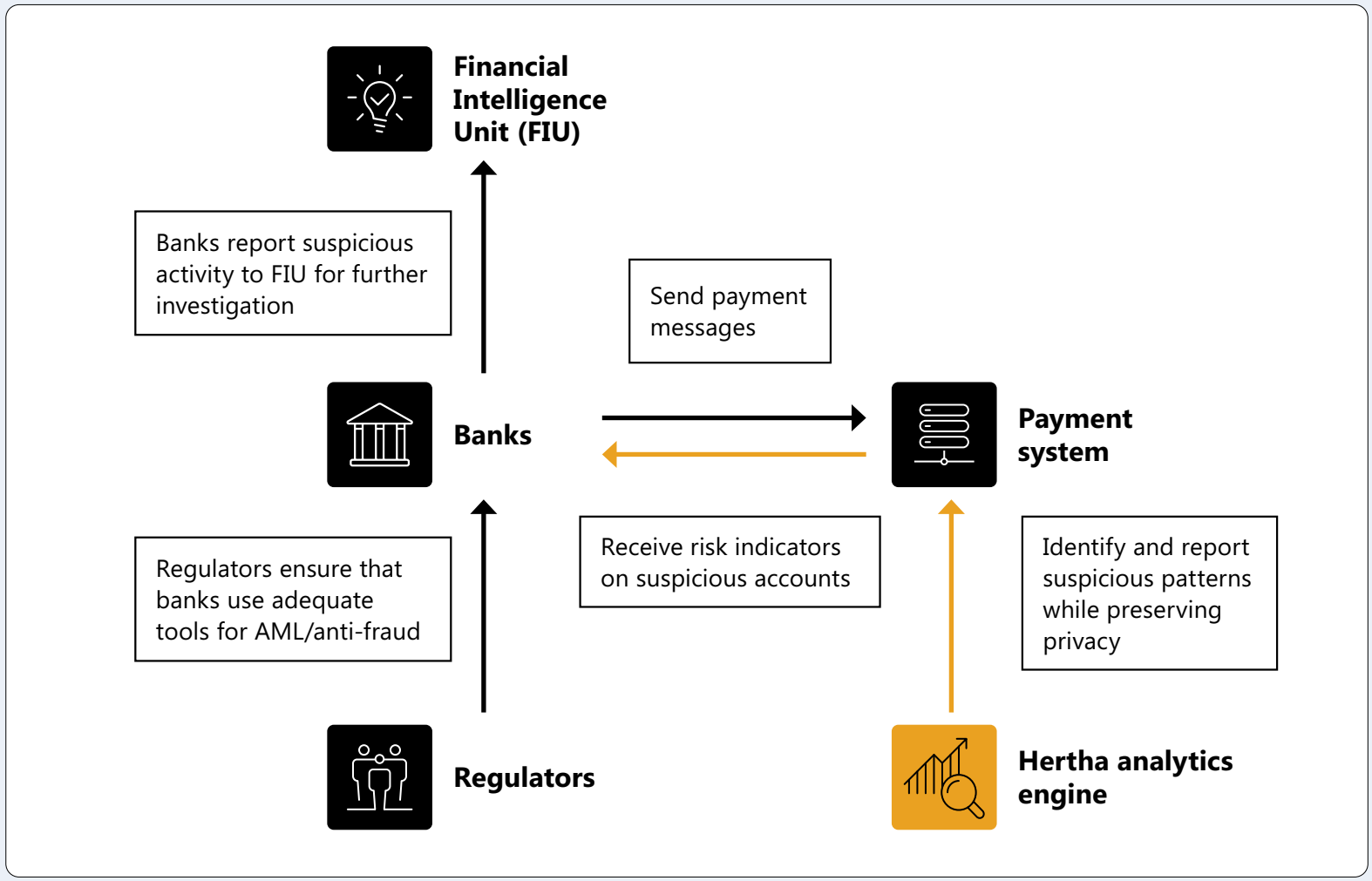
Purpose and approach

In most jurisdictions, banks and PSPs are responsible for monitoring their customer activity for the purpose of financial crime prevention. When suspicious activity is identified, they will typically refer this to an internal investigation team and/or request additional information from customers. Confirmed suspicious cases may then be reported to local financial intelligence units (FIUs).

Project Hertha assumed no changes to these common roles and responsibilities of banks, PSPs and payment system operators.[6] Instead, the project tested how payment systems could support banks/PSPs by identifying network-wide patterns which are not visible to a single institution. Graph 1 illustrates the concept that was tested experimentally in Project Hertha.

6 Analytics on payment system data could also be conducted by the technical infrastructure provider or outsourced to a third party.

Graph 1: Concept tested in project Hertha



Motivation and hypotheses: Research question

The hypothesis is that these network-wide risk indicators can help banks and PSPs target their investigations much more accurately. This could have varied benefits, including for instance:

- **Identifying financial crime:** improving the accuracy of alerts indicating suspicious activity to identify more financial crime.
- **Improved service for customers:** minimising disruption to legitimate customers that may be falsely flagged by the transaction monitoring systems.
- **Reduced cost of compliance:** focusing alerts more efficiently to help banks and PSPs to reduce the time spent on investigating false positives.
- **Safeguarding user privacy:** using as little information as possible to safeguard legitimate customers' privacy rights.

To demonstrate this concept, Project Hertha developed and tested a range of transaction analytics models in a realistic synthetic setting. **Section 2** outlines the results obtained relating to the potential effectiveness of a transaction analytics solution. **Section 3** explains the solution concept and helpful practical insights identified. Finally, **Section 4** flags potential areas for further exploration.

Scope of the project

The project focused on the following scenario, which should be taken into account when interpreting the findings.

- **Retail payment systems.** Business accounts were not explicitly simulated in the project. The findings are expected to be applicable only to retail (low-value) payment systems.
- **Single jurisdiction.** The project tested the application of network analytics on a single jurisdiction's payment ecosystem. There are currently only a few retail payment systems that operate across borders.
- **No external data.** The project has focused on identifying network patterns with the minimum set of data points to preserve user privacy. It did not test linking payment system data to any external information that regulated entities would be expected to use in their monitoring (eg lists of sanctioned entities, politically exposed persons or company beneficial ownership records).

Financial crime modelled

Financial crime is defined as any criminal conduct relating to money, financial services or markets. It can encompass a broad range of activities, such as money laundering, fraud and terrorist financing. These activities are often deeply intertwined. For instance, proceeds of fraud are often laundered through the same criminal networks as the proceeds of other crimes.

Project Hertha focused on detecting money laundering schemes in network-wide data. However, the methods used in the project could be applied to the detection of any financial crimes conducted by criminal networks using electronic payment methods. The money laundering schemes modelled can represent many types of predicate offences, including consumer fraud.

In the report, "illicit activity" is used as a general term to encompass financial crimes that can be identified through transaction and account-level data.

Motivation and hypotheses: Methodology

The concept was tested using a complex and realistic synthetic data set, representing a national payments ecosystem. The data set enabled testing effectiveness of payment system analytics relative to the benchmark of banks/PSPs working in silos.

Synthetic data

To ensure that the findings are representative, it was essential to test the solution in a realistic setting. This required rich transaction data, representing a large number of banks and PSPs, and the inclusion of labels for illicit accounts and transactions. As there is no existing source of equivalent real transaction data, we were required to generate data artificially.

The project team developed a large and realistic simulated synthetic data set. It includes data for 1.8 million accounts and 308 million transactions over one year.[7] The data set seeks to model a national retail (low-value) payments ecosystem.

While the data set focused on a specific country (the United Kingdom), we expect the findings to be applicable to retail payment systems in most jurisdictions.

The data set was developed by applying a combination of generative AI models to anonymised bank transaction and account data. AI models ensured that realistic patterns and complexity are preserved, while fully safeguarding data privacy.

The simulation was conducted in six stages:

1. A model was trained to generate synthetic transaction histories based on account features (eg income).
2. A universe of artificial customers was created to be representative of a wide range of real UK economic and transaction statistics. This included data on consumer income, spending, demographics and use of financial services (eg borrowing).
3. The model then simulated transaction histories for these artificial customers.
4. The customers in the data were assigned to one of eight artificial banks representing different market segments (eg local or digital banks).
5. The resulting data were validated against real payment statistics and iteratively improved to match real-world distributions.
6. An agent-based model was applied to the data to model financial crime schemes and "complete" the network (ensure that each transaction has a counterparty).

As a result, while no real transaction or account data were used in experiments, we expect the findings to be applicable to a real-world scenario. We also expect this data set to be a valuable resource to help the community develop and benchmark improved models (similar to recent efforts by AMLSim[8] and SparNord[9]).

However, it is also important to note the limitations of this data set:

- It does not include any data on corporate accounts or business-to-business transactions.
- Financial crime networks are modelled based on expert input rather than being available in the source data (see page 9).

> "The data set was developed by applying a combination of generative AI models to anonymised bank transaction and account data."

7 The transaction volumes modelled in the data set are still considerably smaller than transaction volumes in a retail payments ecosystem in a major economy eg the United Kingdom. This was to enable more effective experimentation. The methods tested can be scaled to larger volumes.

8 **E Altman et al,** Realistic synthetic financial transactions for anti-money laundering models, IBM Research, December 2023.

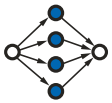
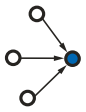
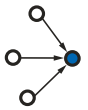
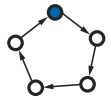
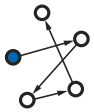
9 **R Jensen et al,** "A synthetic data set to benchmark anti-money laundering methods", Scientific Data, no 10, article number 661, September 2023.

Motivation and hypotheses: Methodology

Modelling financial crime networks

Our synthetic data set includes 2000 simulated money laundering schemes, representing 10 typologies (common patterns, techniques or behaviours). The modelling was based on published reports, past academic work and expert input from our stakeholders. We sought to credibly model diverse ways in which money can be moved through a criminal network in an attempt to evade detection. Graph 2 provides a list of the financial crime typologies included.

Graph 2: Financial crime schemes modelled

	Pattern	Risk factors
Typology 1 Gather-scatter		High amounts from digital wallets Round transaction amounts Rapid fund movement
Typology 2 Scatter-gather		High amounts from digital wallets High volume of cross-border transactions Same counterparties
Typology 3 Stack		High volumes from digital wallets Round transaction amounts Several opposite and similar transactions Large cross-border transactions
Typology 4 Fan-out		High cash deposits New account Abrupt change in behaviour High activity with risky countries
Typology 5 Fan-in		High volume of transactions High volumes cross-border Withdrawal in foreign countries
Typology 6 Fan-in		High amounts transacted with same people Large cash withdrawals Deposit and withdraw similar amount Several opposite and similar transactions
Typology 7 No network pattern		Very large transactions Large cash withdrawals
Typology 8 Simple cycle		High volume of cash deposits High cash deposits Several opposite and similar movements
Typology 9 Stack		High activity with risky countries New account and abrupt change in behaviour Rapid fund movement
Typology 10 Random		High total amount and volume deposited Single very large transaction High amounts transacted with same people

Motivation and hypotheses: Methodology

In our training data, 1% of generated accounts and 0.04% of generated transactions represent financial crime. We expect these to be realistic proportions based on expert input and existing statistics.[10]

Traditional machine learning models struggle when trained on extremely imbalanced data (where target behaviour is very rare). The data set therefore provides a complex and realistic environment in which to test advanced models.[11]

Metrics

The objective of a detection model is to achieve a balance between identifying as many illicit accounts as possible (*recall*) while avoiding false positives, ie not flagging legitimate customers (*precision*).

The models tested can be calibrated to either focus on a small number of the highest risk accounts (maximise precision) or identify as many illicit accounts as possible (maximise recall). The chosen share of the highest risk accounts flagged is defined as the *target rate*.

We measured the percentage of illicit accounts correctly identified under all scenarios using the same target rate of 1%. This ensures a like-for-like comparison between scenarios[12] and means that a higher percentage of illicit accounts identified always implies a corresponding reduction in false positives.

We also used a combined metric called average precision that summarises model performance at any target rate.[13] It ranges from 0 to 100%. Average precision of above 30% is considered high for similar tasks, based on industry feedback.

Experimental setting

Methods used by illicit actors in the real world evolve rapidly to evade detection. This requires financial institutions to continuously adapt their systems to identify new behaviours. We ran two types of experiments in the project to test how this can be done:

1. **Known typologies.** Targets for seven out of 10 typologies were used to train supervised models on first nine months of the year. We evaluated the model performance on spotting these typologies in the remaining three months.
2. **New typologies.** Targets for three out of 10 typologies were not available to the model. We tested the performance of both **unsupervised** and **supervised** approaches to identifying these schemes without past targets to draw on.

While we tested a large number of models, the results presented are from the three best-performing models: XGBoost (supervised machine learning), Isolation Forest (unsupervised machine learning) and UniTTab (supervised deep learning).

Data available to each party

Our synthetic data include both transaction and account data. In our experimental set-up, banks and payment systems have access to different data fields, reflecting their different roles:

- **Banks and PSPs** hold all data relating to their customers across all payment methods.
- The **payment system operator** can only see a limited set of transaction data points within its system: time, amount, purpose and sender/receiver's pseudonymous identifiers.[14]

This enabled measurement of the added value of transaction analytics relative to the baseline of banks/PSPs working in isolation.

10 **European Banking Authority,** Report on Payment Fraud, 2024.

11 Traditional machine learning models perform worse at tasks where the target behaviour is very rare – as they learn primarily from negative rather than positive examples. A primer can be found here Datasets: Imbalanced datasets.

12 This means that all models flag the same number of accounts. Better models will flag a higher number of accounts correctly, and by definition will have fewer false positives.

13 Technically, AP measures the area under the precision-recall (PR) curve into a single value that represents the average of all precisions across different recall levels.

14 In practice, many payment systems will also have access to personal data for the transactions, including names and addresses. However, using these data appropriately would require them to link the data to external sources, which could raise privacy concerns. This was out of scope for the project.

Section 2

Results

Payment system analytics was found to be a valuable tool to help banks and PSPs identify more financial crime patterns.

Results: Overall effectiveness

Payment systems were found to be effective in identifying financial crime schemes but on average, performed worse than banks and PSPs.

First, we assessed how effective banks/PSPs and payment systems can be when trying to identify financial crime schemes in isolation. Table 1 summarises the results that have been achieved while using equivalent machine learning models (XGBoost[15]) for each scenario.

Both banks/PSPs and payment systems achieved good levels of performance. For known typologies, banks correctly flagged 52% and payment systems 48% of illicit accounts. Meanwhile, for new and emerging typologies, banks found 33% and payment systems 31%.

Banks/PSPs were found to be marginally more effective in spotting illicit accounts relative to payment systems. This is an expected finding. While banks and PSPs do not see the whole payment network, they have access to much more data about their customer's identity and their activities across all payment methods.

Table 1: Headline results for detection effectiveness

Result	1. Bank/PSP	2. Payment system	3. Collaboration	% Improvement relative to (1) Bank/PSP
Average precision				
Known typologies	0.52	0.43	0.55	+6%
New typologies	0.12	0.10	0.15	+26%
Overall	**0.44**	**0.40**	**0.50**	**+12%**
Percentage of illicit accounts identified				
Known typologies	0.52	0.48	0.57	+9%
New typologies	0.33	0.31	0.41	+26%
Overall	**0.44**	**0.39**	**0.49**	**+12%**

Payment system operators, on the other hand, rely on limited transactional data.

How do these rates compare with the real world? It is challenging to set an objective benchmark. The vast majority of financial crime is never identified by the financial system. Accordingly, only indirect estimates as to the true scale are available.

Europol has estimated that less than 2% of laundered funds currently get recovered.[16] Reports have suggested that many bank alert systems will have false positive rates as high as 95%.[17]

Relative to these metrics, results achieved experimentally in project Hertha can be considered successful.[18] However, achieving similar real-world results is dependent on a number of conditions, including the availability of high-quality training data.

15 **T Chen and C Guestrin,** Xgboost: a scalable tree boosting system, March 2016.

16 **Europol,** The other side of the coin: an analysis of financial and economic crime, 2024.

17 **B Oztas et al,** "Transaction monitoring in anti-money laundering: a qualitative analysis and points of view from industry", Future Generation Computer Systems, vol 159, October 2024.

18 There are additional complex reasons why similar rates are not currently achieved in real-world settings, which are beyond the scope of this report.

Results: Combining network and transaction analytics

Using payment system insights can help banks and PSPs to be more effective in spotting financial crime.

The main hypothesis of the project was that payment system analytics can add value to each individual bank's monitoring systems. Testing it required assessing how much more financial crime can be detected through collaboration. How could banks and PSPs best utilise network-level insights from the payment system? Can it help them identify additional illicit activity that they do not currently capture?

The project tested three possible options:

1. **Blind trust.** Banks/PSPs could choose to fully rely on payment system risk scores. As shown in Table 1, this would be less effective than relying on their own monitoring (as payment systems overall are less effective than banks).
2. **Combine.** Banks/PSPs could combine[19] risk scores across their internal systems and the payment system.
3. **Active collaboration.** Banks/PSPs could include payment system network-level risk scores as a feature in their models – and iteratively learn where to best apply payment system indicators.

We found that network analytics can help banks and PSPs to be more effective in spotting financial crime. Overall, it helped identify 12% more illicit accounts with a corresponding decrease in false positives. It was particularly valuable for previously unseen patterns, in respect of which it provided an improvement of 26%.

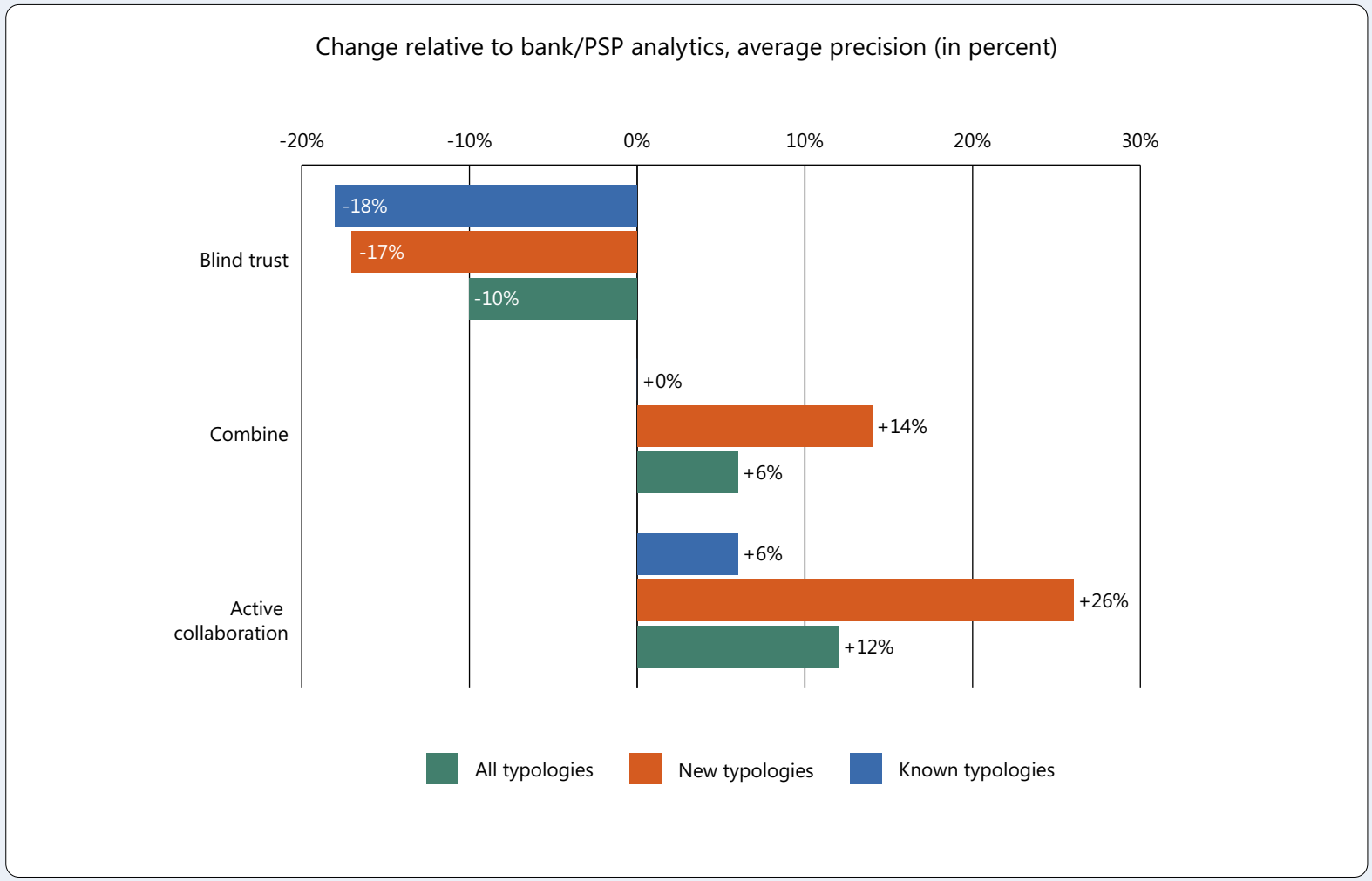
19 In our tests, taking a maximum of two risk scores performed best.

Results: Combining network and transaction analytics

Active collaboration was shown to be the optimal approach. It performed significantly better than simply combining the risk scores across all accounts and typologies. Active collaboration implies that banks and PSPs continuously learn how to best use network analytics findings, by checking them against results from investigations of past cases.

That way network analytics is only used for typologies where it performs best. This approach would also create a virtuous cycle of continuous improvement (see Graph 8 in section 3).

Graph 3: Comparison of options for using the findings of payment system analytics



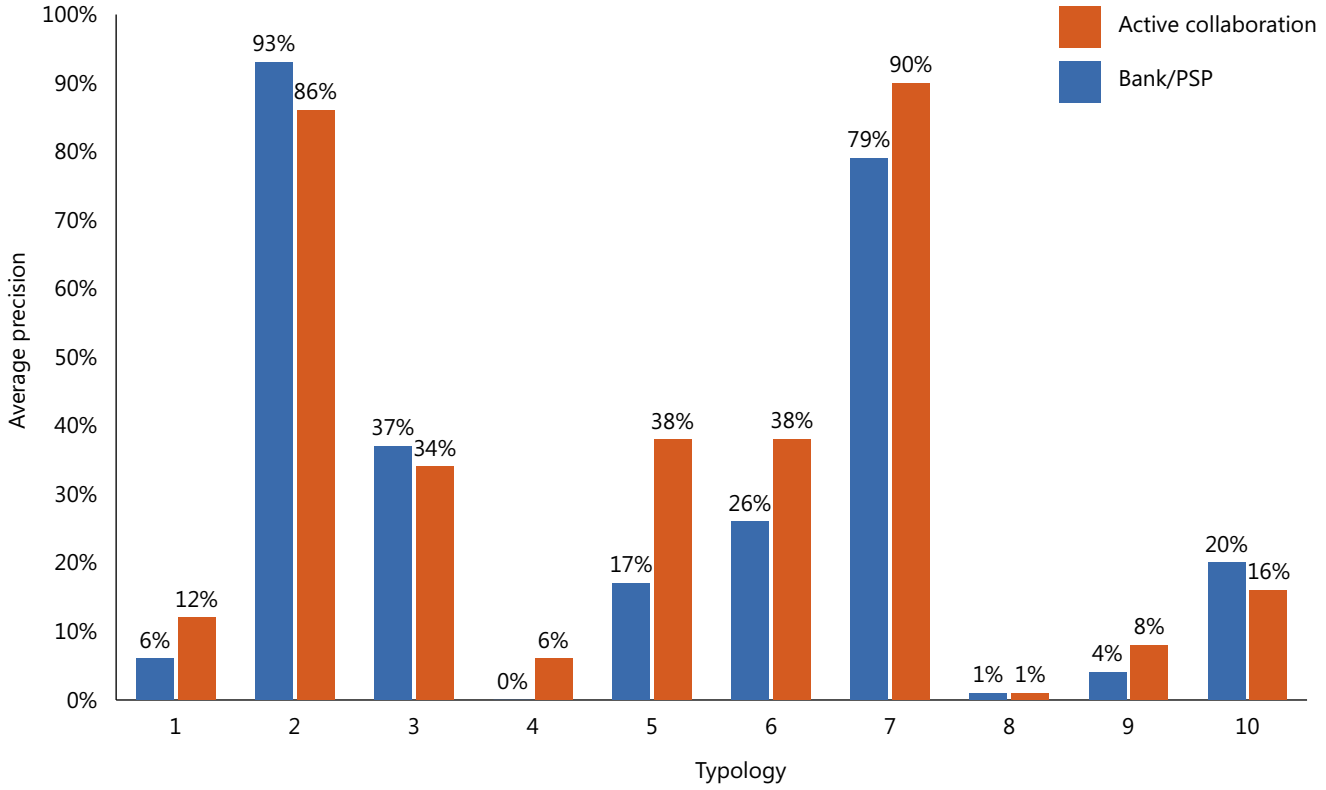
Results: Detection of different typologies

Payment systems were particularly effective in identifying complex typologies involving many accounts across many banks and PSPs.

Network-wide data was found to be particularly valuable for typologies that involve complex transaction chains across many institutions. Graph 4 compares the model results for all 10 financial crime typologies modelled in the data.

Collaboration between banks and payment systems achieved higher results for schemes in which illicit activity includes many bank accounts within the same payment network (particularly 1, 4 and 5). Payment systems were considerably less effective in identifying schemes that involve fewer accounts and rely on a mix of payment methods (eg cash, low-value and large-value payments).

Graph 4: Comparison of effectiveness for different financial crime schemes, average precision



Results: Detection of new typologies

Unsupervised models performed poorly at identifying financial crime patterns.

Detection models need to continuously adjust to new patterns of financial crime. We compared different approaches to spotting financial crime typologies that were not included in training data (typologies 8-10):

- **Supervised**. Using model trained on previously seen schemes (typologies 1-7) to spot new schemes.
- **Unsupervised**. Looking for anomalous patterns without past examples to draw on.

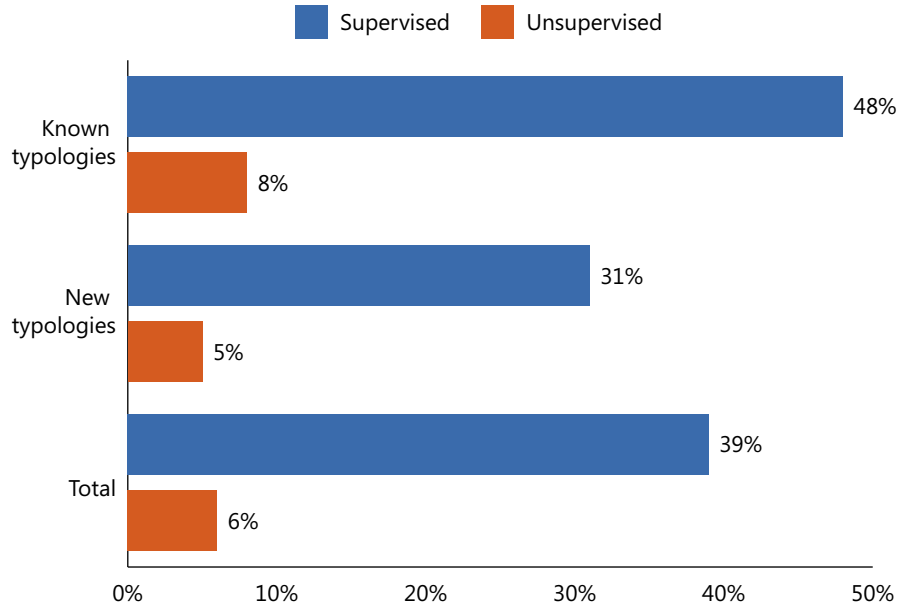
Supervised algorithms were much more effective at spotting new typologies than unsupervised approaches. They correctly identified 31% of accounts involved in previously unseen typologies, relative to only 5% for unsupervised models. Supervised algorithms were similarly superior in respect of known typologies (48% vs 8%) and averaged across all typologies (39% vs 6%).

This shows that having labelled training data is essential for effective detection. In the absence of that, any rare pattern will be flagged as an anomaly, producing extremely large numbers of false positives (92-95%).

In a real-world scenario, labelled training data might not be available. It may also be necessary to update models to keep up with novel patterns. It is also known that the performance of fully supervised models[20] will degrade over time. So unsupervised models have an important potential role. But our results suggest that unsupervised methods should be used with care, and any external and internal intelligence available will help to improve model performance.

20 **D Vela et al,** "Temporal quality degradation in AI models", Scientific Reports, vol 12, article number 11654, July 2022.

Graph 5: Comparison of supervised and unsupervised models, percentage of illicit accounts identified



Results: Calibrating the models

Models can be effectively tuned to focus on a small number of the highest risk accounts.

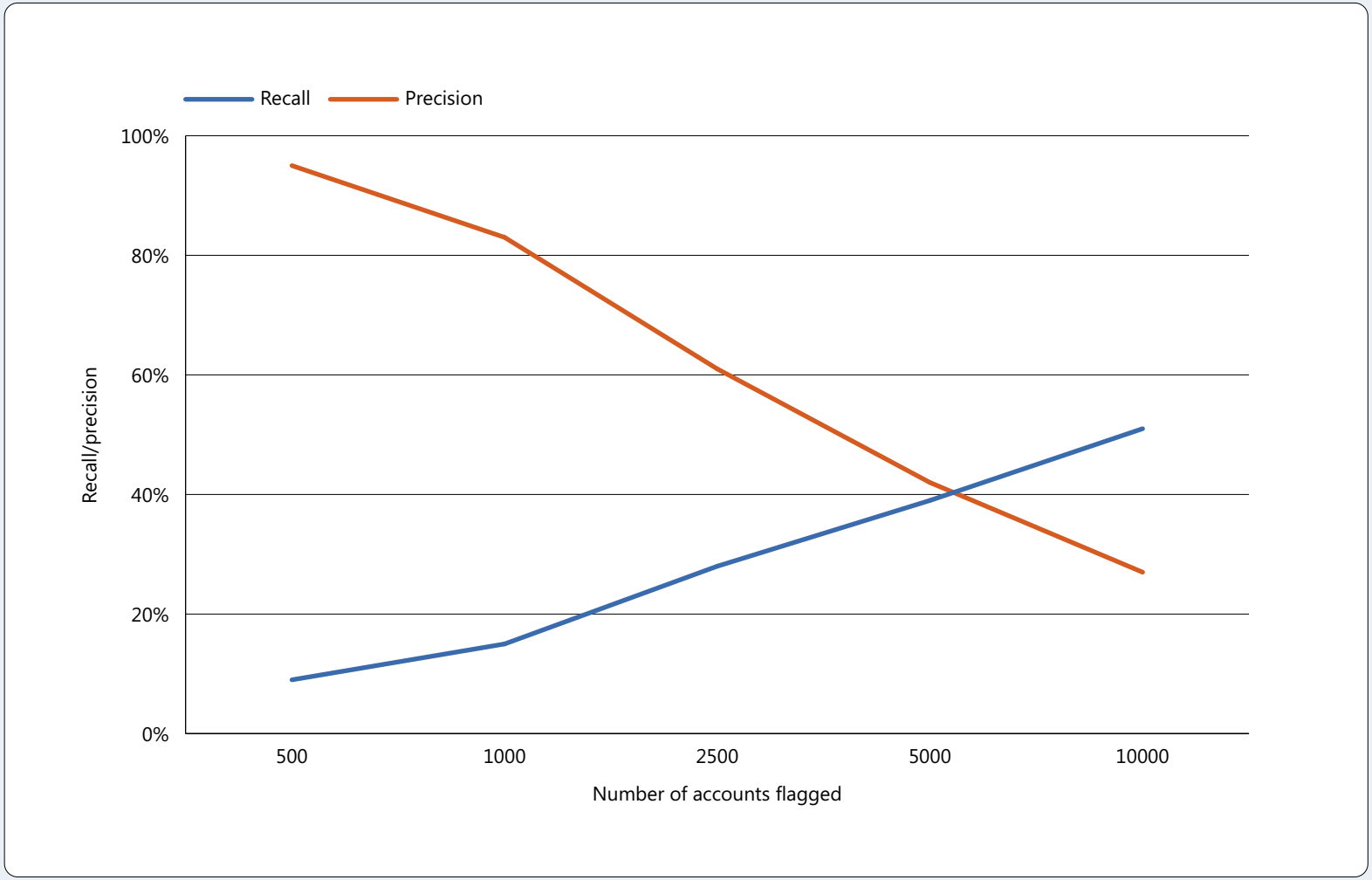
The models used are able to calculate risk scores for every account. The operator can then choose to focus on a certain number of the highest risk accounts.

We found that as more accounts are flagged, the likelihood of false positives (legitimate accounts falsely flagged as illicit increases). Graph 6 demonstrates the trade-off.

To illustrate, if only the top 500 riskiest accounts (top 0.2%) are flagged by the payment system, only 5% of them are false positives. At the other end of the spectrum, payment systems can correctly identify up to 51% of illicit accounts, but at a cost of a false positive rate of 73% (precision of 27%).

This demonstrates that it is possible to tune the models depending on preferences of the payment system operators. For instance, focusing on a small number of alerts could be preferred if there are constrained resources available for review and investigation.

Graph 6: Trade-off between identifying illicit accounts (recall) and avoiding false positives (precision)



Results: Deep learning models

Using cutting-edge deep learning models improved detection effectiveness.

Alongside established machine learning methods, the project tested the performance of novel deep learning models. These models are purpose-built for working with structured tabular data, using a transformer architecture. The main model tested in Project Hertha was UniTTab.[21]

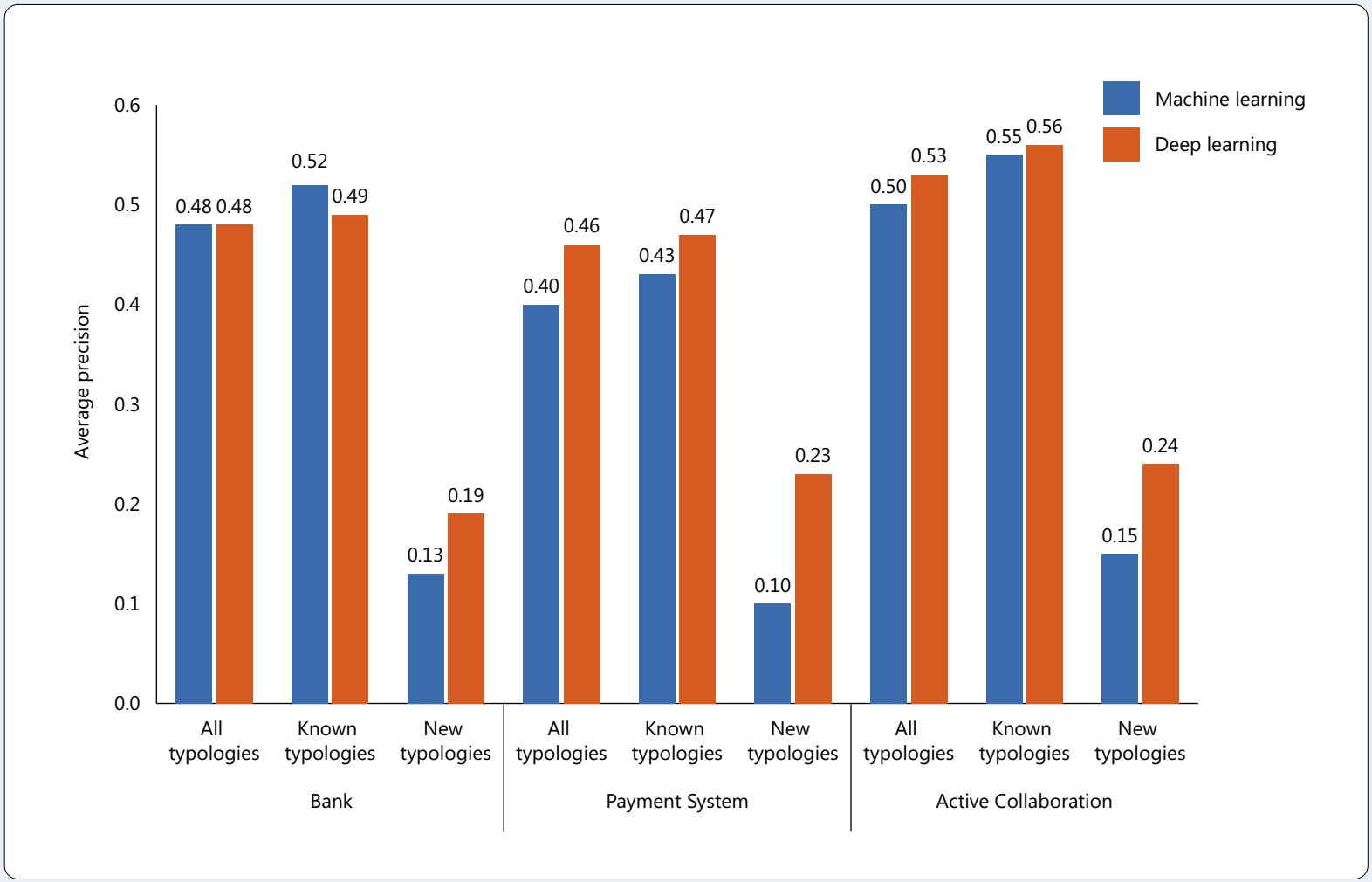
Graph 7 shows a comparison between the best-performing machine learning model (XGBoost) and the deep learning model. There are two key takeaways:

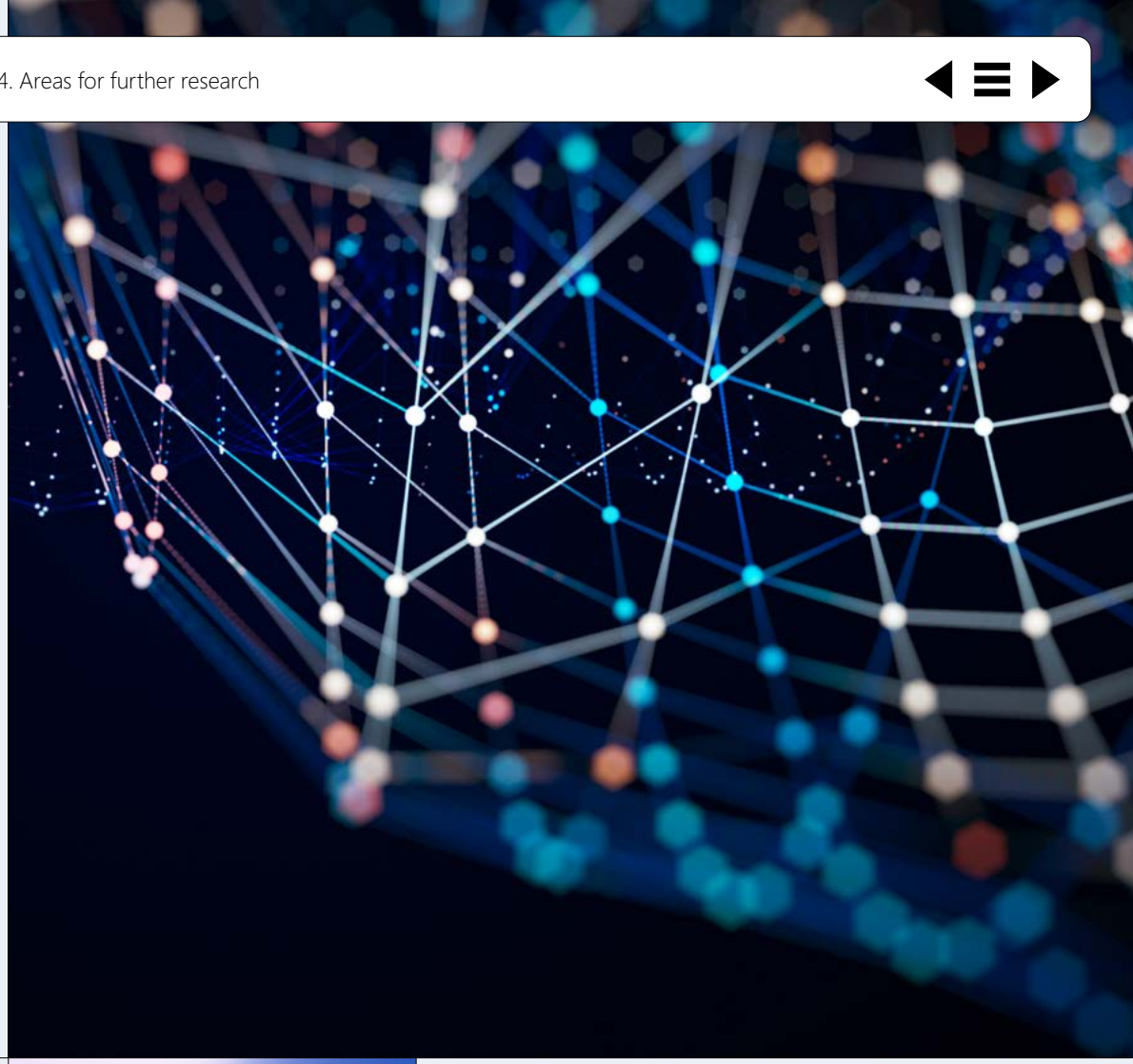
– Deep learning models perform better at spotting previously unseen patterns ("new typologies"): an improvement of 130% for payment systems and 60% in the collaboration scenario.

– They are more valuable for network analytics: improving payment systems' average precision from 0.40 to 0.46 (15% improvement).

21 Introduced in **S Luetto et al,** One transformer for all time series: representing and training with time-dependent heterogeneous tabular data, 2023.

Graph 7: Comparison of machine learning and deep learning models, average precision





Section 3

Key insights

Project Hertha has demonstrated that retail payment systems can identify valuable network patterns in their data. The results of the project highlight the importance of labelled training data, robust feedback loop and explainable AI algorithms.

Key insights: Interpretation of the results

Project Hertha has demonstrated that retail payment systems can identify valuable network patterns in their data, which could be used by banks and PSPs to improve the accuracy of their alerts.

By utilising retail payment system insights, some types of illicit activity could be spotted more accurately and precisely. Banks and PSPs are shown to be well placed to use these network-level insights, and fuse that with their internal detection models.

Overall the results have demonstrated a material, but small improvement in detection rates. The results suggest that retail payment system analytics would be a useful **supplementary tool**. This tool could be used by banks and PSPs to help their internal models and would suggest suspicious activity for further investigation.

Implementing similar solutions in practice would require a robust evaluation of benefits and costs as well as assessing relevant policy, regulatory, practical (eg resource) and legal implications. Exploring these implications is beyond the scope of this report. However, the results provide a few insights that would be valuable to stakeholders considering similar solutions.

Page 25 explains how Project Hertha aligns with other related BIS Innovation Hub initiatives, which suggest components for a potential technology stack to help combat financial crime.

// This tool could be used by banks and PSPs to help their internal models and would suggest suspicious activity for further investigation. //

Key insights: Building a model feedback loop

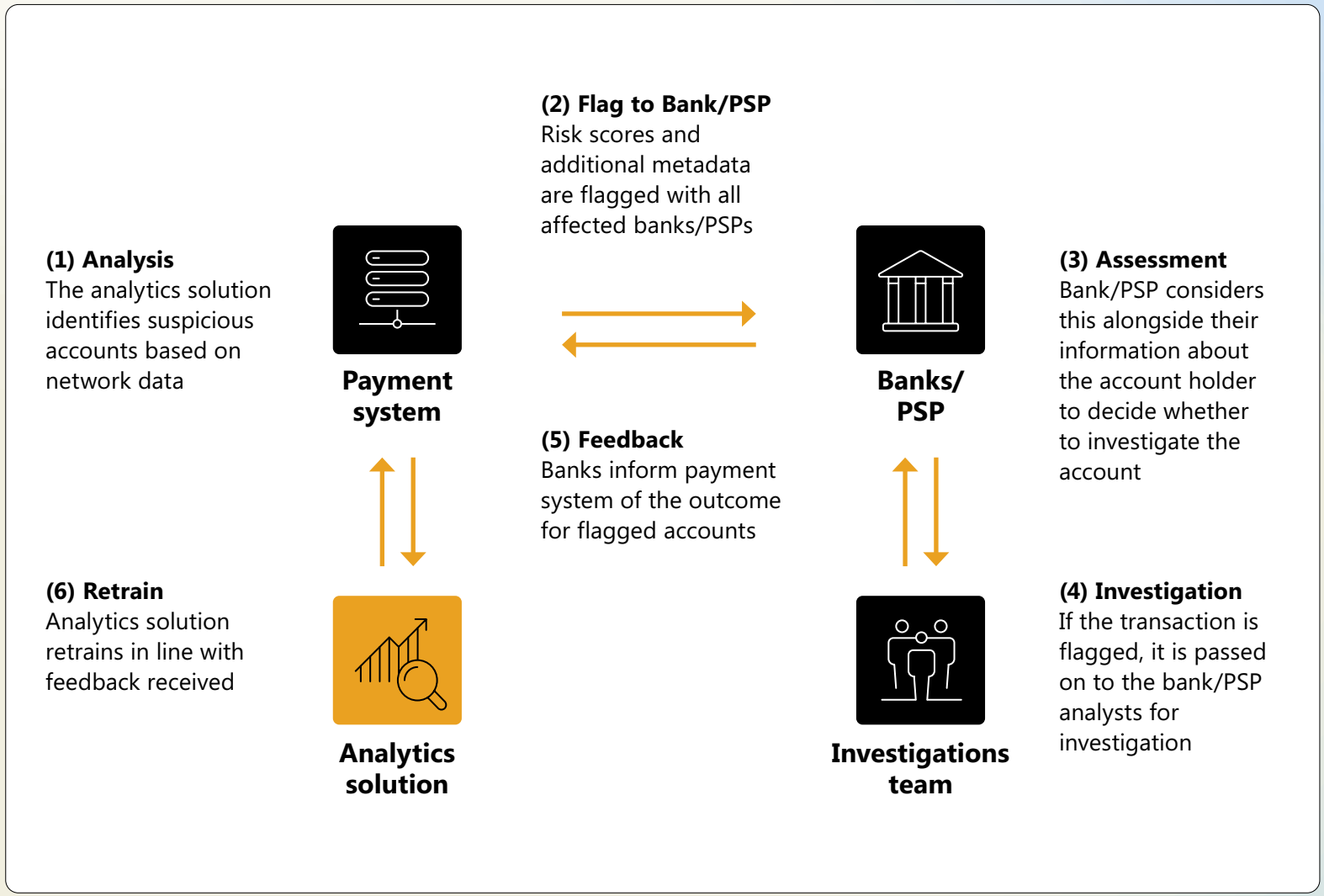
An effective feedback loop is essential to train and continuously improve the analytical models.

Robust machine learning systems require high-quality data to learn, measure effectiveness and continuously evolve. The results of the project highlight the importance of labelled training data: eg examples of past cases or any relevant intelligence. When such labelled data were not available (unsupervised learning), the results were significantly worse and produced large numbers of false positives.

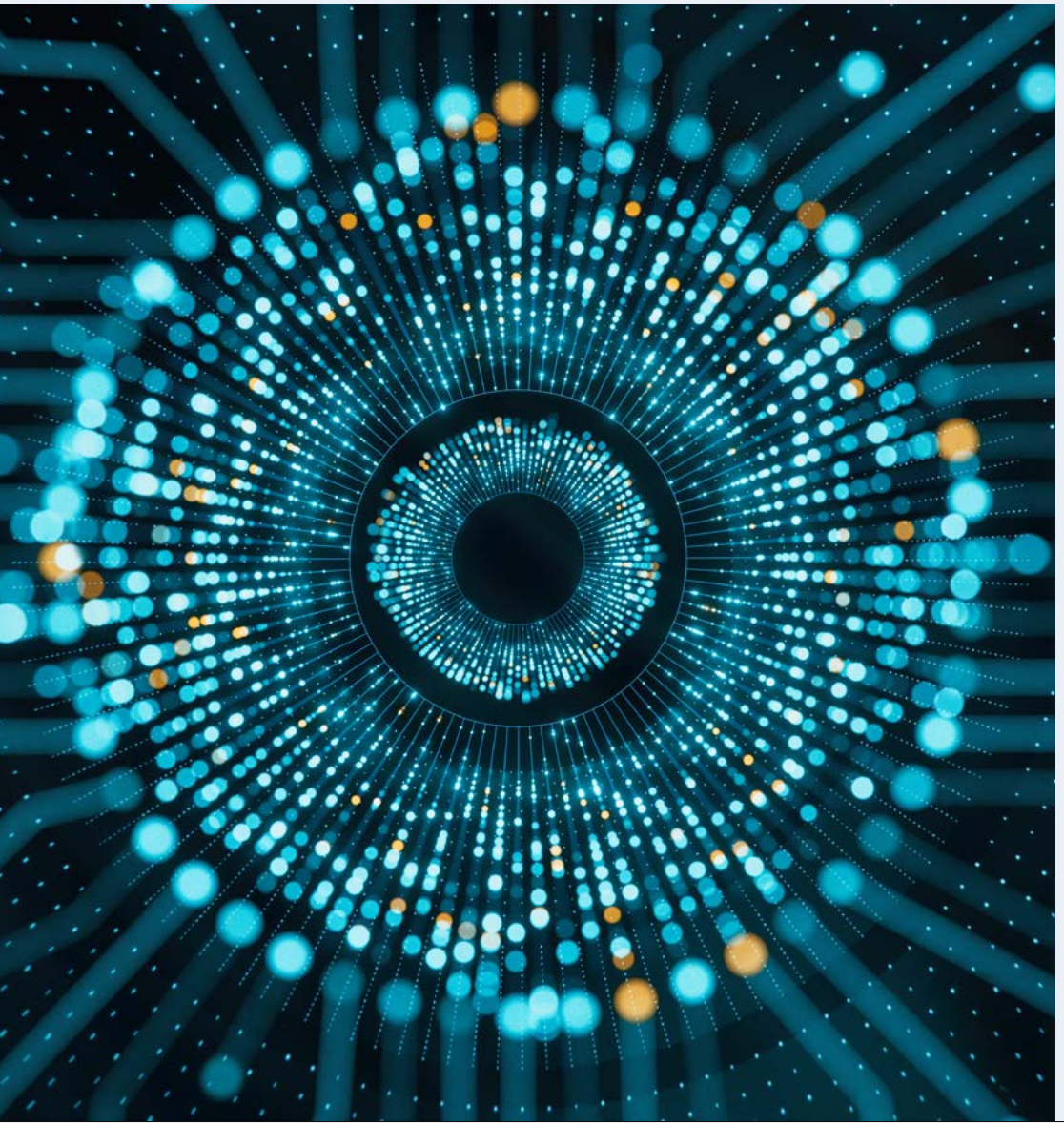
Continuously achieving high performance also requires ongoing feedback on the outcomes. Has the investigation confirmed any suspicious activity? Has the bank or PSP taken any actions (eg closing the account) with respect to this customer?

Graph 8 explains the concept for collaboration between bank/PSP and the payment system, which was tested. Payment systems provide banks/PSPs with supplementary intelligence to help target their alerts. Once investigations are completed, banks/PSPs provide the payment system with feedback on outcomes.

Graph 8: Model feedback loop for payment system analytics



Key insights: Building a model feedback loop



Designing this feedback loop will also involve the consideration of policy and legal issues. It is important to consider how the approach could safeguard customer privacy and avoid tipping off illicit actors under investigation. It is also important to consider the respective requirements of payment system operators and participants to report suspicious activity.

There are a number of possible options to obtain labelled data for the initial set-up.

1. Start with **unsupervised** methods; and build up target data over time.
2. Generate **synthetic training data**, by drawing on expert input and publicly available evidence.
3. Obtain targets from banks/PSPs based on **past reports** (eg reported instances of fraud or suspicious activity reports filed).
4. Obtain targets from **third party organisations**.
5. Collaborate with **law enforcement agencies** to obtain intelligence.

It may also be useful to consider the **infrastructure** for sharing risk scores and feedback. Implementing efficient communication infrastructure (eg via APIs) could help make the process more efficient and reduce the costs of participation. Feedback could differentiate between different outcomes for each account and transaction.[22]

22 Differentiating between different outcomes has been shown to improve effectiveness of money laundering detection. See **M Jullum et al,** "Detecting money laundering transactions with machine learning", Journal of Money Laundering Control, vol 23, no 1, 2020.

Key insights: Using explainable AI methods

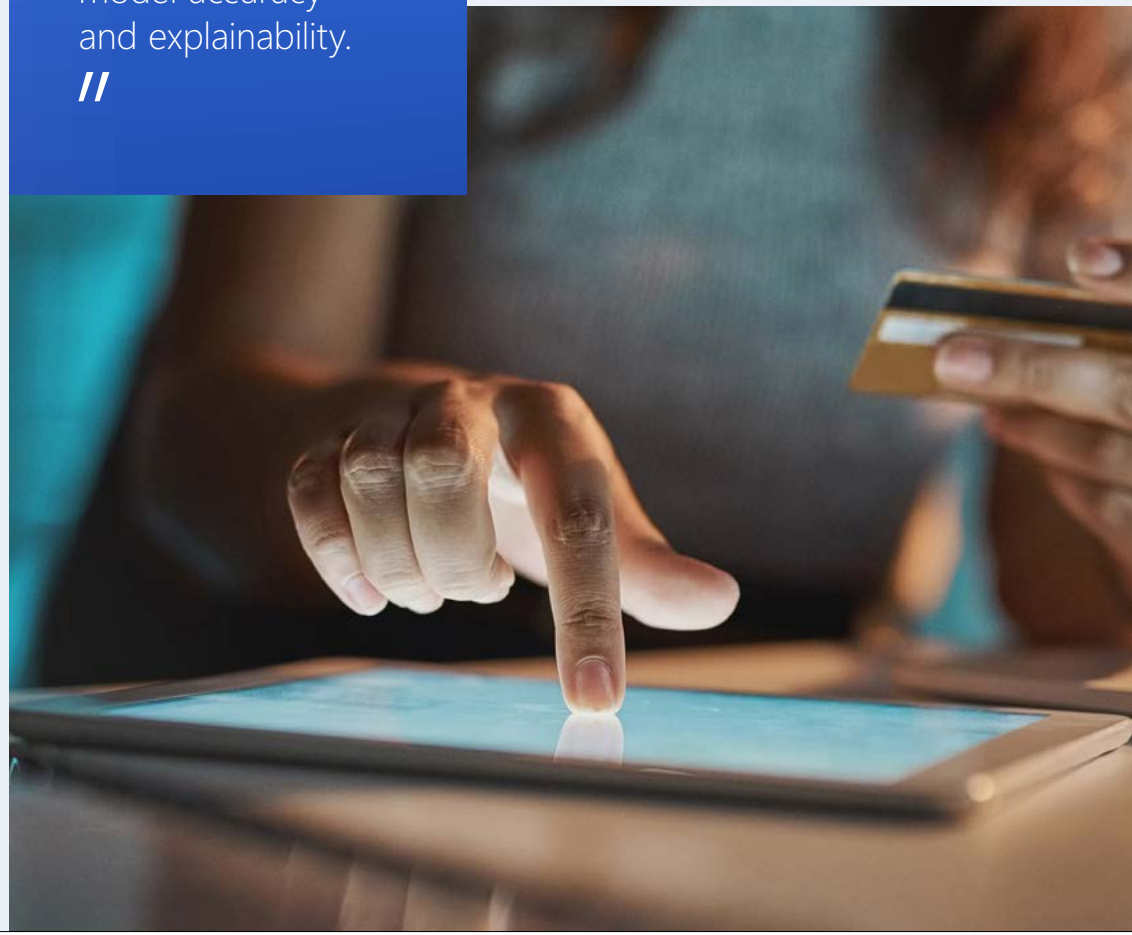
Explainable AI methods can support banks and PSPs in investigations and reporting.

While payment systems are able to identify complex network patterns, the specific rationale may not be obvious to banks and PSPs. Understanding that rationale would help banks and PSPs investigate and report suspicious activity more effectively. They could benefit from learning additional information such as:

- Likely typology identified by the model.
- Variables that influenced the model's decision (eg unusual amounts, timing or counterparties).
- Other banks and PSPs, whose customers are part of the suspected scheme.

There is often a trade-off in AI models between model accuracy and explainability.[23] Some traditional machine learning models (decision trees) tested in Project Hertha are considered to be explainable as it is easier to identify the features that influenced the model's decision. However, the deep learning models tested are more black box as they identify complex relationships from raw transaction data. In this case, explainability could be achieved by training a separate model to identify feature importance.[24]

> // There is often a trade-off in AI models between model accuracy and explainability. //



23 **P Linardatos, V Papastefanopoulos and S Kotsiantis,** "Explainable ai: a review of machine learning interpretability methods", Entropy, vol 23, no 1, 2021.

24 See **S Lundberg & S I Lee,** A unified approach to interpreting model predictions, arXiv:1705.07874, 2017 and **O Sagi and L Rokach,** "Approximating XGBoost with an interpretable decision tree", Information Sciences, vol 572, 2021.

Key insights: Potential industry impacts

Improved transaction analytics can enable better public policy outcomes at a lower cost of compliance for banks and PSPs.

Cost of regulatory compliance

Banks are currently reported to spend over $200 billion[25] every year on compliance with financial crime regulations, while effectiveness remains limited.[26] The results indicate that network analytics could help banks and PSPs comply with their responsibilities more effectively as part of a holistic approach to financial crime prevention.

Banks and PSPs may benefit from reduced operational costs, such as with respect to the manual effort of investigations and customer reimbursements (in the case of fraud). They could also allow bank staff to focus time on higher value activities, such as investigation and reporting.

Benefits for regulatory compliance

International bodies and regulators have highlighted a vision of financial crime regulation to be based on outcomes rather than process.[27] Use of advanced collaborative technology solutions could support improved public policy outcomes at lower cost with reduced effort.

Greater collaboration and use of network-based insights could enable greater efficiency for regulated entities, but that may require support and encouragement from supervisors and policymakers.

Implications for user privacy

Any transaction analytics solutions must safeguard user privacy. The results offer insights on the value of privacy-preserving analytics to improve outcomes. The results are relevant to the following principles for processing personal data:

- **Data minimisation.** Effective analytics is possible with a small number of data points by relying on network patterns rather than personal data.
- **Storage limitation.** Project Hertha utilises data that are already stored within electronic payment systems, minimising data transfer.

Any practical implementation must ensure consent from financial institutions as data controllers, and ultimately from end-users to ensure that data are processed lawfully.

25 **LexisNexis Risk Solutions,** Report: the true cost of financial crime compliance, 2023.

26 **Europol,** The other side of the coin: an analysis of financial and economic crime, 2024.

27 Commonly defined as a risk-based approach, see **Financial Action Task Force,** "Risk-based supervision", FATF Guidance, 2021.

Key insights: BIS Innovation Hub's compliance technology stack

BIS Innovation Hub projects have identified components of a technology stack that could support global efforts to combat financial crime.

BIS Innovation Hub is actively experimenting with technologies that can help safeguard the integrity of the global financial system. Project Hertha sits alongside a wider programme of initiatives, including Projects Aurora and Mandala.

Project Aurora tested the potential for collaborative transaction analytics and information sharing to identify money laundering networks nationally and internationally. It demonstrated the potential for threefold improvement in detection accuracy while reducing false positives by 80%, compared with the existing siloed and rules-based approach.

Project Mandala demonstrated how financial institutions can use privacy-enhancing technologies to prove that they have conducted all necessary cross-border compliance checks.

The solution enhances the efficiency, transparency and speed of cross-border transactions without compromising the quality and soundness of regulatory checks.

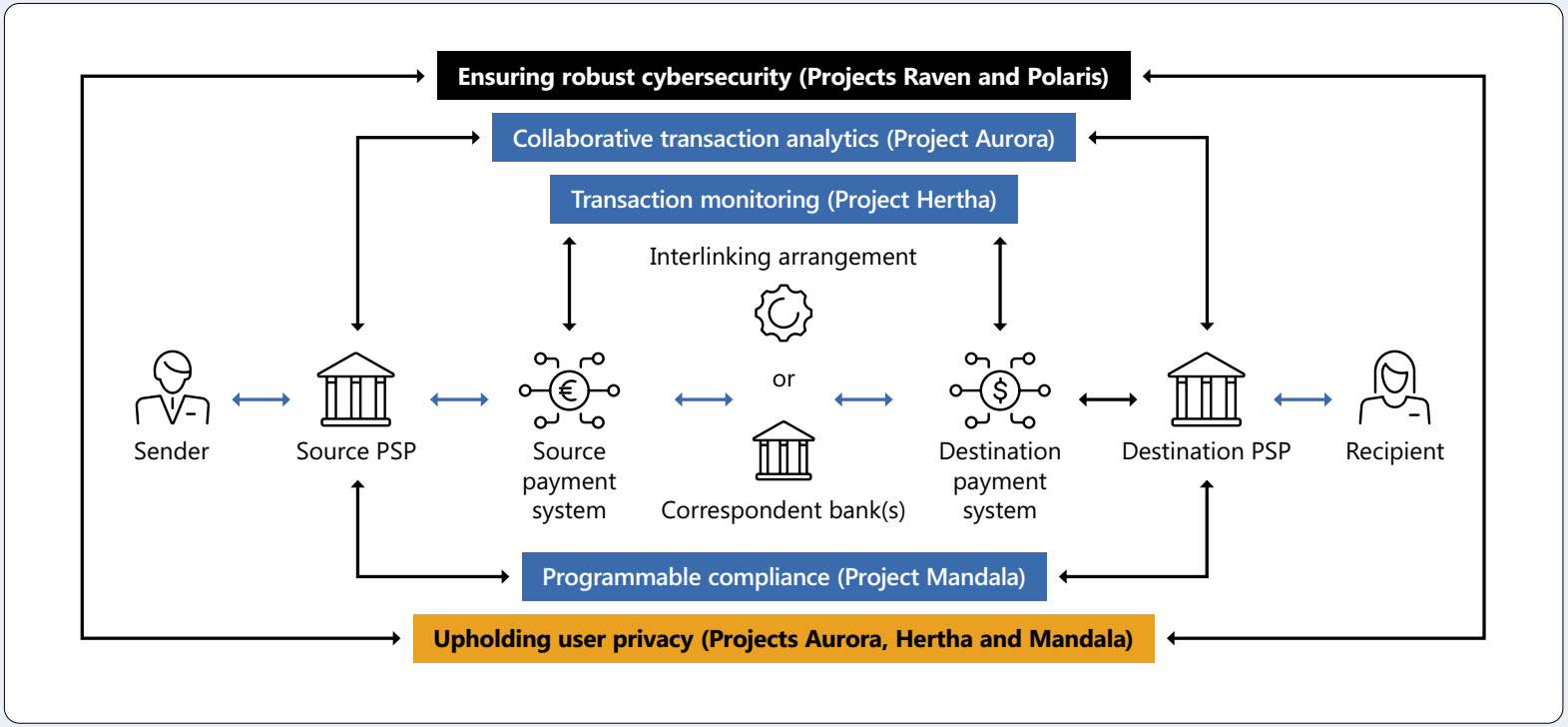
Jointly, our experiments have identified components of a technology stack that could support global efforts to combat financial crime.

It encompasses the following components:

1. Pre-validation and automated compliance checks. (Project Mandala)
2. Transaction monitoring in electronic payment systems. (Project Hertha)
3. Collaborative transaction analytics and information sharing. (Project Aurora)

These projects have a shared focus on safeguarding user privacy through the use of privacy-enhancing technologies. These components also connect to a wider cross-cutting theme of ensuring robust cyber security, as explored in projects Raven and Polaris.

Graph 9: BIS Innovation Hub financial integrity technology stack



Section 4

Areas for further research

Further experiments could focus on transaction tracing, collaborative investigations and applying transaction analytics in other types of payment systems.

Areas for further research: Further use cases and experiments

Project Hertha identified three key areas for further experimentation.

Project Hertha tested the application of predictive analytics to identify patterns indicating financial crime.

However, there are other use cases for analytics, which could be explored in future experiments. This section highlights three examples:

1. transaction tracing
2. collaborative investigations; and
3. extending to other types of payment systems.

Transaction tracing

If a bank account is confirmed to have been involved in financial crime, network data might have clues as to other connected accounts. A **tracing** solution could start from confirmed intelligence (eg reported fraud); and try to identify other connected accounts and transactions. This is different from the approach tested in project Hertha (which focused solely on predictive analytics).

Earlier work[28] has highlighted the potential of this intelligence-led approach to payment system data.

Tracing could have further benefits in identifying criminal networks with greater accuracy, and helping investigations as well as providing early alerts. Tracing solutions have already been widely adopted in respect of cryptoasset networks, to follow the money trail across multiple ledgers.

Collaborative investigations

Detecting suspicious activity is just one of the steps in financial crime compliance. Any alerts then need to be investigated and confirmed. For complex network schemes involving many accounts, there could be further benefits to banks and PSPs collaborating on **joint investigations** and sharing intelligence.

Multiple banks could be alerted about a suspected scheme that involves their customer accounts. Banks/PSPs could then securely exchange intelligence and insights to build a full picture. Advanced technologies (eg differential privacy and data clean rooms) could be used to cryptographically secure data exchange and ensure strict data privacy controls.

Extending to other types of payment systems

Project Hertha tested the application of transaction analytics in national retail (low-value) payment systems. Potential extensions of the project could consider application to **large-value and cross-border payment systems,** recognising differences in transaction patterns, data requirements and legal and regulatory implications.

Similar methods could also be applied to **identifying suspicious patterns in cryptoasset networks** (eg Ethereum or Bitcoin). While they do not have an operator in the same way as traditional payment systems, identifying illicit activity could be valuable to cryptoasset exchanges, financial institutions and supervisors. Transparency of these networks enables analytics to be applied both within a network and across multiple networks.

28 **Deloitte,** Leveraging the payments architecture in the fight against economic crime, 2023.

Glossary

Average precision - a metric that summarises the precision-recall curve, providing a single score to evaluate classification performance.

Deep learning - a subset of machine learning using multi-layered neural networks to model complex patterns in data.

False positive rate – proportion of negative cases incorrectly flagged as positive.

Financial crime – illegal activities involving money or financial transactions, such as fraud, money laundering, or terrorist financing.

Financial crime scheme – a specific plan or method used to carry out a financial crime.

Financial crime typology – specific methods, patterns, and techniques criminals use to commit financial crimes such as money laundering and terrorist financing.

Illicit account – an account that has been involved in financial crime.

Machine learning – a field of artificial intelligence in which algorithms learn from data to make predictions or decisions.

Money laundering – the process of concealing the origins of illegally obtained money, typically by passing it through complex financial transactions.

Payment Service Provider (PSP) – any entity that facilitates electronic payments between consumers and merchants or institutions.

Payment System – a set of instruments, procedures, and rules for the transfer of funds between or among participants. The system includes the participants and the entity operating the arrangement.

Payment System Operator – an entity responsible for managing and maintaining the infrastructure and rules of a payment system.

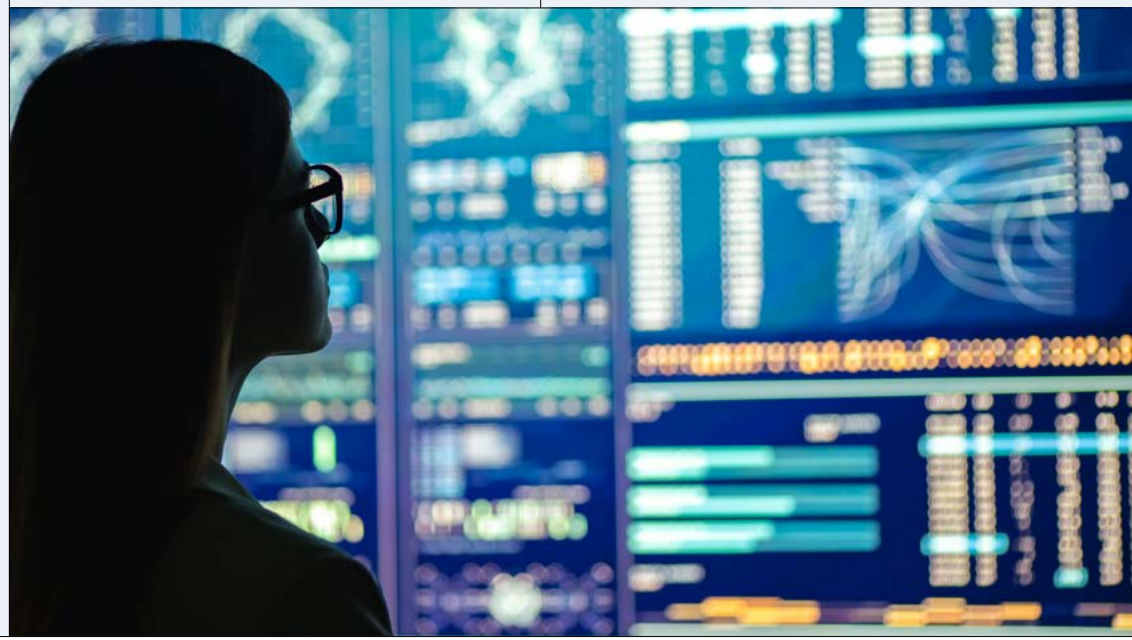
Precision – proportion of positive predictions that are correct (inverse of the false positive rate).

Recall – proportion of all true positives (eg illicit accounts) correctly identified by the model.

Supervised algorithm – a type of machine learning algorithm that learns patterns from labelled data, where the input-output pairs are known.

Synthetic data – artificially generated data that mimics real-world data, used for training or testing models without compromising privacy.

Unsupervised algorithm – a machine learning algorithm that analyses and groups data without labelled examples to identify hidden patterns or structures.



BIS Innovation Hub

Project Hertha
Identifying financial crime patterns in real-time retail payment systems

bis.org