

**OECD AFRICA ACADEMY FOR TAX AND FINANCIAL CRIME INVESTIGATION**  
**Anti-Money Laundering: Current Trends, Prosecutions, and the Challenges**  
**presented by Crypto Assets (Specialty) Virtual Programme**

Analytics and AI as a Tool for Investigators

Kong Yew Hon  
Jun 2025



---

INLAND REVENUE  
AUTHORITY  
OF SINGAPORE

# Learning Objectives

At the end of the course, you should be able to understand and appreciate:

- ✓ Overview of legal framework to combat money laundering in Singapore
- ✓ Interaction between tax and money laundering
- ✓ Identifying potential tax money laundering cases

(1) Combatting Money Laundering in Singapore



- ✓ AI as a Strategic Capability
- ✓ IRAS AI Strategy
- ✓ IFD's Data Journey

(2) Overview of IRAS AI\* Strategy



- ✓ What is Big Data?
- ✓ The different sources of information
- ✓ Data governance and limitations of certain data
- ✓ Application of concept : Case Study
- ✓ Group discussion

(3) All About Data



- ✓ Leveraging on technology & analytics in financial investigation
  - Analytics as a gatekeeper;
  - Analytics for Tax Crime;
  - Analytics for investigative efficiency;
  - Analytics to facilitate funds tracing
- ✓ Application of concept : Case Study
- ✓ Group discussion
- ✓ What's next

(4) Analytics Tool for Investigators

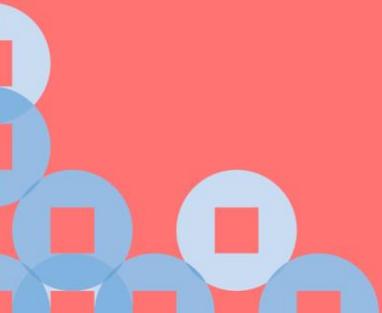


\* AI = Artificial Intelligence

[www.menti.com](https://www.menti.com)

Access via Code :

# (1) Combatting Money Laundering in Singapore



# Overview of legal framework to combat money laundering in Singapore



The **Corruption, Drug Trafficking and Other Serious Crimes (Confiscation of Benefits) Act 1992 ("CDSA")** is the anti-money laundering ("AML") legislation in Singapore, which criminalises the laundering of proceeds derived from corruption, drug offences and other serious offences ("predicate offences").



The **Commercial Affairs Department ("CAD")** is one of the Law Enforcement Agencies ("LEAs") in Singapore which enforces the AML regime through detection of ML activities, investigation and prosecution of ML offences, seizure and confiscation of proceeds from predicate offences.

**Tax ML cases** are under CAD's purview.

# Interaction between tax and money laundering

The following offences investigated by IRAS were designated as money laundering predicate offences (listed under Parts XII and XIII in the Second Schedule of the CDSA):

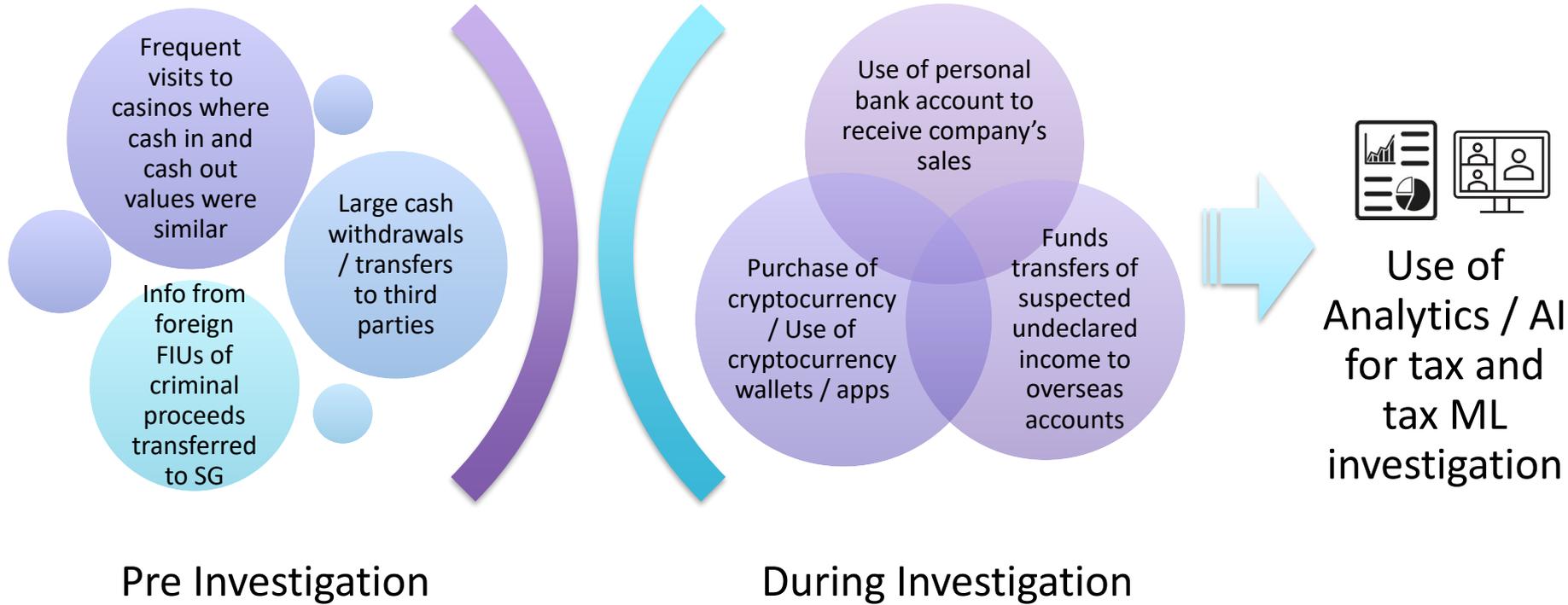
Date with effect from	Predicate Offence
1 July 2013	Section 62, 63 GSTA Section 96, 96A of ITA
1 September 2014	Foreign serious tax offence
3 June 2015	Section 37J(3), (4) ITA



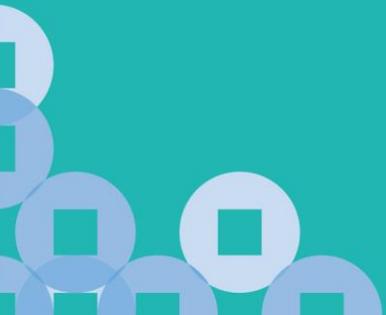
*Serious tax crimes involve omissions, falsifications or fraudulent conduct perpetrated with wilful intent to evade tax/obtain fraudulent refunds and payouts, or to assist others to evade tax/obtain fraudulent refunds and payouts*

This designation enables the powers used to investigate ML to be applied to the proceeds of designated tax crimes.

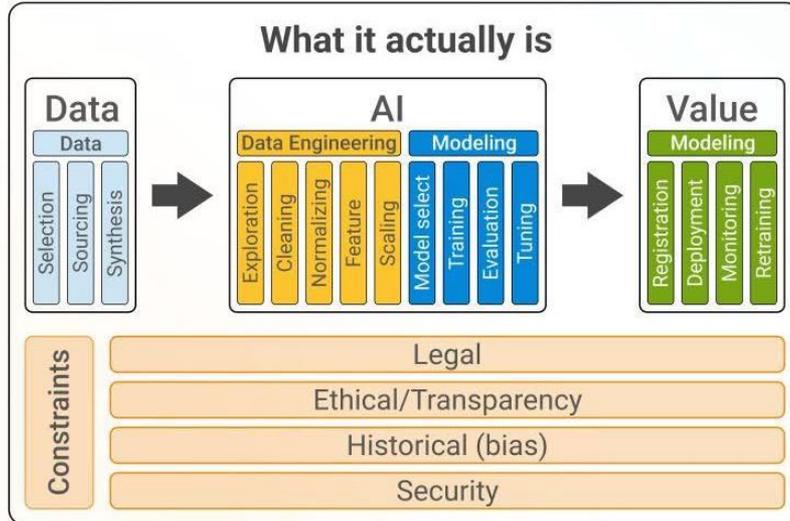
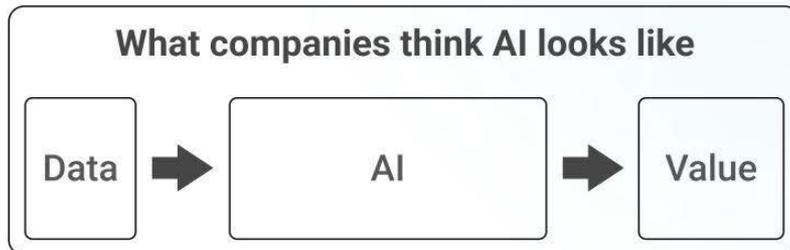
# Identifying Potential Tax ML cases *(indicators are not exhaustive)*



## (2) Overview of IRAS AI Strategy



# What is AI?



# Scaling AI under IRAS Data Strategy

In FY2020, we formulated the **IRAS Data Strategy** to support our LEA:D 2.0 transformation. We seized the opportunity to identify new capabilities required to **scale up AI** developments in IRAS.

**SETTING CONTEXT:**  
Key Shifts in LEA:D 2.0



**WHY STRATEGISE:**  
Data Vision & Desired Outcomes

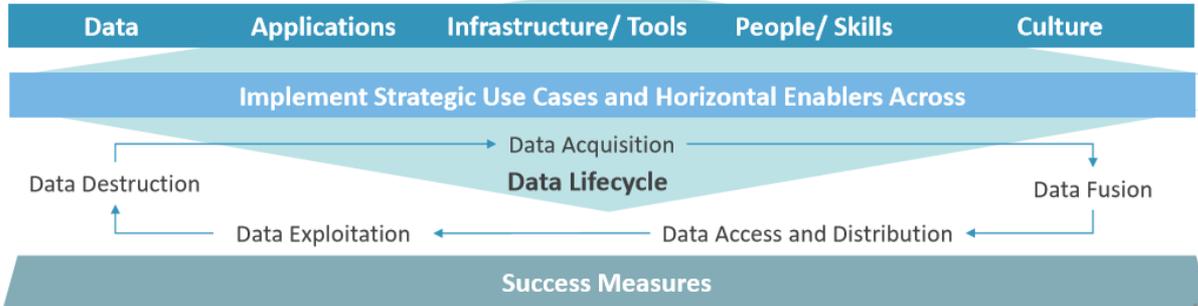
## BY 2025/2030: BIG-DATA & AI-POWERED IRAS



**WHAT PRIORITIES:**  
Key Strategies



**HOW TO ACHIEVE:**  
Implement Strategic Use Cases & Horizontal Enablers across the Key Building Blocks and the Data Lifecycle



# Our DSAI Goals and Vision

Aligning with the vision outlined in the IRAS’ Data & AI Strategies, DDED’s aim is to:

- Develop a set of enterprise DSAI product lines that enhance staff productivity, improve service and compliance processes across IRAS
- Continuously evolve and enhance the horizontal enablers that support or complement the DSAI product lines

## BY2025/2030: BIG-DATA & AI-POWERED IRAS

### AI Answering Assistant (A<sup>3</sup>)

A<sup>3</sup> leverages Generative AI (Gen AI), Large Language Models (LLMs) and its reach to IRAS’ tax knowledge bases to provide context-aware responses and tailored recommendations to augment IRAS workforce

### Service & Content Analyser

The Service and Content Analyser generates taxpayer insights from data collected across different contact channels (e.g., emails, calls and live-chat) and/or tax types to enhance WOI service quality and content understanding

### Compliance Toolbox

The Compliance Toolbox engineers ML/AI models and tools to strengthen IRAS’ capabilities in uncovering tax non-compliance and fraud, which include the identification of potential tax fraud, gaming of payout schemes, and other irregularities

## HORIZONTAL ENABLERS

Data Engineering Tools & Techniques:  
Speech-to-text, OCR, ETL, etc.

Unified Data Platform  
(UDP) + MLOps + Cloud  
DSAI Environment

AI Governance

AI Productivity Tools (e.g.  
Copilot, Pair, Azure AutoML etc)

# Getting the Investigation & Forensic Division (IFD) ready for the future

## IFD's Data Journey\*



### Before 2018 : Early Adoption of Analytics in IFD

- IRIN system containing tax data and key data obtained from other agencies
- Computer Forensics
- E-Discovery Tool
- Network analysis tool ("SNA")
- Digitising bank statements

### 2019 to 2022: Enhancing IFD's Analytics

#### Capability

- Optical character recognition ("OCR")
- Forensics Lite
- Automated Data Extraction (standard request) / Ad hoc Data Extraction (non-standard request)
- Transaction Tracing Tool
- Mobile Evidence Analytics Tool

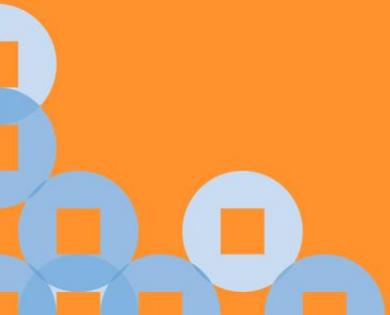


### 2023 onwards: Future of IFD's Analytics Capability

- Next generation network visualiser – Intelligent Network Analysis Tool ("iNAT")
- Cryptocurrencies
- Generative AI

\* Fully implemented for use by all officers including investigators, intel analysts, field officers and auditors

# (3) All About Data *[In the shoe of an Analyst / Investigator]*



# What is Big Data\*?

***“Big Data”** refers to extremely **large sets of data that are so complex and voluminous** that traditional data processing tools and techniques can't handle them efficiently.*

# 5V's of Big Data

- **Characteristics of Big data-** The five V's are volume, velocity, variety, veracity, and value\*.



Video :

[https://youtu.be/bAy\\_rObI7TYE?si=keT0GU\\_C1UsAb5\\_D9](https://youtu.be/bAy_rObI7TYE?si=keT0GU_C1UsAb5_D9)

# 5V's of Big Data



**Volume** : High amount of data points in a big data set e.g. volume of tax returns received by the tax administration are the in millions per year



**Veracity** : Trustworthiness and quality of the data e.g. missing data, bias in data etc



**Velocity** : Speed of data generation and how fast one can collect and process it e.g. posts in social media



**Variety** : Wide range of data types and sources in a big data set, including structured and unstructured data e.g. text, audio, visual etc

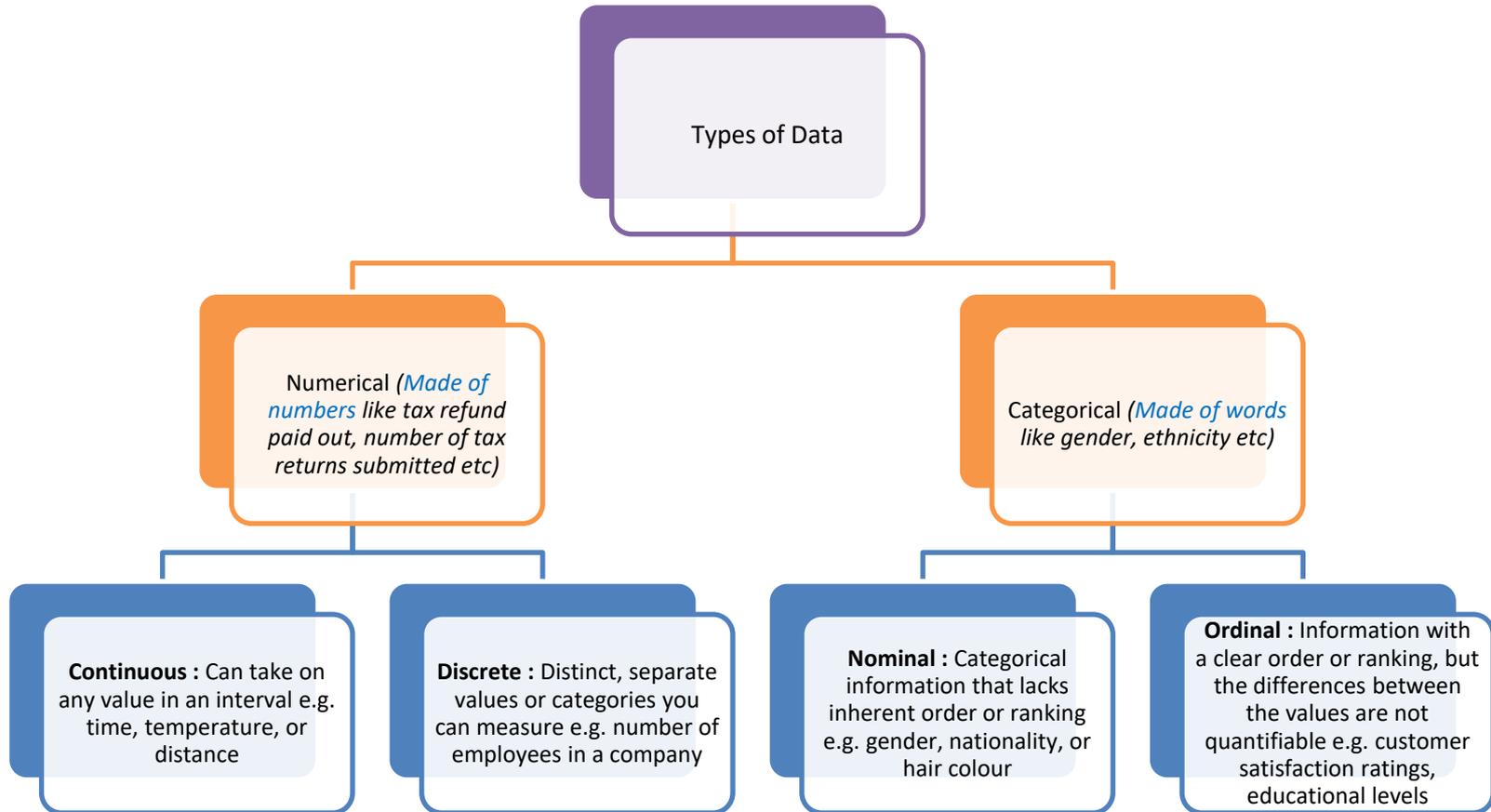


**Value** : The insights and patterns you can find e.g. uncover sentiments and opinions, numerical trends and patterns etc

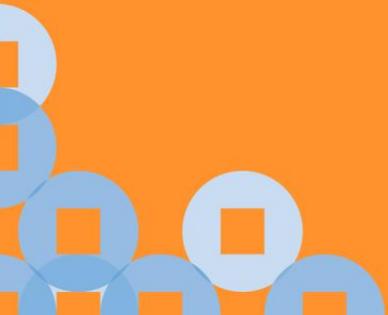
5Vs of Big Data

*When analyzed, big data can provide valuable insights into patterns, associations, behavior, and trends.*

# Types of Data



# Group Discussion



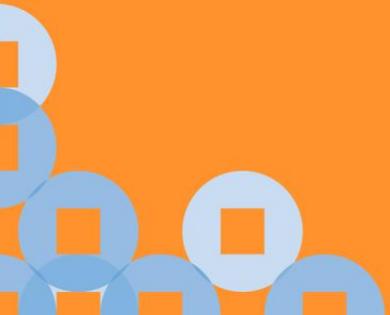
## Group Discussion (20 mins)

What are some use cases of big data for your agency?

Group 1 & 2 : Productivity gain e.g. automate tax filing

Group 3 & 4 : Improve effectiveness e.g. enhance case detection or selection

# Pros / Cons of Big Data



# Pros of Big Data

## 1. Improved Decision-Making Capabilities

Help business leaders make challenging decisions, helping them identify and weigh all the relevant factors that can affect the outcome of their choices

## 2. Better Customer Engagement

Enterprises can gather data from social media, sales records, customer feedback surveys, and other sources to learn more about their buyers.

# Pros of Big Data

## 3. Increased Opportunities for Social Good

Big data can also provide insights to spark positive, lasting changes—researchers have used it to identify domestic violence and homelessness trends

# Cons of Big Data

## 1. Data Privacy and Security Concerns

Collecting and analyzing large volumes of data increases the risk of unauthorized access, data leaks, and cyber attacks, posing privacy and security risks for individuals and organizations.

## 2. Cost of Implementation

Implementing big data solutions often involves significant upfront costs for organizations, including investments in infrastructure, software, and staff.

# Cons of Big Data

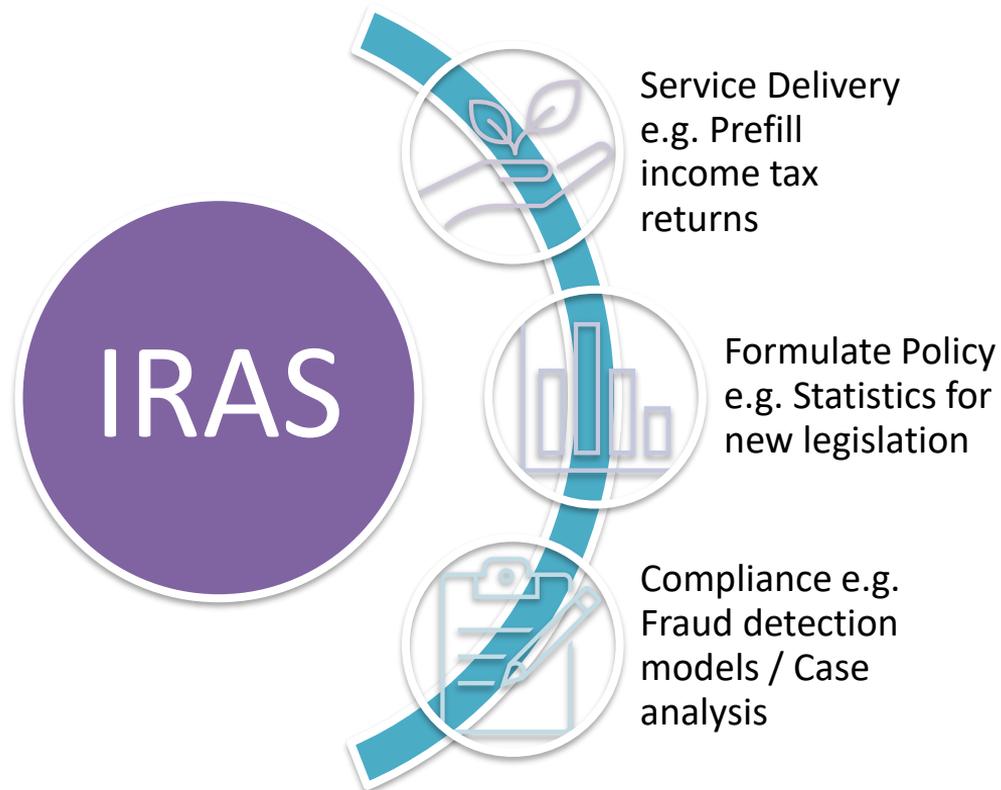
## 3. Bias and Discrimination

Big data algorithms may exhibit bias and discrimination based on factors such as race, gender, and socioeconomic status. Biased algorithms can perpetuate existing inequalities and undermine trust in automated decision-making systems.

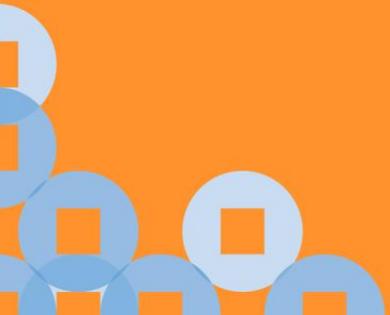
## 4. Data Quality Issues

Duplicate records, inaccurate details, and formatting errors are just a few of the many data quality issues and anomalies that can lead to incorrect conclusions.

# Use of Big Data in IRAS



# The different sources of information



# Sources of Data

## Internal

## External

Structured



Tax Returns e.g. Form B, GST F5



Application Forms e.g. GST registration application form (F1)



Property related – Tenancy / Ownership



Tourist Refund Scheme

Unstructured



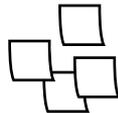
Audit / Investigation reports



Blacklist / Compliance Ratings

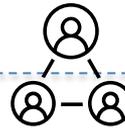


Correspondences with taxpayer



Intelligence

## Semi-Structured



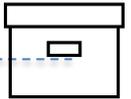
Other Government Agencies e.g. STRs<sup>1</sup>



Other countries e.g. EOI<sup>2</sup>



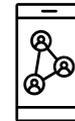
3<sup>rd</sup> Parties – Banks / Telco / Employer - Auto Inclusion Scheme<sup>3</sup> etc



Devices / docs obtained from subjects



Open sources e.g. Internet



Social Media e.g. Facebook / Instagram etc



Informants

<sup>1</sup> STRs = Suspicious Transaction Reports

<sup>2</sup> EOI = Exchange of Information arrangements with foreign jurisdictions

<sup>3</sup> Auto Inclusion Scheme ("AIS"): Under this scheme, employers submit the employment income information of their employees to IRAS electronically

# Sources of Data

Internal Structured Data	Internal Unstructured Data
Tends to be tax related data collected by IRAS from the various tax returns e.g. Income Tax returns - Form B / C, GST returns - GST Form 5 etc	Processed data after analysis performed on the data collected e.g. audit or investigation report / correspondences with the taxpayers
Most of IRAS' tax returns / applications are in electronic form where this facilitates data collection in structured format	Data are digitised and stored in IRAS Document Management System or shared folder
Data are stored in IRAS datawarehouses where it can be extracted for easy analysis and manipulation	

# Sources of Data

External Structured Data	External Unstructured Data
Data provided by other government agencies e.g. Suspicious Transaction Report Office (“STRO”) – STRs or Accounting and Corporate Regulatory Authority (“ACRA”) – Singapore’s equivalent of the Registrar of Business/ Companies	Subject’s Facebook posts generally are collected on adhoc and need-to basis. IRAS also collect data from 3 <sup>rd</sup> parties like banks for deposits / withdrawal data of our subjects during audit / investigation
Data from 3 <sup>rd</sup> parties like employers under IRAS’ Auto Inclusion Scheme – “AIS” e.g. Grab	Data would require considerable “cleaning” before use
Data are provided at a fixed intervals (e.g. annually for Central Provident Fund - “CPF” <sup>1</sup> data)	Users need to ascertain accuracy of data especially if sourced from the Internet due to the prevalence of fake news
	For investigation purpose, we have also employed OCR <sup>2</sup> technology to convert paper document seized from our subjects to digital format for ease of case analysis

<sup>1</sup> Central Provident Fund or CPF is a mandatory social security savings scheme funded by contributions from employers and employees

<sup>2</sup> Optical character recognition or optical character reader i.e. OCR is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene photo or from subtitle text superimposed on an image (Source : Wikipedia)

# Internal Sources of Information – for Tax & ML



Tax Returns e.g.  
Form B, GST F5

Tax returns submitted by taxpayers. Examples of tax returns submitted to IRAS:

- Income Tax Form (Individuals / Partnership / Company)
- GST Form (Form 5 / Form 7 / Form 8)
- Withholding Tax



Application Forms  
– GST F1

Application forms submitted by taxpayers to apply for tax status or IRAS' schemes / incentives

- Income Tax Schemes / Incentives (Job Support Scheme etc)
- GST Registration application (Form 1) or scheme (Major Exporter Scheme etc)
- Overseas entities information (with effect from 1 Jan 2020 - under the Overseas Vendor Registration Regime – “OVR”)<sup>NEW</sup>
- Withholding Tax



Property related –  
Tenancy / Ownership

Property related transactions

- Ownership of property (Sales / purchase price of property, name of buyer / seller etc)
- Tenancy related data (Amount of rental, period of tenancy, name of landlord / tenant etc)

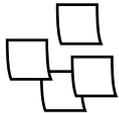
# Internal Sources of Information – for Tax & ML



Correspondences  
with taxpayer

Past correspondences / enquiries from the taxpayers

- Email
- Phone call (Telememo)
- Letter

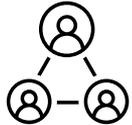


Intelligence

Information concerning the subject of investigation obtained through intelligence channel

- Source can be from internal (IRAS officers) / external parties (e.g. Informant)
- Relationship linkage of subject
- Location of subject
- Modus operandi

# External Sources of Data – for Tax & ML



Other Government  
Agencies

Other government agencies. Some examples:

- Business related data (E.g. Name of company director / shareholders / company registered address)
- Housing data (E.g. Public housing ownership information)
- Vehicle ownership data (E.g. Name of vehicle owner etc)
- Licensees data (E.g. Singapore Food Agency – Hawker licensee information)
- CPF data (E.g. Details of taxpayer's employment record)



Other countries  
e.g. EOI

Other countries:

- Exchange of Information through Avoidance of Double Taxation Agreements (DTAs) & Multilateral Convention on Mutual Administrative Assistance in Tax Matters
- Financial Intelligence Unit (“FIU”)

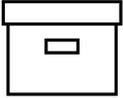


3<sup>rd</sup> Parties – Banks  
/ Telco etc

3<sup>rd</sup> parties. Some examples:

- Financial (E.g. Banks / Shares / Insurance)

# External Sources of Data – for Tax & ML



Devices / docs  
obtained from  
subjects

Obtained from our subjects:

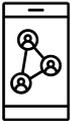
- Physical document e.g. invoices, Purchase Orders etc
- Digital devices e.g. Cellphone, laptop etc



Open sources  
e.g. Internet

Open sources like:

- Internet
- News reports
- Research papers
- Published judgment in courts



Social Media e.g.  
Facebook / Blogs

Social media:

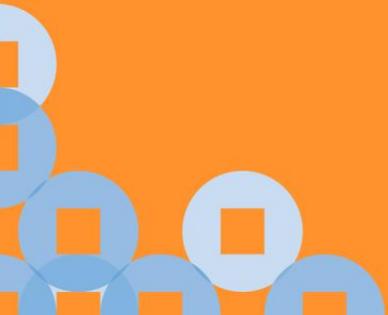
- Facebook
- X (former known as “Twitter”)
- Instagram



Informants

Whistleblowers on tax offence committed by subject

# Group Discussion

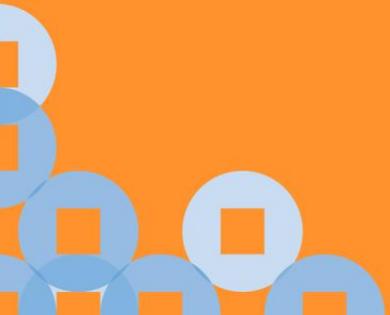


## Group Discussion (30 mins)

**Group 1 & 2 :** [Current] Any source of data that has been helpful to combat financial crime?

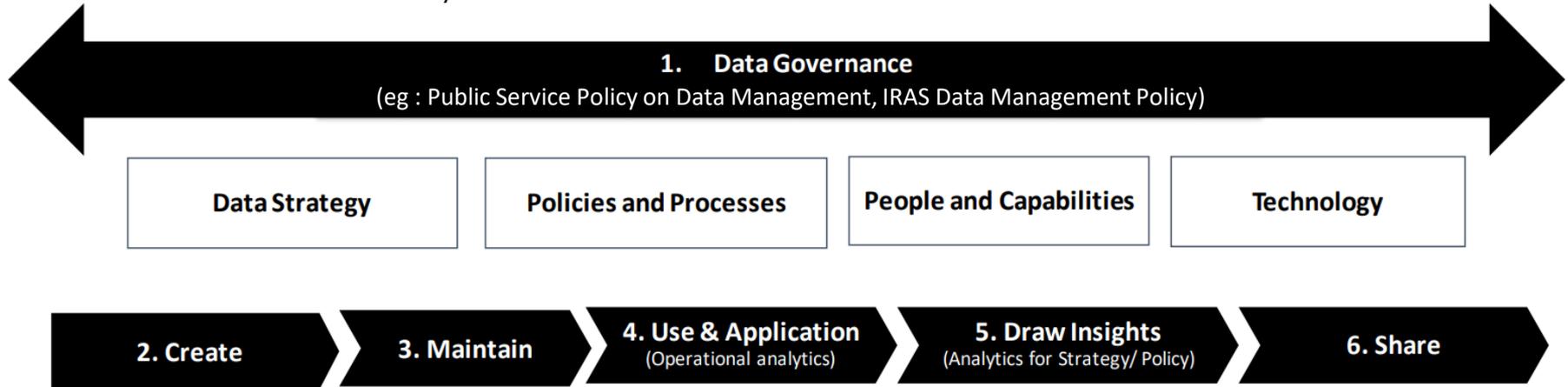
**Group 3 & 4 :** [Future] What new source of data would you like to have to enhance your capability to combat financial crime?

# Data governance and limitations of data



# Data Governance

- Overview of Data Lifecycle\*



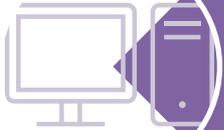
*Data Governance is the exercise of decision making and authority for data-related matters.*



*The purpose of data governance is to ensure that data is managed properly, according to policies and best practices. This enables IRAS and the Government to make the best use of data for decision-making. It allows IRAS to achieve its goal of cultivating data- and insights-driven culture.*

\* Source : Data Steward Reference Materials

# Limitation of Data



**Completeness** : Data completeness describes whether the data you've collected reasonably covers the full scope of the question you're trying to answer, and if there are any gaps, missing values, or biases introduced that will impact your results\*. Example, the data in the employment data table in IRAS system may not be complete as IRAS does not mandate the submission of employment data for businesses below certain staff strength.

*\*Source :*  
<https://www.montecarlo.com/blog/what-is-data-completeness/>



**Relevance** : Data relevance is the degree to which data provides insight into the real-world problem or purpose being addressed and contributes to the overall understanding of the business^. For example, IRAS conducted audit on subject from 2014 to 2015. Due to the long passage of time, the audit findings for the subject may not be relevant for the subsequent investigation in 2024.

*^Source :*  
[www.metaplane.dev/blog/data-relevance-definition-examples](http://www.metaplane.dev/blog/data-relevance-definition-examples)



**Timeliness** : The degree to which data represent reality from the required point in time#. In IRAS' example, certain information are provided at fixed interval e.g. CPF data is provided to IRAS annually. Hence, to get the most up-to-date info, officer would need to write to the relevant authority for the latest information.

*#Source :*  
<https://dsstream.com/introduction-to-data-quality-terms-definitions-examples-of-use/>

# Limitation of Data



Reliability and authenticity : Data reliability is the degree to which data, and the insights gleaned from it can be trusted and used for effective decision-making\*. For example, certain data provided / obtained by IRAS will need to be assessed for its reliability like financial data of Entity A shared by whistleblower. IRAS would not know if the data provided by the whistleblower is authentic.

*\*Source :  
[www.thoughtspot.com/data-trends/analytics/data-reliability](http://www.thoughtspot.com/data-trends/analytics/data-reliability)*



Accuracy : Data is considered accurate if it describes the real world<sup>^</sup>. One example will be for data obtained from the Internet; it would need to be validated against other sources to confirm its accuracy.

*^Source :  
[www.metaplane.dev/blog/data-accuracy-definition-examples](http://www.metaplane.dev/blog/data-accuracy-definition-examples)*

# Evaluation Matrix of Data / Information Provided to IRAS (e.g. Whistleblower)

**TABLE 1 - SOURCE RELIABILITY**

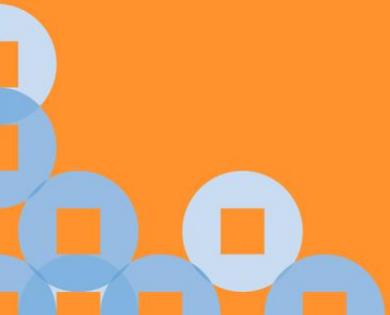
	RATING	DESCRIPTION
A	Reliable	No doubt about the source's authenticity, trustworthiness, or competency. History of complete reliability.
B	Usually reliable	Minor doubts. History of mostly valid information.
C	Fairly reliable	Doubts. Provided valid information in the past.
D	Not usually reliable	Significant doubts. Provided valid information in the past.
E	Unreliable	Lacks authenticity, trustworthiness, and competency. History of invalid information
F	Cannot be judged	Insufficient information to evaluate reliability. May or may not be reliable.

**TABLE 2 - INFORMATION RELIABILITY**

	RATING	DESCRIPTION
1	Confirmed	Logical, consistent with other relevant information, confirmed by independent sources.
2	Probably true	Logical, consistent with other relevant information, not confirmed.
3	Possibly true	Reasonably logical, agrees with some relevant information, not confirmed.
4	Doubtfully true	Not logical but possible, no other information on the subject, not confirmed.
5	Improbable	Not logical, contradicted by other relevant information.
6	Cannot be judged	The validity of the information can not be determined.

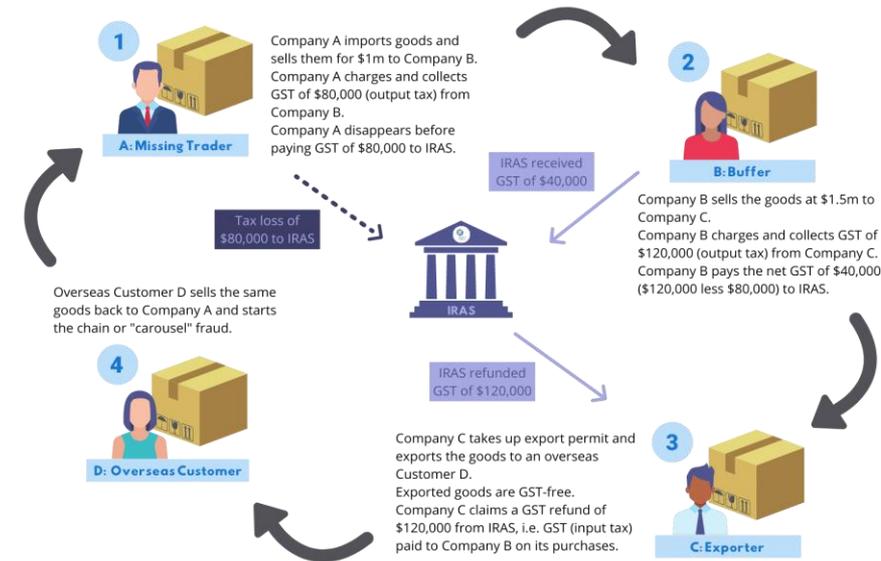
*Question : Can you share how your agency deals with whistleblowers?*

# Case Studies



# What is Missing Trader Fraud?

MTF\* occurs when organised criminal groups abuse the GST/VAT refund system for fraudulent export arrangements and exploit the asymmetry of information in different jurisdictions to avoid being identified/traced.



## What is a Missing Trader Fraud (MTF):

*“Under a typical MTF arrangement, a group of businesses would form a supply chain and the same goods/services would be supplied through the chain. To ensure that the final sales of the goods/services are not subjected to GST, the goods/services would ultimately be exported to an overseas customer. A seller upstream in the supply chain would charge GST on the sale of goods to businesses downstream and instead of paying the GST to IRAS, the upstream supplier would **fail to account** in its GST return the GST it had collected.*”

*This is termed **“Missing Trader” fraud** as the seller disappears with the GST.”*

\* Missing Trader Fraud or MTF is also commonly known as Carousel Fraud in the other parts of the world like the European Union (“EU”)

# What is Missing Trader Fraud?

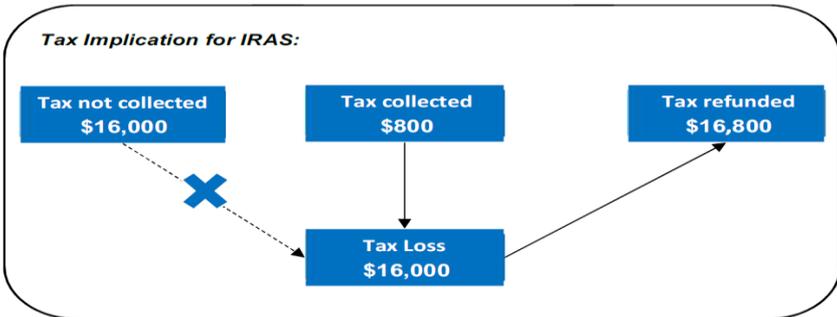
## Illustration of a MTF arrangement

Supply chain:



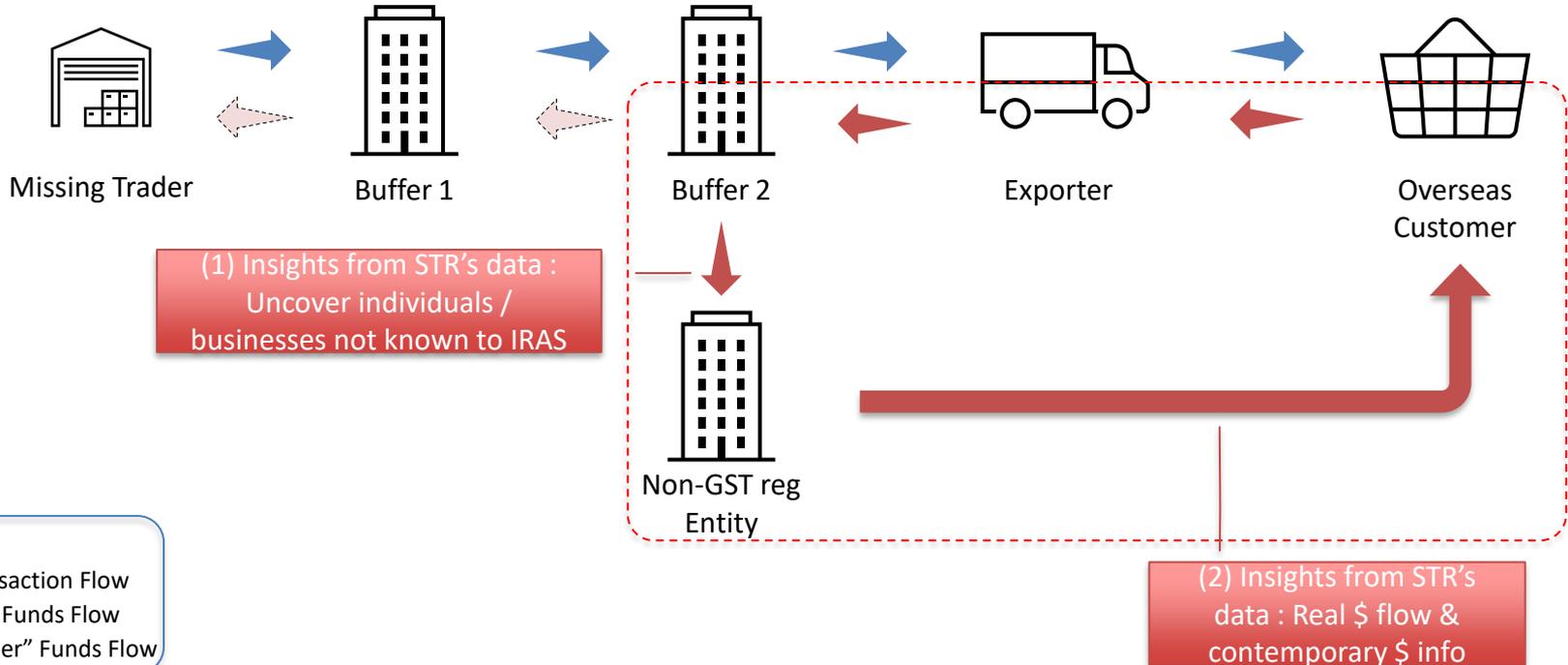
Question : Can share your agency's experience with MTF?

Tax Implication for IRAS:

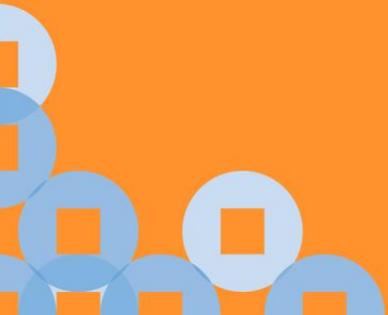


# Operation Wand 2 – Missing Trader Fraud

Example : Use of Suspicious Transaction Report (“STR”) in Operation Wand 2 to combat the tax and money laundering offences



# Group Discussion

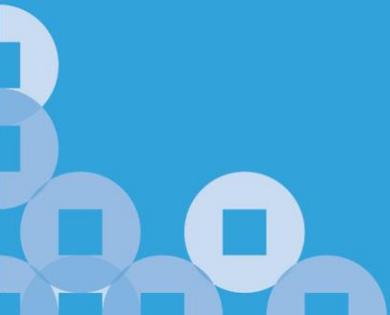


## Country Sharing (20 mins)

All Groups : You are a tax agent from FRAUD CONSULTANT LTD and your client hired you to design a fraud scheme to obtain refund / payout from the government? Pls explain the fraud scheme, possible data that could give your scheme away and what to do to avoid detection / investigation. (20mins : Design the fraud scheme)

The group not presenting : You are the investigator assigned for the above case and explain what the data you need to detect the scheme (Detect the scheme)

# (4) Analytics Tool for Investigators



# What is Data Analytics\*?

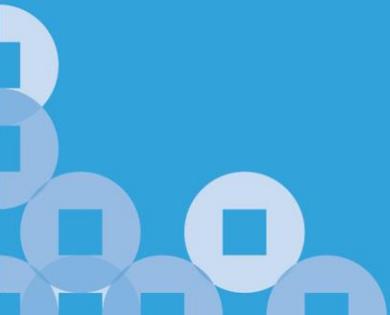
*“Data analytics involves the process of examining **large and varied data sets** to **uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information** that can help organizations make **more informed decisions**. It involves several steps, including data collection, cleansing, analysis, and interpretation. Data analytics utilizes various techniques, such as statistical analysis, machine learning, data mining, and predictive modeling, to extract insights from data. These insights can be used to optimize processes, improve performance, identify opportunities, mitigate risks, and gain a competitive advantage. Overall, data analytics plays a crucial role in modern business operations, scientific research, healthcare, finance, marketing, and many other fields.”*

\*Source : [Big Data Analytics Overview \(openai.com\)](https://openai.com)

## What is Data Analytics (3mins video)?

- <https://www.youtube.com/watch?v=nJa20uMtR7I>

# Group Discussion

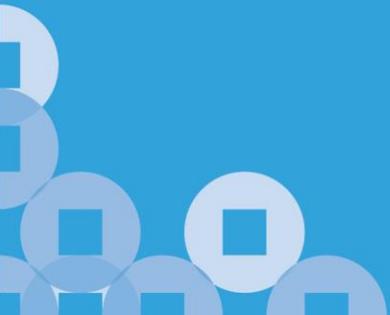


## Group Discussion (40mins)

Discuss within your group :

- (1) Your country's AI strategy;
- (2) Your experience with AI;
- (3) Frauds / crimes that you have encountered involving AI;
- (4) Use of AI to combat the fraud

# Overview of Singapore's AI Strategy



# Singapore National AI Strategy\*

2019

We unveiled our first National AI Strategy, outlining plans to deepen the use of AI to transform our economy.

2023

Launched the **Singapore National AI Strategy 2.0 (NAIS 2.0)** in our ongoing pursuit of setting the pace

Video : <https://www.youtube.com/watch?v=6qHBTi3YQIQ>

# Singapore National AI Strategy 2.0\*



## From Opportunity to Necessity

The expanded landscape of opportunities and risks brought forth by AI today make clear that AI is indispensable for the continued prosperity and relevance as a nation. We must master AI, rather than think of it merely as an accessory or a luxury.



## From Local to Global

Singapore must operate, from the outset, with a global outlook for AI. We must connect to global networks, working with the best and pooling resources among the like-minded to overcome the complex challenges that plague AI today.



## From Projects to Systems

We must further enhance our capability, infrastructure and resource foundations, and accelerate ideas exchange at scale, to amplify and manage the impact of AI across various sectors, and in economy and society at large.

\*<https://www.smartnation.gov.sg/nais/>

# Goals of Singapore National AI Strategy 2.0\*



Excellence

Singapore will selectively develop peaks of excellence in AI, to advance the field and maximise value creation. We intend to direct AI towards addressing the needs and challenges of our time, such as in areas of global importance like population health and climate change.



Empowerment

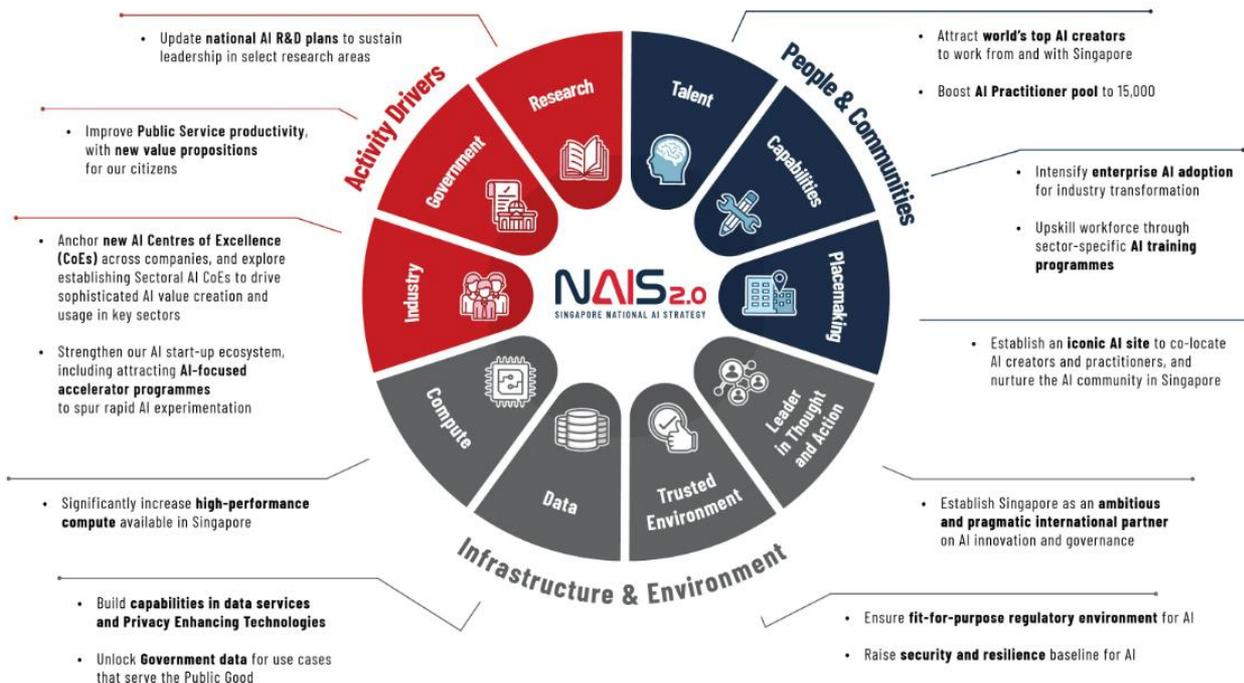
Singapore will raise up individuals, businesses, and communities to use AI with confidence, discernment, and trust. We intend for AI to be the great equaliser, which equips our people and businesses with the capabilities and resources to thrive in an AI-enabled future.

\*<https://www.smartnation.gov.sg/nais/>

# 10 Enablers to achieve our vision and goals\*

## Harnessing AI for the public good

Our second National AI Strategy, or NAIS 2.0, outlines our vision for Singapore to be a place where AI serves as a force for good, and where we harness AI to uplift and empower our people and businesses. To achieve our vision and goals, we will direct efforts under NAIS 2.0 toward 3 Systems, working through 10 Enablers.



\*<https://www.smartnation.gov.sg/nais/>

# WOG DSAI Tools\*



GovTech is the lead agency driving Singapore's Smart Nation initiative and public sector digital transformation. We harness the power of technology to make lives better for citizens, businesses, international audiences and the public service sector.

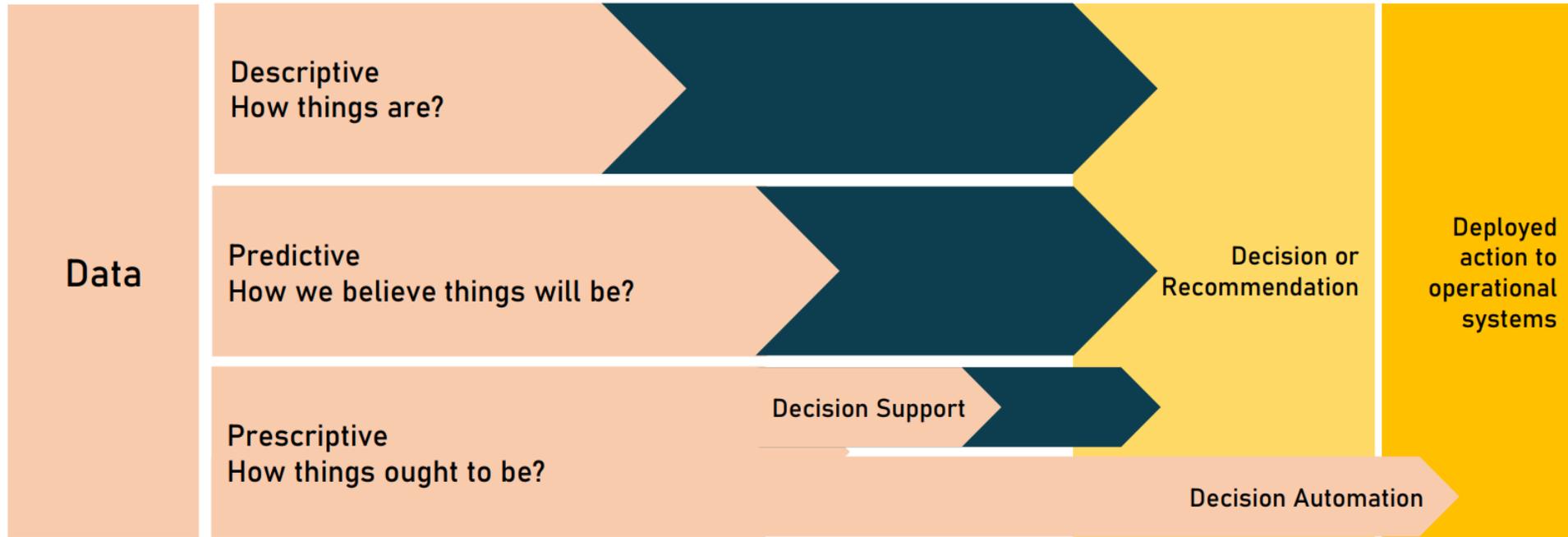
## WOG DSAI Products

WOG AI & Data Product Catalogue curated by DDED:



\*[https://inex3.iras.gov.sg/sites/AnalyticsConnect/DSAI\\_Project\\_Submission/\\_layouts/15/start.aspx#/SitePages/Products.aspx](https://inex3.iras.gov.sg/sites/AnalyticsConnect/DSAI_Project_Submission/_layouts/15/start.aspx#/SitePages/Products.aspx)

# What are Data Analytics\*?

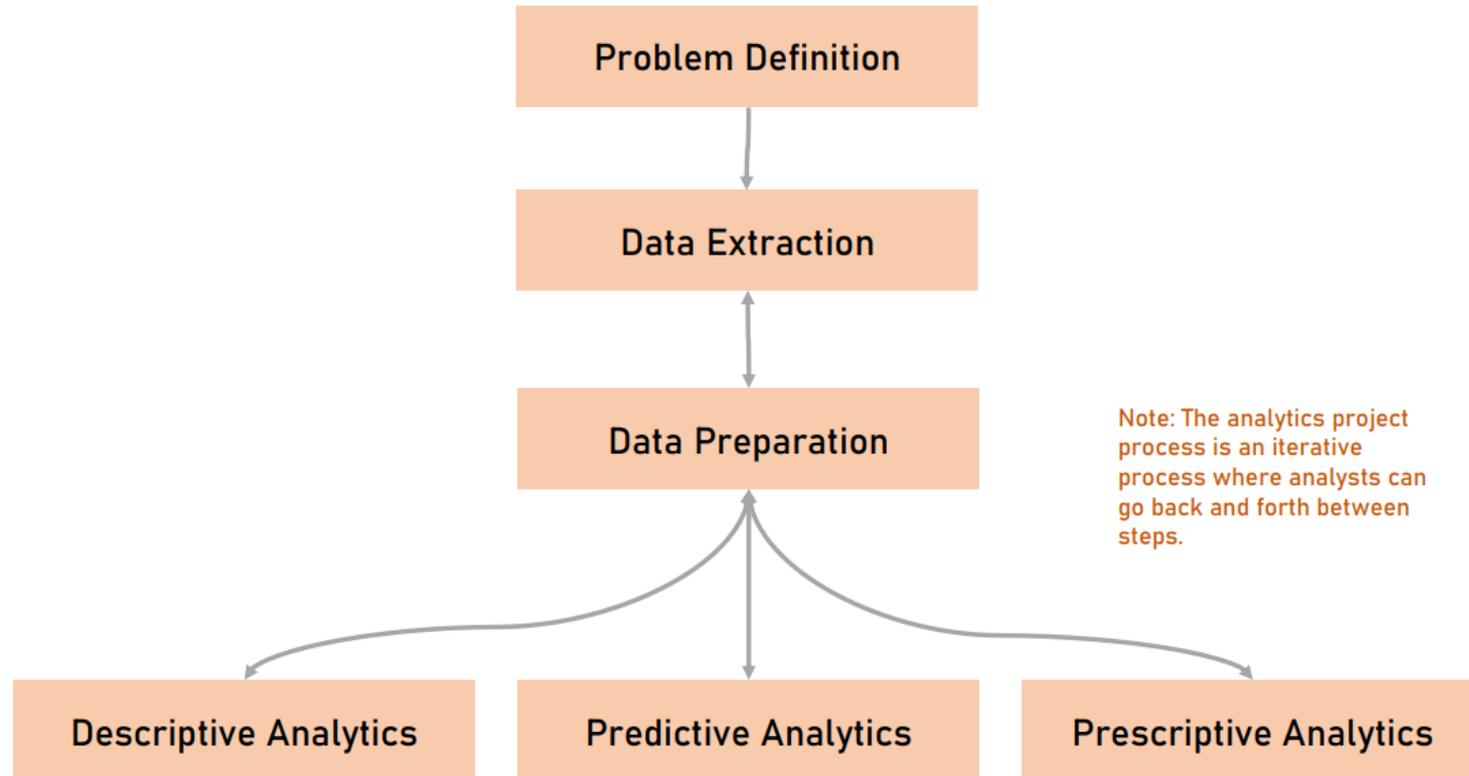


Analytics

Human intervention

\*Source : IRAS' Analytics Toolkit

# Data Analytics Project Process\*



# Data Analytics Project Process\*

## Problem Definition

Problem definition starts with identifying a problem, usually in the form of a symptom. Specifically, there is a need to understand:

1. What is the problem?
2. Who does it affect?
3. What is the impact of the problem, and what happens if left alone?
4. What does a successful solution look like?



## Data Extraction

1. Identify the data required for the project. This involves asking:

- What data do we think we need?
- How will this be useful to achieving our objectives?
- What assumptions?
- How feasible is getting this data?

2. Determine the source of the data, which could be from internal or external sources
3. Evaluate reliability of data source
4. Extract the data from its source(s)
5. Combine the data from the various sources



## Data Preparation

1. Assess quality of data  
Data quality is affected by the way data is collected, entered in the system, stored and managed.
2. Clean the data  
Correct incomplete, incorrect, inaccurate or irrelevant records from a dataset by removing irrelevant data / removing duplicates etc
3. Transform the data
4. Perform feature engineering and selection
5. Handle imbalanced datasets



## Descriptive Analytics

To describe and understand what has happened using visualizations, statistics and unsupervised machine learning. This is also known as data exploratory analysis.

## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Prescriptive Analytics

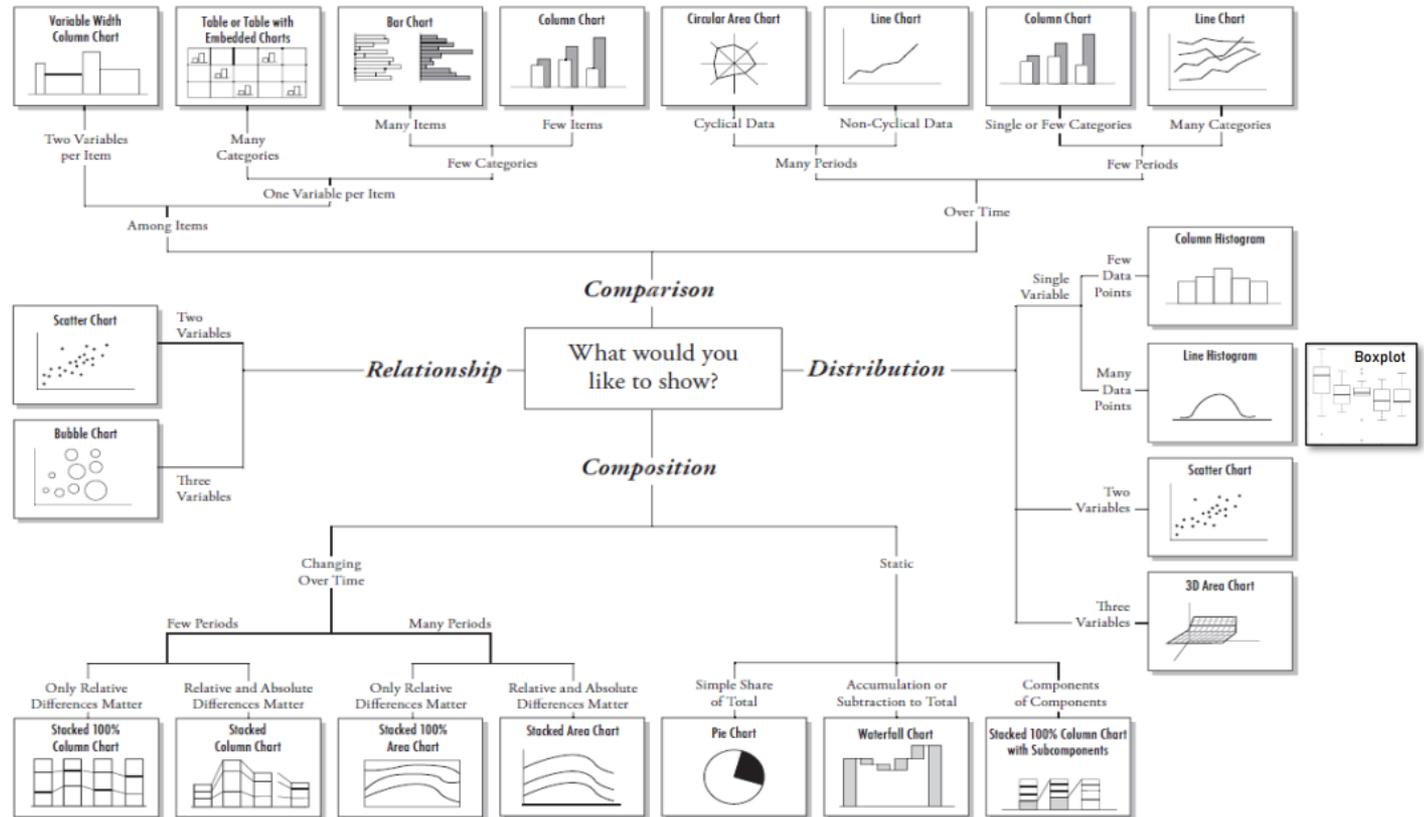
The goal is to automate future decisions which are defined programmatically through an analytic process. Thus, this future decision selected should be the one that gives an optimal outcome e.g. email recommender

# Data Analytics Project Process\*

## Descriptive Analytics

To describe and understand what has happened using visualizations, statistics and unsupervised machine learning. This is also known as data exploratory analysis.

Depending on what you would like to show, you may use the following charts:



# Data Analytics Project Process\*

## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Classification Models

S/N.	Model	Description	Examples
1	Logistic Regression	Logistic Regression predicts the probability of occurrence of an event by fitting data to a logit function	BLP for CTD, PIC cash payout projects
2	Naïve Bayes	Naïve Bayes are a family of powerful and easy-to-train classifiers that determine the probability of an outcome given a set of conditions using Bayes' theorem. It is useful for very large datasets as it is highly scalable.	-
3	Nearest Neighbour	The k-nearest-neighbors algorithm uses proximity as a proxy for 'sameness'. To label a new point, it looks at the labelled points closest to that new point (those are its nearest neighbors).	-
4	Decision Tree	Decision tree builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The result is a tree with decision nodes and leaf nodes. It produces rules in simple English sentences, which can easily be interpreted and presented to senior management.	Potential Registrants Review Programme (PRRP)
5	Random Forest (Bagged/ Bootstrap trees)	Random forests are an ensemble learning method that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. The trees are trained using bagging where each model is trained by a random subset of the data. Random decision forests correct for decision trees' habit of over fitting to their training set.	Reoffender Study - Identifying Variable Importance for Filing Reoffence
6	Adaboost tree	Adaboost tree also uses an ensemble learning method but instead of bagging, it is trained using boosting where models are trained in a sequential way and each individual model learns from mistakes made by the previous model.	-
7	Linear Support Vector Machine (SVM)	SVM is based on the idea of finding a hyperplane that best divides a dataset into two classes	Study on Phoenix companies using SNA - Prioritize Companies Within Group
8	Neural Network	A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output.	-

\*Source : IRAS' Analytics Toolkit

# Data Analytics Project Process\*

## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Classification Models

## Confusion Matrix

Confusion Matrix is a performance measurement for the machine learning classification problems where the output can be two or more classes. It is a table with combinations of predicted and actual values.

*A confusion matrix is defined as the table that is often used to describe the performance of a classification model on a set of the test data for which the true values are known.*

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

\*Source :

<https://www.analyticsvidhya.com/blog/2017/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>

# Data Analytics Project Process\*

## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Classification Models

S/N.	Metric/ Chart	Description
1	Accuracy	The proportion of the total number of predictions that were correct.
2	Precision	The proportion of positive cases that were correctly identified.
3	Negative Predictive Value	The proportion of negative cases that were correctly identified.
4	Recall/Sensitivity/True positive rate	The proportion of actual positive cases which are correctly identified. This metric is useful where we are more concerned with minimizing false negatives.
5	Specificity	The proportion of actual negative cases that were correctly identified. This metric is useful where we are more concerned with minimizing false positives.
6	F1 score	The harmonic mean of precision and recall values. A model does well in F1 scores if the positive predicted are actually positives (precision) and it doesn't miss out on positives and predicts them negative (recall).
7	Gain	The cumulative gain chart tells us the percentage of true positive captured at each decile.
8	Lift	Lift is a measure of the effectiveness of a predictive model calculated. It is the ratio between the percentage of true positives captured with and without the predictive model at each decile. The baseline lift result is 1 using no model.
9	Kolomogororov Smirnov Chart	K-S is a measure of the degree of separation between the positive and negative distributions. The K-S is 100, if the scores partition the population into two separate groups in which one group contains all the positives and the other all the negatives. On the other hand, if the model cannot differentiate between positives and negatives, then it is as if the model selects cases randomly from the population. The K-S would be 0. The higher the value the better the model is at separating the positive from negative cases.
10	AUC - ROC	The Receiver operating characteristics curve (ROC) is used to plot between true positive rate (TPR) and false positive prate (FPR) for various threshold values, also known as the sensitivity and 1-specificity graph. The area under curve (AUC) is utilized for setting the threshold of cut-off probability to classify the predicted probability into various classes. An excellent model has AUC near to the 1 which means it has good measure of separability. A poor model has AUC near to the 0 which means it has worst measure of separability. And when AUC is 0.5, it means model has no class separation capacity whatsoever.

\*Source : IRAS' Analytics Toolkit

# Data Analytics Project Process\*

## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Regression Models

S/N.	Model	Description
1	Linear Regression	<p>The relationship between the features and target is assumed to be linear in nature. Assumptions of linear regression include:</p> <ul style="list-style-type: none"><li>• There must be a linear relation between independent and dependent variables.</li><li>• There should not be any outliers present.</li><li>• No heteroscedasticity</li><li>• Sample observations should be independent.</li><li>• Error terms should be normally distributed with mean 0 and constant variance.</li><li>• Absence of multicollinearity and auto-correlation.</li></ul>
2	Non-linear / Polynomial Regression	<p>The relationship between the features and target cannot be assumed to be linear in nature. It fits a nonlinear equation by taking polynomial functions of independent variable.</p>
3	Lasso/ Ridge Regression	<p>Lasso/ Ridge Regression helps solve the problem of overfitting in linear regression. It adds a penalty term to the objective function using regularization. It is useful where there is:</p> <ul style="list-style-type: none"><li>• Large number of features</li><li>• Low ratio of observations to number of features</li><li>• High multi-collinearity</li></ul>
4	Elastic Net Regression	<p>Elastic Net Regression is preferred over both ridge and lasso regression when one is dealing with highly correlated independent variables.</p>
5	Neural Network	<p>A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output.</p>

# Data Analytics Project Process\*

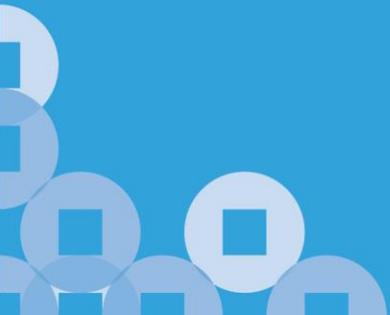
## Predictive Analytics

To predict a future business phenomenon using supervised machine learning techniques. For example, to anticipate which taxpayers are more likely to default.

## Regression Models

S/N.	Metric/ Chart	Description
1	Root Mean Squared Error	RMSE is the most popular evaluation metric used in regression problems. It follows an assumption that error is unbiased and follow a normal distribution. It can range from between 0 and infinity. The lower the value, the better the model.
2	Mean Absolute Error	MAE measures the average of all absolute errors. It answers the question, "How far were you off in your predictions, on average?"
3	R-Squared	R-Squares is the proportion of variance explained by the model. The higher the R-squared, the better the model.
4	Adjusted R-Squared	Adjusted R-squared accounts for the addition of more predictor variables. Adjusted R-squared will only increase with a new predictor variable when that variable improves the model performance more than would be expect by chance. The higher the Adjusted R-squared, the better the model.
5	Mean Squared Error	MSE measures the average of the squares of errors.
6	Range of prediction	Range is the maximum and minimum value in the predicted values. It helps us to understand the dispersion of predicted values between models

# Leveraging on technology & analytics in financial investigation



# Analytics as a gatekeeper

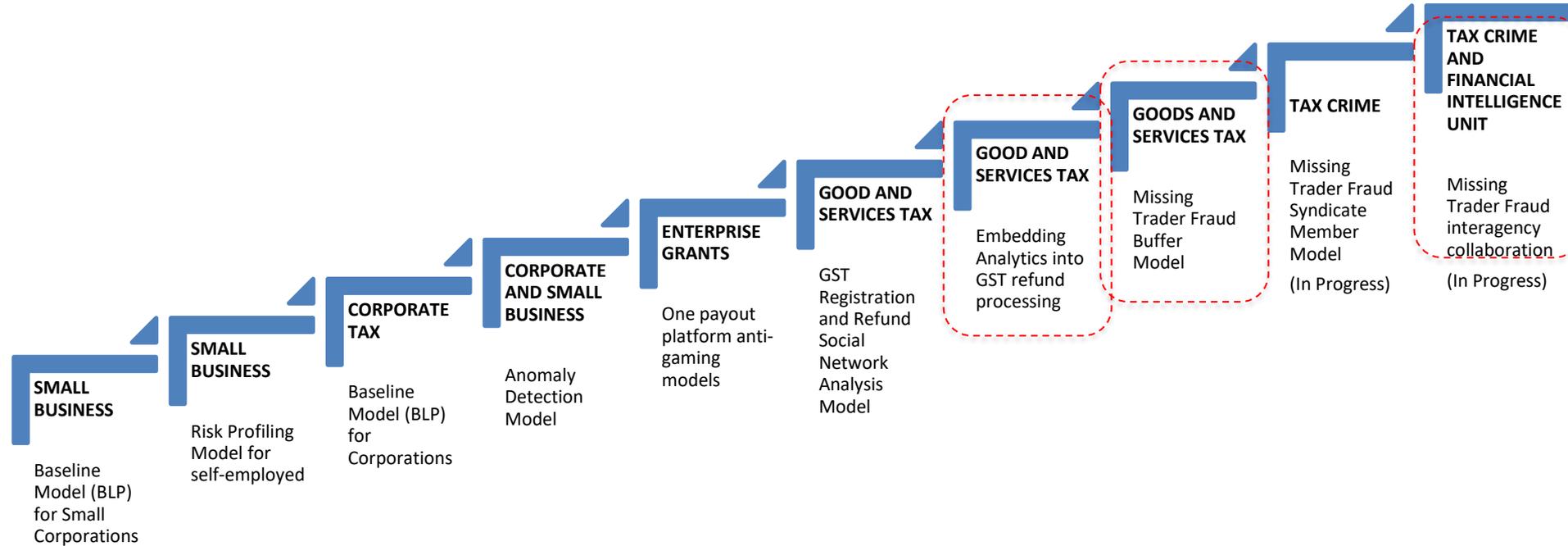


*Over the years, IRAS has deepened our analytics capability where various analytics tools were developed to detect crimes proactively*

# Analytics as a gatekeeper - Overview



## Tax Fraud / Evasion

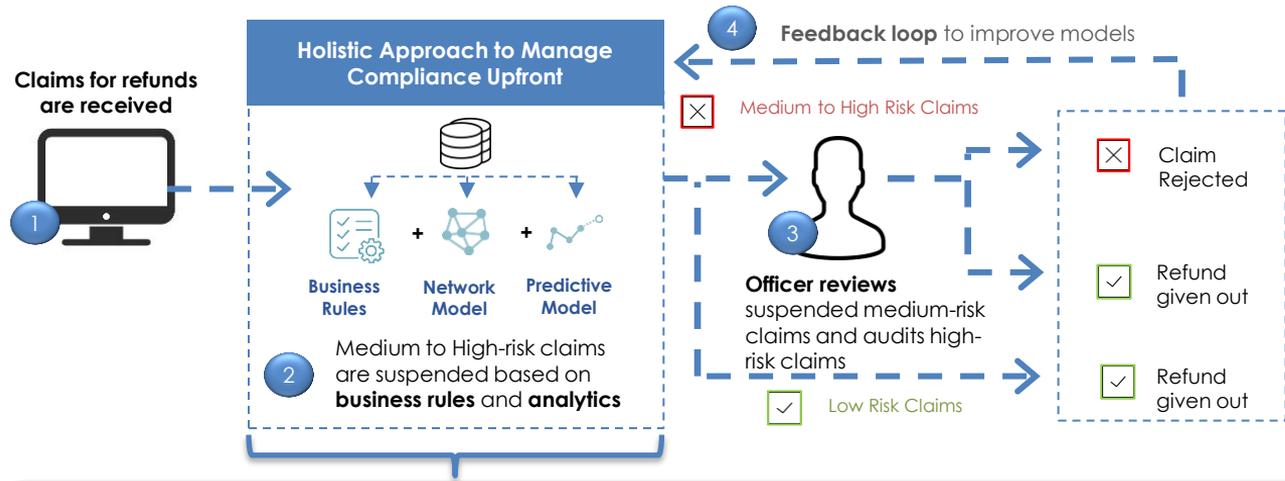


## Tax Non-Compliance

# Embedding Analytics into GST Refund Process

Embedding of **2 analytics models** within the GST Refund Process to identify risky refunds:

- Pre-Refund Analytics (PRA) Model
- Social Network Analysis (SNA) Model



Returns are scored by a mixture of on-the-fly rules that score instantaneously and pre-computed scores

### On-the fly scoring

- SNA transaction rules and PRA model score returns as they are received
- Historical data needed for scoring is aggregated beforehand

Ensures scoring accuracy



### Pre-computed scores

- For SNA entity/network scores that are largely static
- Scores are refreshed in fixed intervals within SNA system

Minimises computing effort



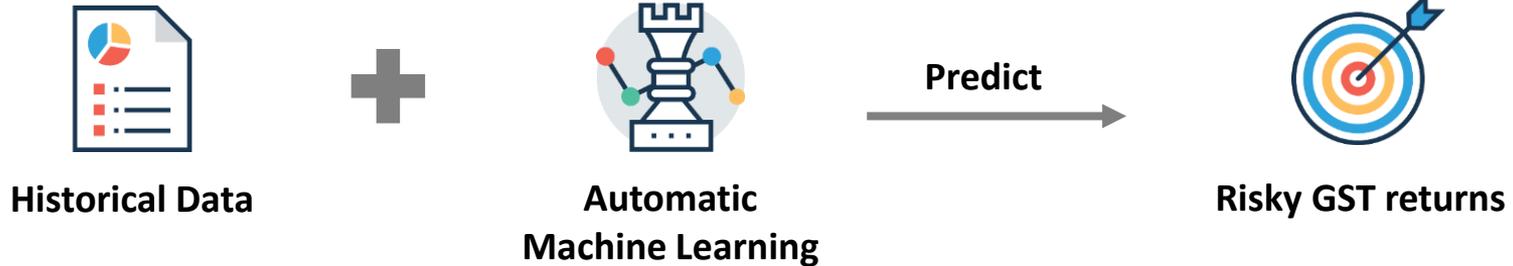
$\Sigma$  Analytics Scores

Final Analytics scores

# Missing Trader Fraud (“MTF”) Buffer Model

The MTF Buffer Model augments IRAS’ analytical capabilities with the use of an Auto Machine Learning solution to aid our fight against Missing Trader Fraud

- Enable IRAS to pre-emptively detect buffer entities engaging in MTF activity and **disrupt the supply chain early**
- **AutoML model** trained on known buffers and non-buffers
- Scoring is performed on GST returns to generate a score of **the likelihood that the return displays buffer-like attributes.**

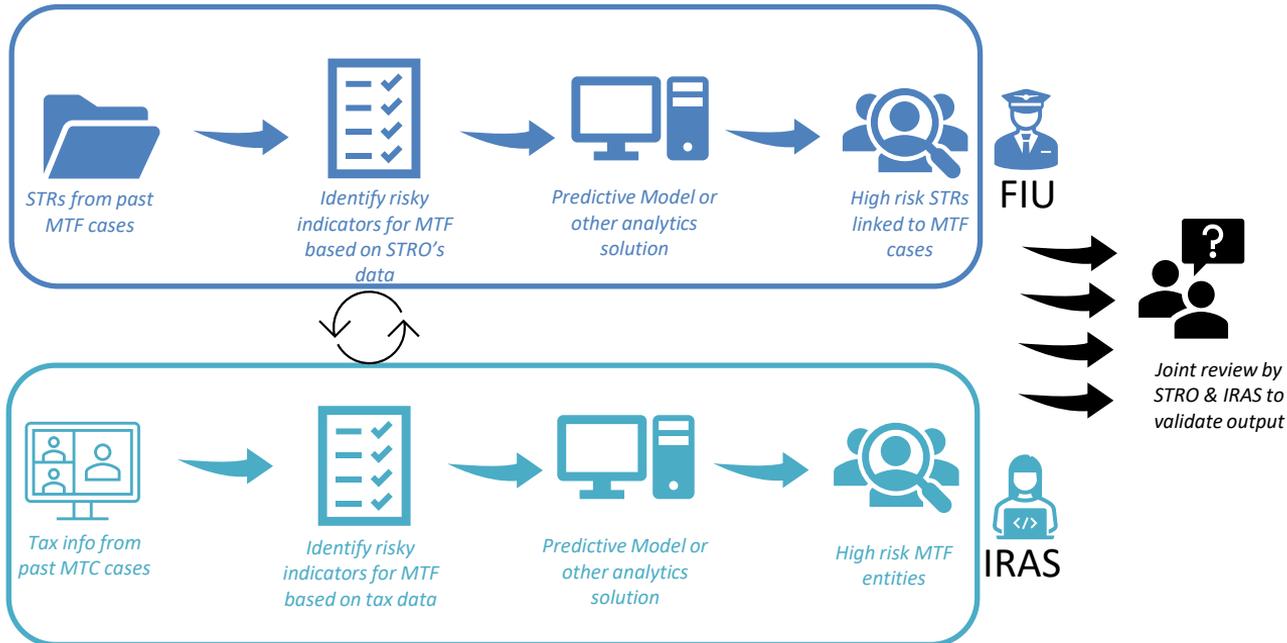


## Results

Model was **3 times more effective** in detecting MTF buffer entities compared to traditional approaches, allowing IRAS to detect **>\$30m of potential tax fraud.**

# IRAS / FIU Interagency Analytics Collaboration

The IRAS-FIU interagency analytics collaboration adopts a federated data analytics approach to better identify Missing Trader Fraud (“MTF”) in IRAS and STRO’s respective systems



## Progress to date

- Collaboration with our Financial Intelligence Unit (“FIU”) from the Singapore Police Force to develop analytics models using past STRs filed / tax info from MTF entities to identify unknown MTF cases / clusters.
- Currently, project is at the model deployment stage where both agencies will analyse the cases generated to evaluate the model effectiveness.

# Analytics for operational analysis



*IRAS employs analytics to amplify intelligence for better sense-making and targeting*

# Enterprise Network Visualising Tool - Intelligent Network Analysis Tool (“iNAT”)

iNAT helps officers easily connect suspicious entities, even if they are linked by a complicated network of entity ownership, business transactions, and personal relations over multiple hops.

## Intelligent Network Analysis Tool

A holistic view of a taxpayer's relationships

For IRAS auditors & investigators to conduct effective risk identification and investigations

17

Entities & Relationship Types

25,000,000

Entities

70,000,000

Relationships

### Investigate Missing Trader Fraud

Missing Trader Fraud involves a group of traders performing organised activities to claim fraudulent tax refunds from IRAS. Using the iNAT, investigators can plot and find related parties of the suspected trader(s).

### Evaluate related party transactions

Taxpayers are to apply the arm's length principle to ensure that the pricing of their transactions with their related parties reflects independent pricing. Using the iNAT, auditors can find the relationship between two parties.

#### Real-time Graph Analytics



Map networks

Find linkages between entities

Identify influential entities

Within seconds

#### Customizable



Upload self-collected data

Add/remove nodes/links

Colour & design your network

At your own convenience

#### Intelligent Risk Scoring



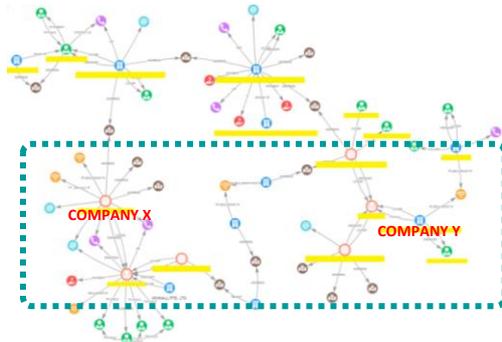
Coming soon

# Enterprise Network Visualising Tool - Intelligent Network Analysis Tool (“iNAT”)

iNAT helps officers easily connect suspicious entities, even if they are linked by a complicated network of entity ownership, business transactions, and personal relations over multiple hops.

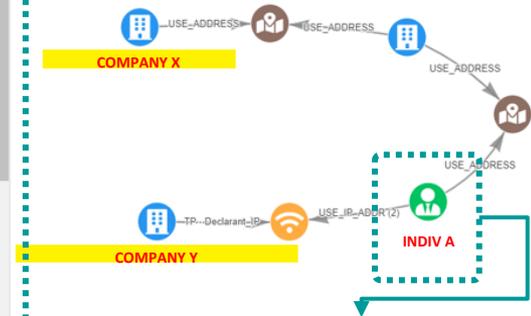
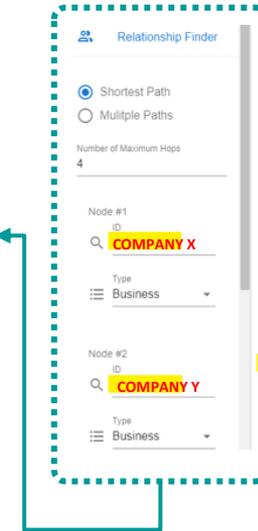
## Example of iNAT applied to Missing Trader Fraud (“MTF”) investigations

- 1 X and Y are missing traders in a Missing Trader Fraud scheme that involved **\$114 Million in fictitious sales** and over **70 entities**.



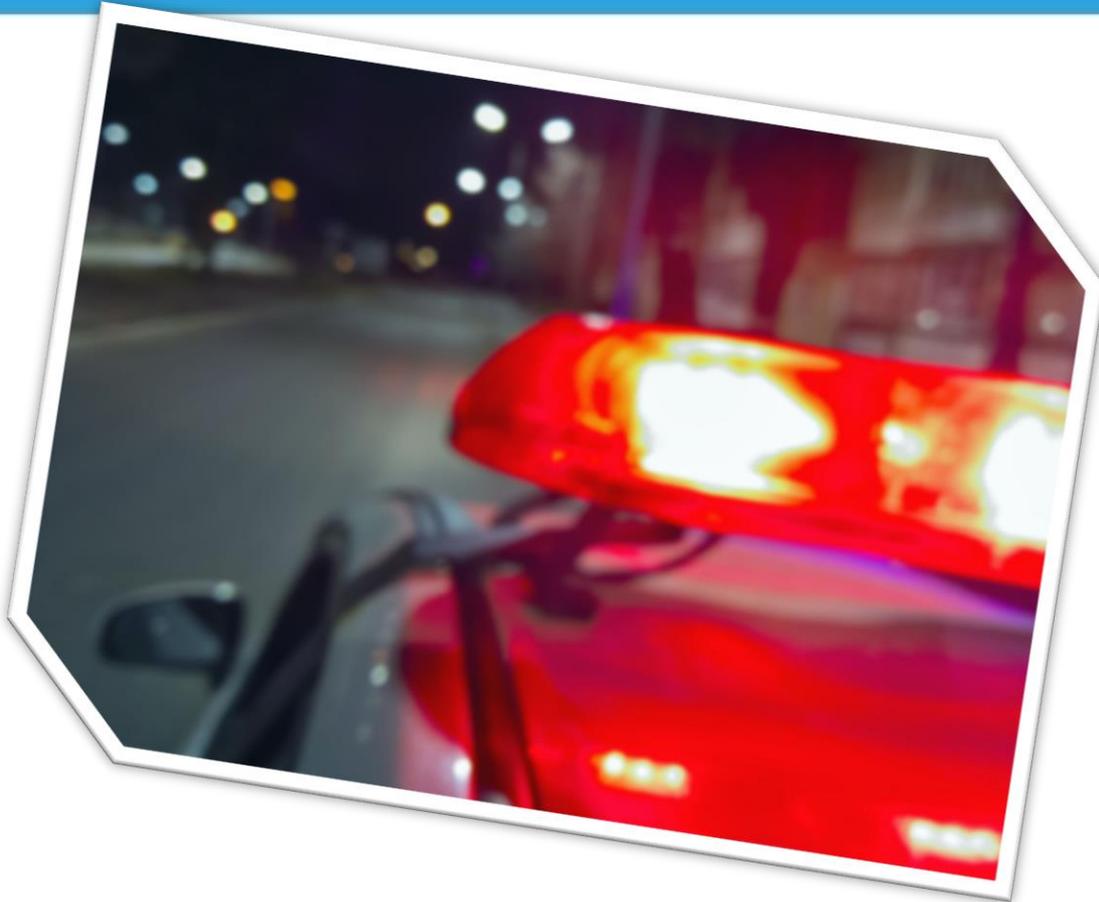
Part of MTF network involving X and Y

- 2 With iNAT’s Relationship Finder feature, IOs can access rich **relational databases** and visualize the links between X and Y



- 3 A is quickly uncovered as a suspicious individual who shares location and IP address with X and Y. IOs can **expand targeted investigations on A**.

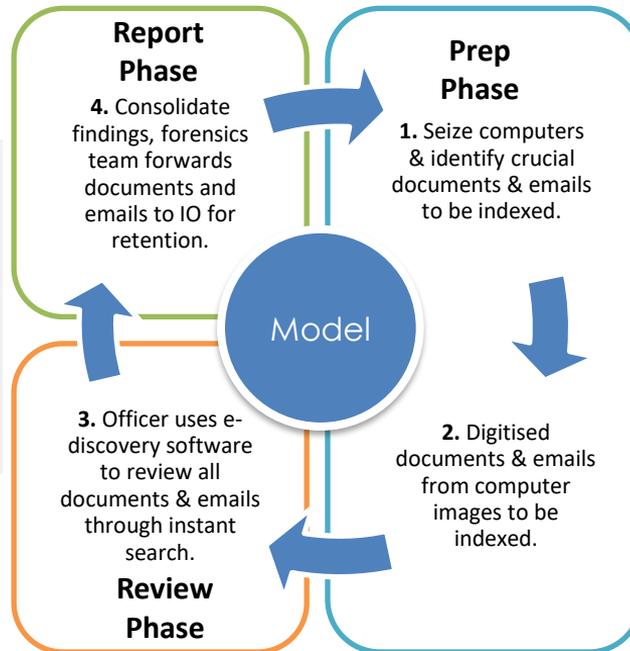
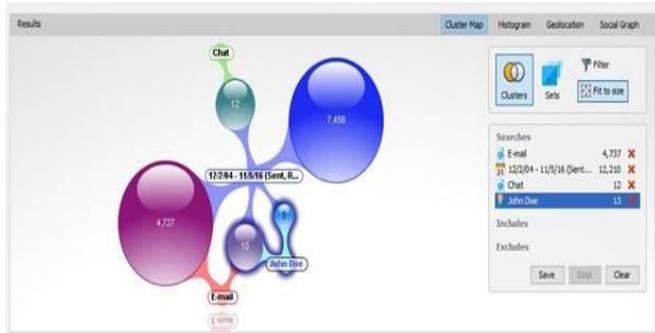
# Analytics for investigative efficiency



*Democratising  
analytics to all  
officers for greater  
speed and  
effectiveness of  
casework during  
investigation*

# e-Discovery : Self-review Workstation

Investigation and eDiscovery software tool for single users who need to process, search, filter and produce amounts of digital evidence.

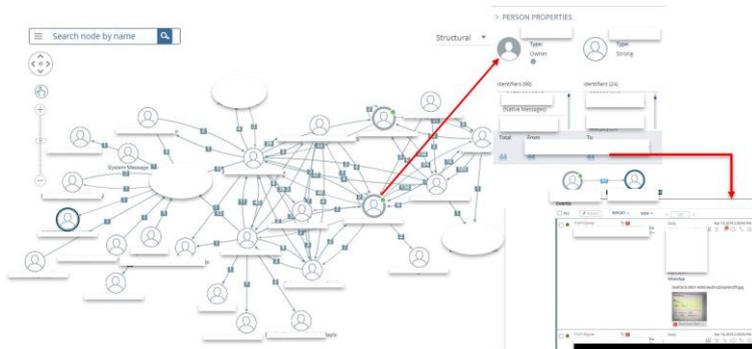


## Benefits

Users can quickly and easily process, search, review, and analyze digital evidence obtained. The easy-to-use interface and Cluster Map allows analysts to quickly find critical data, visualise relevant relationships, and drill down to the most pertinent data.

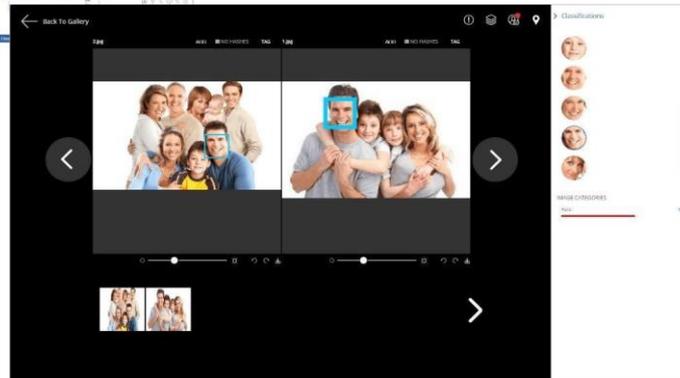
# Mobile Evidence Analytics

Allows multiple mobile devices to be reviewed simultaneously and provides a contextual view of the relationships between devices



## Relationship/Graph View

## Face/Image Similarity Recognition



## Benefits

- Enable multiple users to access at once to look at different facets of a syndicated case
- Other capabilities like facial recognition also allowed users to identify new subjects / relationships

# Automated Data Extraction project

Improve productivity of users when reviewing case by automating some of the data extraction process from IRAS' various databases and charts / tables are provided for ease of analysis



Close to **300** data fields were extracted from various databases

*(> 150 data fields for business & > 100 data fields for individual)*



> **10 Key financial ratios** (e.g. Industry Gross Profit Margin) & network analysis are incorporated to supplement background check



Up to **30%** of First Level Report (“FLR” – report submitted by Intel Analyst) is pre-filled with requisite information from extracted data



> **10 charts and tables** are created using POWER BI to facilitate user's review and analysis

# Automated Data Extraction project

## Key Benefits



Data from IRAS database is extracted automatically by running a code



Report template is pre-filled with information extracted via DW



The report template is linked to POWER BI for data visualization

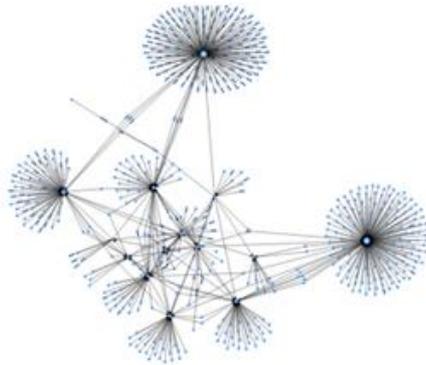


Productivity gained from manual to automated data extraction process (1-3 mandays to 1-2 hours)

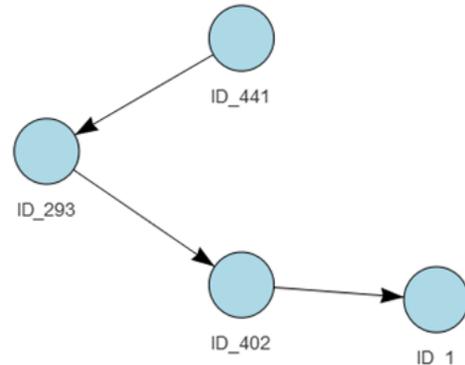
# Transaction Tracing Tool (“TTT”)

Facilitate mapping of the supply chains for syndicated cases by providing officers with the entire flow of transactions based on a selected transaction

- Some Missing Trader Fraud (“MTF”) clusters involved **large number of entities** (up to 50 entities per cluster) with **high transactional volume** (may be more than 10,000 transactions per cluster).
- After ingesting the transactional data, TTT will provide the supply chain by mapping out the predicted transaction flow for the users



Network Visualisation of transactions



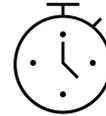
Predicted Transaction Flow in Supply Chain

# Transaction Tracing Tool (“TTT”)

## Key Benefits



Productivity gained by analysing large amount of transactions effectively – can quickly identify the fraudulent supply chains



- Saved effort of reading and matching invoice no & date, long text descriptions, quantity of goods
- Saved time from manually drawing supply chains



Officers can complete the tracing within 6.5 to 9.5 manhours by using TTT compared to 9.5 to 12.5 manhours when tracing transactions manually

# Analytics to facilitate funds tracing



*Application of  
technology for  
funds tracing  
(money laundering)*

# Automated Bank Deposit Analysis Tool (NEW – Oct 2024)

- With increased use of digital payment over cash<sup>^</sup>, we also see spike in the volume of banking transactions. The Automated Bank Deposit Analysis Tool was developed to analyse voluminous banking info accurately and quickly.
  - At intel level – to determine if there is prima facie case of tax evasion;
  - At investigation level - in the absence of accounting records, to compute the actual income and turnover of an entity;
  - For money laundering – to trace the flow of the proceeds of crime

## Step 1 - Collect

Collected data from banks.

## Step 2 - Insert

Place the raw data files into the appropriate subfolders of the Input Files folder in File Explorer, based on whether the data is from POET or ABBY.

## Step 3 – Run

Execute the automation using the data from the Input Files folder using Python. Summary statistics are also provided in the Python notebook for initial analysis.

## Step 4 –Generate Pivot Table with Excel

Save the CSV file in another folder and open the output file with Excel to generate the Pivot Table.

4 Steps to run Tool

# Automated Bank Deposit Analysis Tool – Result

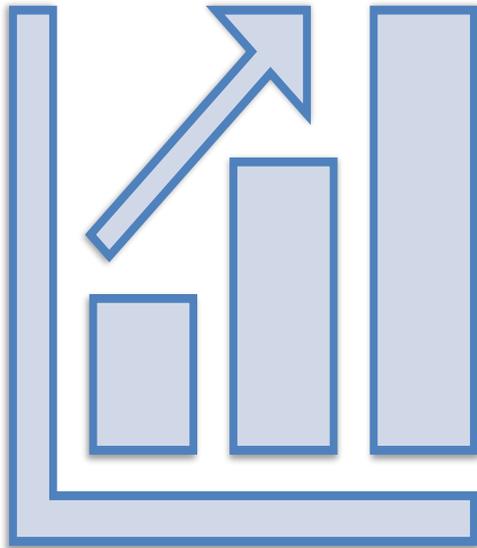
Pivot table manually generated with Excel

Year	(All)	
Row Labels	Sum of Deposits	Sum of Withdrawals
Advice	123.90	
Bill Payment	-	704,627.69
Bonus Interest	10.26	
CARE AND SUPPORT GOVT PAYI	300.00	
Cash Deposit Machine	374,810.00	-
Cash Withdrawal	-	94,858.75
CashCard Refund-ATM	9.91	
CashCard/FlashPay Top-up		250.00
Cheque Deposit	70,940.77	-
CR Adjustment-IATM	3.14	-
Debit Adj-IATM		13.28
Debit Card transaction		6,442.78
FAST PAYMENT		635,052.06
FAST RECEIPT	446,024.08	
FUND TRANSFER - PAYMENTS	-	1,831,365.71
FUND TRANSFER - RECEIPTS	324,080.93	-
GIRO	107,474.27	36,466.44
GOVT PAYOUT	600.00	
GST VOUCHER	600.00	
Interest Credit	48.67	-
Inward Credit-FAST	1,796,793.33	-
Misc CR-Debit Card	2,566.19	-
Misc Credit	5,150.00	
NETS CashC	140.00	2,945.00
NETS Debit-Consumer	-	10,157.98
NTUC SIRSPAYOUT	9,000.00	
Payment - Purrier	17,566.52	
Payment - Transferwise	47,951.80	
PAYNOW-FAST		673.37
Point-of-Sale Transaction		3,582.00
SALARY		21,700.00
Service Charge	0.50	168.00
Transfer - Chong	23,903.00	196,075.00
Transfer - Jayden		226,528.31
TT	568,991.35	150.24
<b>Grand Total</b>	<b>3,797,094.62</b>	<b>3,771,056.61</b>

Pivot Table generated with automation using Python

	Deposit	Withdrawal
<b>Category 1</b>		
<b>FAST</b>	\$433,662	\$374,429
<b>Inward Cash</b>	\$374,820	\$87,109
<b>Inward Transfer</b>	\$204,696	\$0
<b>Misc</b>	\$332	\$597,844
<b>NA</b>	\$57,632	\$23,818
<b>Outward Cash</b>	\$0	\$3,570
<b>Outward Transfer</b>	\$115,121	\$1,649,330
<b>Payment</b>	\$367,128	\$852,239
<b>Point of Sale</b>	\$0	\$3,582
<b>Receipt</b>	\$2,243,703	\$179,135
<b>All</b>	<b>\$3,797,095</b>	<b>\$3,771,057</b>

# Automated Bank Deposit Analysis Tool – Result



Boost Productivity & Accuracy



All dates are converted to the MM/DD/YYYY format to enable easy sorting of data by date.



Transaction types, which differ for each bank, are categorized into a few groups to facilitate cross-bank data analysis. For example, "Cash" encompasses both Cash and Cash Deposit Machine transactions.

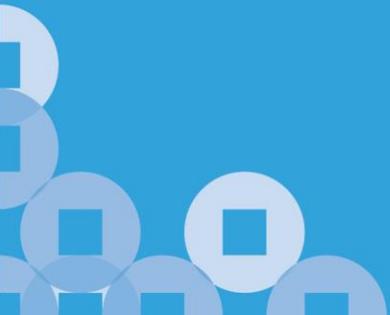


Previously required 3 to 4 weeks (i.e. 120 to 160 hours) to complete for over 1000 entries, can now be completed in less than 30 minutes.



Boost productivity- reduction of 15 man-days per person to just 30 minutes per person. If 20% of all new cases uses the tool, this would translate to annual savings of close to 100 mandays.

# Case Studies

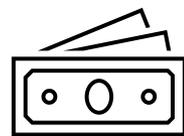


# Prosecution Highlight for Tax & Money Laundering Case

## BIGGEST CASE (IN TERMS OF IMPRISONMENT) FOR IRAS

**THE STRAITSTIMES**  
**Ex-company director who evaded tax, laundered money gets 8½ years' jail, must pay almost \$9.3m**

逃税、欺骗政府津贴和转移赃款，被判坐牢八年半，还被下令缴交罚款和罚金共927万元。  
**被告李之雄 (41岁) 面对127项罪名，他承认其中41项，余项交由法官下判时一并考虑。**  
《联合早报》报道，综合控方提控的罪状，2014年至2017年间，



**> \$9m**

in fines and penalties



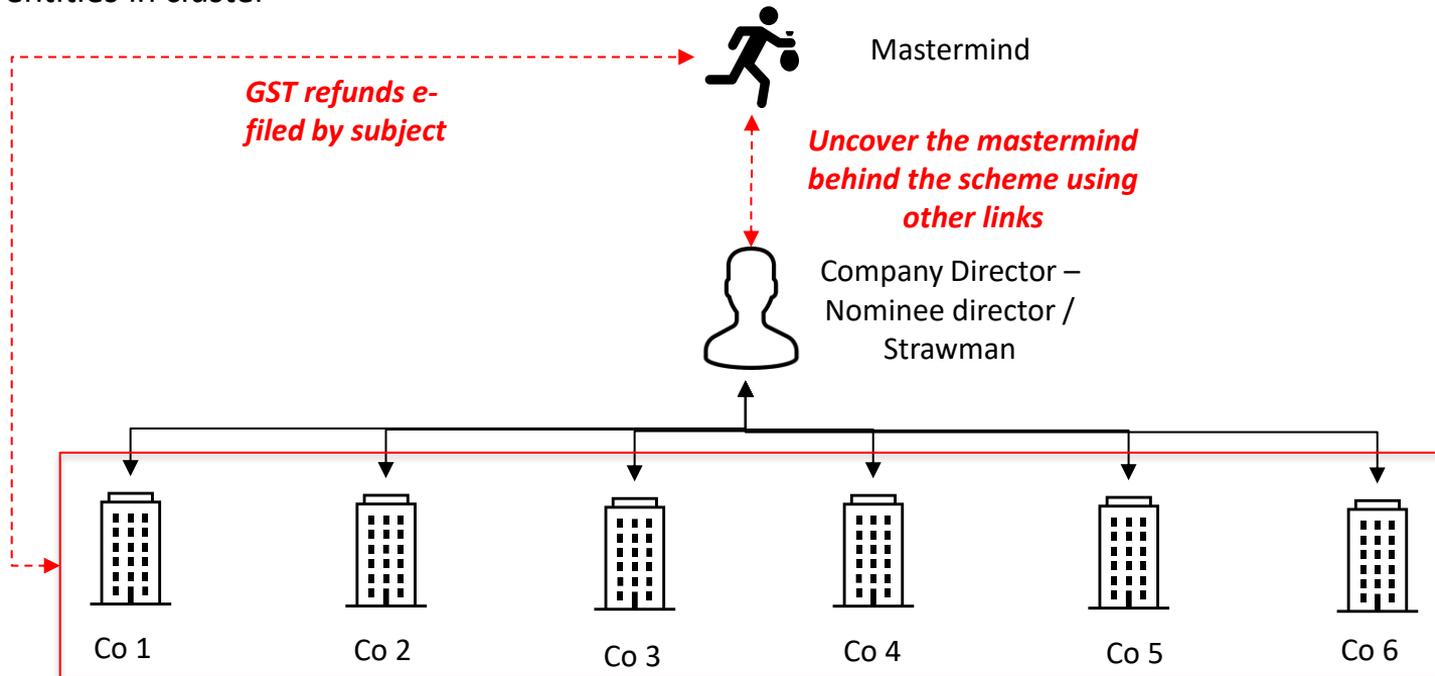
**8½ yrs** of imprisonment term

Close collaboration with CAD allowed for swift investigations into money laundering activities carried out by Thomas Lee using the illegally obtained funds, in addition to IT and GST offences

# Case Sharing – Using Network Tool to Link Entities

## Thomas Lee

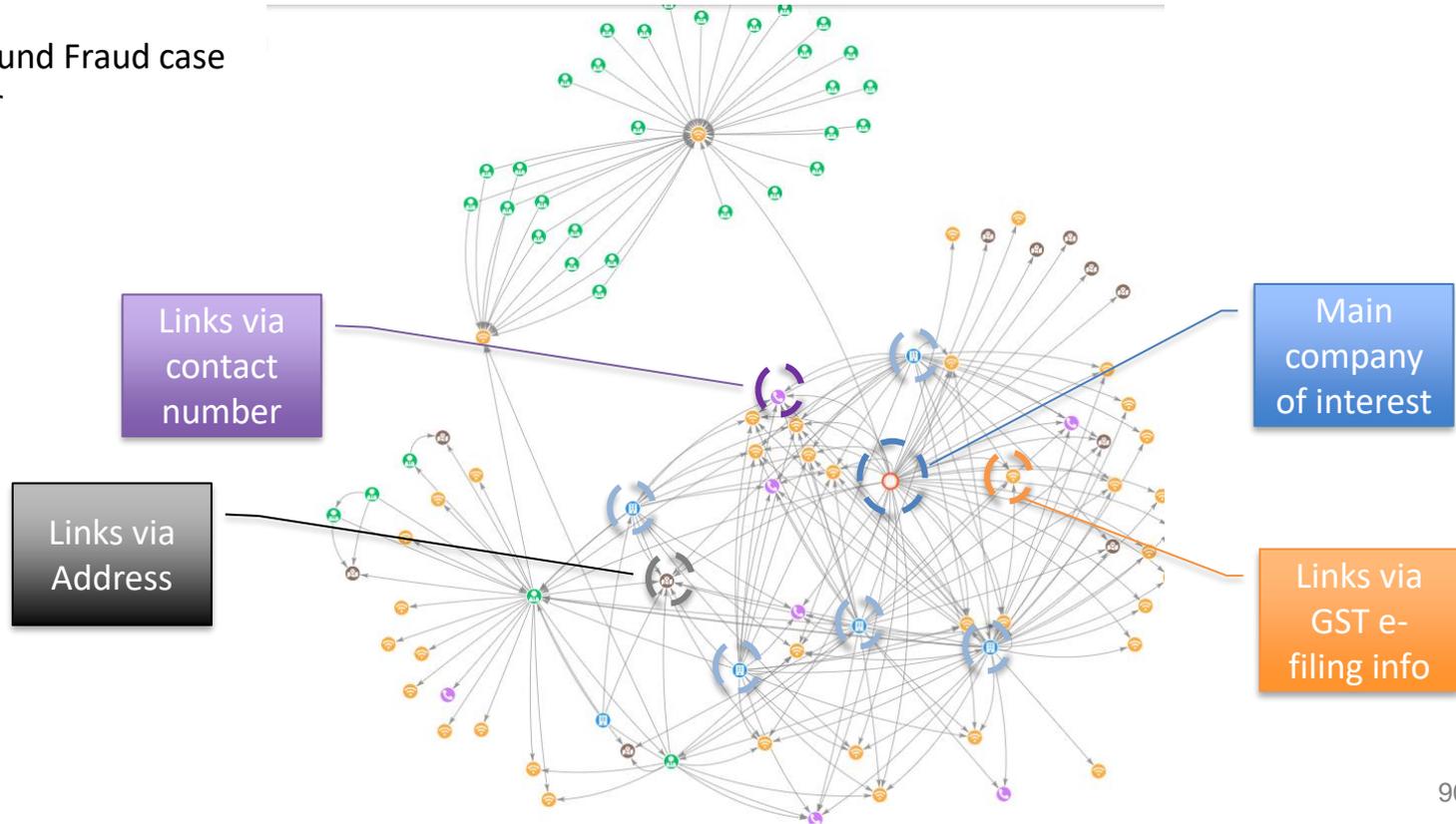
- Suspected GST Refund Fraud case
- 6 entities in cluster



# Case Sharing – Using Network Tool to Link Entities

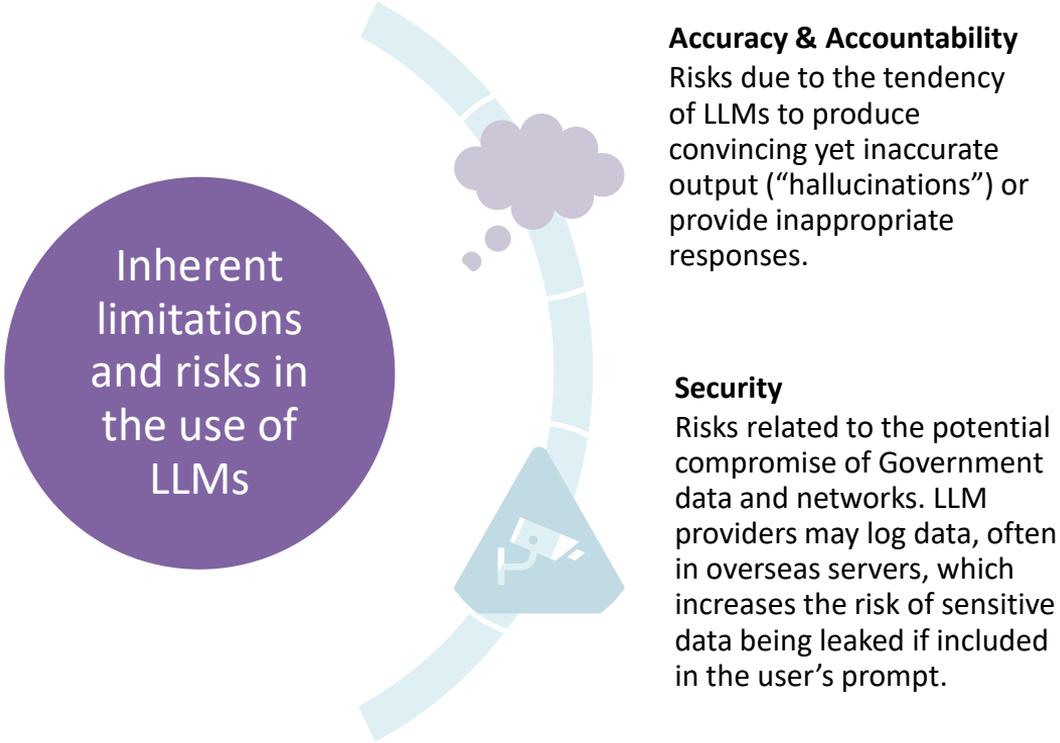
## Thomas Lee

- Suspected GST Refund Fraud case
- 6 entities in cluster



# Large Language Models (LLMs)

## Advisory from the Prime Minister Office (“PMO”) on use of LLMs for Government Agencies



Inherent limitations and risks in the use of LLMs

### **Accuracy & Accountability**

Risks due to the tendency of LLMs to produce convincing yet inaccurate output (“hallucinations”) or provide inappropriate responses.

### **Security**

Risks related to the potential compromise of Government data and networks. LLM providers may log data, often in overseas servers, which increases the risk of sensitive data being leaked if included in the user’s prompt.

# IRAS' Experiment of Large Language Models (LLMs)

IRAS are currently limiting the exploration of LLMs in the domain of productivity aid and information retrieval while managing the risks involved. One potential use case for investigation is to use LLM to extract key names mentioned in statement / document.

## Extract Key Names mentioned in the Statement

Entity Name	Email	Contact Number	Address
POI 1	d@lone2u.com	N/A	N/A
POI 2	@lone2u.com	N/A	N/A
COI A Pte Ltd	@lone2u.com	N/A	221 Address jS159557
POI 3	N/A	N/A	N/A
COI B	N/A	N/A	N/A
POI 4	N/A	N/A	N/A
COI C	N/A	N/A	N/A
COI D	N/A	N/A	N/A
COI E	N/A	N/A	N/A
POI 5	com.sg	N/A	N/A

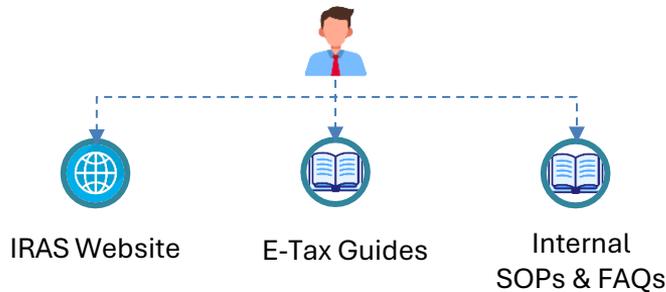
- **Pair Chat** is a free, fast and secure version of ChatGPT, and is currently available and free to use for all public officers. It is currently powered by the same Large Language Model underlying ChatGPT.
- Pair Chat has been cleared by the Smart Nation and Digital Government Office (SNDGO) for public officer use (up to certain data classification).
- Depending on the prompt, the no. of extracted entities extracted differ.

# IRAS' Experiment of Large Language Models (LLMs)

- IRASearch (Beta Version)

## Challenges

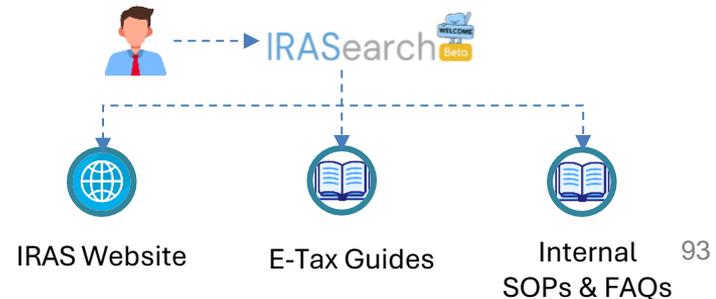
Currently, tax officers need consult multiple sources of information to answer tax inquiries. This could be onerous and time consuming, especially for new tax officers.



## Approach

In 2023, IRASearch was developed to assist staff in addressing tax inquiries more effectively. It is an intelligent search engine that leverages Gen AI/LLMs and references authoritative data sources.

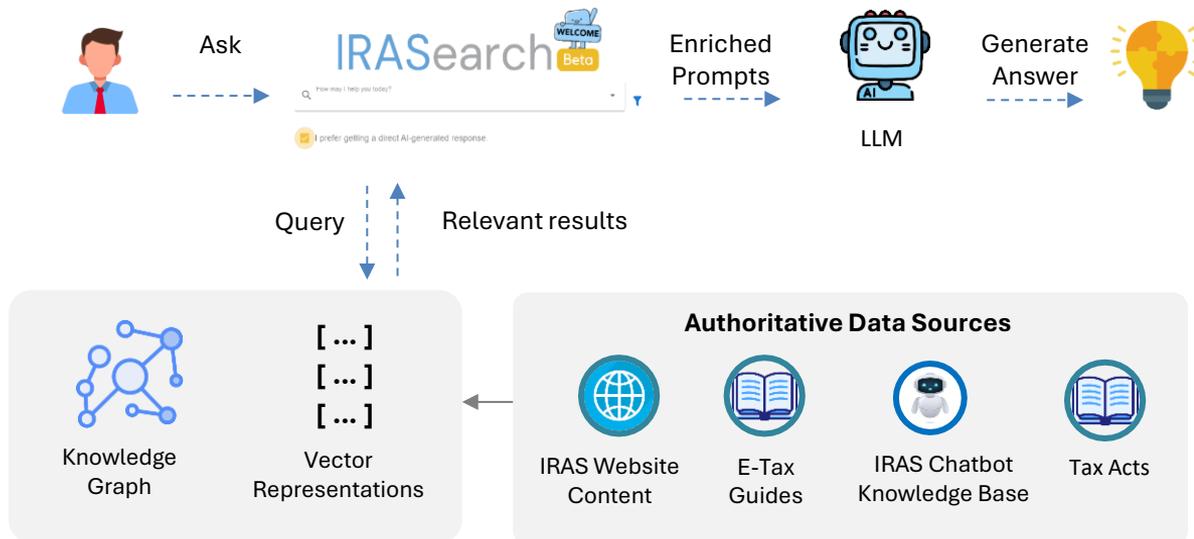
### In-House IRASearch



# IRAS' Experiment of Large Language Models (LLMs)

The IRASearch is an in-house search engine and Q&A system that delivers more relevant and reliable responses.

## Harnesses the power of Gen AI and LLMs



1

Transforms text queries into intent and enhances search with a hybrid methodology, using keyword search and semantic search.

2

Utilises Retrieval-Augmented Generation (RAG) and references authoritative data sources to deliver more reliable and contextually relevant results.

3

Provides customised content generation and assists frontliners with tax inquiries to improve service delivery and operational efficiency.

# IRAS' Experiment of Large Language Models (LLMs)

Pilot testing showed that the Q&A functionality in IRASearch is effective in generating accurate and relevant responses

## Integrating Retrieval & Generation Models for Improved Q&A

IRASearch  Hello,

\* Content is AI Generated. Please be mindful of hallucinations and never use the output without proper check.

**Generated Answer** ★★★★★

No, single mothers cannot claim Working Mother Child Relief (WMCR). WMCR is only available for working mothers who are married, divorced, or widowed and are maintaining an unmarried Singapore Citizen child in the preceding year. Single taxpayers, including single mothers, are not eligible for this relief. However, unmarried single parents who are maintaining a child can still claim for Qualifying Child Relief (QCR) as an administrative concession. QCR applies to both single mothers and single fathers, including those with illegitimate children. To claim QCR, you can email your claim request via myTax Mail in myTax Portal within each Year of Assessment.

**Citation**  
Can single mum claim Working Mother Child Relief (WMCR) in respect to her child born out of wedlock (i.e. the child is illegitimate)

[Q&A] FAQ on Child Related Relief for Single Parents  
[Q&A] FAQ on Working Mother Child Relief (WMCR)  
Tax savings for married couples and families - Working Mother's Child Relief (WMCR)  
Tax savings for married couples and families - Claiming WMCR on the first child

Search results were passed on to generation model along with the original query to produce a coherent and informative response

**ENHANCE ANSWER**

The Q&A functionality is tailored for tax-related inquiries via customised OpenAI backend system prompts.

**94%**

**Found Relevant Answers from Generated Responses**

In Sep 2023, pilot testing showed that 94% of users found relevant answers when trying out the new Q&A feature in IRASearch. Enhancements were subsequently made to improve the relevancy and reliability of the AI-generated responses.

# IRAS' Experiment of Large Language Models (LLMs)

Deployment of IRASearch for 2024's IIT Filing Season proved useful for newer staff; enhancements needed to increase adoption and productivity gains

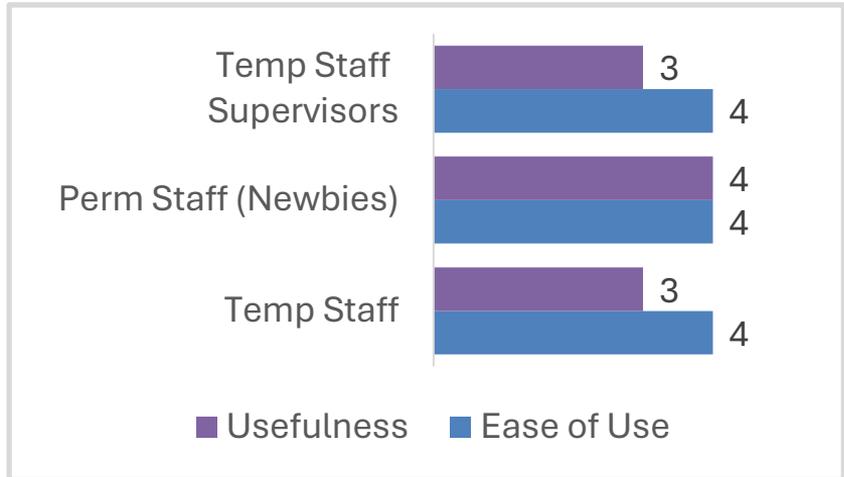
In Feb 2024, we rolled-out IRASearch to a group of 202 frontline staff at Service Experience Centre (SVE), including new perm staff, temp staff for the IIT Filing Season and temp staff supervisors.

Overall, **Perm Staff (Newbies)** found it to be **most useful** (rating 4 out of 5), while most staff found it to be easy to use.

For **experienced frontline staff**, they prefer to **look for their supervisors** as they require end-to-end advice quickly while on the line – the current IRASearch could not handle their enquiries fully due to the lack of data sources covering internal processes and SOPs.

Further enhancement: To Incorporate internal SOPs and other documents into data source

Usefulness & Ease of Use (Average Rating 1 to 5)

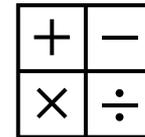


# New Development in LLM

- New development in LLM\*



*Snap a picture of a landmark while traveling and have a live conversation about what's interesting about it.*



*Help your child with a math problem by taking a photo, circling the problem set, and having it share hints with both of you.*

# Use of Aibots by IRAS

AI chatbots have been developed to provide quick and reliable responses to IO's questions, making it easier to navigate through the investigation manual and tax legislations.

## 1. GST -

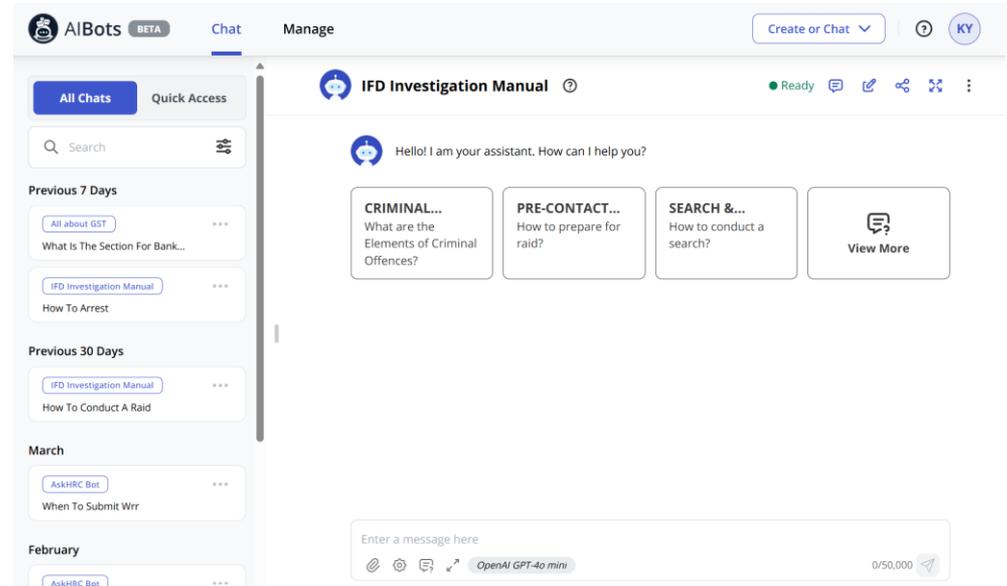
<https://aibots.gov.sg/chats/gst-act>

## 2. Income Tax -

<https://aibots.gov.sg/chats/all-about-income-tax>

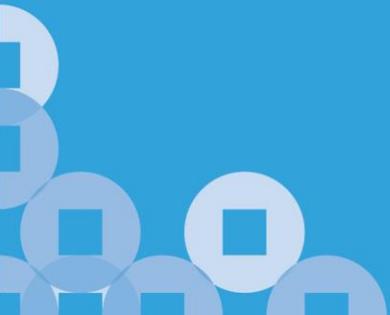
## 3. IFD Investigation Manual -

<https://aibots.gov.sg/chats/ifd-investigation-manual>



The screenshot displays the AIBots chat interface. The top navigation bar includes the AIBots logo, a 'BETA' badge, and tabs for 'Chat' and 'Manage'. A 'Create or Chat' button is visible on the right. The main chat area is titled 'IFD Investigation Manual' and shows a 'Ready' status. The chat content begins with a greeting: 'Hello! I am your assistant. How can I help you?'. Below the greeting are four interactive cards: 'CRIMINAL...' (What are the Elements of Criminal Offences?), 'PRE-CONTACT...' (How to prepare for raid?), 'SEARCH &...' (How to conduct a search?), and a 'View More' button. At the bottom, there is a text input field with the placeholder 'Enter a message here' and a character count '0/50,000'. The left sidebar shows a list of chat categories: 'All Chats' (Quick Access), 'Previous 7 Days' (All about GST, What is The Section For Bank...), 'Previous 30 Days' (IFD Investigation Manual, How To Arrest), 'March' (AskHRC Bot, When To Submit Wrr), and 'February' (AskHRC Bot).

# Group Discussion



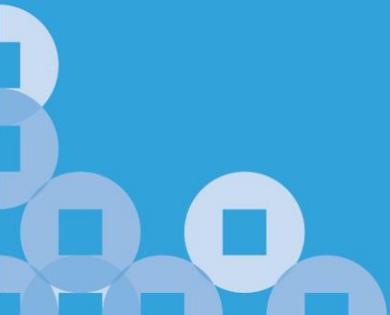
## Country Sharing (20 mins)

What are some possible applications for tools like ChatGPT for your agency?

Possible uses of tools :

- Gatekeeper
- Operational analysis
- Investigative efficiency

What's next



# What is next in IRAS AI Strategy

## Rich Entity-centric Data Portfolio

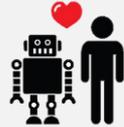


Enrich understanding of taxpayers using data/big data

### New Data Sources

- To continue to expand our sources of data (e.g. InvoiceNow)
- To tap on new data to refine existing solutions (e.g. fraud detection using bank transaction data) or explore new solutions (e.g. sentiment analysis on audio data)

## AI for Everyone



Leverage AI/data science for intelligent tax administration/ enterprise disbursement

### Expand AI Applications

- By identifying strategic use-cases with high impact and readiness, to increase the effectiveness and efficiency of our service delivery
- By enhancing and scaling existing AI applications

## Future-Ready Data Technology with Optimal Connectivity

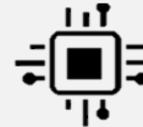


Strengthen capabilities to receive, store and use big and real-time data and scale AI solutions

### Modernise Data Architecture

- By migrating the UDP to the Cloud to exploit Cloud capabilities and continuously improving it to strengthen AI capabilities
- By exploring and identifying technologies required to enhance governance in the ML/AI development lifecycle

## Democratise Data Science and AI Capabilities



Drive pervasive data usage across all staff

### Develop Expertise

- By continuously investing in our people and relevant tools
- By exploring partnership opportunities with external stakeholders (e.g. short-term attachment/exchange)

## Trusted Data Ecosystem



Use and share AI and data responsibly

### AI Ethics & Governance

- By establishing and formalizing a set of principles and measures to inculcate the responsible adoption of AI in IRAS without stifling digital innovation

# Thank You

The information presented in the slides aims to provide a better general understanding of taxpayers' tax obligations and is not intended to comprehensively address all possible tax issues that may arise. This information is correct as at the date of presentation. While every effort has been made to ensure that this information is consistent with existing law and practice, should there be any changes, IRAS reserves the right to vary its position accordingly.