AML/CFT INDUSTRY PARTNERSHIP

Industry Perspectives – Adopting Data Analytics Methods for AML/CFT

November 2018

Table of Contents

1. INTRODUCTION

1.1. BACKGROUND

There is wide consensus about the promise of new data analytics methods in the field of AML/CFT. Through the leveraging of data, existing and rapidly developing technology, and data analytics models, FIs could potentially improve the effectiveness of their AML/CFT measures and address some key weaknesses with the current AML/CFT approaches. These include:

I. High false positive rates in both name screening and transaction monitoring

Current approaches to name screening and transaction monitoring render false positives at a high rate. This is particularly problematic for screening of names in North Asian or other non-Roman character languages. High false positive rates result in inefficiency as resources are needed to review and dispose hits. They can also cause delays in onboarding or processing of transactions, resulting in customer dissatisfaction. High false positive rates could also cause alert fatigue in analysts and therefore impact institutions' effectiveness in addressing these hits. Data analytics solutions have the potential to reduce false positive rates and facilitate or, where appropriate, automate the disposition of hits.

II. Difficulty of applying rule-based transaction monitoring where rules are complex and rigid, and typologies change rapidly

Transaction monitoring is traditionally rule-based, and depends on systems being configured to detect known ML/TF typologies through pre-set parameters and thresholds. However, financial criminals do not always behave in established and predictable patterns that can be detected through these means. In addition, criminals are able to circumvent such transaction monitoring and evade detection by avoiding the relevant thresholds. Further, this approach requires the establishment of mature typologies and therefore often does not account for, or is otherwise slow to react to, new and emerging risk typologies. Data analytics solutions could help to improve rule-based monitoring methods, as well as present alternative approaches to detection.

III. Manual processes and decision-making, which result in inconsistency and human error

AML/CFT processes and decision-making are largely manual. As a result, AML/CFT processes can be inefficient, and headcount is spent on low value work. Manual processes are also vulnerable to inconsistency and human error. While some of these are judgment-based and require skilled human input, many can be automated or improved through technology and the use of data analytics.

Data analytics has already demonstrated potential in an AML/CFT context. For example, one member bank reported that in a proof-of-concept conducted on an AI machine learning solution encompassing customer name screening and transaction monitoring modules, the name screening module resulted in a 50-60% reduction in false positives, while the transaction monitoring module resulted in a 40% reduction in false positives, and in addition demonstrated capability to detect new suspicious patterns which resulted in a 5% increase in true positives. Another member bank reported that the use of data analytics applications produced double-digit efficiencies across certain operational tasks.

Despite the potential for data analytics to add value to or solve problems within the AML/CFT functions of the financial services sector, adoption is, at the time of this paper, still relatively immature across the industry. This may in part be attributable to a reluctance to confront legacy approaches to AML/CFT or a fear of diluting focus on what are perceived to be core processes and controls. In certain FIs, there may also be a misconception that data analytics solutions are invariably novel, highly complex, require high risk appetite and heavy investment, and are therefore only suitable for larger FIs. However, the incorporation of these tools is necessary in the fight against criminals, and in the context of wider banking infrastructure which is highly technology-driven and often automated.[1]

For these reasons, the AML/CFT Industry Partnership (ACIP) has identified data analytics as a key area of interest. ACIP, a private public partnership established in April 2017, and co-chaired by the Commercial Affairs Department

[1] Contributed by Exiger.

of the Singapore Police Force and the Monetary Authority of Singapore, was formed to bring together selected industry participants, regulators, law enforcement agencies and other government entities in Singapore to collaboratively identify, assess and mitigate key ML/TF risks in Singapore. To focus on data analytics, ACIP set up the Data Analytics Working Group for member banks to share their respective journeys of adoption and implementation of AML/CFT analytics, and to provide practical insights for FIs at different stages in their own journeys. This paper is a product of this collaboration amongst WG members, with inputs from invitees and contributors.

The Data Analytics WG is chaired by the Head of Group Legal, Compliance and Secretariat, DBS. The Data Analytics WG members (representatives from commercial banks operating in Singapore) and other organisations which contributed to this paper are listed in Appendix B. Please note that the incorporation and attribution of contributions by vendors is not intended to represent any endorsement of any such vendor(s) by ACIP, the WG, or any of its members.

Unless otherwise stated, the defined terms used in this paper shall have the meanings set out in Appendix C.

1.2. Objective

The objective of this paper is to provide information and perspective on the use of data analytics for AML/CFT purposes, thereby starting or taking forward conversations on the adoption and implementation of such solutions, both within individual FIs at varying stages of the analytics journey and across the industry. It is hoped that this discussion will facilitate increased adoption of data analytics by the industry in AML/CFT functions and decision-making where appropriate, increase the leveraging of existing tools, and foster greater collaboration between analytics tools and crucial human input/judgement to strengthen AML/CFT frameworks.

To this end, Section 2 of this paper provides a scan of possible use cases for analytics solutions, ranging from simpler commonly adopted solutions to more cutting edge and experimental use cases, and maps these according to degree of adoption or exploration by banks.

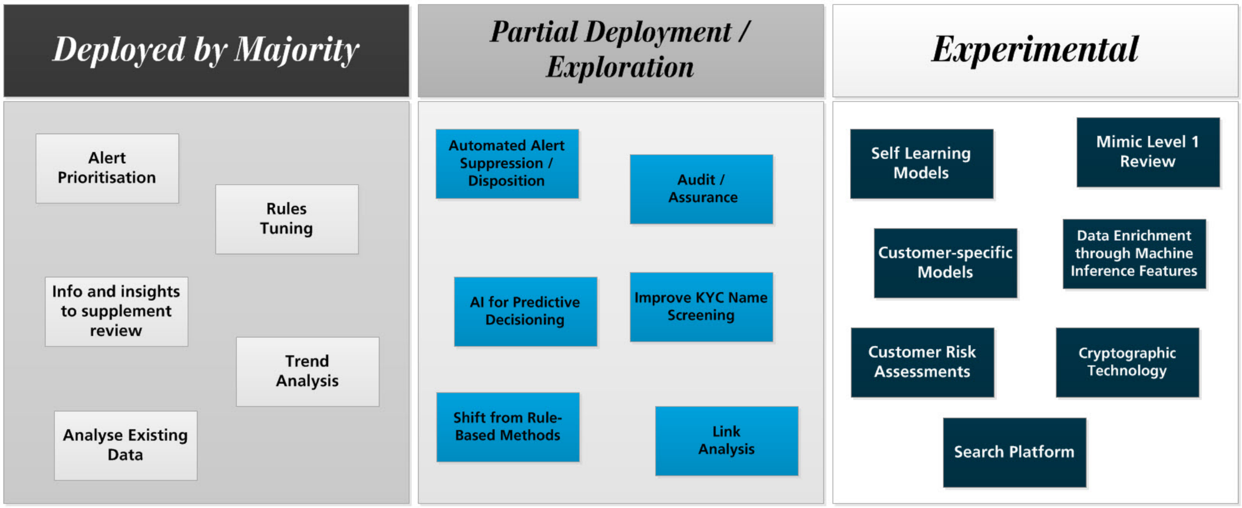
Section 3 of this paper addresses challenges that FIs may face in adopting analytics for AML/CFT purposes and shares some of the ways these may be solved or managed by FIs keen on driving the use of analytics.

With the benefit of the examples and findings in Section 2 and Section 3, Section 4 will recommend several focus areas for partnership between the private and public sector that the WG believes will help to drive the adoption and increase the effective use of data analytics for AML/CFT purposes.

2. Overview of AML/CFT Analytics Use Cases

Across the industry, there is a spread in degrees of adoption of AML/CFT analytics by FIs. Some FIs are further along the journey, and may already be using data analytics throughout their AML/CFT programmes or exploring cutting edge solutions. On the other end of the spectrum, some FIs are at a preliminary stage and may be unsure as to how to appropriately start incorporating data analytics into their AML/CFT functions. This Section gives an overview of use cases, ranging from simpler applications (which can be very effective) to more advanced, cutting edge solutions. This Section will also indicate which of these are commonly deployed, and which are only being explored or experimented with by a few FIs. It is hoped that this mapping of use cases will give readers at varying stages of the analytics journey some perspective on how analytics might be able to contribute to their AML/CFT programmes, and how they can start or progress through their respective journeys.

The WG has observed various themes running through data analytics use cases for AML. The graphic below maps out some use cases according to degree of adoption:


Deployed by Majority
Alert Prioritisation
Rules Tuning
Info and insights to supplement review
Trend Analysis
Analyse Existing Data
Partial Deployment / Exploration
Automated Alert Suppression / Disposition
Audit / Assurance
AI for Predictive Decisioning
Improve KYC Name Screening
Shift from Rule-Based Methods
Link Analysis
Experimental
Self Learning Models
Mimic Level 1 Review
Customer-specific Models
Data Enrichment through Machine Inference Features
Customer Risk Assessments
Cryptographic Technology
Search Platform

I. To enhance current rule-based transaction monitoring

a. Rules tuning

Transaction monitoring rules are designed to capture known typologies or scenarios through monitoring of fixed parameters and applying thresholds beyond which alerts are triggered. However, these thresholds may be set at levels which trigger (too many or too few) alerts which do not reflect the actual risk environment. Absolute rules may also fail to take into account different contexts (business lines, products) and environments.[2] In addition, risk environment and typologies may change over time, rendering previously established thresholds less appropriate or effective.

Data analytics can be used to leverage historical information to systematically evaluate the productivity of these rules and thresholds. Banks can then tune rules optimally for detection (increasing true positives) and reduction of false positives. This includes tuning rules to optimise alert volumes with greater risk focus, including adjustments that are appropriate in response to a changing risk environment.

Maturity: Based on the feedback provided by the member banks, the use of analytics for rules-tuning in transaction monitoring is fairly established and commonly implemented.

[2] Contributed by SAS.

Example: One member bank shared that it uses data analytics to define the optimal grouping of clients for transaction monitoring. Data analytics are applied to model client risk and transactional behavior to determine peer grouping and scenario thresholds for such groupings.

b. Analyse existing data

Typically employed methods of analysing individual data points with respect to established thresholds to determine unusual behavior may not effectively detect activity that could be deemed unusual if analysed across multiple data points. For example, transaction monitoring rules can be easily circumvented by criminals and terrorists through various means, including transacting just below widely known thresholds. Another example in the context of sanctions screening is the implementation of detection capabilities that use the previously alerted transactions to identify attempts to bypass a rule-based screening system through stripping techniques. With the use of data analytics, the analysis of such existing data from traditional sources can be improved.

Maturity: Based on the feedback provided by the member banks, the use of analytics for analysis of existing data in transaction monitoring is fairly established and commonly implemented.

Example: A member bank shared that it applies machine learning models to detect anomalous activity by taking into account multiple factors and variables in outlier detection, rather than using frequency distribution graphs (histograms) to identify thresholds to determine outliers, which only considers data points from one dimension.

Example: A vendor shared its observation that many FIs are implementing additional detection capabilities that leverage data based on past alerts and use data memorisation techniques to learn from the "fingerprint" of transactions which are the subject of these alerts. This means that characteristics and patterns found in these transactions are extracted automatically and stored in memory and then used to capture transactions which bear similar characteristics and patterns, but which may have been stripped of a piece of data in order to avoid detection.[3]

c. Alert prioritisation

As discussed above, transaction monitoring systems raise a high number of alerts, many of which are false positives, generating work on the processing of hits, inefficiencies, alert fatigue, and delays to transactions. To address this, many banks have used analytics for risk scoring of alerts. Analytics are applied to investigation metrics, including client type, client risk, scenarios flagged and historic productivity, to assign risk scores to alerts. These risk scores may be reflective of a prediction on how likely they are to result in STR filing. Alerts are then prioritised for review based on their risk score; this prioritisation can take the form of reviewing "higher-risk" alerts first, or having "higher-risk" alerts reviewed by more senior analysts.

Maturity: Based on the feedback provided by the member banks, the use of analytics for risk scoring or other prioritisation of alerts in transaction monitoring is fairly established and is implemented by almost all member banks surveyed.

Example: One bank applies a filter model using various inputs relating to the customer (age, citizenship, occupation, business segment, number of party to party relationships), the account (account age and type), and transactions (transfer frequency, transfer count, and channel types). The model applies both machine learning and network analysis to determine a risk score for each alert. Alerts with higher risk scores are then prioritised during investigations. In a blind test, it was found that the model's judgement was aligned with that of a human financial crime expert. The model was able to risk score the alerts to improve efficiency, leading to time savings, lower costs, lowered alerts backlog, and more timely escalation of high risk alerts to authorities.

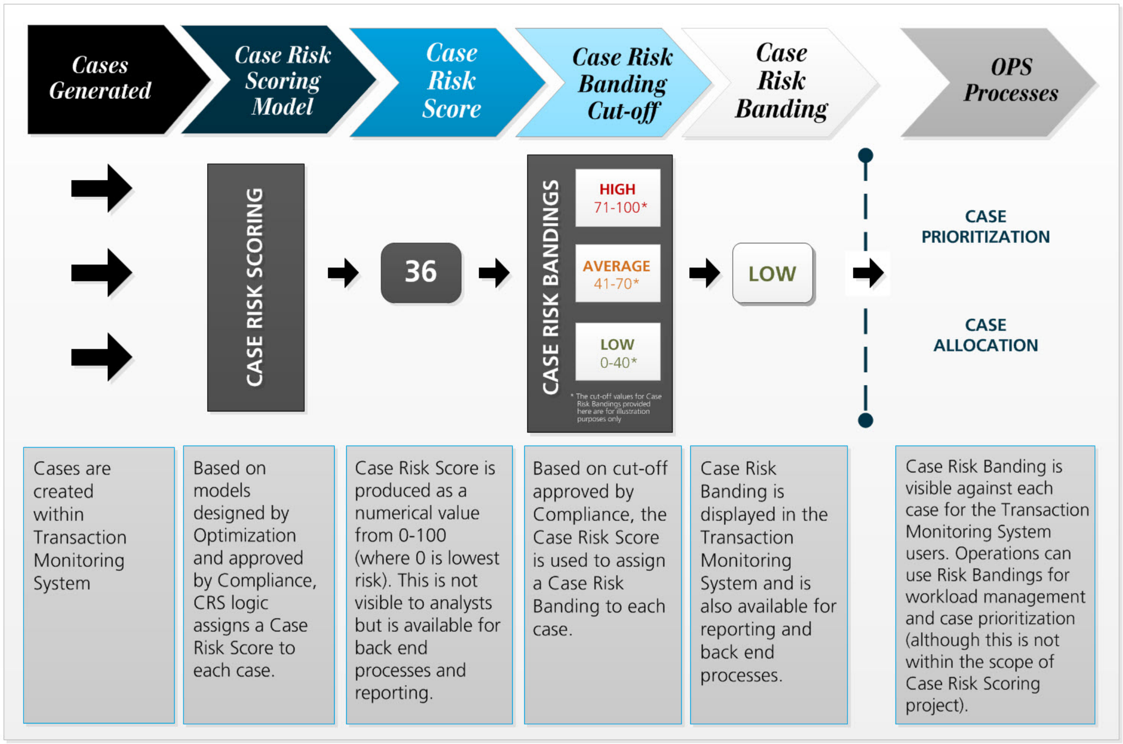
Example: One bank is exploring a partnership with a FinTech firm to co-create a machine learning solution it hopes will produce accurate risk scoring for alert classifications (so as to minimise mis-classification) without any human input. The model is also intended to provide explainability as to the

[3] Contributed by Fircosoft.

alert scoring in order to ensure consistency of risk treatment, facilitate review, and allow for the necessary accountability for decisions from a regulatory perspective.

Example: A bank is exploring the use of AI in post internal bank screening systems to categorise alerts for review. The solution is intended to also generate a corresponding percentage of accuracy and the methodology behind the categorisation, so that the method is verifiable by analysts. This solution is intended to bring about benefits in analyst productivity and time taken for investigations. The explanation of methodology and assessment of accuracy would also help to provide a trail for audit and assurance purposes.

E.g. of an alert prioritisation operating model:



* The cut-off values for Case Risk Bandings provided here are for illustration purposes only

Cases are created within Transaction Monitoring System

Based on models designed by Optimization and approved by Compliance, CRS logic assigns a Case Risk Score to each case.

Case Risk Score is produced as a numerical value from 0-100 (where 0 is lowest risk). This is not visible to analysts but is available for back end processes and reporting.

Based on cut-off approved by Compliance, the Case Risk Score is used to assign a Case Risk Banding to each case.

Case Risk Banding is displayed in the Transaction Monitoring System and is also available for reporting and back end processes.

Case Risk Banding is visible against each case for the Transaction Monitoring System users. Operations can use Risk Bandings for workload management and case prioritization (although this is not within the scope of Case Risk Scoring project).

d. Information and insights to supplement alert review

Data analytics can also be applied to data from a wide range of sources, including non-traditional internal sources (data available within other financial crime applications or across the bank) and external sources (publicly available data). In particular, analytics and artificial intelligence can be applied to unstructured data, being data that is typically text-heavy and not organised into a pre-defined format, from these sources.[4] For example, some vendors are focused on providing semantic technology, which helps analysts to derive meaning and usable data points from reams of unstructured data.[5] Natural language processing, which allows computers to interpret, understand and manipulate human languages, can be applied to process non-standard documentation. Information derived can then be used to enrich and supplement alert review.

Maturity: Based on the feedback provided by the member banks, the use of analytics for analysis of existing data in transaction monitoring is fairly established and commonly implemented.

[4] Exiger has defined this type of technology as "Feature Creation" technology, a quintessential example being Optical Character Recognition (or OCR) technology.
[5] Contributed by IBM.

Example: One bank shared that it was exploring the possibility of applying a semi-supervised machine learning model (being a model which utilises a combination of labelled and unlabelled data) to provide additional insights/dimensions to further supplement the alert review process. Labelled data refers to a set of data for which the target answer is already known, such as "fradulent" or "true positive".

e. Trend analysis

In addition to use of rule-based monitoring for detection of specific suspicious or unusual transactions or behavior, reporting from rule-based monitoring can also be leveraged for the analysis of ML/TF trends, and detection of new trends and emerging typologies. Such trend analysis techniques can also help to highlight and monitor the risks associated with certain activities (e.g., data quality indicators for payments received from the bank's counterparties). In addition, machine learning models can be leveraged to detect complex and non-linear relationships (i.e. relationships that are not easily apparent because the variables are not directly correlated), as well as to quickly, and with minimal human input, identify complex trends not readily noticeable by non-experts or without time consuming manual investigation.

Maturity: Based on the feedback provided by the member banks, the use of analytics for trend analysis is fairly established and is implemented by almost all member banks surveyed.

Example: A vendor shared its observations that many institutions are implementing systems to generate "non-blocking" alerts. Analytics performed on these alerts enable FIs to analyse trends and risk areas, for example, identifying risk areas from data on correspondents who have not been providing accurate or complete information on senders or receivers. This allows the FI to then adjust its correspondent relationships accordingly if appropriate.[6]

f. Predictive decisioning and automated alert suppression or disposition

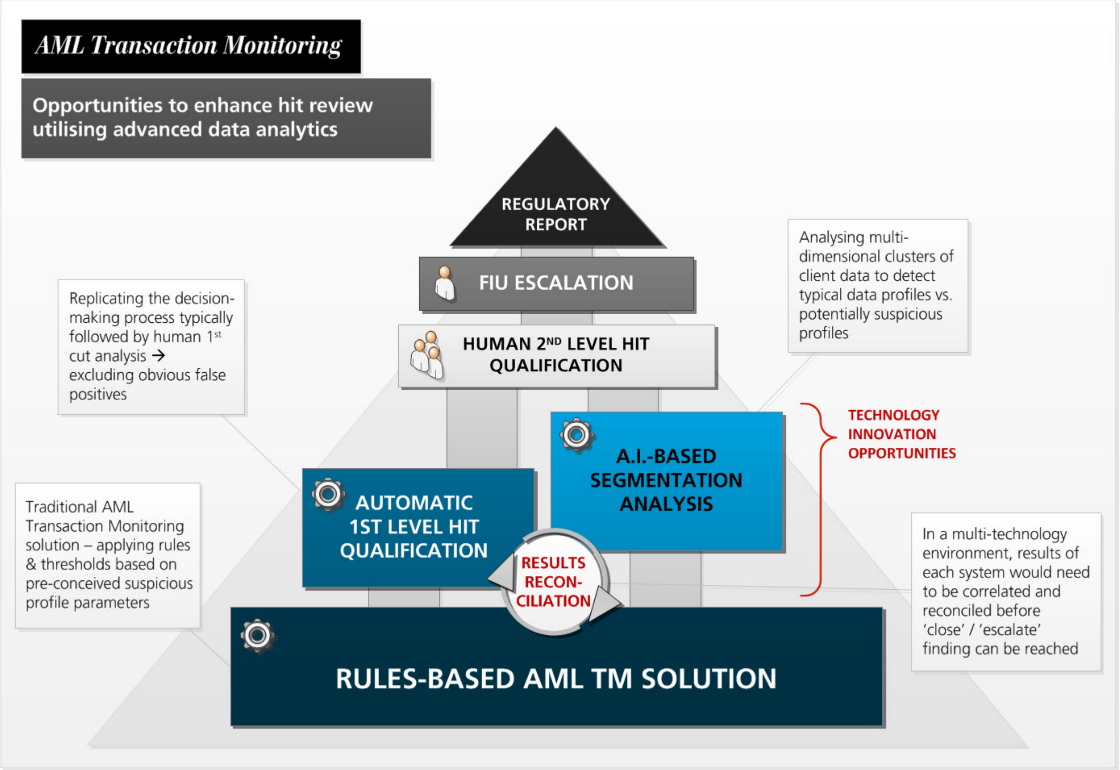
Based on large volumes of data from past alerts and decisions, some FIs are exploring the application of advanced analytics algorithms to automatically determine a recommendation for an alerted transaction, including whether an alert is true or false, and what the model's degree of confidence is. FIs and vendors are also exploring the use of analytics models for automated suppression or disposition of alerts. These analytics can be applied to alerts generated from both sanctions screening tools (for example when a customer is homonym with a sanctioned entity) and threshold-based transaction monitoring systems.

Maturity: While most FIs recognise the potential for such new detection methods, the methods are at a relatively early stage of adoption. Other FIs are currently only exploring or experimenting with these methods.

Example: Through memory of hit-level decisions and analytics applied for pattern recognition of raised hits, combined with an understanding of the context in which they are raised, models are able to suggest the exceptions that should be added to the white lists used in the systems to automatically dispose hits. This is particularly useful in the context of trade transactions, which may vary slightly and therefore create a high number of hits. Reapplying hit level decisioning will allow for operators to see hits automatically cleared by the system, and to concentrate their attention only on new hits, thereby much accelerating the review time needed for the alert.[7]

[6] Contributed by Fircosoft.
[7] Contributed by Fircosoft.

Graphic showing how solutions can be layered to uplift transaction monitoring:


AML Transaction Monitoring
Opportunities to enhance hit review utilising advanced data analytics
REGULATORY REPORT
FIU ESCALATION
HUMAN 2ND LEVEL HIT QUALIFICATION
A.I.-BASED SEGMENTATION ANALYSIS
AUTOMATIC 1ST LEVEL HIT QUALIFICATION
RESULTS RECON-CILIATION
RULES-BASED AML TM SOLUTION
TECHNOLOGY INNOVATION OPPORTUNITIES
Replicating the decision-making process typically followed by human 1st cut analysis → excluding obvious false positives
Traditional AML Transaction Monitoring solution – applying rules & thresholds based on pre-conceived suspicious profile parameters
Analysing multi-dimensional clusters of client data to detect typical data profiles vs. potentially suspicious profiles
In a multi-technology environment, results of each system would need to be correlated and reconciled before 'close' / 'escalate' finding can be reached

II. For building new detection methods

a. Shifting from rule-based methods

Aside from using data analytics for the enhancement of these traditional rule-based approaches, FIs can also explore new detection methods powered by data analytics. Rules-based monitoring makes potentially broad assumptions[8] on ML/TF, which are then applied to entire segments of customers without catering for whether such behaviour is normal or consistent for the customer. On the other hand, these new detection methods use bottom-up construction of profiles and behaviours based on available data, without needing to make such assumptions.

One example of such a detection method is clustering, an unsupervised learning technique that is applied to data sets to segment them into clusters. The underlying assumption is that outliers from these clusters are more likely to represent suspicious or high-risk scenarios. Such clustering can be done based on clients, products, location and/or even alerts generated through rule-based monitoring. This method can enable FIs to identify complex patterns of money laundering through such outliers, without reference to specific rules or typologies which are inevitably non-exhaustive and imperfect. This has the further benefit of facilitating the potential discovery of previously unknown money laundering risks (see paragraph on trend analysis above). Other methods that similarly focus on analysis of the available data rather than depending solely on fixed rules and human assumptions include topological data analysis (which aims to study and represent the shape of available data) and other means of dimensionality reduction (which is the process of reducing the number of dimensions to help models perform better or to allow humans to visualise the data). In addition, such methods may reduce false positive rates, thereby improving efficiency and reducing alert fatigue amongst analysts.

[8] IBM has commented that one issue is how to remove biased perspectives or preconceptions regarding financial crime from critical financial crime decision processes. SAS has also commented that such assumptions are generally not empirically based.

Maturity: While most FIs recognise the potential for such new detection methods, the methods are at a relatively early stage of adoption and do not yet give FIs significant confidence for them to move away from the traditional rule-based approach entirely. As such, some FIs using or exploring these new detection methods may use clustering approaches on an *ad hoc* basis or to layer over rule-based monitoring. Insight garnered through these new detection methods can also be fed back into transaction monitoring. Other FIs are currently only exploring or experimenting with these methods.

Example: One bank shared that it applies unsupervised machine learning to analyse client transactions and to achieve advanced segmentation of client transaction data. This machine learning solution is expected to both reduce the number of false positive alerts generated, and to detect suspicious client activity not detected by existing rule-based solutions.

Example: A vendor shared that it has developed an unsupervised anomaly detection engine based on an ensemble of two unsupervised algorithms which work to detect outlier behaviour, and to generate a score to provide an explanation of the anomaly score per feature, which helps to pin-point the key data-point(s) responsible for anomalous behaviour within an account.[9]

b. Link analysis for network detection

Financial criminals often work through organised networks and the layering of transactions to conceal the origin of illicit funds. Money launderers form groups of accounts that may appear to be unrelated, and use these to perform complex series of transactions which are not flagged by rule-based monitoring, and which make the tracing of such funds by law enforcement difficult. The use of such complex series of transactions and hidden relationships was explored in detail in the ACIP Legal Persons – Misuse Typologies and Best Practices paper published in May 2018.

Link analysis or "social network analysis" refers to the use of graphing technology, which specialises in the storing, processing and analysis of nodes and relationships between the nodes, to model and detect multiple relationships and networks between entities and accounts. This information is often represented visually, allowing for quicker processing of relationships and interdependence by analysts.

Graph analytics is broken down into five key pillars:

(i) Path analysis: Provides analysis of common traits between entities.

(ii) Connectivity analysis: Analyses the strength of links between entities.

(iii) Centrality Analysis: Provides a mechanism to identify important entities within a cluster of data

(iv) Community Detection: Analytics to identify communities of people or groups of entities within a larger cluster of entities.

(v) Sub-graph isomorphism: This is a validation technique used in hypothesis testing to show unusual elements within the dataset.[10]

In addition to facilitating an FI's compliance with regulatory requirements to pay special attention to complex or unusual patterns of transactions, such analysis can also help in detection of networks operating across accounts and FIs, and therefore money-laundering which poses a risk to the larger financial system. In addition, link analysis models are dynamic[11], facilitating the detection of new relationships and networks, and enabling FIs to detect changing or new relationships and criminal behaviour which may be quickly evolving.

[9] Contributed by Delta Capita.

[10] Contributed by IBM.

[11] Contributed by IBM: Graph analysis is being combined with Ensemble Techniques and Graph Powered Machine Learning, to look at dynamically generating entity graphs associated with compliance incidents. The key for compliance is to provide macro perspectives of entities and their activities across institutions, to provide a more dynamic risk profile of entity actions with a compliance focus.

Maturity: While most FIs recognise the potential for link analysis to improve network detection, the methods are at a relatively early stage of adoption. Many FIs are therefore exploring or experimenting with link analysis, and may on an *ad hoc* basis use insights generated through link analysis to enrich their core transaction monitoring.

Example: One bank shared that it is exploring the use of link analysis for the detection of links and collusion across parties for transaction surveillance purposes. The bank is assessing the capability of this model for the detection of hidden relationships and hidden networks which may not be readily noticeable by non-experts or without significant time-consuming investigation.

III. To improve name screening

A key component of AML/CFT measures is the screening of the names of customers, as well as the connected parties of customers and counterparties to transactions, against money laundering and terrorism financing information sources, including sanctions watchlists, politically exposed persons lists, and adverse news sources. Data analytics can be used to improve these screening and name matching capabilities. One of the more common uses is fuzzy logic, which are algorithmic processes used to detect and evaluate near matches. Another use is natural language processing technology for screening against adverse news, as this often involves the analysis of unstructured data from multiple sources. These use cases have the potential to improve the accuracy of matching, thereby reducing false positives and increasing detection rates.

Also with a view to addressing the false positives problem, more advanced data analytics solutions can also be used for the automatic disposition of false hits, together with explanation of how the hits are disposed which will also enable institutions to have a level of explainability in their use of analytics.

Another potential use of analytics with regard to name screening is the monitoring of screening processes from a process improvement perspective. This includes monitoring the operational impact of hit rates and the tuning of parameters.

Maturity: Most FIs recognise the potential for data analytics solutions to improve name screening. However, the methods are at a relatively early stage of adoption, with many FIs exploring or experimenting with such solutions. Some banks have already reviewed and have decided not to pursue this at this time.

Example: A vendor shared that it has built a model which leverages features extracted from historical name matching data from customers, vendors and banks to train an ensemble of supervised machine learning models. This model was tested on screening data and the output used to improve screening rules and policies.[12]

IV. For audit/assurance

Instead of traditional judgmental or statistical sample-based testing using data from individual systems within the bank, analytics solutions can allow for evaluation of complete populations of data sets for testing and identification of weaknesses, for example, assessing whether rules are performing as intended and evaluating the alert backlog. Other data analytics uses are use of data mining and data visualisation tools for deriving and presenting fact-based insight into risks and data driven representation of the control environment.

The use of such tools allow FIs to perform ongoing continuous monitoring and assessment rather than cyclical assessment. This also facilitates ongoing continuous feedback to management. In addition, these solutions may represent cost savings in processing of information and collection of evidence for the purposes of audit.

Maturity: Most FIs recognise the potential for data analytics solutions to be used for audit/assurance purposes. However, the methods are at a relatively early stage of adoption, with many FIs exploring or experimenting with such solutions.

[12] Contributed by Delta Capita.

Example: A vendor shared that it has observed amongst FIs the use of data analytics to monitor the screening process real-time with a view to overseeing operations, identifying breaches in service level agreements, or implementing up to date watchlists.[13]

In addition to the themes above, some new and experimental use cases or solutions the WG may have seen in the market include:

1. Self-learning models

   We have discussed above how data analytics can be used to enhance rule-based transaction monitoring through the refinement and tuning of applied rules or parameters. Self-learning models are models that are able to re-weight rules and manage applied thresholds automatically and unsupervised over time. Such models have the potential to save time otherwise spent on the manual tuning of models. Further, models can be constantly and automatically updated and tuned to meet changing AML/CFT risks or typologies with much shorter turnaround time than a periodic tuning approach.

2. Customer-specific models

   Most models used for transaction or account monitoring function with reference to typologies or rules established for the customer type. Some vendors are experimenting with the building of models on a per-customer basis, reducing dependency on these typologies or rules, with a view to enabling FIs to detect outlier behavior for that specific customer based on its unique profile.

3. Analytics models to mimic "level 1" review

   As discussed above, models currently used for transaction monitoring generate alerts which are then reviewed by analysts, and if more complex or requiring escalation, by more senior analysts. There are new models that seek to mimic or replicate this "level 1 review" by analysts.

   Using supervised machine learning techniques which memorise past analyst decisions, advanced systems are able to automate low risk or recurrent payments decisions to reduce the number of level 1 alerts routed to analysts, allowing them to focus on higher risk transactions or hits. FIs have seen a range of outcomes when experimenting with this approach, and one vendor shared that it had in certain cases seen up to 70% reduction in the number of such alerts routed to analysts.

4. Use of analytics to conduct customer risk assessments

   We have discussed above how current solutions attempt to assess and score alerts for risk, allowing for higher prioritisation of higher risk alerts. Vendors are also experimenting with the use of analytics for the assessment of customer risk, which could then feed into bank-wide risk assessments, customer selection, and ongoing monitoring of customers' accounts and transactions. These customer risk assessments can be scored or unscored.

   FIs and vendors are also exploring the use of data analytics tools for the improvement of customer risk assessment through the aggregation of client risk metrics (e.g. name screening, adverse news, transaction monitoring, trade surveillance) into a singular view, to enhance holistic client risk assessment and monitoring. These solutions also seek to integrate different modules of AML/CFT risk programmes into a single customer view with multiple risk profile dimensions.

[13] Contributed by Fircosoft.

5. Search platforms

   One member bank shared that it is working with vendors to develop a platform powered by data analytics technology which is also used to underwrite the Google search engine, in order to explore the possibility of searching customers across the member bank as easily as current Google searches.

6. Data enrichment by building machine inference features

   This refers to the creation of additional information derived from the raw information stored in the systems. This can be used, for example, to compute whether a transaction occurred during high or low peak based on the timestamp of the transaction.

7. Cryptographic technology to address data protection considerations

   Section 3 will discuss how data protection regulations and restrictions can pose challenges to the adoption of data analytics use cases by FIs. One vendor has suggested that FIs can explore whether new and developing cryptographic technology may enable them to share data for analytics purposes in a way that is compliant with relevant regulations.[14]

As demonstrated, there is a wide range of data analytics use cases and solutions – some may be more suitable for advanced players, while other simpler solutions may be appropriate for some FIs to consider when starting their analytics journeys. In each case, this will depend on a number of factors, including the individual FI's broader strategy, existing in-house skillsets, risk profile, and appetite for use of AI or automation.

The WG highlights that the above organisation of data analytics use cases by themes is not the only useful approach to thinking about the available technology. Exiger has suggested the alternative of categorising technology into four overarching types: (i) applications that define features for modelling (or "feature creation"); (ii) applications that facilitate the execution of data assessment or transformation (or "data transformation"); (iii) applications that allow you to replicate or model decisions (or "decision analytics"); and (iv) applications that monitor for data or model issues (or "witness models").[15] Most of the use cases described above may utilise more than one of these types of technology.

[14] Contributed by IBM: Fully Homomorphic Encryption (FHE) schemes, allow data to be processed in arbitrarily complex ways while it remains encrypted. Cryptographic multilinear maps have been used for the construction of cryptographic program obfuscation schemes, a major breakthrough that had been thought to be impossible. Unlike FHE, cryptographic multilinear maps and cryptographic program obfuscation are currently quite slow to be feasibly implemented. Further research and work is being carried out to bring it to the desired computational speed.

[15] Contributed by Exiger.

3. Some Key Considerations for Implementation

This Section highlights key considerations FIs may need to address when seeking to implement an analytics programme or specific solutions, and shares practical ways to address these challenges.

These key challenges are interconnected and sometimes overlapping. In addition to these key impediments, there is a range of other challenges that may arise in relation to each of three possible models for adoption: "Buy" (FIs purchasing solutions from vendors wholesale), "Build" (FIs independently and internally building solutions) and "Co-Create" (FIs working with vendors to build solutions). These challenges and their potential solutions or mitigants are explored in Appendix A.

3.1. Data quality, adequacy and availability

A significant challenge to the adoption of data analytics is that FIs may not have adequate data for effective development, training, and validation of models. While FIs may sit on large quantities of data, the data may be low-quality in that it is incorrect, inconsistent, incomplete, or outdated. FIs may struggle with remediating existing low-quality data, as well as putting in place the necessary frameworks and models to ensure that data is accurately recorded and updated in future. Further, the data that an organisation collects may also not be machine-readable (scanned copies, images), resulting in difficulties with the processing of data. The data may also be in a language (e.g. North Asian languages) that existing technology may not yet be able to accurately process.

Such data is also often disaggregated in that it is stored across multiple systems or locations. Aggregating and harmonising this data across an organisation can be difficult, particularly where data sets have been collected and maintained in disaggregated ways over a long time. In these cases, FIs may struggle to ensure the availability of and access to the data necessary for AML/CFT analytics models to function.

To enrich the data set, FIs could also benefit from more and better "labelled data" – in this context, financial crime data and adverse information that may already be known to other FIs or to government authorities, including for example data on convictions. Such labelled data is useful for the development, training and validation (including back-testing) of models. FIs may also use data from their own STRs to train and validate models; however, there is a risk that such data may affect the accuracy and effectiveness of such models, and would therefore need to be minimised in the data sets used for model training.

Solutions/Mitigants: FIs will need to put in place data governance policies, frameworks and controls that are designed to ensure:

(a) the quality of data, including completeness of data, how recently data was collected or updated, whether the data is structured in a machine-understandable form, and whether the source of the data would affect the interpretation of or reliance on the data. This also would include means of assessing the veracity of data;[16]

(b) the provenance of data is tracked, including keeping data provenance records or audit trail of data;[17]

(c) consistency of approach to data across the FI, including consistency of systems architecture and data formats. This is particularly so because AML/CFT departments are usually consumers of upstream data provided by various departments within the FI; and

(d) availability of and access to data as is appropriate to support its data analytics projects and as well as its broader data driven strategy. To address problems of access to internal data, FIs should consider adopting an enterprise approach to data, ensuring that data access frameworks are in place which allow AML/CFT functions to leverage FI-wide data as far as possible.

[16] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

[17] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

On enriching "labelled data", please see Section 4 for a discussion on potential public-private projects to address this.

3.2. No "blackbox" – enhancing explainable model outputs

Another significant challenge is vendor reluctance to explain "blackbox" models, i.e., where a model's internal workings are opaque to the user, as a result of which the output of the model is not explainable by the user. "Blackboxes" may arise where solutions are built externally by vendors who are unwilling to disclose how the models work on the basis that the technology or algorithms are proprietary.

The inability by FIs to explain these models could raise challenges such as:

(a) Model risk management – difficulty in ensuring that models are correctly applied, that biases are identified, and in monitoring model performance post-implementation.

(b) Audit and quality assurance – difficulty in reviewing and auditing the model's output.

(c) Ethical implications – FIs may struggle to satisfy themselves that the models are aligned with the FI's policies and standards relating to the ethical or responsible use of data.

(d) Regulatory considerations – FIs may not be able to sufficiently explain these models in order to give regulators the requisite comfort.

Solutions/Mitigants:

Therefore, consistent with accepted principles of risk management[18], FIs would take steps to understand and verify the performance of such models, and would not adopt the solution solely based on a vendor's representations as to the model's internal workings and capabilities.[19]

FIs may find the following approaches shared by the WG members useful when evaluating vendor models:

1 For appropriate risk management, FIs are seeking to understand a product's components, design and intended use, as well as how the system functions in principle. In addition to requiring vendors to provide such information, FIs have also obtained contractual assurances as to the model's features and aspects of its performance.[20] FIs are also seeking to understand the potential limitations of models used in the vendor solution, including both the limitations inherent in the model's assumptions, design and training (where applicable), as well as the risks and likelihood of model failure. Such understanding can help FIs to establish appropriate processes to address such models' limitations, and mitigate and, where possible, guard against the risk of model failure. FIs would also seek to ascertain the input data used by such models, particularly where this data comes from external sources that the FI may not know well. To assist with this, FIs could implement "black box recorders" that capture all input data streams.[21]

2 FIs are also seeking to understand and be able to explain the technical workings of the model, including the algorithms employed within it. However, obtaining full understanding of the model's technical workings may not always be possible – for example, if the model uses algorithms proprietary to a vendor, if the FI lacks the computing/IT technical expertise, or if it would require the investment of resources which are manifestly disproportionate to the extent of deployment of, benefit afforded by, and risk associated with the model/solution in

[18] See http://www.mas.gov.sg/Regulations-and-Financial-Stability/Regulatory-and-Supervisory-Framework/Risk-Management.aspx for MAS guidance on risk management practices.

[19] Paragraph 5.4.1 of MAS Guidelines on Outsourcing, 27 July 2016 sets out MAS' expectation that FIs should subject service providers to appropriate due diligence processes to assess the risk associated with outsourcing arrangements when considering such arrangements.

[20] These principles are consistent with MAS Technology Risk Management Guidelines, June 2013 which make reference to FIs ensuring that outsourced service providers grant the FI access to their systems, operations, documentation and facilities in order to carry out review or assessment for regulatory, audit or compliance purposes. The Guidelines also discuss FIs including contractual terms and conditions including performance targets and service levels of IT outsourced service providers.

[21] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

question. In view of this, where an FI is unable to achieve such complete understanding of a vendor solution, consideration of the following factors could be helpful in assessing the risks of adoption:

(a) Autonomy: This refers to the extent to which there is human intervention in the model's output or the FI's application of such output. Generally speaking, the lower the degree of human intervention, the higher the risk associated with not understanding the use of a "blackbox" solution. This is likely to be particularly relevant to proprietary vendor solutions with automated or predictive functions. For example, with reference to the use case themes outlined in Section 2, if an FI applies the output of an automated alert disposition solution with little or no further review, the risks that would arise from flaws in the model are higher, and correspondingly the threshold for explainable outputs and validation of the model would also be higher.

(b) Criticality: The extent of work undertaken by an FI to "unpack" the inner workings of vendor models/solutions should be proportionate to the impact and criticality of the model/solution in the FI's broader AML/CFT framework and ecosystem. In general, where the failure of a vendor model/solution may affect the effectiveness of an FI's key AML/CFT controls, the FI should invest more effort to understand and explain the technical workings of the model. On the other hand, less unpacking of the model could be acceptable where an FI is using the model in parallel with, or to supplement, more established and well-understood solutions.

3.3. Model validation and risk management

In addition to being able to explain model outputs, model validation also encompasses model performance and risk management. The challenge is assessing whether models perform effectively in accordance with their role and purpose, and verifying whether they subsequently continue to perform as intended. Model validation is also needed to identify the limitations of and assumptions underlying models, and to assess the impact of these limitations and assumptions. Model validation is a component of a wider model risk management framework that an FI needs to have in place.

Solutions/Mitigants: FIs have included the following elements in their model validation frameworks[22]:

(a) To the extent possible, appropriate testing of model for accuracy, robustness (i.e., resilience to outliers) and stability (i.e., ability to perform when underlying assumptions are altered), which testing assesses the model's potential limitations and evaluates its behavior over a range of input values. In particular, models are tested for "repeatability", i.e., whether the model performs consistently in same or similar scenarios. Otherwise, there should be design consideration for how exceptions should be identified and handled.[23]

(b) Framework for validation processes and techniques, including optimal thresholds and validation methods that ensure accuracy and efficiency, such as sensitivity analysis and benchmarking.

(c) Validation framework that includes evaluation of conceptual soundness and post-deployment monitoring and analysis:

i. identifying metrics that are useful in monitoring of model's performance; these metrics on underlying data should be monitored as well;

ii. continuous monitoring of models' performance after deployment to ensure that the model continues to operate to specification;

[22] See MAS guidance on management of operational risk at http://www.mas.gov.sg/Regulations-and-Financial-Stability/Regulatory-and-Supervisory-Framework/Risk-Management/Operational-Risk.aspx, which states that *"[a]n institution should establish sound and appropriate operational risk management strategies, policies and processes to identify, measure, assess, monitor, report and control or mitigate operational risk. These operational risk management strategies, policies and processes should be consistent with the institution's risk profile, risk appetite and capital strength…, and address all relevant aspects of operational risk prevalent in the business of the institution on an institution-wide basis*".

[23] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.[24] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

iii. analysis of cases where model output is ignored, altered or reversed by model users;

iv. active monitoring and tuning where models developed in static environments display inability when deployed in dynamic environments;[24]

v. periodic independent review of model against objective standards;

vi. benchmarking model's inputs and outputs to estimates against alternative internal or external data or surrogate models which provide similar statistical insights but are more easily interpreted; and

vii. to the extent possible, analysis of model output compared against actual outcomes, including back-testing.[25]

(d) Both at development stage and on an ongoing basis, model validation is performed by a team with clear governance, including staff with appropriate incentives, competence, experience and influence, with a degree of independence from model use.

(e) Review of not only individual models, but the suite of models used across the AML/CFT functions to ensure coordination and compatibility.

Organisations using data analytics models also need to have appropriate model governance frameworks and structures such that:

(a) there is appropriate identification of opportunities for analytics models;

(b) there is alignment of new models with broader AML/CFT framework and data strategy;

(c) there is appropriate design of models to suit the relevant use case or purpose, which must be clearly stated, and accommodating all stake-holder input. The model's design should be conceptually sound and mathematically and statistically correct;

(d) there is "traceability", i.e., there is audit trail or ability to trace input data streams;[26]

(e) there is review of model validation by independent auditors. There is in this regard an additional question of whether it is the model validation framework and process, rather than individual models, which should be subject to audit. It has been suggested that this is likely to be more agile and risk-based in view of rapidly developing technology and landscape;

(f) there is effective execution and implementation of models in accordance with the above model and broader AML/CFT framework;

(g) any model uncertainty and inaccuracy should be accounted for appropriately in the use of the model, including by supplementing the model's output with other models, approaches, analysis and/or information;[27]

(h) there are programmes for adjustment and refining of models to meet changes in requirements of time, including identifying a methodology for such review (metrics, sample size etc), and conducting parameter optimisation;

[24] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

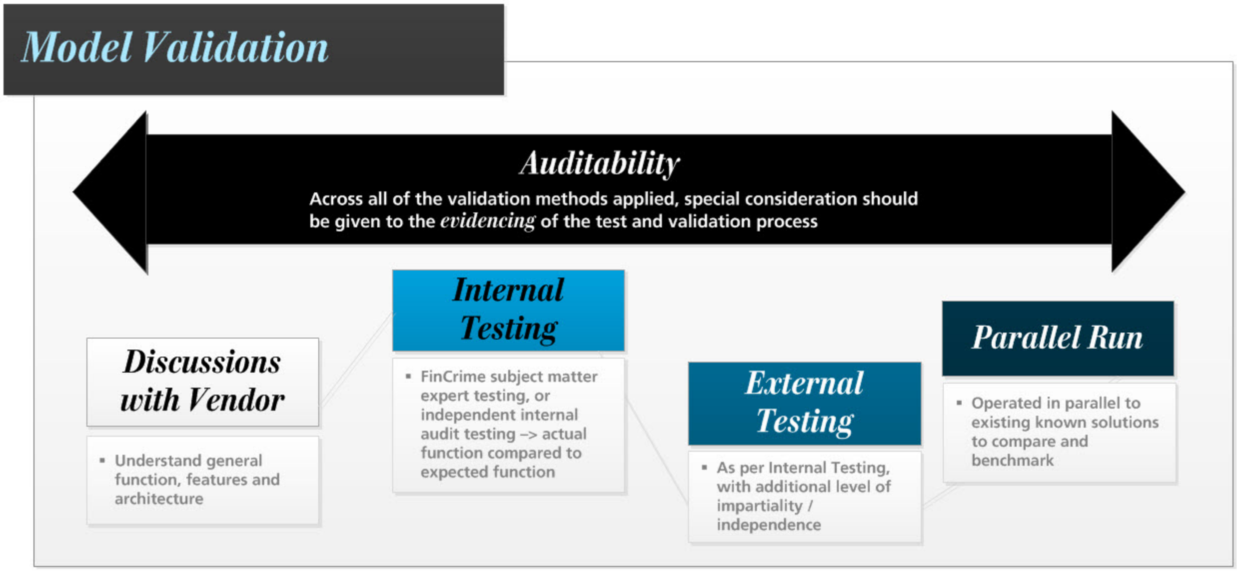
[25] Aspects of the MAS Guidelines on Risk Management Practices – Credit Risk, which discuss the back-testing of models used in the context of credit risk modelling, may provide guidance. However as discussed above, the back-testing of models in an AML/CFT context may be more challenging due to the lack of labelled data.

[26] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

[27] See MAS Guidelines on Risk Management Practices – Credit Risk, which in the context of credit risk models discusses the comparison of model output with other models and supplementing models with other methodologies in order to assess model accuracy and consistency.

(i) there is a framework for identification of possible misinformed decisions that taken based on outcome of a misunderstood model, and feedback of these learnings into model development and deployment; and

(j) to the extent access to labelled data is not possible and model validation limited as a result, an FI's model governance framework should provide a framework for decision-making and risk management in relation to such models.


Model Validation
Auditability
Across all of the validation methods applied, special consideration should be given to the *evidencing* of the test and validation process
Discussions with Vendor
▪ Understand general function, features and architecture
Internal Testing
▪ FinCrime subject matter expert testing, or independent internal audit testing –> actual function compared to expected function
External Testing
▪ As per Internal Testing, with additional level of impartiality / independence
Parallel Run
▪ Operated in parallel to existing known solutions to compare and benchmark

3.4. Regulatory considerations

FIs wishing to embark on analytics projects will need to have a clear understanding of the relevant regulatory considerations, particularly in respect of ever developing and evolving legal and regulatory requirements. FIs would need to evaluate any apparent inconsistencies between their proposed models and regulatory requirements, then design or adjust models in order to work within regulatory parameters. In addition to AML/CFT regulations, data privacy and protection laws are also relevant as they govern the collection, disclosure and use of data, especially personal data, in the jurisdictions within which the FIs operate or have customers.[28]

Solutions/Mitigants: To address regulators' concerns around the use of novel models or approaches, FIs need to demonstrate that they have adequately assessed the risks associated with their proposed use of such models (see Sections 3.2 and 3.3 above); and in view of the risks assessed, taken steps to achieve an appropriate level of "explainability" in relation to these models, as well as undertaken reasonable model validation and model risk management.

For example, in the context of sanctions screening systems, an FI could demonstrate that it has:

(a) Devised test cases covering a range of variations in names;

(b) Specified expected behavior and performance of the screening system in each test case;

(c) Created a simulation environment in which the test is run;

[28] In Singapore, the Personal Data Protection Act 2012 of Singapore (the "PDPA") specifies requirements for organisations to obtain consent for the collection, use or disclosure of personal data. The PDPA also imposes requirements in respect of the transfer of personal data outside of Singapore. Do however note that MAS Notice 626 on Prevention of Money Laundering and Countering the Financing of Terrorism – Banks (last revised on 30 November 2015) provides that "*[f]or the purpose of complying with this Notice, a bank may, whether directly or through a third party, collect, use and disclose personal data of an individual customer, an individual beneficiary of a life insurance policy, an individual appointed to act on behalf of a customer, an individual connected party of a customer or an individual beneficial owner of a customer, without the respective individual's consent*".

(d) Analysed the results of the simulation, identifying any gaps in the hits generated; and

(e) Explained and documented the rationale for any differential in hits.[29]

As part of broader engagement and partnership between FIs and regulators, where there is ambiguity as to regulatory or risk management considerations, FIs should engage regulators to arrive at a clearer position. In so doing, FIs should engage regulators with as much transparency as possible in order to collaboratively address any issues, achieve policy alignment on the industry's use of data analytics for AML, and give regulators appropriate comfort around the FIs' use of analytics. The participation of regulators and a commitment to engage with FIs in the search for effective solutions to appropriately address risks would encourage a legal and regulatory environment that facilitates innovation and help to realise the potential for uplift in AML/CFT capabilities. Increased engagement between FIs and regulators on the implementation challenges will also be beneficial for FIs when performing internal audits or testing on analytics-supported AML/CFT processes.

3.5. Organisational change management and a broader data strategy

One of the most significant challenges to adoption of data analytics solutions is organisational change management within FIs.

In particular, organisations may struggle with:

(a) legacy mindsets and fear of uncertainty. This includes a reluctance to confront legacy approaches that may not be working well, resistance to new approaches and the risk these may entail, and difficulty in prioritising these in parallel with enhancements to traditional programs. It also includes reluctance to change business processes to incorporate analytics insights into the workflow. FIs may also need to grapple with staff concerns that analytics, technology and AI will rationalise or diminish the significance of their roles;

(b) data analytics initiatives in different parts of the organisation creating duplication and inconsistency, and consequent inefficient use of organisational resources;

(c) insufficient or inappropriate skillsets; and

(d) in certain cases, a preconception that analytics projects are only appropriate for "big players".

Solutions/Mitigants: One key solution for organisational change resistance is a broader data-driven strategic vision and management buy-in. FIs without these are more likely to struggle with legacy mindsets, lack of underlying data architecture, and inadequate frameworks for making data available across business, compliance and financial crime applications where possible. For such FIs, management buy-in for AML-related analytics can be improved through education and increased discussion on the business case for data analytics solutions. See Appendix A for more detail on this.

As mentioned, another driver for organisational change resistance may be employee concerns that they will be replaced by such models. In order to address this, FIs could invest in education to reinforce to employees the potential for innovative measures to augment existing processes, and to allow for staff to spend their time on valuable tasks and activities requiring high value human judgment which cannot be replicated by technology solutions, rather than low value, repetitive tasks. This refers not just to practical "upskilling" but to instilling a culture of openness to change, improvement, and the leveraging of available tools, the benefits of which could include a smoother transition and less employee resistance as an FI undergoes a technological and AML/CFT processes shift.

To address the specific challenge of skillsets in the FI's workforce being insufficient or unsuitable for data analytics projects, FIs need to invest in 'up-skilling' staff with training on technology and analytics, so that existing staff and roles can be adapted to absorb incoming technological innovations and the disruptions (and improvements) that accompany them.

[29] Contributed by Fircosoft.

Analytics solutions can and should be proportionate to risk profile and operations of the organisation in question, its risk profile, and its operations. Such tools are not exclusive to large organisations, as is borne out by the range of solutions presented in Section 2 above. For the purposes of ongoing monitoring, MAS regulations require banks to "*put in place and implement adequate systems and processes, commensurate with the size and complexity of the bank, to monitor its business relations with customers; and detect and report suspicious, complex, unusually large or unusual patterns of transactions*".[30] In a similar vein, FATF recommendations guide FIs to adopt a risk-based approach in their conduct of AML/CFT measures.[31] In this regard, it should also be emphasised that most major algorithms, AI and machine learning methods are free or public tools. In addition, simpler data analytics can be performed with readily available tools (e.g. Microsoft Access/Excel) and member banks have found that these can often provide very effective results.

[30] See MAS Notice 626 on Prevention of Money Laundering and Countering the Financing of Terrorism – Banks (last revised on 30 November 2015).

[31] FATF Recommendations, recommendation 1 and interpretive note to recommendation 1.

4. Focus Areas for Industry and Private-Public Collaboration

In view of the key impediments to analytics adoption highlighted in Section 3, this Section discusses a number of focus areas for industry and private-public collaboration that the WG believes may be worth exploring further.

4.1. Skills and training

We discussed in Section 3 how the acquisition of appropriate skillsets is not only a significant component of organisational change management that is necessary for an FI to drive adoption of analytics solutions, but also crucial in order for FIs to be equipped to handle complex issues around model validation and model risk management. On a national level, the relevant talent pool in Singapore needs to be grown significantly to meet this demand. While this growth of the talent pool can be supported through the importing of talent from outside Singapore, steps should be taken to develop the necessary capabilities in the local workforce.

Existing infrastructure can be leveraged to achieve this – relevant skillsets can be added to the IBF Standards promulgated by the Institute of Banking and Finance (the "IBF"), which set out the functional skills required for various industry segments and guide IBF accreditation of structured skills training programmes. Specifically, careers in data analytics for AML/CFT can be added to the career roadmaps published by the IBF in order to educate practitioners on the career opportunities and core competencies they will need to develop to pursue these.

The IBF also administers programmes for professional conversion under Workforce Singapore's Adapt and Grow initiative, which could support FIs in re-skilling or up-skilling mid-career professionals for data analytics roles. Other sources of support for such mid-career up-skilling include the SkillsFuture Study Award for the Financial Sector which provides funding for training programmes in areas including data analytics.

Aside from career roadmaps and "upskilling", acquisition of appropriate skillsets can also be managed through the creation and promotion of talent pipelines from universities, polytechnics, and specialised vocational training institutes such as coding schools. These pipelines can be nurtured through increased collaboration between educational institutes and the financial services industry, with additional support from government initiatives and funding.

4.2. Regulatory engagement and dialogue

As discussed in Section 3, FIs will need to take regulatory considerations into account when adopting analytics for AML/CFT, including the need to demonstrate the effectiveness of new solutions in detecting and mitigating risks. Given the nascent adoption of data analytics solutions in the AML/CFT domain, FIs should continue to engage in early and continued discussions with regulators on specific use cases and areas of concern, so that new solutions are set up within a facilitative and sound framework.

One focus area for private-public partnership is the setting up of workshops involving both the public and private sectors, to discuss key issues around the use of data analytics for AML/CFT purposes. Such engagement could facilitate alignment, common learning, and increased clarity on issues such as model governance, modelling techniques and how data analytics solutions can target particular ML/TF risk areas.

4.3. Increasing data availability and standardising data quality

Section 3 also discusses the need for more and better labelled data and resultant challenges FIs face in model training, validation, and risk management. Problems of data availability present even broader challenges to the AML/CFT framework, as data silos are vulnerable to exploitation by criminals. Opportunities may therefore exist for cross-industry and private-public partnerships to address or mitigate these problems.

For example, information sharing between FIs could be facilitated through industry utilities or data repositories. However, such projects are highly complex, involving issues such as the disclosure of customer information or personal data, and would require significant government and regulatory support to resolve. Another complexity is the consolidation of data from multiple FIs, which would require the alignment of these data sets so that they can be effectively shared and used. In this regard, the ACIP paper on Legal Persons – Misuse Typologies and Best Practices has suggested that regulators and relevant authorities share standardised data sets and risk analytics with the industry to help FIs enhance their risk-based AML/CFT programmes and to facilitate industry collaboration. This could include a discussion on the critical data such authorities require in order to uplift their intelligence capabilities. Government authorities and regulators could then also facilitate the dissemination of industry-wide information such as financial crime data through government-led initiatives supported by FIs as information providers. Additionally, regulator-industry partnerships aimed at facilitating the sharing of specific financial crime data and adverse information from government authorities to FIs could increase the supply of labelled data to FIs. Such projects could potentially be supported through technology-based solutions for data sharing and privacy preservation, including cryptographic technology referred to in Section 2.

These information-sharing initiatives can enable FIs to obtain a range of information from credible sources and perform more comprehensive risk assessments of customers. This in turn could improve detection of risk factors by FIs, and on the other hand reduce unfair de-risking by FIs based on incomplete and less credible adverse information.

An additional area for private-public collaboration that would support such information sharing projects is the development of a common industry approach towards and standards for data governance to improve consistency of data quality across organisations.

4.4. POOLED, NATIONAL LEVEL NAME SCREENING

Another opportunity for partnership across the industry in the AML/CFT analytics space is pooling KYC name screening at a national level. Given that there is likely to be material overlap in the customers and names being screened by different FIs, pooled screening has the potential to bring about the following benefits:

(a) efficiencies and cost savings across the industry from reduced duplication of work;

(b) enlargement and enrichment of database against which screening can be conducted;

(c) collective resources and knowledge of the industry can be applied to uplift screening standards, source and incorporate best-in-class and cutting-edge analytics and technology; and

(d) establishment of an industry "gold standard" for KYC name screening to provide clarity and facilitate coordinated uplift of standards.

5. Future Vision

The focus areas highlighted in Section 4 are areas where the WG believes the industry and government can leverage private-public collaboration and collective resources to enable FIs and the industry to progress in their journeys of adopting analytics for AML/CFT.

However, the potential for data analytics to transform the fight against ML/TF lies not only in increasing adoption, but in the interconnectivity of data sources and analytics applications. Just as this paper has discussed how the adoption of analytics can be driven through the use of simple and widely available analytics tools, much of the value to be unlocked lies not in cutting edge technology but in the interlinking and integration of existing data and capabilities. At the time of writing, there are already private sector projects with the ambition of building platforms which leverage and interlink analytics capabilities to enable data aggregation providing a single view of customers and cross border coordination across the Financial Intelligence Units (FIUs) of multiple jurisdictions.[32] Given that money laundering, sanctions evasion and terrorism finance are by their nature complex and cross border, and that current data and operational silos hinder effective AML/CFT measures as well as result in significant inefficiency, these projects propose to leverage existing initiatives and technology to eliminate such silos as far as possible, reduce unnecessary de-risking, and improve efficiency across the AML/CFT framework.

While such projects are likely to be extremely challenging[33], if successful, they could point to a future state in which the global AML/CFT framework is much more tightly linked and responsive to financial crime threats.

With an eye on this future state, and the awareness that increased adoption of analytics and growth in analytics capabilities are necessary incremental steps in the journey towards this future state (and beyond), it is the WG's hope that its findings and suggestions in this paper will be useful in guiding and furthering readers as well as the broader industry in taking these next steps.

[32] Information on project contributed by Barclays Bank PLC, Singapore Branch.

[33] Likely challenges include governance and ownership (given the wide range of stakeholders and interests involved), funding, commercial viability and business model, legal liability for errors or failures, data aggregation challenges, data sharing and protection/privacy challenges, legal/regulatory challenges, and technology requirements.

6. APPENDICES

APPENDIX A – CHALLENGES

A. COMMON CHALLENGES IMPEDING ADOPTION

	Challenges	**Solutions**
1.	Organisational change management See Section 3.	See Section 3.
2.	Funding and building a business case FIs may face challenges in building the funding or budget models needed to obtain funding for the adoption or exploration of data analytics solutions. These challenges include difficulty in quantifying the benefits of such programmes and building of a business case.	FIs should build a business case using a standard cost-benefit analysis, quantifying benefits to the extent possible. Some of the more quantifiable benefits include the following: (a) Time savings in terms of man hours. These can be measured in man hours or through number of full time employees dedicated to specified existing functions. These savings can be realised through either redeployment or rationalisation. (b) Cost savings in terms of licence fees and costs through decommissioning of legacy systems and platforms. Some of the other benefits FIs can anticipate reaping which may be less quantifiable include the following: (a) Enhancing client experience through quicker onboarding and processing, less iteration within this process and less back and forth with clients. (b) Generating additional data through these systems, which data can then be used in other business applications and potentially to support revenue generating activities. Alternatively, FIs could assess the expansion of existing or planned data analytics initiatives for business application to also cover ML/ TF risks. This is particularly relevant for FIs adopting a broader enterprise-wide data strategy. (c) Risk mitigation. Quantifying this benefit may be particularly challenging. Some options for quantifying include extrapolating risk mitigation savings from estimations of costs avoided, including for example regulatory fines, estimated with reference to fines previously imposed on firms, and remediation costs, similarly estimated with reference to internal project costs for

	Challenges	**Solutions**
		previous regulatory remediation projects.
3.	Data security FIs adopting a data strategy will encounter challenges in maintaining the security of their data, including preventing data loss and theft, and unauthorised access. While these risks can be mitigated through encryption of data at rest and in transmission, analytics processes which generally require data to be decrypted will result in additional exposure.	FIs should invest in the enhancement of their data security frameworks, as well as focus on the operationalisation of these frameworks which can be especially challenging. These frameworks should include data loss prevention, data access controls and encryption requirements.
4.	Ownership and use of data and data-derived assets Another factor that may impede FIs looking to implement data analytics for AML/CFT is lack of clarity around the ownership of and rights to use data and data-derived assets.	An FI should establish clear policies and frameworks for its position on ownership and rights to use data and data-derived assets. This would support not only the FI's analytics projects for AML/CFT purposes, but the FI's broader data driven strategy.
5.	Model risk management In addition to the considerations highlighted in Section 3, FIs may face the following risk management challenges in employing data analytics models: (a) the potential misapplication of models (where models are applied inappropriately to data or problems); (b) the possibility of models becoming biased (for example, due to the human input or underlying assumptions being biased, or due to bias in the training data used); (c) models overfitting (where the analysis produced by the model "overfits" the particular data set such that it cannot reliably generate observations from other data sets); and (d) the FI's difficulty in monitoring model performance.	See Section 3.
6.	System integration FIs considering data analytics solutions and projects may be deterred by challenges in the integration of such models with their current systems.	An FI should consider which of the "Buy", "Build" or "Co-Create" models would be most effective for it to integrate new analytics solutions. See below for a discussion on the challenges and their potential solutions or mitigants of such models.
7.	Post-implementation challenges After the initial implementation of analytics solutions, FIs may face challenges arising from the	As these challenges are closely related with model risk and data governance risks identified above, the solutions for managing these challenges are also linked with some of the solutions identified above.

	Challenges	**Solutions**
	following: (a) changes in IT architecture which may not be compatible with the models built; (b) changes in the risk and regulatory environment which impact the models or their implementation; (c) use and interpretation of models by non-technical stakeholders and new joiners who may be less familiar with aspects of the model, resulting in risk of misapplication; and (d) data security issues that may be present during production. Many of these challenges arise from and/or overlap with the risks identified earlier in this table in relation to model risk and data governance considerations.	In particular, organisations should consider the following: (a) changes in data structure and IT architecture should be closely monitored to ensure that risks are anticipated and managed. To minimise the impact of changes in IT architecture, the initial design and build of analytics models should be part of the organisation's broader data and IT strategy, and should therefore take into account any proposed future changes to the IT architecture. Models should also be designed and built with maximum flexibility and using open architecture where possible; (b) model risk management frameworks should include ongoing review of the regulatory and risk environments to determine the impact of any changes to these environments on the existing models and to identify measures for the management or mitigation of these risks; (c) guidance as to a model's design, purpose and application should be incorporated and documented as part of the design and build process for a model, and such guidance should be made available to non-technical stakeholders and new joinders. Appropriate processes should be put in place for the clarification of these matters as necessary to non-technical stakeholders; and (d) FIs should develop appropriate frameworks for data governance, including data security.
8.	Ethical or responsible use of data Some of the key risks associated with extensive use of data analytics relate to potential ethical issues. These issues may arise, for example, where sensitive or discriminatory data (e.g. gender, ethnicity, religious views etc) is used, or where the use of data may be discriminatory or intrusive. Such issues may also give rise to legal and reputational risks on the part of FIs. These ethical implications and attendant risks may be particularly difficult to monitor or manage in the case of "blackbox" models i.e., where the model's internal workings are opaque to the user, as a result of which the output of the model is not explainable by the FI. We address the challenges associated with such "blackbox" models in more	Organisations intending to embark on data analytics solutions or projects should put in place appropriate governance and controls frameworks for the responsible use of data analytics and handling ethical issues that may arise. These frameworks should be designed and developed with reference to regulatory guidance in this area, in the context of Singapore: (a) MAS' proposed FEAT (Fairness, Ethics, Accountability & Transparency) guide which is to set out key principles and best practices for the use of AI and data analytics; and (b) The PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018. Some of the key components and principles for the

	Challenges	**Solutions**
	detail below.	said frameworks:[34] (a) Internal governance: constructing or adapting corporate governance to ensure clear roles and responsibilities, oversight mechanisms, monitoring or reporting systems, and periodic review of governance; (b) Risk and/or harm mitigation: conducting and documenting risk and impact assessments which incorporate ethical considerations; (c) Data accountability; (d) Model risk; and (e) Policies for disclosure, transparency and explainability *vis-a-vis* customers, where appropriate. Institutions with global operations will need to ensure that analytics programmes supporting those operations are also aligned with guidance on ethical or responsible use of data that may from time to time be published in relevant jurisdictions. In relation to "blackbox" models which are bought, and where vendors are unwilling to share proprietary information regarding the model, FIs will need to consider alternative means of satisfying themselves that the models are acceptable to the FI from an ethical or responsible use of data perspective. This may include asking what types of data fields are used in the algorithm, paying special attention to potentially sensitive or discriminatory fields. FIs may also consider asking for confirmations that such fields are not inferred or reverse engineered either.
9.	Legal and reputational risk management Legal and reputational risks may arise from analytics projects, including potential data use or data protection breaches. These challenges are closely linked with the challenges identified above in relation to data governance, model risk management, and implications in respect of the ethical or responsible use of data. Beyond data protection and security, cross-jurisdictional issues may arise in relation to data sovereignty or data portability. For example, where data is collected in a jurisdiction where applicable laws mandate that such information must be held onshore within that jurisdiction, or include strict	See paragraph on regulatory considerations under Section 3. As mentioned under "data security" above, data analytics projects may also result in a heightened risk of data leakage, not only from cyber attacks, but from unintentional leakage that can arise during the sharing or transmission of data. As such leakages may result in legal and reputational risks to FIs, FIs should consider putting in place additional controls to monitor for and guard against these leakages. Aside from strict legal requirements, FIs should also consider the ethical implications of their data analytics projects as these may give rise to reputational risks notwithstanding that strict legal

[34] PDPC's Discussion Paper on AI and Personal Data – Fostering Responsible Development and Adoption of AI, published on 5 June 2018.

	Challenges	**Solutions**
	rules as to the portability of that data, such restrictions may affect the viability of data analytics programmes or solutions.	requirements may be met. Having in place clear and considered frameworks for handling these implications and comprehensive documentation of decision making will help to mitigate such risks. As suggested under item 8 above, the alignment of the FI's internal governance and frameworks with guidance or standards from regulators or government authorities should also serve to mitigate potential reputational risks.

B. CHALLENGES TO "BUY" IMPLEMENTATION

	Challenges	**Solutions**
1.	Explainability of models or the "blackbox" problem See Section 3.	See Section 3.
2.	Capabilities Although bought solutions will have been designed and built externally, FIs will require specialised skillsets in order to adequately review and evaluate the solutions and the vendors' claims, evaluate the interaction of models with the FI's own systems, and customise solutions for the FI's purposes. There is an additional concern that over-reliance on "bought" models will not encourage the necessary building of in-house expertise in the FI to evaluate or adapt to new "bought" models when necessary.	FIs may consider getting dedicated human resourcing either via "upskilling" of current staff or external hire. Please refer to the paragraphs on Organisational Change Management in Section 3 for more detail. FIs should also ensure transfer of knowledge as far as possible, and ensure that it builds in-house knowledge. FIs should also have contingency plans for situations where the "bought" model becomes unavailable.
3.	Suitability and customisability of pre-built solutions FIs looking to buy pre-built solutions may find that certain analytics solutions are not specifically built for AML/CFT purposes, or are otherwise unsuitable for the FI's unique needs. This is particularly so where vendors only offer "off the shelf" solutions and are unwilling to customise these for FIs. This may result in models being not "fit for purpose", and analytics models being forced on use cases with sub-optimal results.	FIs should determine the suitability of vendor models, including obtaining information on the data used to develop the model and assessing whether and to what extent that data is reflective of the FI's situation. Instead of forcing pre-built models on cases where they are not fit for use, FIs could consider developing partnerships with vendors who are willing to tailor and customise solutions to optimise effectiveness for the FI's intended use. FIs could also require vendors to build in specific features such as explainability-related features to assist FIs in meeting regulatory or audit requirements. The FI should document its customisation decisions. The above approach may be relatively easier for bigger FIs undertaking larger scale projects to negotiate. Smaller FIs may need to invest proportionately more to customise a solution.
4.	Flexibility of models	In addition to customising solutions, FIs should also

	Challenges	**Solutions**
	FIs who buy analytics solutions "off the shelf" may also find that the models lack flexibility and are unable to respond to emerging risks or changing environments.	consider the flexibility of solution when choosing which to buy and implement. FIs could require vendors to demonstrate such flexibility, or pre-build features that would allow for such flexibility. This may represent an additional cost to vendors which may result in increased pricing. FIs could consider engaging with vendors who provide other applications or services for business requirements, as such vendors may have more appetite for investment with the FI.

C. CHALLENGES TO "BUILD" IMPLEMENTATION

	Challenges	**Solutions**
1.	Model development and requisite skillsets The key advantage of internally developed models is the capacity for total customisation to suit the organisation's needs, including its unique control framework and risk footprint. However, some main challenges that FIs would expect to face are that of independently developing potentially-complex analytics models, how to start the process and sustain it, and concerns around whether the FI has the necessary skillsets for this within its workforce. For this reason, such internal solutions may take much longer at the initiation stage and may also require more resources than 'Buy' or 'Co-Create' approaches.	FIs intending to build their own analytics solutions will need to align their human resource strategy with these plans in order to ensure their teams have the requisite skillsets and experience to develop analytics models for AML/CFT purposes. On a national level, the relevant talent pool in Singapore needs to be grown significantly to meet this demand. This can be achieved through the importing of talent from outside Singapore as well as through the "upskilling" of Singapore's existing workforce, as discussed in more detail in Section 3. In this regard, one member bank shared its experience that in certain cases, "Build" models have been a conducive environment for such "upskilling" of staff, as well as for increased cross-team collaboration and improved inter-organisational model awareness and governance.
2.	Staffing A related challenge is that of staffing analytics projects appropriately to ensure the requisite skillsets and backgrounds are represented.	FIs should ensure that such projects are staffed with the appropriate mixture of data scientists and analysts as well as AML/CFT subject matter specialists.
3.	Legacy IT architecture An FI's existing IT architecture may not be suited for the building of an analytics solution in-house.	FIs will need to adopt a broader data and technology driven strategy, which should encompass and drive investment appetite for constructing and transitioning to new IT architecture that support such a strategy.
4.	Availability and labelling of training data This relates to the challenges highlighted in relation to data governance and data adequacy. These challenges will be more acute in "Build" and "Co-Create" approaches where data may be required for training of models.	FIs will need to focus on building up a sufficient database of use cases and labelled training data. See Section 3 and 4 for more details.

	Challenges	**Solutions**
5.	Model validation The challenge of model validation is likely to be more acute in "Build" and "Co-Create" approaches where FIs will themselves have played a significant role in the model development.	See Section 3.

D. CHALLENGES TO "CO-CREATE" IMPLEMENTATION

	Challenges	**Solutions**
1.	Intellectual property ownership One key challenge FIs may face in co-creating models with vendors is how to resolve questions around the ownership of intellectual property which is generated through the co-creation process and which supports the eventual model. These may arise where the FI has invested heavily in the design and refinement of a technology solution with a third-party vendor.	FIs who have invested significantly in the co-creation of models should at the outset negotiate and settle issues relating to the intellectual property which is generated through co-creation. FIs should consider: (a) agreements to ensure that FIs will continue to have the rights to use such intellectual property in the context of AML/CFT analytics models, including future models the FI may wish to build or co-develop; and (b) their rights in relation to the product of co-creation, including the commercial value of such product. The industry and government could also consider a partnership to discuss and determine standard co-creation ownership conditions.
2.	Model development and requisite skillsets As with the "Build" approach, a "Co-creation" approach will also require FIs to engage in the building and development of models, raising similar concerns around whether the FI has the necessary skillsets for this within its workforce (although to a lesser degree than in a pure "Build" scenario).	FIs will need to align their human resource strategy with these plans in order to ensure their teams have the requisite skillsets and experience to develop analytics models for AML/CFT purposes. On a national level, the relevant talent pool in Singapore needs to be grown significantly to meet this demand. See paragraphs on Organisational Change Management in Section 3 for more detail.
3.	Legacy IT architecture An FI's existing IT architecture may not be suited for the co-building of an analytics solution in-house.	FIs will need to adopt a broader data and technology driven strategy, which should encompass and drive investment appetite for constructing and transitioning to new IT architecture that support such a strategy.
4.	Availability and labelling of training data This relates to the challenges highlighted in relation to data governance and data adequacy. These challenges will be more acute in "Build" and "Co-Create" approaches where data may be required for training of models.	FIs will need to focus on building up a sufficient database of use cases and labelled training data. See Section 3 and 4 for more detail.
5.	Sustainability	FIs should drive the transfer of knowledge from vendors to their own personnel. To this end, FIs need

	Challenges	**Solutions**
	FIs which have opted for a "Co-creation" approach because they may have originally lacked the skillsets required for model development under a "Buy" approach may be concerned with the sustainability of the solution and over-reliance on a third-party vendor.	to ensure that they have an appropriate mixture of data science/analytics expertise and subject matter expertise across their internal teams to acquire such knowledge.
6.	Model validation The challenge of model validation is likely to be more acute in "Build" and "Co-Create" approaches where FIs will themselves have played a significant role in the model development.	See Section 3.

APPENDIX B – DATA ANALYTICS WG MEMBERS AND OTHER CONTRIBUTORS

WG Members

Firm	Representative(s)
BNP Paribas	Andrew Chow
DBS Bank Ltd	Lam Chee Kin Yaw Tan
Citibank N.A.	Rashmi Dubier Khor Boon Keng
The Hongkong and Shanghai Banking Corporation Limited	Victor Eng Antony Lee Samuel Ong Karen Yeo
Oversea-Chinese Banking Corporation Limited	Terence Gue
Standard Chartered Bank	Jodie Arthur Zubin Chichgar Stuart Christmas
UBS AG	Andrew Barker Mabel Ha
United Overseas Bank Limited	Eric Ang Boon Hin

Invitees

Firm	Representative(s)
KPMG Services Pte. Ltd.	Lem Chin Kok Elisa Ang Eric Poh
Credit Suisse AG	Rayson Tan
Barclays Bank PLC, Singapore Branch	Richard Carrick

Other Contributors

Firm
Delta Capita
Exiger
Fircosoft
IBM
SAS

ACIP Secretariat Representatives

Commercial Affairs Department
Monetary Authority of Singapore

The WG thanks Dentons, Rodyk & Davidson LLP (Sarah Chan) for drafting and project management of this paper.

APPENDIX C – GLOSSARY

Terms	Description
ACIP	Anti-Money Laundering and Countering the Financing of Terrorism Industry Partnership
AI	artificial intelligence
AML	Anti-Money Laundering
AML/CFT	Anti-Money Laundering/Combating the Financing of Terrorism
CAD	Commercial Affairs Department
FATF	Financial Action Task Force
FI	financial institution
FIU	Financial Intelligence Unit: a national centre for the receipt and analysis of STRs and other information relevant to money laundering, associated predicate offences and the financing of terrorism, and for the dissemination of the results of that analysis.
IT	information technology
IP	intellectual property
KYC	know-your-customer
labelled data	a set of data for which the target answer is already known, such as “fradulent” or “true positive”
machine learning	an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed
MAS	Monetary Authority of Singapore
PDPC	Personal Data Protection Commission
STR	Suspicious Transaction Report
STRO	Singapore’s Suspicious Transaction Reporting Office
WG	Working Group