

DOS AI and Generative AI Playbook

Discover the world of AI and Generative AI with this playbook, where we explore the benefits, limitations, and latest technologies relating to these game-changing fields.

Updated Feb 2026

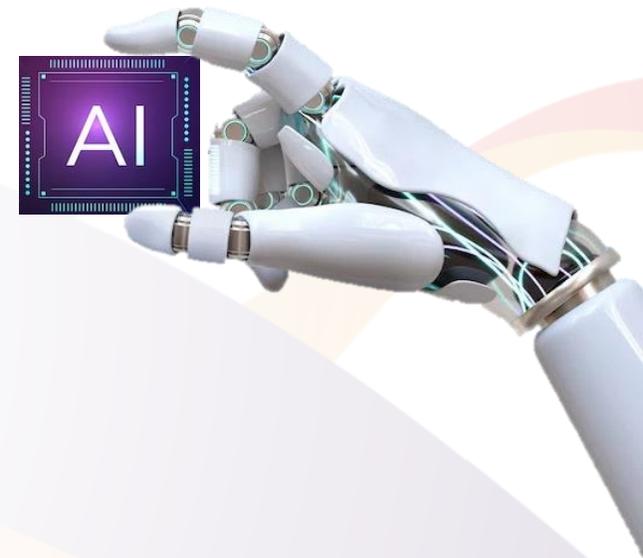


Table of Contents

01. **Introduction**

02. **Understanding AI and Generative AI (GenAI)**

03. **Leveraging AI**

04. **Governance and Safeguards**

05. **Learning Resources**



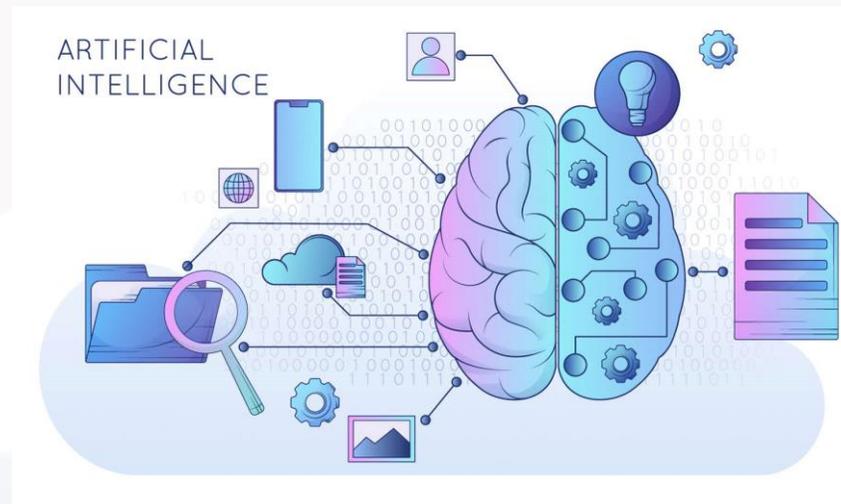
01. Introduction

- Purpose of DOS AI Playbook
- AI/ML in DOS Digitalisation Strategies

Purpose of Playbook

This playbook provides a brief overview of artificial intelligence (AI) and generative AI (GenAI), showcases some of the current and potential uses of (Gen)AI in DOS, as well as highlights the principles and governance of the use of (Gen)AI.

The examples and use cases presented in this playbook are curated to showcase the practical AI applications in the statistical and data analytical fields useful for DOS officers.



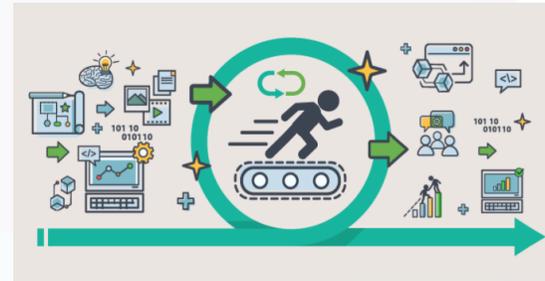
DOS Digitalisation Strategies

Leveraging statistics and AI to create value is one of four key digitalisation strategies in supporting DOS to be the first stop for data and data services.

1 - Leverage combination of expertise in statistics, Artificial Intelligence (AI) and Machine Learning (ML) to create value



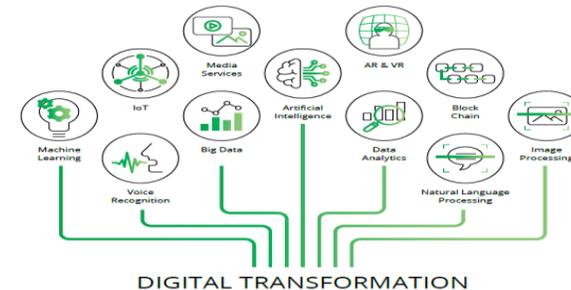
2 - End-to-end service journeys to improve customer experience

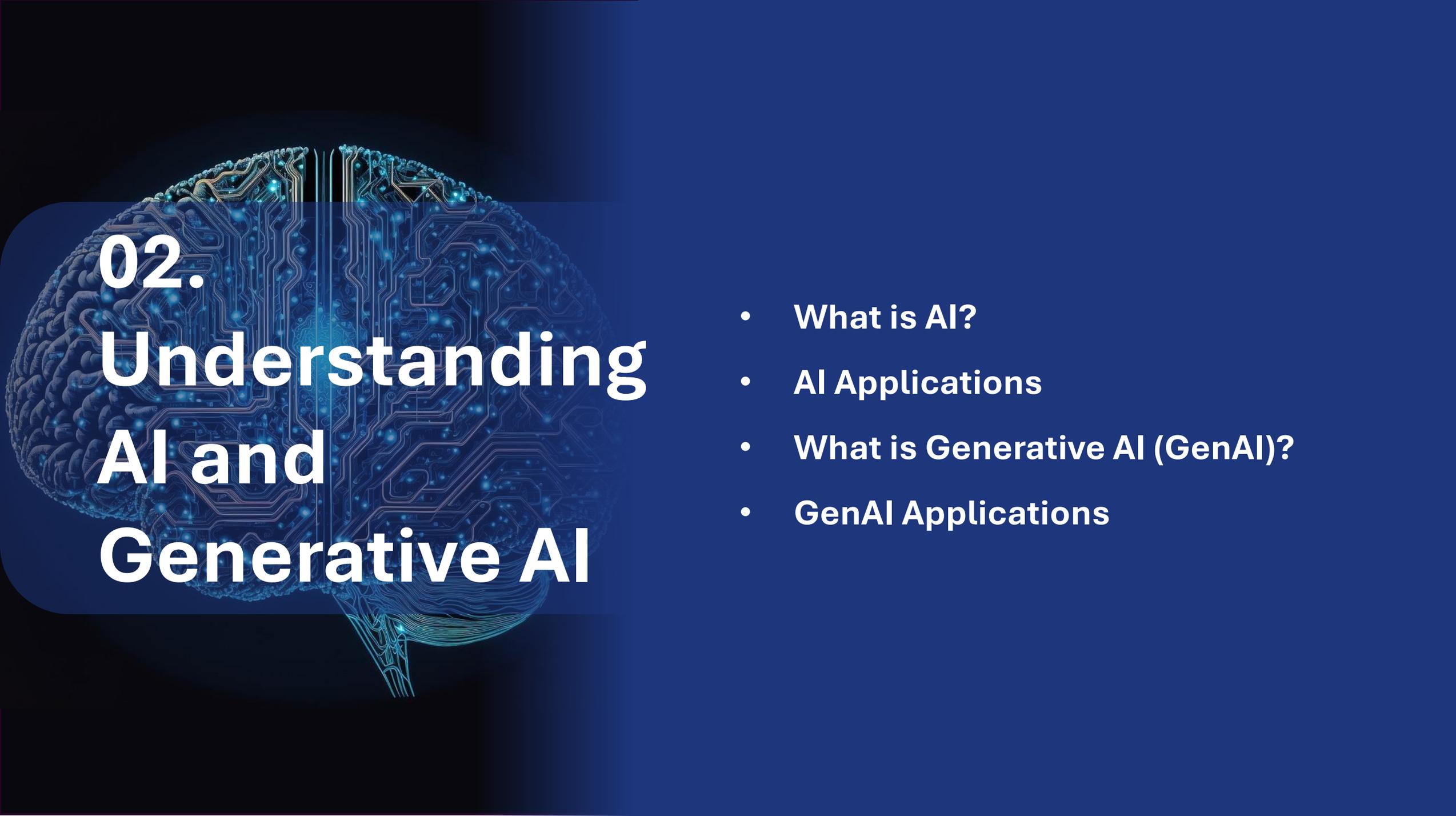


3 - Adopt an “Agile-first” approach



4 - “Cloud-first” approach to deliver new capabilities





02.

Understanding AI and Generative AI

- What is AI?
- AI Applications
- What is Generative AI (GenAI)?
- GenAI Applications

What is AI?

AI is a field referring to the collective advancements around building systems that mimic what humans can do – see, speak, hear, move and most importantly, learn. This means that AI systems are able to adapt their approach to solving a problem based on new data without the need for manual reprogramming of the system.

Example:

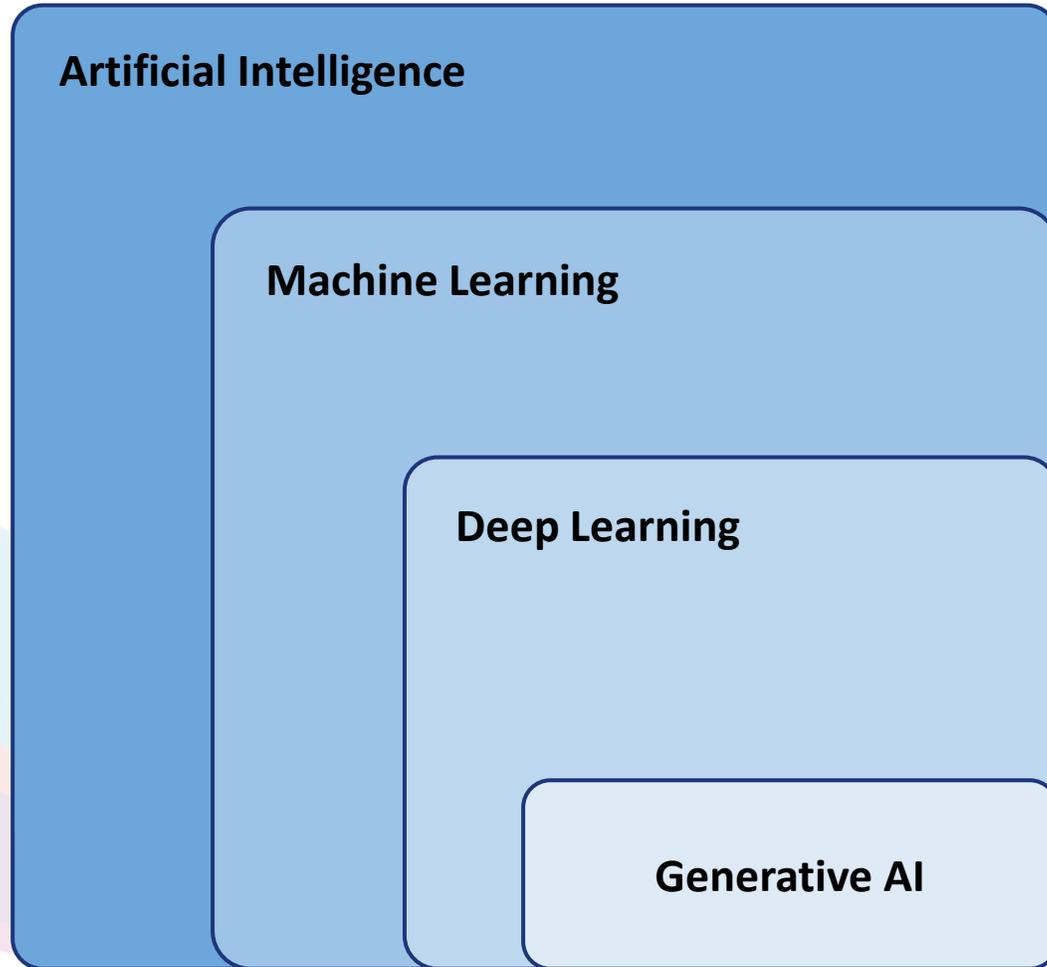


PlateSnap is a **video analytics** tool developed through a collaboration between GovTech and URA for **detection and identification** of vehicles and parking behaviour.

It has 3 key features:

- Detect parking behaviour by unique vehicles
- Detect and identify license plates through optical character recognition (OCR)
- Provide insights on dwell time and parking utilisation rates within region of interest

What is AI?



1950s

Birth of **Artificial Intelligence**

The field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence

1959

Machine Learning was coined

Subset of AI that enables machines to learn and/or improve from data to make decisions or predictions

2017

Deep Learning became mainstream

A machine learning technique in which layers of neural networks are used to process data and make decisions

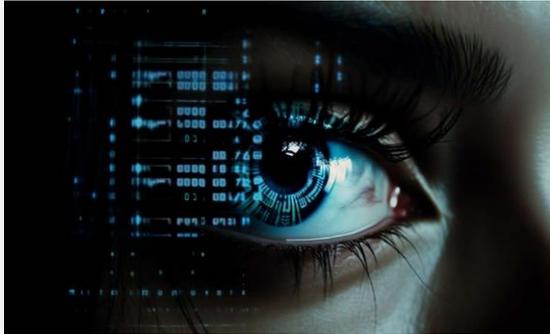
2021

Exponential growth in **Generative AI**

A rapidly growing field of AI that creates new written, visual, and auditory content given prompts or existing data

AI Applications

AI can be used for many applications. The following is a list of non-exhaustive examples:



Computer Vision

Enables computers and systems to derive meaningful information from digital images, videos and other visual inputs through deep learning and convolutional neural networks.

Key areas:

- Image classification
- Object detection
- Facial recognition
- Optical Character Recognition (OCR)



Natural Language Processing

Combines computational linguistics with statistical, machine learning, and deep learning models to enable computers to understand text and spoken words in much the same way human beings can.

Key areas:

- Translation
- Classification
- Summarisation
- Question and answering



Time Series Forecast

Predicts future data based on analyses of historical data.

Key areas:

- Forecasting
- Anomaly detection

What is Generative AI (GenAI)?

Generative AI (GenAI) is a field of AI development that focuses on models capable of generating new content and data such as text, images, or sounds that mimic the styles, patterns, and features of the original training dataset.

W

What is Generative AI? In less than 100 words

P

Generative AI is a type of artificial intelligence that can create new content, such as images, music, or text, that is similar to content it has been trained on. It uses deep learning algorithms to learn patterns and relationships within large datasets, and can be used for a variety of applications, such as creating realistic images for video games or generating personalized recommendations for users. However, it also raises ethical concerns, as it can be used to create fake content, such as deepfakes, that can be used to spread misinformation or manipulate public opinion.



DALL-E 2 creates realistic images and art from a description in natural language.

[Pair, an alternative to ChatGPT](#), generates human-like text and holds human-like conversations.

GenAI Applications

GenAI can be used for many applications. The following is a list of non-exhaustive examples:



Create new text and image content

Eliminate writer's block and boost productivity and creativity by generating writing and art instantaneously



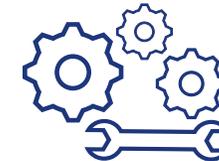
Discover trends and gain insights from data

Discover key information from your data while working with complex data models



Summarise contents for fast decision-making

Search, comprehend and extract information from large volumes of content



Automate solutions and processes

Automate information retrieval and categorisation or recommendation steps for a recurring process



03. Leveraging AI

- **DOS AI Journey**
- **AI@DOS**
- **AI Tools Test Drive @ DOS**
- **AI@GovTech and Other Governments**
- **AI in the Near Future**
- **What's Next for AI?**

DOS AI Journey



In 2018, DOS embarked on an AI journey to ride on the rapid evolution of data analytics and digitalisation. To stay ahead, DOS established the Digitalisation and Analytics Workgroup to spearhead and co-ordinate DOS-wide analytics and digitalisation efforts. This was an enterprise-wide endeavour involving operational unit, HR, and IT representatives, who combined their statistical expertise, staff competency developmental support and a robust computing infrastructure, to handle the surge in data volume and complexity as well as demanding computing needs.

In 2019, DOS identified R and Python as key open-source data science tools, and embarked on a plan for a new computational infrastructure to support the emerging field of AI/ML. In 2020, thematic sharing on data science and analytics at the semi-annual DOS-RSUs¹ meetings facilitated the exchange of ideas and experience of using AI for statistical work. In Census 2020, DOS used ML to automate the coding of SSOC and SSIC² codes, achieving an estimated 5600 man-hours savings.

¹ RSUs refers to Research Statistical Units.

² The SSOC is the national standard for classifying occupations while SSIC is for classifying economic activities undertaken by economic units.

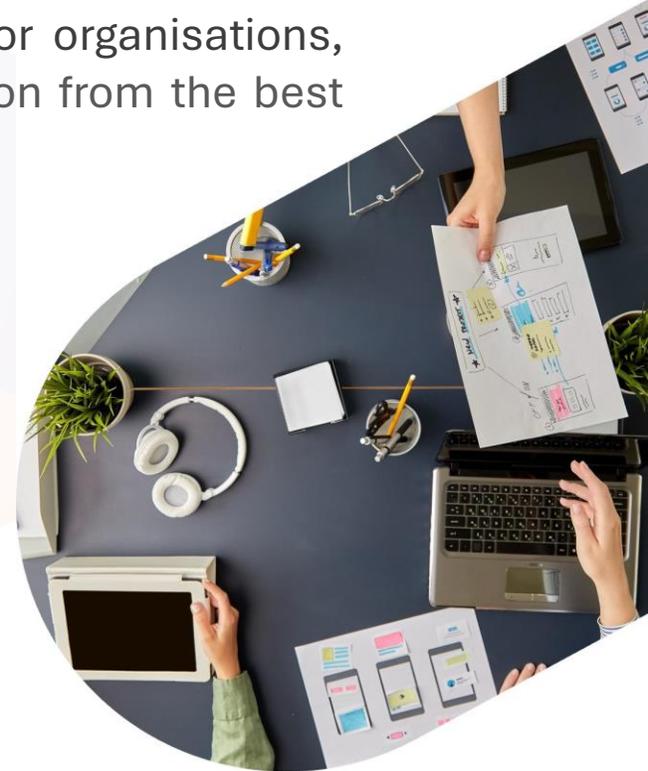


DOS AI Journey

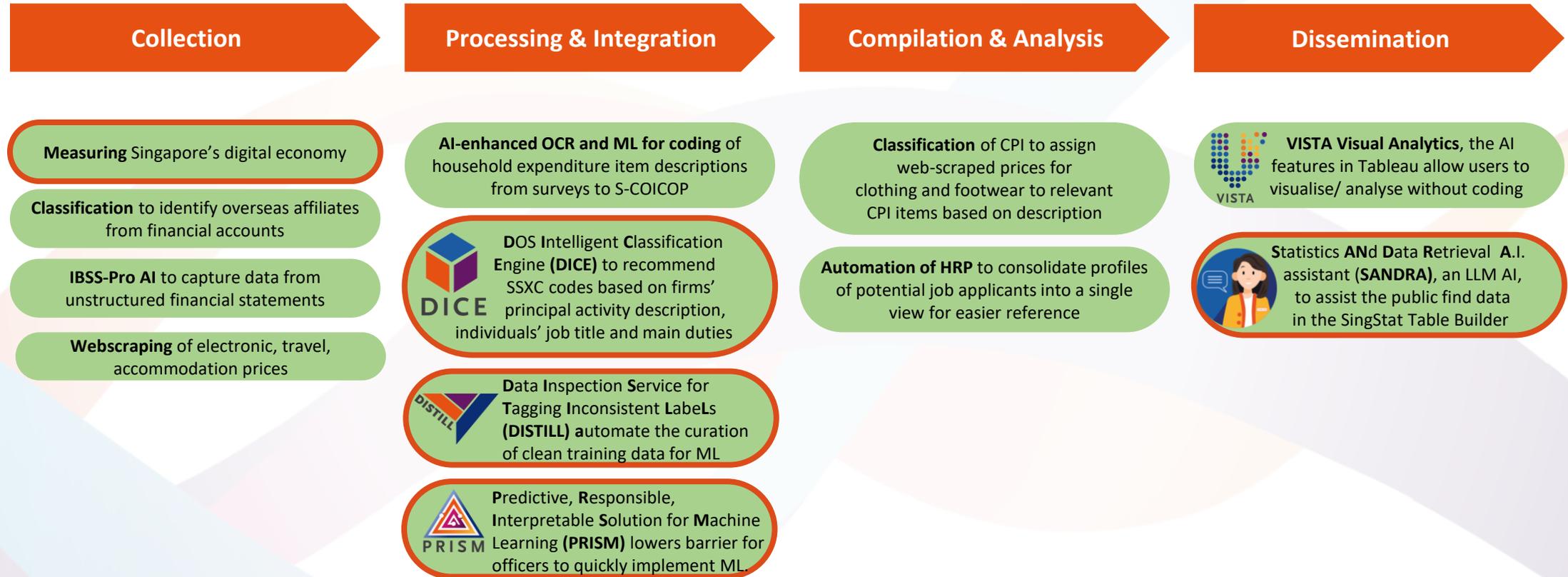


2021 was a milestone year with the inauguration of DOS AI/ML Day, an event which promotes wider AI/ML use and offers valuable insights from industry and WOG agencies on utilising AI/ML to address business challenges. Additionally, the Microsoft Machine Learning Server was deployed onto DOS Nutanix servers, providing a high-speed computing infrastructure for our AI/ML and Analytics projects. A few selected officers were also attached to innovative private sector organisations, such as Carousell, Expedia and Grab to draw inspiration from the best industry practices.

In 2022, the DOS Digital Transformation Unit (DTU) was formed to provide more dedicated support for strategic initiatives and accelerate time-to-market. The formation of DOS Community-of-Practice also enabled business teams to tap into a wider pool of talents within DOS, and curated training workshops helped to uplift officers' technical knowledge in strategic AI/ML and automation domains.



Selected AI/ ML use case in Statistical/ Data Production Processes



AI @ DOS - DICE

Simplifying standard classification through natural language processing

The Singapore Standard Classification (SSXC) is used in the collection, compilation, and dissemination of national statistics. For instance, the Singapore Standard Industrial Classification (SSIC) is used to classify firms according to their principal economic activity. Incorrect classification can adversely affect statistical compilation (e.g., GDP reporting), research studies (e.g., sectoral heterogeneity analysis), and policy implementation (e.g., level of support to promote a sector's growth). However, users—whether business firms or public officers—might find it challenging to choose the most appropriate SSIC code from a range of possible codes without clear knowledge of their nuanced differences. Similarly, the Singapore Standard Occupational Classification (SSOC) is used to classify occupations; incorrect classifications can affect population censuses, household surveys, and administrative databases.

The DOS Intelligent Classification Engine (DICE) was developed to automatically identify the most relevant SSXC code through machine learning, thus improving consistency and reducing manual effort in the process. DICE processes textual inputs (e.g., a firm's economic activity or a job's title and description) and recommends suitable SSXC codes as an output, along with the engine's confidence score.

Input:

Firm's self-declared activity description

Wholesale of Cars



Prediction	Score
SSIC 46561: Wholesale of motor vehicles except motorcycles and scooters	92%
SSIC 46563: Wholesale of parts and accessories for vehicles	5%
SSIC 47311: Retail sale of motor vehicles except motorcycles and scooters	2%

AI @ DOS – Statistics ANd Data Retrieval A.I. Assistant (SANDRA)



- ❑ Statistics ANd Data Retrieval A.I. assistant ([SANDRA](#)) is an artificial intelligence (AI) chatbot that uses **natural language** to assist users with their data enquiries on Singapore's economy and socio-demography.
- ❑ Through semantic search capabilities and use of Large Language Model, SANDRA **understands the context of users' queries and confines to retrieving data available** in the [SingStat Table Builder](#).
- ❑ SANDRA presents the most relevant data retrieved in both **chart and table formats**, and facilitates data exploration by **suggesting other related data from the SingStat Table Builder** which may be of interest to users.

Chat with SANDRA

View SANDRA in action at
go.gov.sg/sandravid

DEPARTMENT OF STATISTICS SINGAPORE

About SANDRA Publications & Methodology Glossaries Data FAQ

Chat with SANDRA ^{BETA}

Hi there!
I'm SANDRA, your Statistics ANd Data Retrieval A.I. assistant. I'm here to help you find data on Singapore's economy and socio-demography.

To get started, kindly ask your question in this manner: "[data you need], [data frequency you require (monthly, quarterly, or yearly)], [period you're interested in]".
For example, you can ask: "Inflation rate, annual, 2020 to 2024".

Type your message here

Look out for the chart and table for your data query here.
Happy exploring!

Here are some popular queries which you may find interesting:

- GDP growth in the last 5 years
- Monthly inflation rate in the past year
- Household employment income in the last 5 years
- Latest monthly household expenditure on goods and services
- Current population
- Tourist arrivals since 2020
- Employment by occupation in the last 2 decades
- Latest greenhouse gas emissions data

AI @ DOS – IBSS Pro-AI

Leveraging AI for Data Extraction of Unstructured Data from Financial Statements

Unstructured data from enterprises' financial statements are also a rich source of financial information and new insights. However, considerable manual effort is required to read, analyse and extract the relevant information. This curtails the number of financial statements and data points that can be captured and processed for statistical use.

IBSS Pro-AI was developed for data extraction and processing of unstructured data from enterprises' financial statements to supplement existing financial information in the Statistical Business Register (SBR) for compilation of economic and business indicators.

Two examples are presented below:

- Detailed assets information extracted from the notes of the financial statements can be used to support in-depth analysis on enterprises' asset structure and investment.
- More detailed shareholding information supplement existing machine-readable data available in the SBR, for ownership analysis.

DETAILED ASSETS INFORMATION²

	Note	Group		Company	
		2020 \$'000	2019 \$'000	2020 \$'000	2019 \$'000
Non-current assets					
Property, plant and equipment	4	1,929	2,165	247	213
Intangible assets	5	769	990	637	773
Investment properties	6	2,730	2,981	–	–
Subsidiaries	7	–	–	86,663	86,163
Other investments	9	18,819	25,096	54	14
Loans, advances, hire purchase and leasing receivables	10	82,332	83,092	75,837	69,368
Deferred tax assets	12	3,692	3,856	–	–
Right-of-use assets	38	2,525	2,839	1,834	2,020
		112,796	121,019	165,272	158,551

Extract value '247000' based on interpretation of column names (i.e., year 2020 and units ('000) and row name (i.e., Property, plant and equipment)

SHAREHOLDING INFORMATION

As at 31 December 2019, the Company's immediate holding company is **AB Limited**, a company incorporated in the **Republic of Singapore**. The Company's intermediate holding company is **ABO**, a company incorporated in **Denmark**, and the ultimate holding company is **AB Foundation**, an enterprise foundation registered in **Denmark**.

Extract name and country of the immediate, intermediate and ultimate companies (highlighted)

AI Tools Test Drive @ DOS

Hear the feedback from DOS colleagues who tried out AI tools!

Bing Image Creator – an AI image generation tool that creates images based on text.

The tool can **quickly generate images and provide image variations**, which can be used as references when starting on design work.

It is notable that the generated square images are in JPG format, which can lose its quality when scaled and are **not editable for customisation**. The styles of the images vary, so to achieve greater style consistency, prompts need to be lengthened progressively with increasing details without rewriting the prompts entirely. --- Feedback from CED



Prompt: Vector art of groceries in bag



Prompt: Lady eating, illustrations



Prompt: Vector art of laptop in monochromatic blue



Prompt: Flat graphics of money

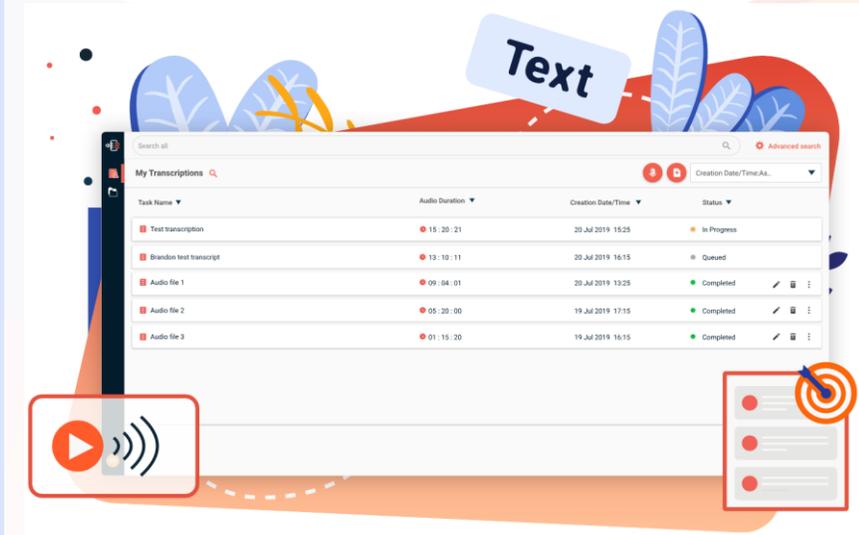
AI Tools Test Drive @ DOS

Hear the feedback from DOS colleagues who tried out AI tools!

Transcribe – a platform with localised speech-to-text services using speech models to produce transcripts

For meetings or training sessions conducted via Microsoft Teams, the platform has an **integrated Transcribe feature that enables live transcription in real-time**, including the speaker's name and timestamp. The transcription can also be downloaded and shared with the participants after the session.

For previous video recordings where the transcription feature was not activated, GovTech's Transcribe is a useful tool for transcribing these videos, and **viewers can refer to the transcript for the corresponding timestamps in the video**. In addition, the "summariser" feature integrated in GovTech's Transcribe is a helpful feature that allows users who have access to the transcription to **generate a summary of the transcribed text**. However, it would be **more beneficial if the transcription could be directly converted into video captions** and added to the video recording, making it easier for viewers to follow the video.



Potential GenAI@DOS

Content generation

E.g., Social media content, organisational planning, consideration metrics for data strategy for the organisation

Text summarisation

Extract key information and simplify long documents, instead of trudging through long documents for the main points

Writing assistance

Let the chatbot know any context you have (e.g., the email you are replying to), or get help in code writing (e.g., debug code, explain code, convert natural language to code)

Data analytics

E.g., summary statistics, explanations of complex data relationships, recommendations for further analysis

Transform unstructured data into structured data

E.g., entity extraction of company names mentioned in an article, skillsets mentioned in a CV

Question answering

E.g., suggest answers to simple methodological or technical questions, personalised learning experience for internal staff based on wealth of knowledge accumulated within the organisation

Exploring

In progress

Potential GenAI@DOS

Average Monthly Household Income Deciles	Real Change in Average Monthly Household Income from Work Per Household Member (Including Employer CPF Contributions) Among Resident Employed Households by Deciles, 2022	Average Household Size of Resident and Resident Employed Households by Deciles, 2022
Total	1.3	3.27
1st (Lowest)	10.1	3.41
2nd	4.8	3.45
3rd	2.9	3.43
4th	2.5	3.52
5th	2.3	3.53
6th	2.6	3.48
7th	2.2	3.35
8th	2.4	3.21
9th	1.1	2.95
10th (Highest)	-1.3	2.34

Example of ChatGPT used for data analytics

Given the tabular data input (above), ChatGPT returns a description of the input data (right). You can even raise questions relating to the data.



This table provides information on the real change in average monthly household income from work per household member (including employer CPF contributions) among resident employed households by deciles, as well as the average household size of resident and resident employed households by deciles for the year 2022.

The total real change in average monthly household income from work per household member for resident employed households is 1.3.

Looking at the data by deciles, the lowest 10% of resident employed households experienced a much higher real change in average monthly household income from work per household member of 10.1, compared to the other deciles.

The 10th decile (highest) experienced a decrease of -1.3 in their real change in average monthly household income from work per household member.

Regarding the average household size, it ranges from 2.34 for the 10th decile to 3.53 for the 5th decile. The lowest 10% of households also have a relatively larger average household size of 3.41.

AI @ GovTech

AI tool	Data classification	Other requirements	Pricing*
<u>Transcribe</u> : platform with localised speech-to-text (STT) services to produce transcripts of interviews, speeches and meeting minutes etc.	Up to Restricted/ Sensitive Normal	-	Free
<u>Video Analytics System (VAS)</u> : platform for development, deployment, and utilisation of video analytics capabilities	Up to Restricted/ Sensitive Normal	Require Techpass account (part of SEED platform)	-
<u>GovText</u> : natural language processing (NLP) platform that can perform text summarisation and topic modelling	Up to Restricted/ Sensitive Normal	-	Free
<u>Pair</u> : ChatGPT for Public Officers	Up to Restricted/ Sensitive Normal		Free

AI @ Other Government

Countries that have embraced (Gen)AI

France, May 2023. The French National Institute of Statistics and Economic Studies uses neural network to classify the French classification of Activities (NACE Rev. 2) based on company-provided activity descriptions. It has been deployed in production since Nov 2022. [[Source](#)]

Australia, Feb 2020. The Australian Bureau of Statistics built a convolutional neural network to classify aerial imagery based on the address use. This reduces the amount of manual intervention required during regular address register maintenance processes. [[Source](#)]

UK, Jul 2023. The UK police used an AI-equipped van to identify drivers using mobile phones, drivers or passengers not wearing seat belts, and mechanical offences relating to issues with the vehicles. [[Source](#)]

UAE, May/Jul 2023. The UAE government launched a unified AI-powered chatbot for government services, “[U-Ask](#)”, which allows users to access information about government services in one place in both Arabic and English. The UAE is also employing AI in various fields to accelerate digital transformation. [[Source](#)]

Portugal, Jun 2023. Portugal is trialing a ChatGPT-based chatbot to answer emergency calls during busy times. The AI system will respond to the calls first to assess the reported situation before directing the calls to a human dispatcher. If the pilot trial goes well, the system could be rolled out by 2025. [[Source](#)]

Countries that have blocked ChatGPT

15 countries in total, including China, Iran, North Korea and Russia, have banned ChatGPT. [[Source](#)]

Countries investigating GenAI/coming up with AI regulations

EU, Jun 2023 – The members of the European Parliament have approved the negotiating draft of the European AI Act. [[Source](#)]

China, Jul 2023 – The Chinese regulators jointly passed the Interim Measures for the Management of Generative Artificial Intelligence Services, which comes into effect on 15 Aug 2023. [[Source](#)]

AI in the Near Future

Generative Business Intelligence (BI)



Use case

With continuously growing data volumes, the demand for BI is growing.

Traditional BI tools require human experts in operational processes such as data preparation and processing, as well as visualising insights in an appropriate form. Such manual processes require time and human resources. There is a risk of making mistakes due to human factors, such as human fatigue or human bias.

Solution

Generative BI enables the building, discovering, and sharing of meaningful insights in seconds, turning insights into impact faster. Business analysts can perform common tasks using natural language, including describing the visuals in natural language to quickly generate visuals and creating calculations without needing to know specific syntax.

Business users who consume and interact with dashboards can generate, customise, and share compelling visual narratives using natural language prompts. This would help users to make informed business decisions.

AI in the Near Future

Use case

Augmented Analytics

Data analysts are required to review multiple data sources and formats, such as CSV/PDF files, online articles, and data provided by other agencies.

With many different data sources, officers may find it difficult and time-consuming to consolidate the data and generate meaningful insights.

Solution

GenAI automates the rote aspects of data analysis, freeing humans to focus on higher-level considerations. It generates and explains insights based on the provided data sources, or even searches data on the Internet.

Question Answering

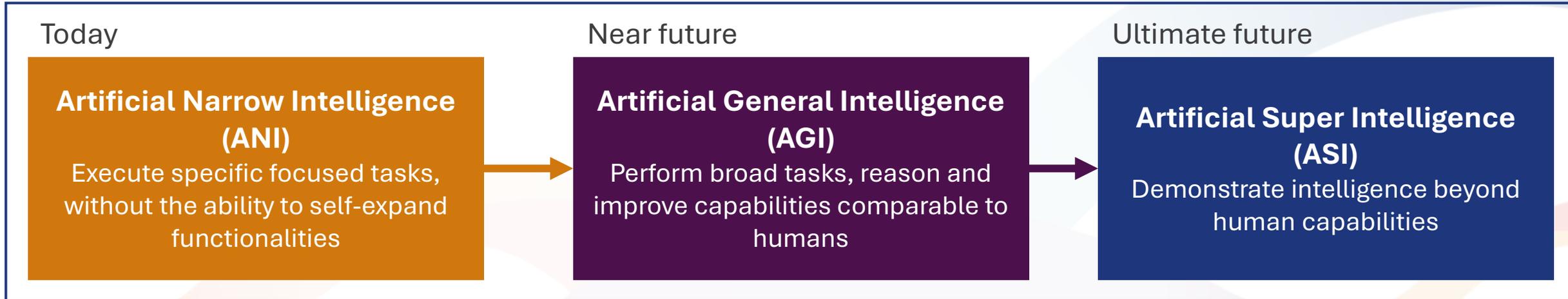


In high-profile meetings, a significant amount of effort is required to ensure quick and accurate responses to questions asked.

Relatedly, general enquiries directed to feedback mailboxes require officers to manually reply.

By utilising a Q&A chatbot, the reply process can be expedited. The chatbot accesses documents, performs web searches, and allows users to interact with the documents, hence streamlining information retrieval.

What's Next for AI?



With ChatGPT being a global sensation, the world now anticipates Artificial General Intelligence (AGI). AGI refers to intelligence that can understand or learn or perform any intelligent task exactly as human beings do. To achieve AGI, AI must apply knowledge across a range of domains and process a large amount of data along with the necessary computational technology. While AI has made great progress in recent years, experts believe achieving AGI is still a long way off since most AI systems today are trained for specific tasks and struggle to adapt to new situations.

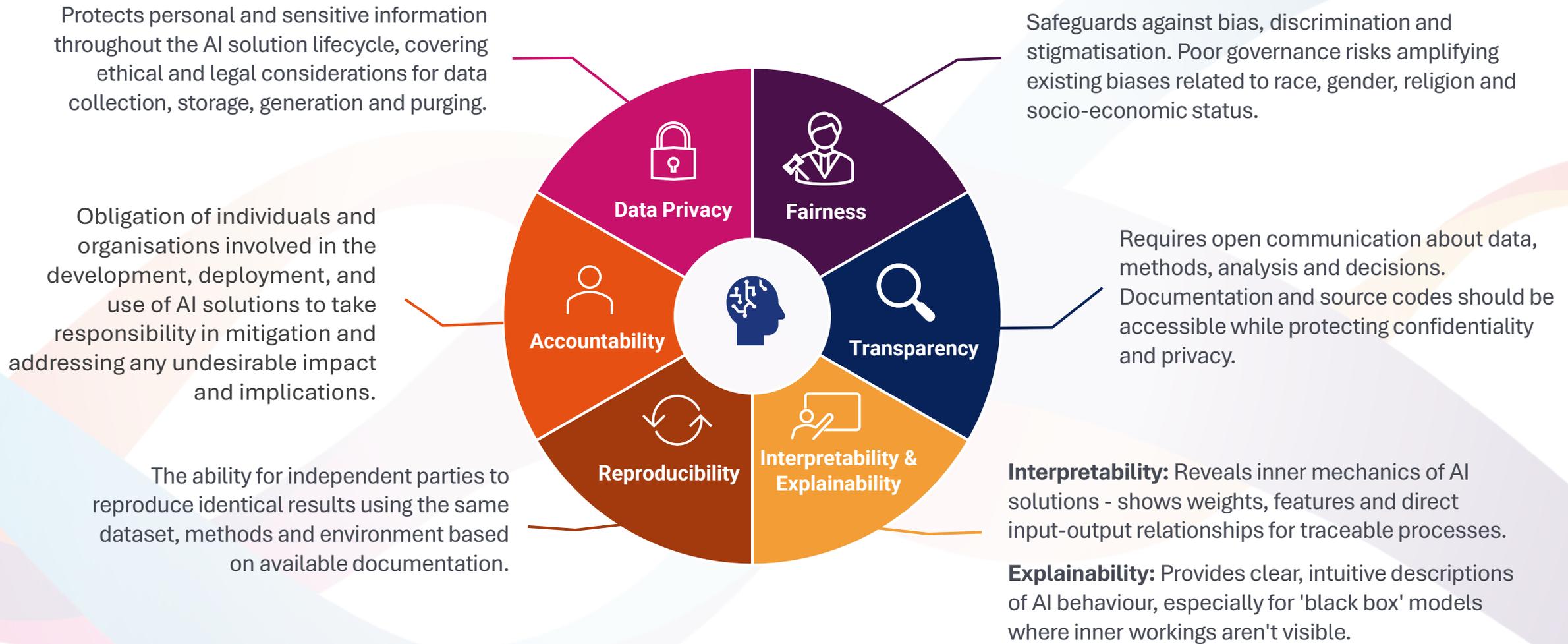
In the next couple of decades, we would likely see the realisation of AGI and eventually, Artificial Super Intelligence (ASI), which may have the potential to transform our world in ways we cannot yet imagine. Nonetheless, in the coming few years, we will surely see wider adoption of AI within the workplace to boost employee productivity and improve business operations.



04. Governance and Safeguards

- **Key Principles in Ensuring Responsible Use of AI**
- **Safeguards within Singapore Government**

Key Principles in Ensuring Responsible Use of AI



Safeguards within Singapore Government

ASEAN Regional Guidelines, Jun 2023 [Source]

Singapore is working with other ASEAN nations to **produce a set of guidelines on responsible use of AI**, which will be released in early 2024. Singapore has been working with industry players to discuss AI standards, best practices and regulations.

AI Verify Foundation, Jun 2023 [Source]

Singapore has launched AI Verify Foundation to harness the collective power and contributions of the global open-source community to **develop AI testing tools for the responsible use of AI**. The Foundation will look to boost AI testing capabilities and assurance to meet the needs of companies and regulators globally.

AI Verify is now available to the open-source community and will benefit the global community by **providing a testing framework and toolkit that is consistent with internationally recognised AI governance principles**, e.g., those from EU, OECD, and Singapore. The AI Verify toolkit provides an integrated interface to generate testing reports that cover different governance principles for an AI system. It enables companies to be more transparent about their AI by sharing these reports with their stakeholders.

Artificial Intelligence Government Cloud Cluster (AGCC), May 2023 [Source]

Singapore partnered with Google Cloud to launch AGCC. Through AGCC, agencies can take advantage of Google Cloud's AI technology stack which includes explainable AI and data governance toolkits – a suite of built-in tools that can help government agencies deploy AI securely and responsibly. This includes features for access control and content moderation, unique mechanisms for incorporating human feedback to improve model performance, and the ability to **audit the sources behind AI models' outputs to detect and resolve potential biases and ensure that model behavior is compliant with regulations**.

A photograph of a library aisle with tall bookshelves filled with books. Several light bulbs are hanging from the ceiling, some of which are lit, creating a warm, ambient glow. The scene is captured from a low angle, looking down the length of the aisle.

05. Learning Resources

- **Reading Materials**
- **Learning Paths by Cloud Service Providers**

Learning Resources (1)

Reading Materials

- [DeepLearning.AI Resources](#)
- [A list of the Campus' Open Source tools • Data Science Campus](#)
- [MAS Veritas Toolkit: Open-source Toolkit for Responsible Use of AI in the Financial Industry](#)
- [AI resources by IMDA](#)

Learning Paths by Cloud Service Providers

- [Google: Machine Learning Engineer Learning Path](#)
- [Google: Introduction to Generative AI Learning Path](#)
- [AWS Skill Builder: Machine Learning](#)
- [AWS Educate: Machine Learning & AI Foundation Courses \[Sign-up required\]](#)
- [Azure: AI, ML and NLP Learning Path](#)

Learning Resources (2)

Internet AI tools (Note: Only for data classification up to Official (Closed)/Non-sensitive)

AI image/video generation tool	Pricing*
Bing Image Creator	Free
MidJourney	Subscription-based
Adobe Firefly	Free account with limited features
Stable Diffusion	Free
Synthesia	Free under Basic plan with limited usage
AI coding assistant tool	Pricing*
GitHub Copilot	Free account with limited usage
Amazon Q Developer	Free for individual tier with limited usage
IBM watsonx Code Assistant	Free under Lite plan with limited features
Multipurpose AI tool	Pricing*
ChatGPT	Free for basic version
Google Gemini	Free for basic version

* Accurate as of 10 Feb

2026. | 33

Thank You

Our Vision

National Statistical Service of Quality, Integrity, and Expertise

Our Mission

We Deliver Insightful Statistics and Trusted Statistical Services that Empower Decision Making



www.singstat.gov.sg



info@singstat.gov.sg



[@SingStat](https://t.me/SingStat)



[@singstat_dos](https://www.instagram.com/singstat_dos)



[@SingStatvideo](https://www.youtube.com/SingStatvideo)



[@sg-department-of-statistics](https://www.linkedin.com/company/sg-department-of-statistics)