



SUPREME AUDIT INSTITUTION OF INDIA
लोकहितार्थं सत्यनिष्ठा
Dedicated to Truth in Public Interest



Compendium on Responsible Artificial Intelligence 2023





Compendium on Responsible Artificial Intelligence



SUPREME AUDIT INSTITUTION OF INDIA
लोकहितार्थ सत्यनिष्ठा
Dedicated to Truth in Public Interest

Comptroller and Auditor General of India
2023



SUPREME AUDIT INSTITUTION OF INDIA
लोकहितार्थं सत्यनिष्ठा
Dedicated to Truth in Public Interest



Contents

Chapter	Subject	Page No
	Foreword by CAG of India	7
	Preface	10
1	Case Studies on Use of Artificial Intelligence in Audit	12
2	Case Studies on Use of Artificial Intelligence in Public Sector	56
3	AI - Tools, Technology and Framework	78
4	Voices	88

Foreword

Artificial intelligence (AI) is rapidly transforming the way we live and work. Digital Transformation and AI driven technology platforms are replacing the repetitive manual processes and bringing in efficiency. Globally, AI has seen a surge in applications due to its immense potential to provide efficient and cost-effective solutions. Many public sector entities have started using AI in the delivery of public services. Though many countries have started regulating AI, they are at various stages of evolving their approach to regulate it. Realising the importance of AI



in delivery of Public Service as well as in everyday life of citizens, various regulations, strategies, policies have been brought out by countries to regulate AI in a responsible manner. For instance, in India NITI Aayog has prepared a National Strategy for Artificial Intelligence – AI for all. European Union has come up with General Data Protection Rules (GDPR) 2016, US has introduced Algorithmic Accountability Act of 2022, Russia has adopted a National Strategy for AI by 2030, China has Next Generation AI Development Plan, UAE has formulated Strategy for AI, etc.

The overwhelming belief at the recently held Seminar on Responsible AI organised by SAI India was that democratisation of AI technologies is inevitable. Today we have reached a level where, AI could contribute up to \$15.7 trillion to the global economy by 2030. It has the potential to lead socio-economic growth and can be used to benefit citizens and the country through targeted and timely interventions. Healthcare, retail, finance, agriculture, food, water resources, environment and pollution, education, special needs, transportation, energy, public safety, disaster management, judiciary, etc. are just a few of the areas that AI has the potential to solve.



While AI offers many opportunities, it also raises concerns related to transparency and fairness. These issues include the impact of AI on privacy, bias and discrimination in AI systems, and inadequate understanding of AI algorithms by the general public. These problems are complex and interconnected, highlighting the need for responsible AI practices, where the fairness of solutions is ensured.

Cybercriminals are also leveraging the same AI systems to create more advanced attacks against the security infrastructure and solutions. The issue of AI-generated 'Deep Fakes' is making it increasingly difficult to curtail propagation of false information. In recent times, ChatGPT (generative pre trained transformer) has emerged as a front runner among the various Open AI tools with more than 100 million users. Necessary caution must be ensured while using such tools.

As organizations increasingly rely on technology to manage their business processes, there is a strong perspective among the auditing fraternity too to review our role in this change and to align ourselves with the changing technological environment. The Supreme Audit Institutions have the mandate to provide independent and objective examination of public services and serve as the much needed feedback mechanism for policymakers in identifying the areas for improvements and opportunities. Hence, SAIs need to adopt a two-pronged approach to integrate AI into their operations, one as auditors of AI systems and the other, as users.

SAIs must adapt their skills and techniques to be able to audit AI algorithms in audits. This will require development of new tools and techniques that can effectively assess the integrity of financial statements and other critical information. SAIs also need to explore ways in which AI can be harnessed into building auditing tools that can improve the quality of our audit, improve our efficiency and help us stay relevant.

It then becomes essential to draw principles for appropriate handling of data, establishing data protection frameworks & sectoral regulatory frameworks and promotion of adoption of international standards. Significant emphasis is being placed on adoption of Audit Data Standards. Technical committee 295 of the International Standards Organisation is working on standards for Audit Data Services.

As we strive for a world of inclusive and sustainable development, and prosperity, SAI20 with its collaborative commitment to an accountable, economic, efficient and effective governance has a crucial role to play in promoting responsible AI in governance. Hence, the SAI20 engagement group this year seeks to promote responsible use of AI technology,

fostering public trust and confidence in its applications and optimizing its true potential.

As chair of the Knowledge Sharing Committee of INTOSAI and the Working Group on IT Audit, I recommend working towards setting guidance for adopting AI in audit and also developing auditing framework for optimising the use of AI in audit and for promoting responsible AI for good governance.

In order to appreciate diverse perspectives on these highly complex areas of AI, SAI India has interacted with the SAI Community domain experts from government and the private sector. We conducted seminars to understand global experiences and initiatives and also how external stakeholders view the role of Audit in these areas. These learnings have fructified in this Compendium of case studies of experience of different SAIs & Government entities. The case studies range from use of AI to automate traditional audit procedures, to developing new approaches to risk assessment and fraud detection.

Each case study offers valuable insights into the strategies and solutions implemented by public sector organizations to overcome challenges and achieve their goals. The real-world examples will help gain a deeper understanding of the potential benefits and risks of using AI, as well as best practices for implementing these technologies effectively.

I strongly believe that this compendium will be a valuable resource for the Auditors and other professionals to leverage AI technologies in their work and stay ahead of the curve in the rapidly evolving field of AI in governance. I would like to express my sincere gratitude to all the SAI20 member countries for their valuable contributions to the compendium. The insights, experiences and lessons learned from these case studies will prove to be informative and useful in developing strategies to address the unique challenges of AI applications.



Girish Chandra Murmu
Comptroller and Auditor General of India



Preface

Welcome to this compendium of case studies, where we explore the strategies, and outcomes of use of Artificial Intelligence in Audit by various SAIs and implementation of Artificial Intelligence by the Government across various countries.

Artificial Intelligence (AI) has been making significant advances in recent years and is being increasingly utilized in various industries, including auditing. AI has also been adopted by various governments for delivery of public services. However, there are two major challenges that effect any AI driven service platform – bias induced due to data and bias induced due to the opaque algorithm which makes AI a 'black box'. Governments have introduced various regulations/strategy to regulate AI to control such biases and improve transparency.

SAIs world over have recognized that engagement with AI is inevitable both in terms of using AI for audit or auditing AI for promoting responsible AI in governance, However, as is with countries, SAIs are also at different stages of maturity in these AI engagements. Responsible AI has been chosen as one of the themes for SAI20 summit this year to share experience and expertise among SAI20 countries to build on the knowledge in this emerging area.

Implementation of AI in the audit process can improve the efficiency and effectiveness of auditing while reducing the risk of errors and fraud in transactions. With the ability to process large amounts of data and to make predictions, AI can be applied to various areas within audit, such as risk assessment, fraud detection and continuous auditing. These AI driven audit processes should be closely monitored to ensure that the auditors use it in a responsible way.

With this vision, SAI India decided to capture footprints of AI both in SAI20 and invited guest countries chair. These have been incorporated in this compendium.

This compendium comprises four Chapters. Chapter-1: Case Studies on Use of Artificial Intelligence in Audit, Chapter-2 : Case Studies on Use of Artificial Intelligence in Public Sector, Chapter-3 : AI - Tools, Technology and Framework and Chapter-4 : Voices.

The case studies provide a unique opportunity to delve into the complexities of real-world situations and understand the nuances of decision-making, problem-solving, and leadership. Analyzing these cases would help the SAI members to gain insights into best practices, learn from successes and failures and apply these lessons to the various Audit assignments in future.

This compendium also contains a diverse range of case studies from various industries, sectors and disciplines, highlighting the breadth and depth of challenges that SAIs and public sector face. We hope that these case studies will inspire critical thinking, foster creativity, spark discussion and debate.

We encourage readers to approach these case studies with an open mind, to consider different perspectives, and to engage in thoughtful reflection. Ultimately, the goal of this compendium is to contribute to a deeper understanding of Artificial Intelligence and explore the means of using Artificial Intelligence responsibly.



Chapter-1

Case Studies on Use of Artificial Intelligence in Audit



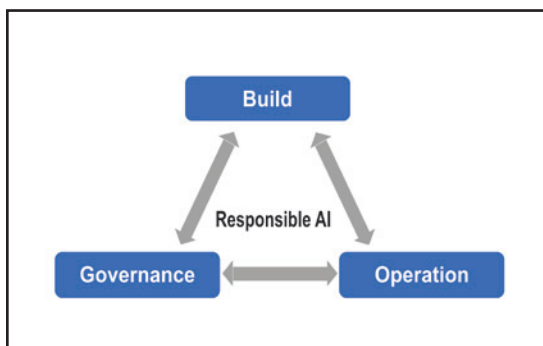
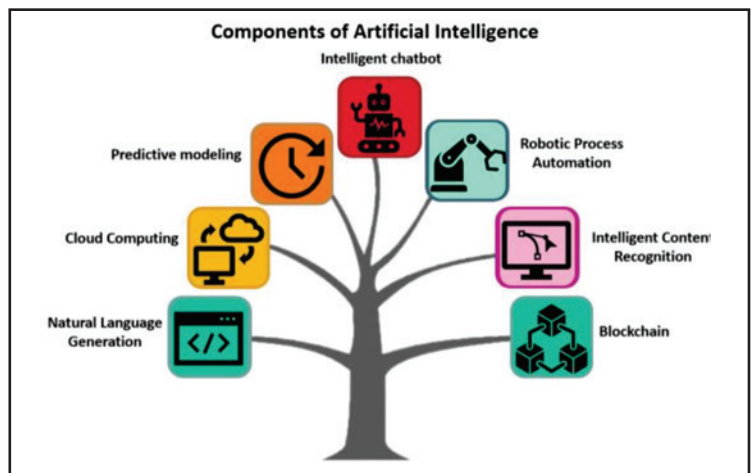
Case studies on Use of Artificial Intelligence in Audit

Introduction

Artificial Intelligence (AI) has been making significant advances in recent years and is being increasingly utilized in various industries, including auditing. The implementation of AI in the audit process can improve the efficiency and effectiveness of auditing while reducing the risk of errors and fraud in transactions.

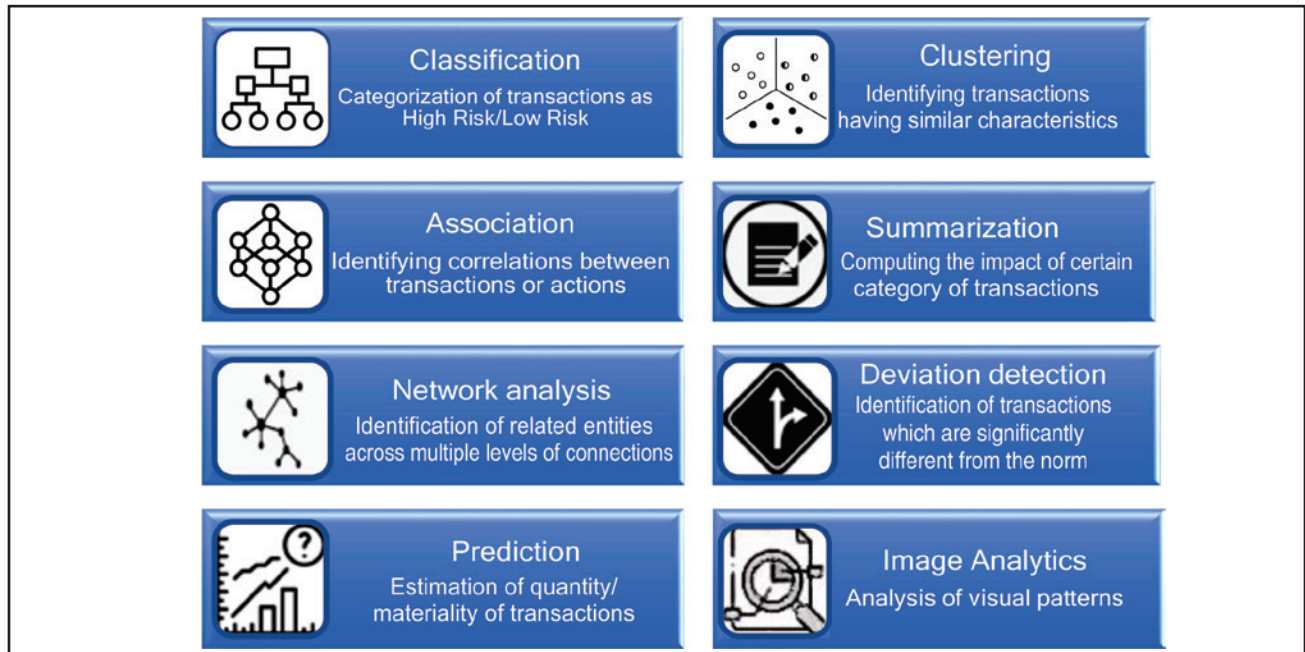
Artificial Intelligence (AI) has the potential to revolutionize the way audit is performed by improving accuracy, reducing manual effort, and increasing efficiency. With the ability to process large amounts of data and to make predictions, AI can be applied to various areas within audit, such as risk assessment, fraud detection and continuous auditing.

AI can be used at each step of the auditing process, removing repetitive tasks common in the process and enabling analysis of large volumes of data to have an in-depth understanding of the business operation. AI based tools in auditing makes detecting high-risk transactions easier. Manual auditing may not detect such risk due to sample population testing, unlike AI technology that allows for full population testing.



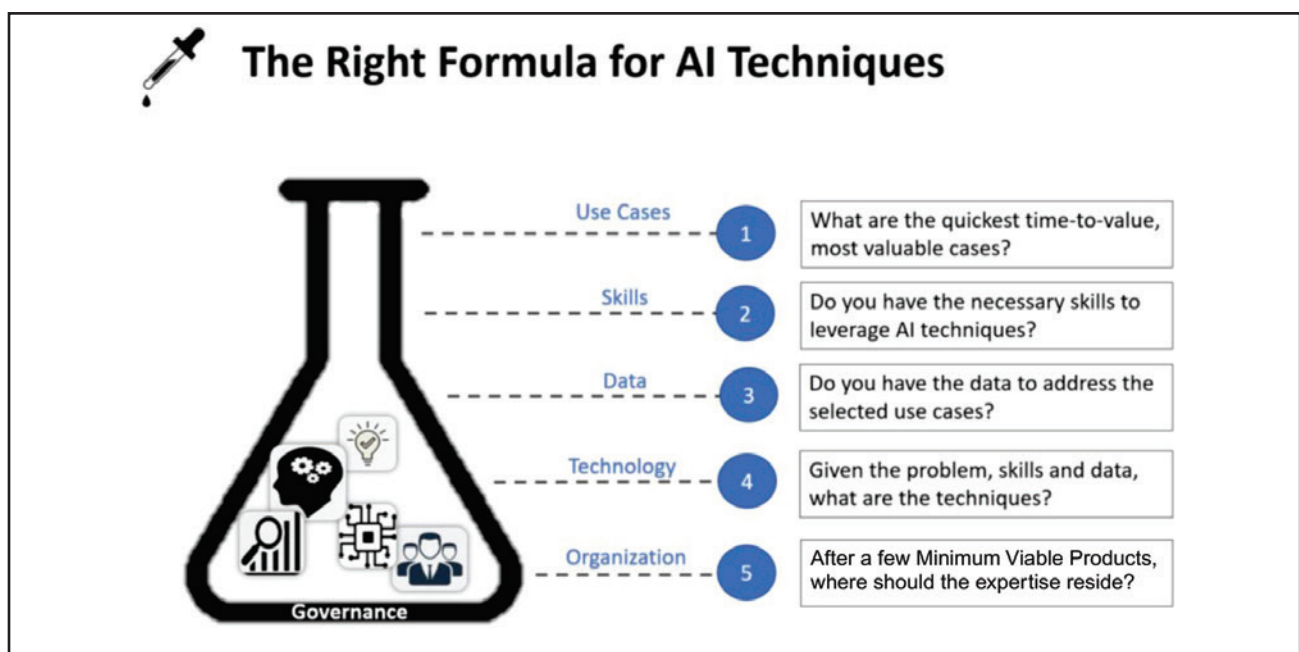
In order to implement AI in SAI, three basic activity groups can be formed for implementation and use of AI in the Audit process as shown in the diagram. The Build team can concentrate on the technological part and team of developers, Governance team can govern the over all project of AI and operation team can work on the day to day operational difficulties faced.

Artificial Intelligence techniques can be used for various analysis on high volume transactions during audit as shown in diagram below :



There are many ways auditors can leverage automation capabilities throughout the audit life cycle, including risk assessments, audit planning, fieldwork, and reporting. It is important to see automation, analytics and AI for what they are: enablers, the same as computers. They will not replace the auditor; rather, they will transform the audit and the auditor's role.

The right formula for using AI Techniques as per Gartner¹ is as shown below :



¹ <https://www.gartner.com/en/newsroom/press-releases/2021-08-04-data-and-analytics-summit-india-day1-conference-highlights>

CHINA

National Audit Office of China

Case Study A1 : Creating AI based Archives and use of OCR

- Nanjing City Audit Office of China's Jiangsu Province have transformed digitalized archives using AI.
- Auditors use optical character recognition- OCR technology to convert paper documents to digital data.

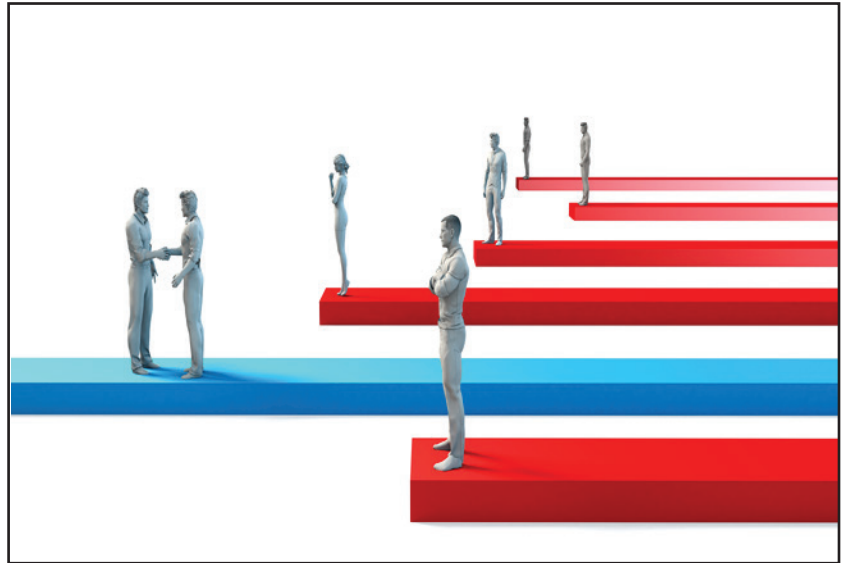


PROBLEM STATEMENT & SOLUTION

1. The auditor processes, refines and transforms the original digital audit archives into standardized data that can be combined and analyzed on demand. This will facilitate the management departments of audit institutions to formulate sound plans, accurately manage audit quality and carry out post evaluation of audit projects, which is conducive to better control of the whole audit process, such as pre-audit investigation, developing plans, audit verification, problem characterization, making audit recommendations and drafting audit reports.



2. Various documents viz. meeting minutes and business documents were converted into electronic data, and use word segmentation and view point extraction technology to analyze all kinds of text information of the entity, construct high-frequency vocabulary and keyword cloud maps, collect articles with similar themes, which helps auditors to focus on the main business of the audited entity and on the audit priorities.



SOLUTION

1. Collecting all data on government bidding and tendering;
2. Using Python to build data mining models to analyze the probability of simultaneous occurrence of bidders. The main applied algorithms include association rule algorithm, such as the Apriori algorithm and FP-growth algorithm, etc.
3. Conducting in-depth research on the correlation between suspicious enterprises and officials to identify collusion issues by using Graph database.

CONCLUSION

We continuously adjust the model parameters through verification of the analysis results, in order to detect abnormalities better. Through data mining, we successfully discovered illegal cases such as bidders and bid-invitees act in collusion with each other, involving enterprises and officials.

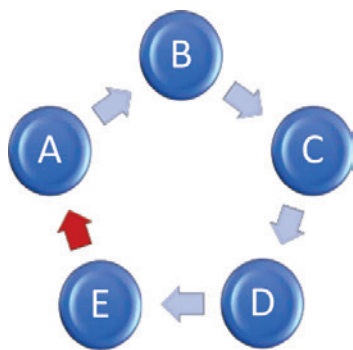
INDIA

Comptroller and Auditor General of India

Case Study A3 :

Use of AI in Identifying Circular Trading transactions in Taxation

Circular Trading is normally used for creating a flow of fake sales transactions. These fake invoices are used for claiming input tax credit in the taxation which helps the companies to inflate their turnover, gain larger loans from financial institutions and avail tax credits on every lap of transactions done, causing loss of money to the exchequer.



Circular Trading

PROBLEM STATEMENT

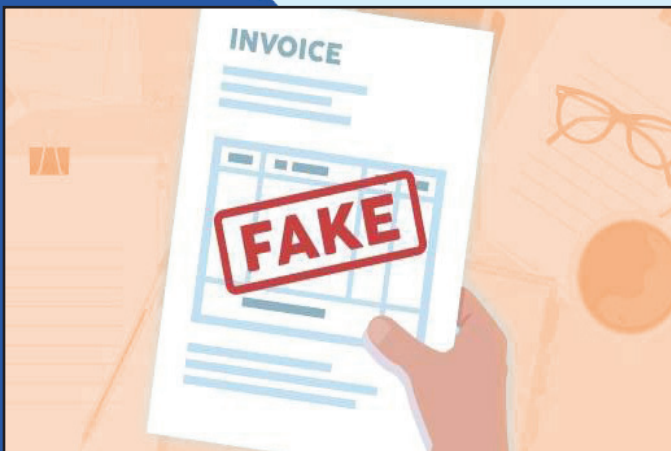
Circular trading normally is used to issue fake invoices in transactions among multiple parties without actual supply of goods.

The mapping shows that this problem requires advanced analysis methods and techniques to detect such cases.

SOLUTION

This problem is addressed by developing an algorithm that detects circular trading transactions to further verify the illegitimate cycles to uncover the suspicious transactions.

Using Artificial Intelligence Algorithms, specific types of circular transactions up to 8 iterations were identified. The model was trained and tested on the selected E-Way Bill data set related to Taxation in India and several circular trading transactions/ patterns were discovered.



CONCLUSION

Circular Trading needs to be established further by evaluating the type, quantity and value of Goods transacted through circular trading.

Using AI, the transactions that are forming circular trading can be flagged as high risk for further field level verification in detail to evaluate the effects of the circular trading i.e. fraud, excess input tax credit, fake invoices, inflated sales, etc.

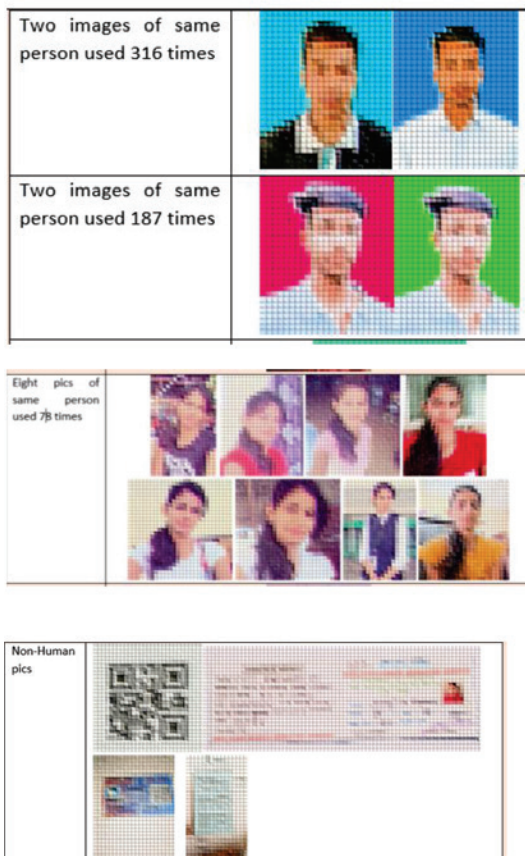
INDIA

Comptroller and Auditor General of India

Case Study A4 :

Use of AI in Detecting similar images used by multiple applicants

The data related to Digital Literacy programme by Government of India was analysed. Images of the beneficiaries of the scheme was analysed. The cases detected helped in identifying the risky transactions, duplicate beneficiaries, fake and ineligible beneficiaries.

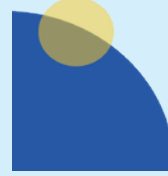


PROBLEM STATEMENT

The beneficiaries under the scheme were spread across India and the training were imparted by designated training centers. The photo of the beneficiaries needs to be uploaded for each training conducted. The analysis of the photo images cannot be done manually and an intelligent model needs to be implemented to detect various anomalies in the images.

SOLUTION

The model was developed using open source platform – Python. Different in built



libraries were used to detect the following cases :

- (1) Same Images used for claiming the training cost.
- (2) Different images of same beneficiaries used for claiming the training cost.
- (3) Non-human images used for beneficiaries.

The solution helped in detecting such anomalies which were not possible manually.

CONCLUSION

The cases detected helped in identifying the risky transactions, duplicate beneficiaries, fake and ineligible beneficiaries.

INDIA

Comptroller and Auditor General of India

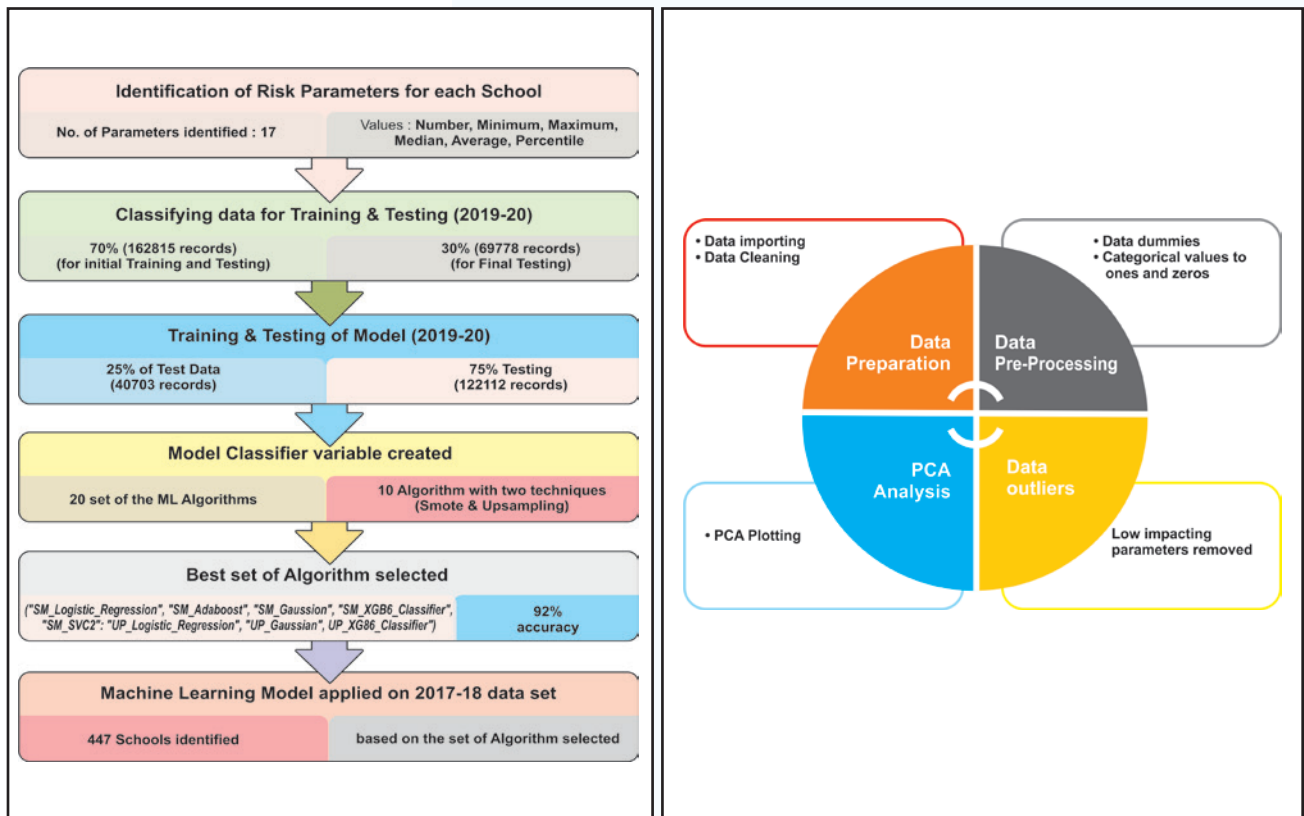
Case Study A5 :

Use of ML/AI to detect non-existent schools claiming scholarship benefit

Under the National Scholarship Scheme, various schemes are operated for the students belonging to low income group. The scholarship is paid to students after validation by the respective recognised schools where the students are enrolled for scholarship.

PROBLEM STATEMENT

One of the main objective of the audit of the scholarship data was to assess the genuineness of the students and the schools through which the scholarship is claimed, and identify high risk/ suspected fake schools. The huge data size and geographical distribution of the schools across the country meant that it was not possible to physically verify all the existing schools. Hence, in order to achieve the above objective, a model was required to detect suspected fake schools based on risk parameters.



SOLUTION

Machine Learning model was developed in Python to identify suspected fake schools which claims scholarship in the year 2017-18, based on pre-defined risk parameters identified from the data pertaining to the year 2019-20.

A total of 17 parameters at school/ institute level were identified for the model. A set of 10 different Machine Learning Algorithms was used with two different techniques (SMOTE and UPSAMPLING).

CONCLUSION

The model achieved above 92% accuracy. This helped in identifying the risky samples for the field level verification.

INDIA

Comptroller and Auditor General of India

Case Study A6 : Use of AI to read watermark data from Images

Rural housing scheme is being implemented by Government of India. It require capturing of the photos of house/land with or without beneficiaries at different stages of the scheme. These photos must be captured by an application designed by Government. Audit attempted to find duplicacy/irregularities in the images using AI.

PROBLEM STATEMENT

Cases were noticed where Photos were having watermark text/datetime in the database. This can be because of following reasons:

- Photos are uploaded from gallery against the guidelines of the scheme.
- The users may be capturing the photos from other application.
- It may involve underlying fraud to bypass scheme guidelines or system flaws.

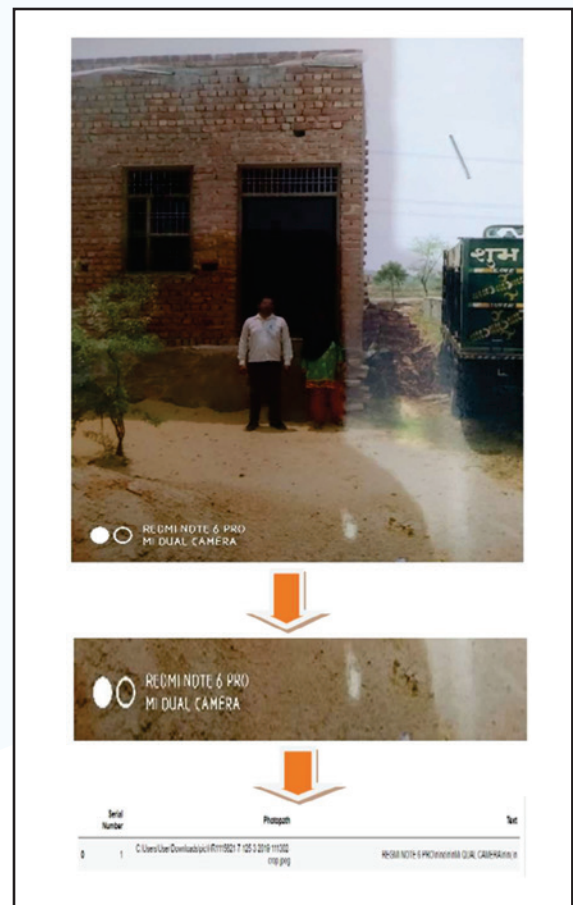
SOLUTION

Using open-source libraries of python implementing OCR technology, the watermark data was converted to text data for further analysis.

After removing Noise and background data, the enhanced and cropped image data for each image was processed to find watermark text using pytesseract module of python and stored in the data frame to identify the photos with watermark text.

CONCLUSION

The Model was able to detect text using OCR from images and classify millions of photos having watermark text. It has scope of further augmentation to better enhancement and accuracy of the watermark text.



INDIA

Comptroller and Auditor General of India

Case Study A7 :

Use of AI to detect Students claiming scholarship under ineligible quota

National Scholarship Scheme is a flagship scheme of the Government of India for students belonging to marginalised communities. Empowerment through education, which is one of the objectives of this scheme, has the potential to uplift the socio-economic conditions of the marginalised communities.

PROBLEM STATEMENT

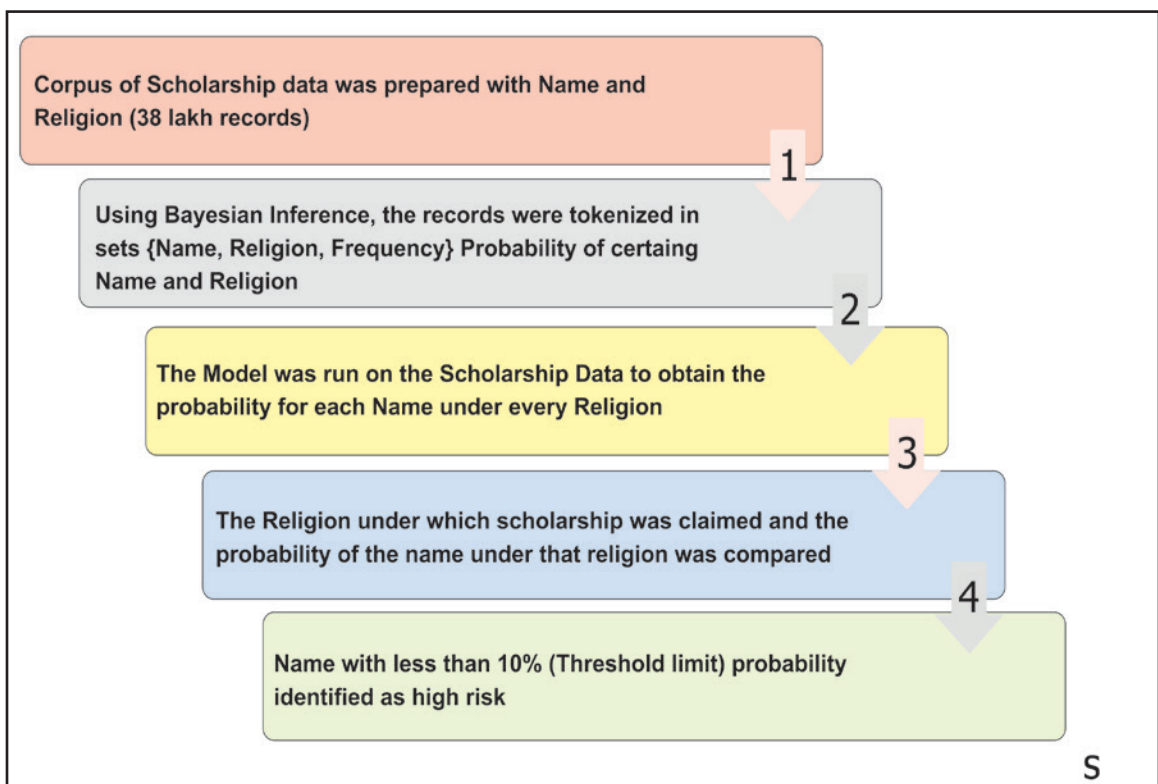
Cases were noticed where ineligible beneficiaries were claiming scholarship in Scholarship scheme for marginalised communities.

SOLUTION

A model capable of predicting the community of the beneficiary using the name was developed using python. The model was based on the Bayesian Inference Algorithm.

CONCLUSION

- The Model helped in detecting fraudulent cases.
- Those cases with threshold having less than 10% probability under a specific quota were identified as outlier for further analysis.



INDIA

Comptroller and Auditor General of India

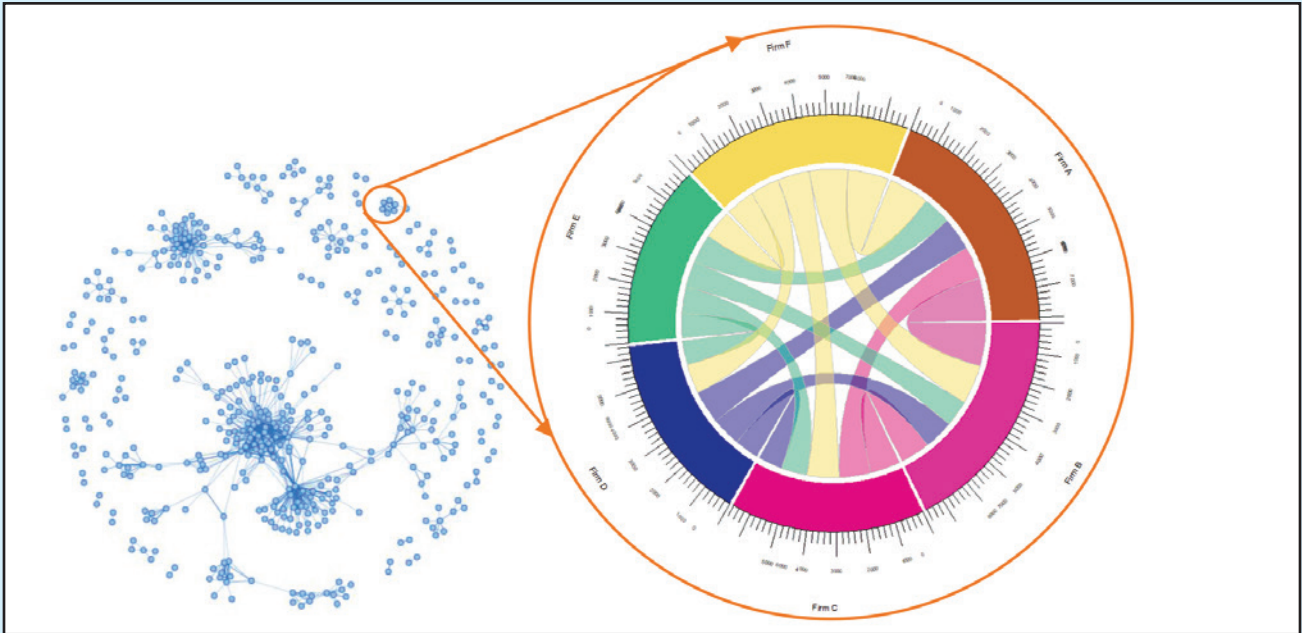
Case Study A8 :

Detection of collusion in the Government Procurement System using Graph Theory

Government procurement is done through the E Tender Portal wherein the bidders register on the portal for submitting their bids for various tenders. During the audit of the tendering process, the data of the bidders and bids processed under various tenders were analysed and it was noticed that various bidders were bidding in collusion, resulting in anti-competitive practices impairing the transparency of the bidding process. Graph algorithm was used to detect such collusion cases.

PROBLEM STATEMENT

Code of integrity for Public Procurement stipulates that bidders should not be engaged in any collusion, bid-rigging and anti-competitive acts or any other practice coming under the purview of the Competition Act, 2002 that may impair the transparency, fairness and the progress of the procurement process or to establish bid prices at artificial, non-competitive levels. Thus, one of the objectives of data analytics was to ensure that any of such anti-competitive practice has not been adopted by the identified network of bidders.



One of the challenges to identify network of bidders who may bid in collusion was that data may grow exponentially by forming combinations of bidders.

SOLUTION

The solution adopted was an integrated approach in R programming using 'igraph' package¹. The bid data was converted into graph with nodes and edges. Further, required weights were also allocated to these edges according to the number of tenders in which the bidders has participated together. Plotting this graph object we were able to detect connections between each of the bidders in the procurement system.

Further as part of case studies to be analysed in detail, we found complete subgraphs² in the igraph object.

CONCLUSION

Network Analysis helped in processing huge amounts of data to detect collusion among bidders, which was not possible through traditional data analytics approach.

¹ https://www.researchgate.net/publication/221995787_The_igraph_Software_Package_for_Complex_Network_Research

² A complete subgraph is a subgraph in which all the nodes (or verices) are connected with each other.

INDONESIA

The Audit Board of Republic of Indonesia
Badan Pemeriksa Keuangan Republik Indonesia

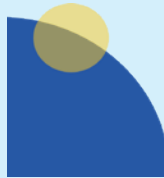
Case Study A9 : Detecting anomalies in Government spending using AI

The use of Big Data Analytics in identifying anomalies in Government spending trends including procurements, social aid fund, and salary costs. SAI of Indonesia has considerable experience developing Big Data analytics for financial audits, such as identifying anomalies in government procurement processes and government spending behaviour.

PROBLEM STATEMENT & SOLUTION

1. Although the SAI is still in its initial stage of experimenting and innovating with AI, it has applied AI in creating government supplier profiles to support financial audits. Furthermore, it has developed AI-powered analytics to identify alleged frauds in payments of government subsidies to unqualified recipients during the Covid pandemic. The use of web surveys such as LimeSurvey and Survey Monkey in Performance and Financial Audit in collecting information and data from populations and creating the results, not a sample.

- Sentiment analysis of the SAI's social media posts. The opinion mining technique was used to measure whether social media users respond positively to specific issues posted by SAI.



INDONESIA

The Audit Board of Republic of Indonesia
Badan Pemeriksa Keuangan Republik Indonesia

Case Study A10 : Use of AI in Auditing Government Procurement

The Government of Indonesia requires all government agencies to use standardized procurement applications. However, with over 600 e-procurement portals hosted separately, identifying anomalies, irregularities, and bid-rigging possibilities in procurement processes presents challenges to auditors. To address this, the Big Data Analytics (BIDICS) team developed an AI system using a rule-based system and association rules to identify anomalies, irregularities, and bid-rigging possibilities. As the source information is publicly available, no ethical measures or access limitations are necessary, but auditors should communicate the AI result to the auditee as audit evidence.

PROBLEM STATEMENT

Indonesian governmental agencies must manage their procurement processes using a standardized application on locally-installed portals that allows any company to participate in any procurement. As a result, auditors encounter difficulty to examine procurement data distributed at more than 600 portals and identifying whether there exist anomalies, irregularities, and bid rigging.

SOLUTION

Initially, the Big Data Analytics (BIDICS) Team created a program to collect procurement data from over 600 government portals and store it in a single database. Subsequently, the team developed and implemented a rule-based AI system on the BIDICS Platform using Association Rules to identify potential Bid-Rigging in all government procurement processes. As government procurement information is publicly available, no access restrictions or ethical considerations are necessary. However, according to the standard, the auditee must be informed of the AI results to ensure they can be considered audit evidence.

CONCLUSION

The use of AI solutions has helped auditors identify potential issues and discrepancies in the preliminary stage, before the actual field audit. In return, this allows auditors to verify and confirm the preliminary indications during the field audit phase, which may result in audit findings.



INDONESIA

The Audit Board of Republic of Indonesia
Badan Pemeriksa Keuangan Republik Indonesia

Case Study A11 : Use of AI in Auditing Government Social Assistance Fund

The Indonesian government has allocated funds for social assistance programs, but strict rules prohibit individuals from receiving multiple programs. To identify any violations, auditors have requested preliminary information on the distribution of funds to recipients. A Big Data Analytics (BIDICS) team has developed an AI system by rule-based system to analyze the data. The AI result is presented on the BIDICS dashboard under the Social Assistance Program Audit Package. As for the AI result, security measures are in place to limit access to authorized individuals due to the classified nature of the information.

PROBLEM STATEMENT

The Indonesian government has set aside funds for social assistance programs such as Direct Cash Assistance, Family Hope Program, Pre-Employment Program, and Basic Food Support Program. In order to ensure fairness, strict rules prevent individuals from receiving multiple assistance programs. However, because there is a large number of recipients, auditors experience difficulty in identifying any violations in the distribution of funds. To address this, auditors have requested preliminary information

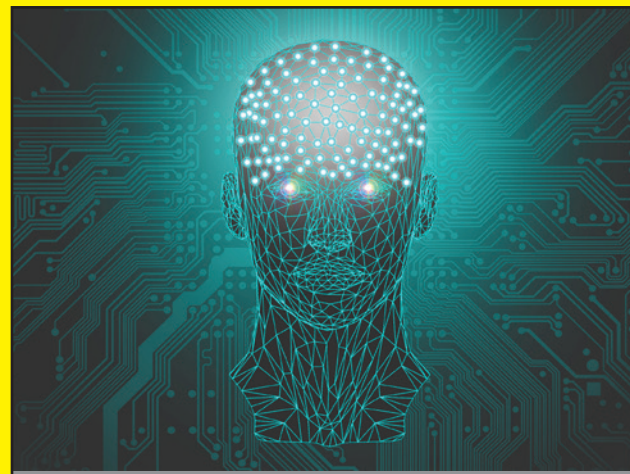
on the distribution of funds to recipients, including their name, address, profession, and other relevant information that could potentially violate the distribution criteria and rules. The auditors will use this information to create an audit sample for further investigation.

SOLUTION

The BIDICS team developed an ETL process that combines data collected from various sources by auditors into a single large database. The team then applied a rule-based system to the BIDICS Platform as an AI implementation, embedding criteria, terms, and conditions into the platform. The results of the analysis are presented visually through several visualizations on the BIDICS Dashboard, specifically under the Social Assistance Program Audit Package. However, as the information related to fund distribution is classified, access to analytics results is restricted only to authorized individuals with official audit assignments. Additional security measures are also in place to prevent illegal access.

CONCLUSION

Auditors are now able to utilize AI solutions to identify cases where individuals are receiving more than one assistance funds, which is not allowed. This preliminary information helps auditors conduct their audits more efficiently and effectively by informing their selection of interview subjects for further investigation.



MEXICO

Superior Auditor of the Federation
Auditoría Superior de la Federación

Case Study A12 : “Machine learning” Applied to defining audit sampling in social programs

SAI Mexico conduct financial and compliance audits to different social programs. These programs are distributed to millions of beneficiaries nationwide. To leverage audit resources, it is required to determine a sample audit to visit beneficiaries to identify possible program deviations.

Applying cluster analysis techniques with multiple variables, it was possible to determine the sample for the audit.

PROBLEM STATEMENT

Social programs are distributed to millions of beneficiaries nationwide in 2,474 municipalities in the 32 Federal Entities of Mexico.

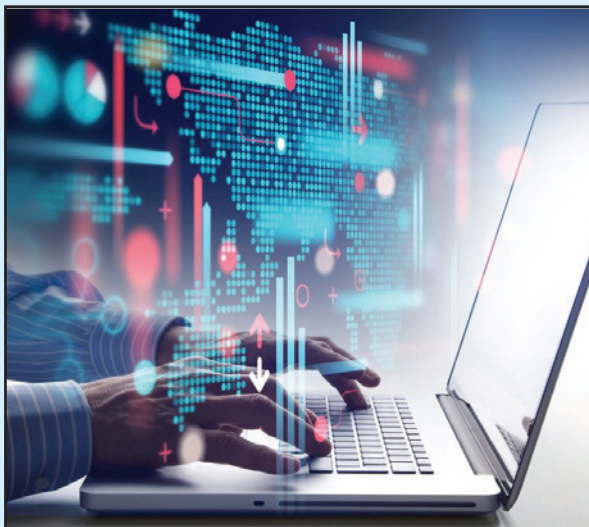
Information provided by the entities in charge of social program distributions consist of multiple variables such as data of the beneficiaries, payments, locations, and program rules.

To leverage the resources (human and monetary) assigned to audit social programs, the definition of the sample of beneficiaries to visit takes significant relevance to identify possible program deviations.

SOLUTION

To define audit sample for social programs SAI Mexico developed the following method:

First Data quality. applying data profiling and quality tool to detect errors, invalid information, duplicate registers in the information of social programs.



Data cluster analysis with K-means method considering variables to define:

- First Federal Entity.
- Second a specific analysis to identify municipalities to visit.
- Third selection of beneficiaries in the municipalities.

CONCLUSION

Machine learning techniques such as cluster analysis provides the possibility to handle multi variable analysis that social programs require, giving accurate and specific results that facilitate the selection of the municipalities and beneficiaries to visit during the audit activities.

Also open the opportunity to make cross information analysis between different social programs.



NETHERLANDS

Netherlands Court of Audit

Case Study A13 : Audit of AI Algorithm

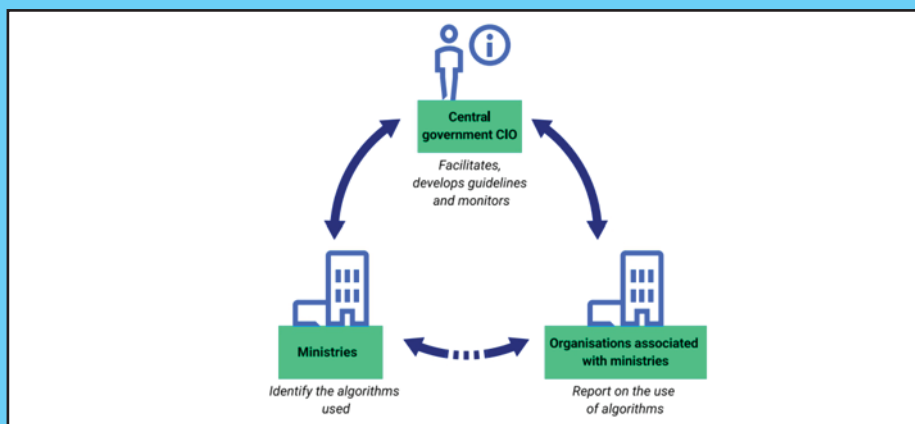
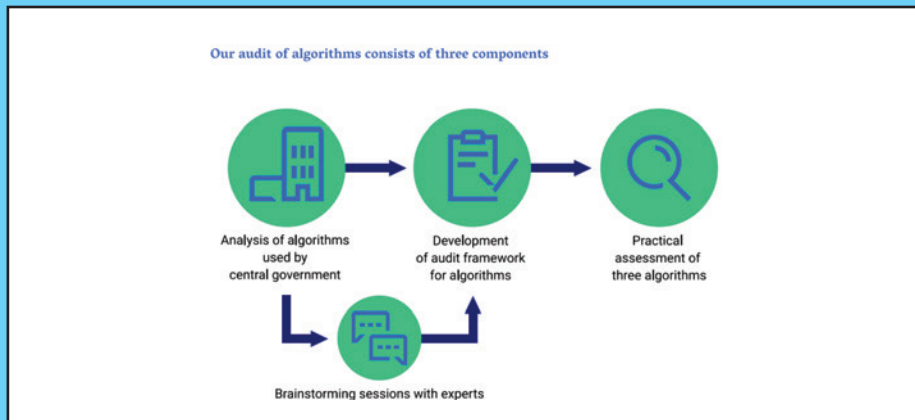
The Netherlands Court of Audit (NCA) has investigated the Dutch government's use of algorithms in 2020.

PROBLEM STATEMENT

The government, like private companies, is making increasing use of algorithms to automate its work, solve problems and make forecasts. The responsible development of more complex automated uses requires better scrutiny than ministers currently exercise and better-quality control.

SOLUTION

We found that about one third of the predictive and prescriptive algorithms listed by the ministries use automated decision-making. Our analysis did not identify any fully self-learning algorithms in central government, only learning ones. Automated decision-making is used only by algorithms that perform simple administrative tasks that have no effect on private citizens. But there is scope for improvement as the algorithms become more complex. To address the significant gaps in the transparency, explainability, and accountability of these AI systems, the NCA developed an audit framework in the absence of a specific auditing tool for AI systems which also considers ethical aspects of algorithms.



Source: <https://english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms>

CONCLUSION

Our audit revealed that three of the nine algorithms that we assessed satisfy the requirements of our audit framework. This shows that algorithms can be used both fairly and judiciously. We also found that six of the nine algorithms tested do not comply with all the requirements set out in our audit framework. There is great scope for improvement here and the three main improvement areas are (1) Clear agreements for checking that Algorithm was operating in accordance with its intended purpose, (2) IT controls for safety of data and (3) Checking for Bias in the Algorithm. Along with the audit framework used for auditing the algorithms, the NCA made several recommendations to the Dutch government to improve the transparency, explainability, and accountability of AI systems used in public services.

NIGERIA

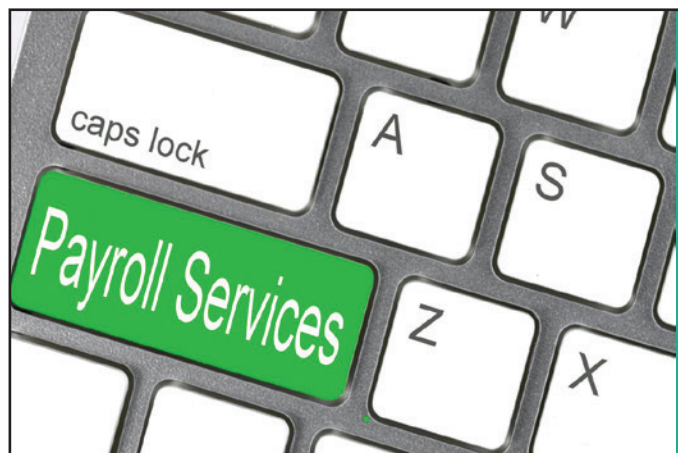
Office of the Auditor General for
the Federation, Nigeria

Case Study A14 : IT Audit of Integrated Payroll and Personnel Information System

The IPPIS project was conceived as an ICT solution to centralize the payment processes for Federal Government employees, and to store and maintain their personal information.

ISSUES IDENTIFIED

- i. Higher cost to government
- ii. Poor/lack of return on investment
- iii. Loss of privacy and confidentiality of sensitive employee data due to un-restricted access by vendors and other third parties to these data





OBSERVATIONS RAISED

The following major objections were raised during the audit :

1. Ineffective governance, project and operations management of the IPPIS
2. Data Input and Integrity Issues
3. Improper System development and Configuration
4. Absence of HR policy
5. URL not owned by Federal Government
6. Inadequate Business Continuity and Disaster Recovery Plan
7. Unauthorised changes to critical data of employees.
8. Outdated software version

Oman

State Audit Institution, Oman

Case Study A15 : Digital Technologies in combating corruption

Digital technologies have had a significant impact on the development and implementation of anti-corruption practices and measures.

It provides new tools and approaches for combating corruption. These technologies have made it possible to increase transparency and accountability, improve reporting mechanisms, and reduce the opportunities for corruption in government and business practices.

PROBLEM STATEMENT & SOLUTION

Some ways in which digital technologies have been used to combat corruption are :

1. **Transparency and Accountability:** Open Data and Big Data project coming in Oman for receiving the complaints of citizens and residents against for the Gov Services that are not meeting the expectations.



2. **Whistleblowing:** Individuals report cases of corruption anonymously and securely. SAI Oman has been receiving the complaints through online portal.
3. **E-procurement:** Electronic procurement systems allow for greater transparency in government procurement processes, reducing the opportunities for corruption. These systems allow for open bidding, automated evaluations, and greater tracking and monitoring of procurement activities.
4. **Data Analytics:** Advanced data analytics tools can help to identify patterns and anomalies in government data, making it easier to detect potential cases of corruption. SAI Oman we have a good practices with ACL tool.
5. **Digital Payments:** The adoption of digital payments can reduce the opportunities for corruption in the financial sector. By reducing the need for cash transactions, digital payments can help to eliminate the use of bribes or other forms of illicit payments. The Government in Oman is enforcing now all commercial and government service to use e-payments.



Republic of Korea

Board of Audit and Inspection

Case Study A16 :

BAI Audit of an AI-based System: Audit of KEIS's Job-Matching System TheWork AI

This Audit put forward recommendations to improve the job recommendation criteria of the 'Job-Matching System', which uses the DeepFM model

PROBLEM STATEMENT

The Korea Employment Information Service (KEIS) has developed and is operating 'TheWork AI', a 'Job-Matching System' which learns job application data through the DeepFM model and recommends job postings to applicants through its database. However, the actual job application rate (matching rate) of the 'Job-Matching System' was relatively low- not only when compared to the predictions that were made previously for it, but also lower than the matching rate of the existing big data-based job-recommendation system, titled the TheWork. Hence, the information being recommended by the matching system was not being utilized to the extent that was expected.

SOLUTION

To determine why the information being recommended by the matching system was not being utilized to the expected level, we analyzed using the following methods based on the data of the number of job postings recommended by the matching system (total 1,676,634 cases) and job application records (total 206,262 cases):

- (1) Correlation analysis: The point-biserial correlation was used to derive the correlation coefficient to check whether there is a correlation between the matching score, which is the likelihood of a job seeker

applying for a job posting recommended by the matching system, and the application for employment.

- (2) Analysis of recommendation criteria (thresholding value): To check whether the matching score criteria for recommending jobs to job seekers is appropriate, we extracted data of those recommendations which were given to users despite having low matching scores and calculated the job application rate.
- (3) Criteria for selecting recommended work district: To analyze the distance between the job seeker's desired work district and the district of the job posting recommended by the matching system, we selected a specific district as a sample and analyzed cases where no job applications were made for job postings viewed for that area.

CONCLUSION

The results of the analysis through the above three methods are as follows:

- (1) The correlation between the matching score, which is the result of the algorithm of the matching system, and the application for employment was low, and considering this point, it was necessary to optimize the algorithm through various simulations by adjusting the composition of AI learning data.
- (2) It was necessary to prepare measures to adjust the threshold value so as not to recommend job information from the matching system when the matching score is below a certain range, to reduce the search cost of job seekers and increase the application rate.
- (3) It was confirmed that the algorithm is set to expand the range of recommended job areas too broadly compared to the job seeker's desired area, and it was necessary to prepare measures that can increase the application rate by limiting the physical distance between the desired district of work and the recommended districts to a certain level.

We pointed out these problems and recommended that the KEIS improve the Job-Matching System. In response, KEIS stated that they would review the recommendations and improve the system to increase the matching rate who found their desired workplaces through the system.

Republic of Korea

Board of Audit and Inspection

Case Study A17 :

Use of AI in BAI's Audit: Welfare Blind Spot Identification System

BAI requested SSiS to fix problems in setting variables of the Welfare Blind Spot Identification Management System which uses the XGBoost model.

PROBLEM STATEMENT

To efficiently identify and support recipients of social security services, The Social Security Information Service (SSiS) is currently operating the 'Welfare Blind Spot Identification System (WBSIS).' The WBSIS enables the SSiS to decide on 48 identification variables such as income data, using a wide range of information on risk signals, and to predict the probability of risk for people (subjects for analysis) who fall into 27 types of risk signals through the XGBoost Algorithm-based risk probability predictive model. In this way, the WBSIS allows for identifying people (targets for identification) potentially at high risk.

However, as we learned that the WBSIS disproportionately advantages people who have received basic livelihood benefits in selecting recipients of social security services, we deemed it necessary to analyze how past records pertaining to receiving of welfare services affect the WBSIS's predictions and verify if the System effectively identifies households who are in situations of risk.

SOLUTION

To see if the WBSIS identifies households at risk effectively, we analyzed the system in the following ways:

- (1) Based on the information on risk signals collected from 4,217,826 people subject to analysis, we analyzed and compared potential recipients according to whether or not the history of receiving welfare services was reflected.
- (2) And we verified if any data was mistakenly left out of the information on risk signals during the data analysis process.

CONCLUSION

The results of the analysis using the above three methods are as follows:

- (1) As the past record of receiving welfare services is of great significance (24.5%) among the 48 identification variables, it is functioning as the variable causing the disproportionate advantage in receiving welfare services. We realized that the identification system needs to be fixed because it is hard for people without a history of receiving welfare services to be selected as recipients even though their livelihoods are at risk in reality.
- (2) By comparing the resident registration information to risk signal data, we found that 1,015,546 people were not considered as subjects to analysis. To see whether these missed people were identified as recipients, we analyzed their risk probability. As a result, it was found that 3,704 people were not selected as potential recipients because their related risk signal data was not reflected. As such there is a need to utilize not only risk signal data but also resident registration information so that omission of eligible people is prevented in advance.

BAI pointed out the above problems and recommended fixing such loopholes to the audited institution, who replied that they will accept the recommendations and improve the system thoroughly.

RUSSIA

Accounts Chamber of the Russian Federation

Case Study A18 : Digitalization and Standardization of Government records for use of AI

The Accounts Chamber has completed a two-year analytical project on the prospects of public administration digitalization in Russia, the key subject of which was the availability of reliable and necessary information stored in state resources. The Accounts Chamber has analyzed a set of documentation from more than 25 thousand files.

PROBLEM STATEMENT & SOLUTION

Within the framework of the event, a number of factors that affect the possibility of making managerial decisions based on data had to be analyzed:

- quality of the description and structure of data;
- the degree of systematization and harmonization of data compositions and structures.



RUSSIA

Accounts Chamber of the Russian Federation

Case Study A19 : Auto detection of violations using AI

Artificial intelligence is used in various process in procurement for detecting violations; checking the timing of entry into the budget list by the chief administrators of budget funds; checking the timing of entry into budget estimates; checking payments to WWII veterans and widows of WWII veterans.

PROBLEM STATEMENT & SOLUTION

Artificial intelligence is used in the process of searching for procurement violations; checking the timing of entry into the budget list by the chief administrators of budget funds; checking the timing of entry into budget estimates; checking payments to WWII veterans and widows of WWII veterans; controlling the accuracy of budget planning; documentary checks of log orders and payment orders; comparison of





standards and accounting data; analysis of anti-corruption activity plans; extraction of performance audit criteria from reports of international SAIs.

Among the used tools: PolyAnalyst, JupyterLab, Label Studio.

Employees of the Accounts Chamber managed to develop a script in Python, which automatically downloads scans of state contracts from the Unified Information System (UIS), after which PolyAnalyst turns them into text files, and AI detects signs of violations.



Currently, the employees managed to test the model on a sample and the result was 88% accuracy.

Further plans include the refinement of the system to 100% accuracy of the output results.

TÜRKIYE

Turkish Court of Accounts

Case Study A20 : Implementation of a Business Intelligence Software - VERA

Turkish Court of Accounts use VERA which is TCA's data analysis system. It is a Business Intelligence Software as one of the big data tools. In TCA, there is Data Analysis Group which was formed in 2017 and has the mission of designing a methodology for using computerized audit systems and improve the implementation of CAATs into the auditing process.

IMPLEMENTATION OF VERA

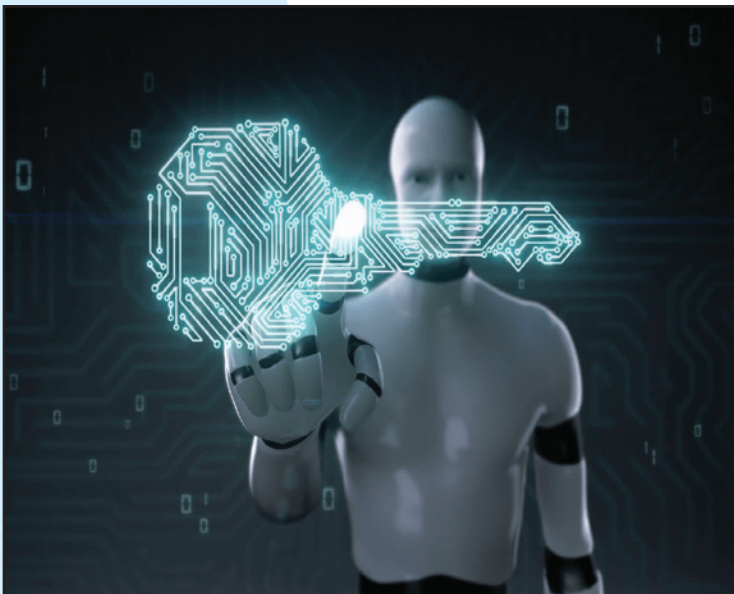
VERA is a web based system and the data is being kept in central database. The auditors are accessing the data via the system independent of their current location by using VPN (secure internet connection).

The data of the Administrations within the Scope of the Central Government, Social Security Institutions, Municipalities and their Affiliates, Special Provincial Administrations and Departments of Investment Monitoring and Coordination , Development Agencies, Revolving Funds, Companies, and Other Public Administrations are collected and put into TCA's data warehouse.

In 2022, auditors could access approximately 250 central government institutions' data via VERA. In 2022, approximately 1100 municipalities uploaded their data to the TCA database by using web portal. Also in 2022, auditors could access 81 Special Provincial Administrations' and Departments of Investment Monitoring and Coordination data via VERA.

Furthermore, The Turkish Statistical Institute, Ministry of Treasury and Finance and the Central Bank of the Republic of Türkiye publish some data open to the public. TCA collects some of these data for instance demographic data in order to use in analyzing accounting data, financial data, payroll data and other data.

With VERA, auditors examine big amounts of data thoroughly, identify mistakes or errors in data processing on an earlier step, thus the system enables automation of analysis and continuous monitoring.





Chapter-2

Case Studies on Use of Artificial Intelligence in Public Sector



Case Studies on Use of Artificial Intelligence in Public Sector

Introduction

The use of Artificial Intelligence (AI) has the potential to transform the public sector by enabling faster, more efficient, and more effective delivery of services. AI can be used in a variety of ways, from automating routine tasks to predicting and preventing future events. As governments around the world seek to improve the delivery of public services and increase citizen engagement, many are turning to AI as a solution.

In the public sector, AI can be applied to a wide range of areas, such as healthcare, transportation, education, public safety, and social services. For example, AI can help healthcare providers diagnose diseases more accurately, predict and prevent traffic congestion, personalize learning experiences for students, and identify patterns of criminal behavior to prevent crime.

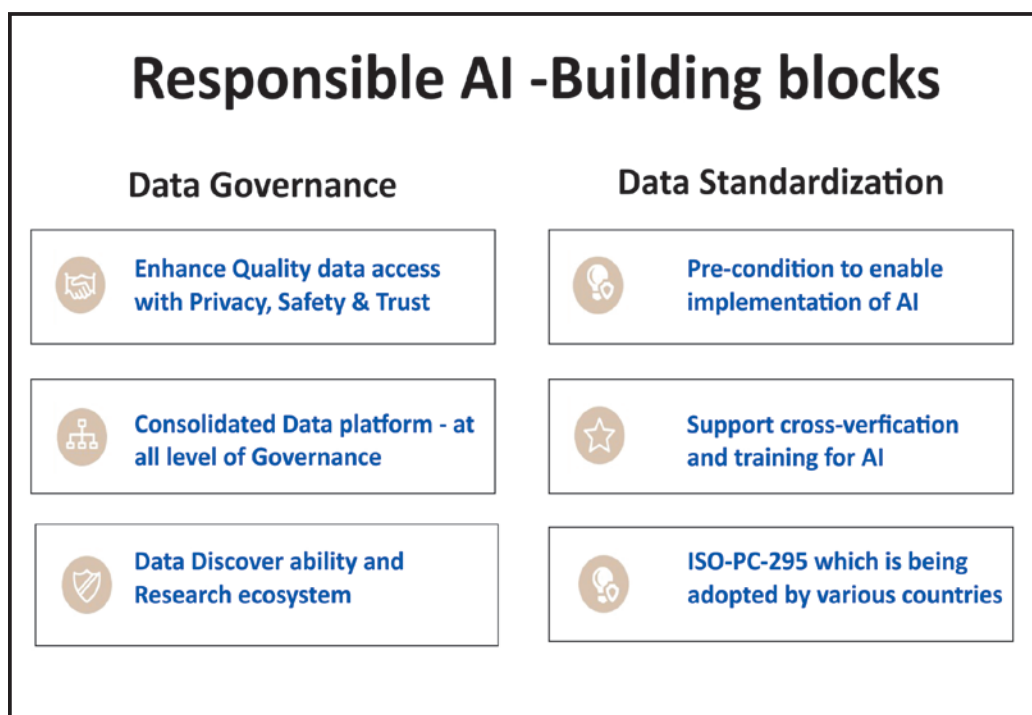
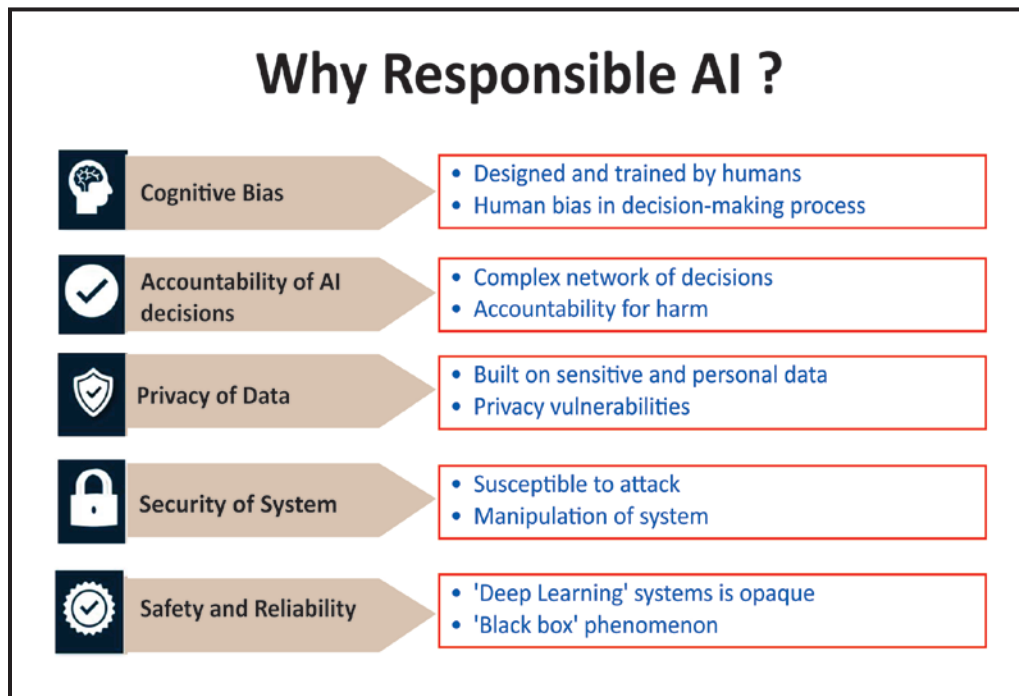
The use of AI in the public sector also has the potential to increase transparency and accountability. By analyzing large amounts of data, AI can identify patterns of corruption and fraud, detect errors and anomalies, and provide insights that can inform policy decisions.

However, the use of AI in the public sector also raises important ethical, legal, and social issues. These include concerns about privacy, bias, accountability, and transparency. It is essential that governments and other stakeholders consider these issues carefully and develop policies and frameworks that address them effectively.

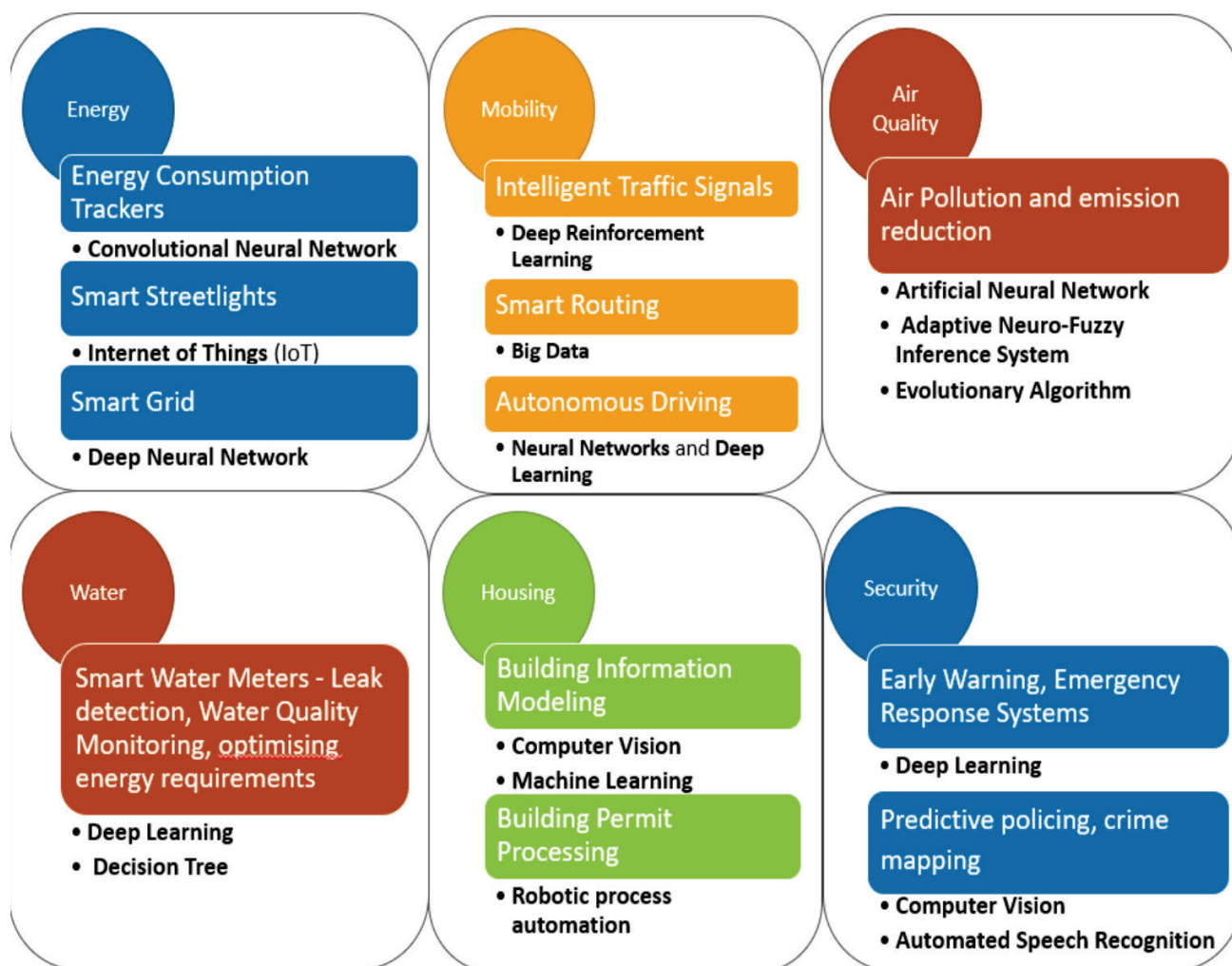
Overall, the use of AI in the public sector has the potential to transform the way that governments deliver services and engage with citizens. By leveraging the power of AI, governments can improve efficiency, effectiveness, and accountability, while also addressing complex social and economic challenges.

Why Responsible AI ?

AI systems have gained prominence over the last decade due to their vast potential to unlock economic value and help mitigate social challenges. Thus, not only the development but also adoption of AI has seen a global surge in recent years. Responsible AI is designed to help recognize, prepare for, and mitigate potential harmful effects of AI.



Some of the commonly identified use of AI in improving Public Service Delivery are shown below :

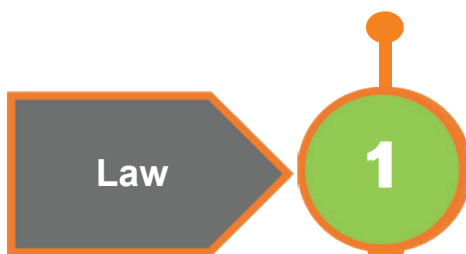


Some of the case studies in use of AI in public sector in various countries are as follows:

ARGENTINA



Key Initiatives – Argentina's AI National Plan is to develop policies that contribute to sustainable growth and the improvement of equal opportunities through AI technologies, ultimately positioning the country as a regional AI developer.

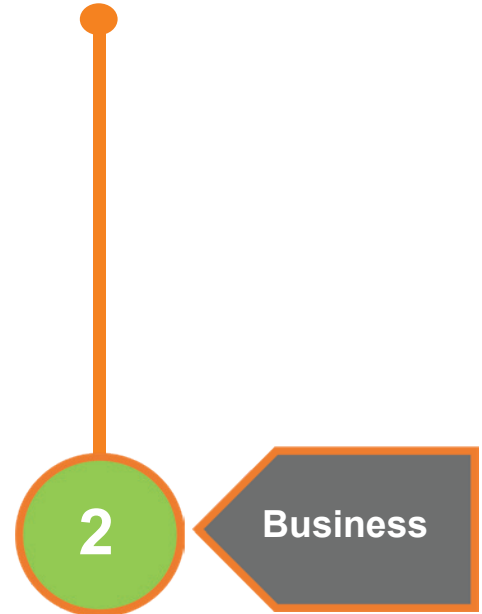


Prometea

Prometea is an AI system designed to streamline the work of the justice service. It was developed in 2017 jointly by the Public Prosecutor of the Autonomous City of Buenos Aires and the Innovation and Artificial Intelligence Laboratory of the Faculty of Law of the University of Buenos Aires (IALAB). Prometea's objective is to free judicial officials from repetitive tasks and allow them to focus their work on complex cases where human labor is necessary. The system acts as a virtual assistant that predicts case solutions (based on previous cases and solutions) and assists in providing the information required to build the file. Prosecutors then decide whether the solution that was predicted is worth considering. In Buenos Aires, between October 2017 and mid-2020, Prometea helped resolve 658 cases.

Business Opportunities Map

The Government of the City of Buenos Aires offers a business overview to entrepreneurs interested in starting or expanding their activities in the city. The Business Opportunities Map requests two pieces of information (geographic area and business category) and provides detailed information about the characteristics of the market in each area with four indicators: the opening and closing of stores, the level of risk, population indicators, and real estate indicators. The platform works through an algorithm that collects different data sources and summarizes business opportunities by neighborhood, ultimately allowing direct investments to be made where they are presumed to generate the most value. It is planned to include predictive models in future versions of the platform.



Tina

"Tina" is a chatbot with artificial intelligence that provides assistance for more than 200 procedures, available on the Mi Argentina platform, WhatsApp and Argentina.gob.ar

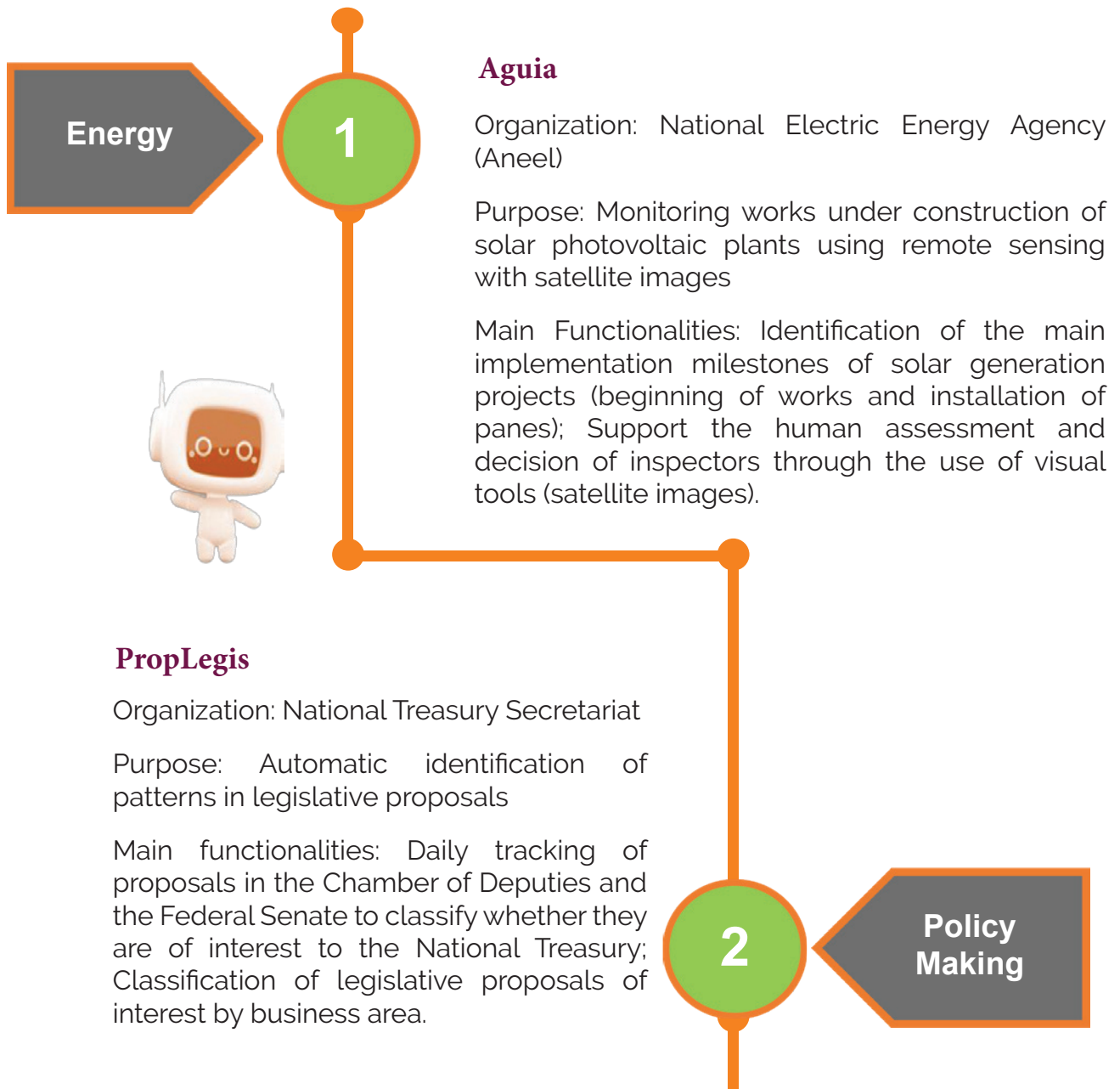


BRAZIL



Key Initiatives –

Brazil's National AI Strategy (EBIA)





Assistente Conjur

Organization : Federal Court of Accounts (TCU)

Purpose: Suggestion of thesis to respond to judicial inquiries and generation of draft opinions

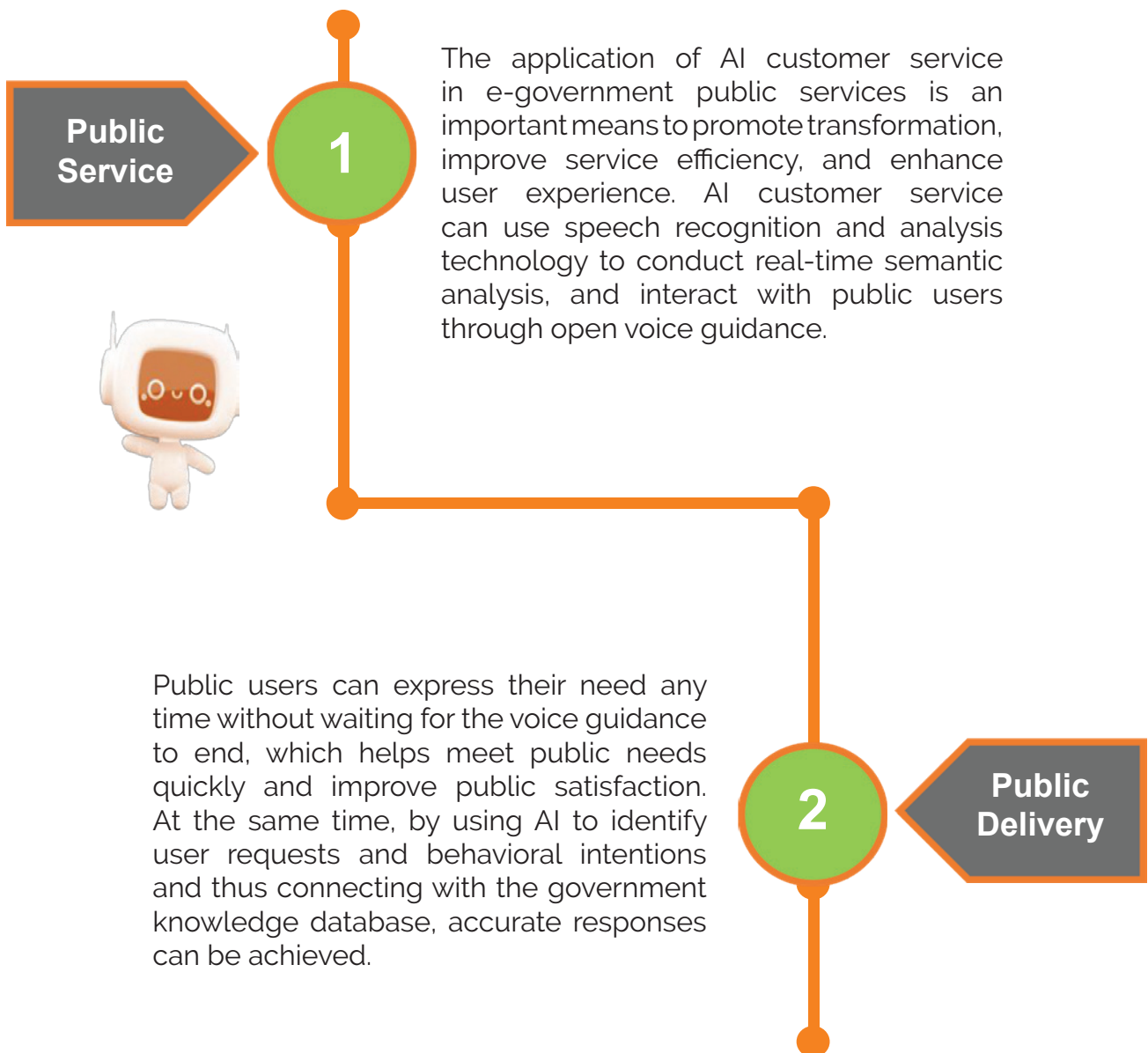


CHINA



Key Initiatives – In 2019, China issued Governance Principles for the New Generation Artificial Intelligence: Developing Responsible Artificial Intelligence.

AI development should follow the 8 principles, namely, Harmonious and Human-friendly, Fair and Just, Inclusion and Sharing, Respect for Privacy, Safe and Controllable, Shared Responsibility, Open and Collaboration, and Agile Governance.



EGYPT



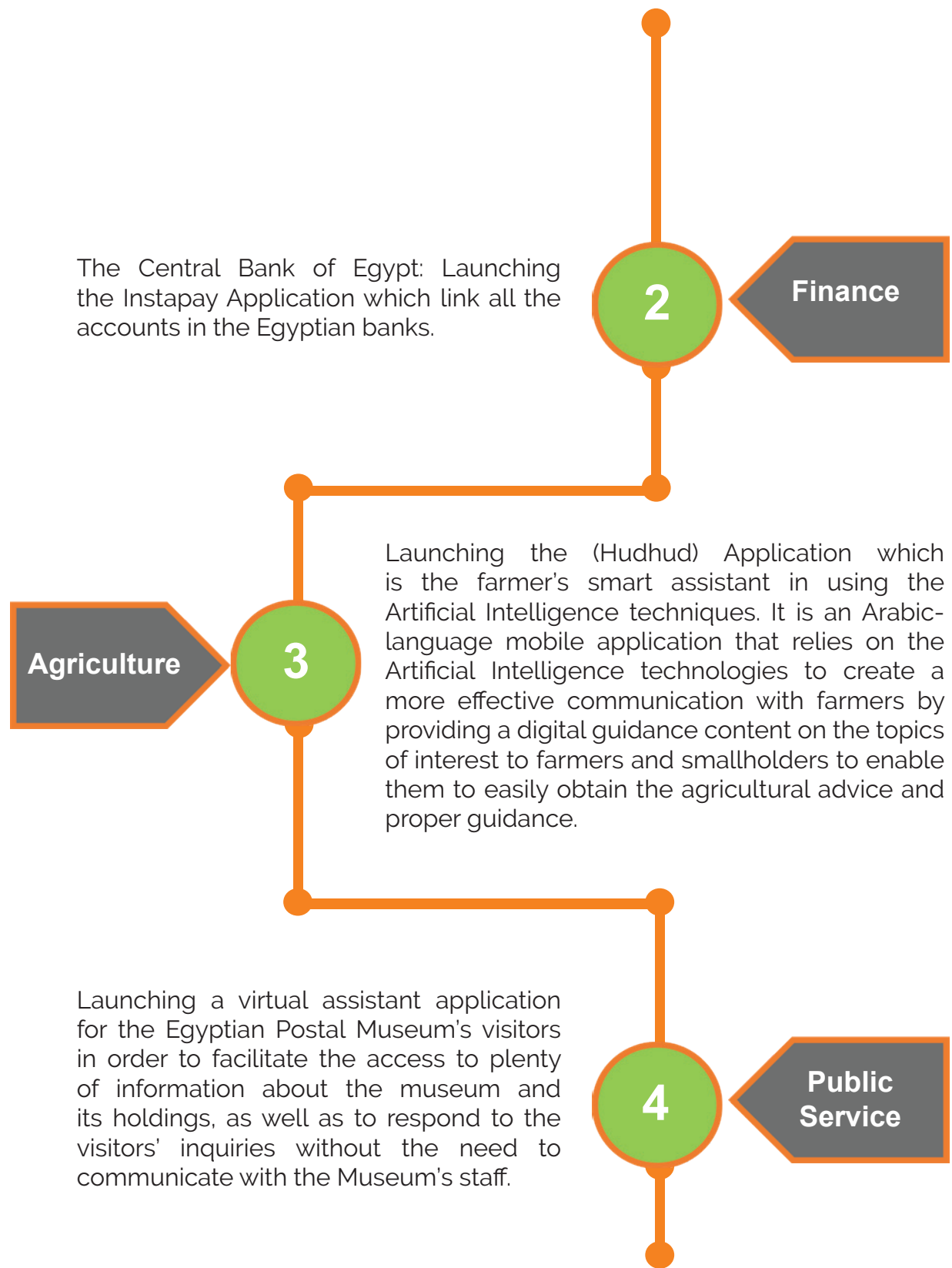
Key Initiatives – National Council for Artificial Intelligence (2019)

AI National Strategy (2021)

National Charter for Responsible AI (2023)



Egypt's Digital Portal: It includes all the services of digital justice, health, education, universities, social protection, civil services, travel etc., in addition to the e-government portal services.



The Central Bank of Egypt: Launching the Instapay Application which link all the accounts in the Egyptian banks.

2

Finance

Agriculture

3

Launching the (Hudhud) Application which is the farmer's smart assistant in using the Artificial Intelligence techniques. It is an Arabic-language mobile application that relies on the Artificial Intelligence technologies to create a more effective communication with farmers by providing a digital guidance content on the topics of interest to farmers and smallholders to enable them to easily obtain the agricultural advice and proper guidance.

Launching a virtual assistant application for the Egyptian Postal Museum's visitors in order to facilitate the access to plenty of information about the museum and its holdings, as well as to respond to the visitors' inquiries without the need to communicate with the Museum's staff.

4

Public Service

INDIA



Key Initiatives – National Strategy for AI, Draft Personal Data Protection Bill (2022) (PDP), The Data Empowerment and Protection Architecture (DEPA), National Programme for AI – key components include Data Management Office, National centre for AI, Skilling and Responsible AI.

Transport

1



iRASTE

iRASTE (intelligent Solutions for Road Safety through Technology and Engineering) - The Government of India, through a unique partnership with industry and academia, has launched Project iRASTE to reimagine road safety in India using the predictive power of AI. This is a step closer to the government's target of 50% reduction in road fatalities on Indian roads by 2030.

Aadhaar Mitra

Aadhaar Mitra: The chat-bot comes with enhanced features like - checking Aadhaar enrolment/update status, tracking of Aadhaar PVC card status, information on enrolment center location, etc. People can also register their grievances and track them using Aadhaar Mitra.



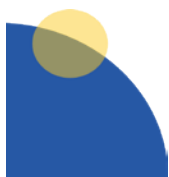
Diabetic Retinopathy

Detection of diabetic retinopathy in primary care using AI - NITI Aayog is working on using AI for early detection of diabetic retinopathy (DR). The objective is to develop an AI algorithm for the detection of signs of DR in diabetic patients and to scientifically validate the algorithm to be used as a screening tool in primary care.



Bhavishya

Bhavishya: It is an AI-supported portal for pension payment and tracking. It will send automatic alerts to pensioners. It will also regularly receive their inputs, suggestions as well as grievances for a prompt response.





Security

5

Installation of AI-enabled CCTV cameras by the Indian Railways: The ministry has assigned the RailTel with the task of completing the installation of a Video Surveillance System (VSS) equipped with high-resolution CCTV cameras. CCTV cameras of stations and video feeds will be monitored at three levels to ensure enhanced safety and security at railway premises. The system comes with Artificial Intelligence (AI) enabled Video Analytics Software and Facial Recognition Software, which will help to detect known criminals triggering an alert when they enter station premises.

Kisan Suvidha disseminates relevant information to farmers regarding weather, market prices, plant protection, inputs, weather alerts, go-downs and storages and market-related information such as conducive markets, market rates, quantity demands, etc.

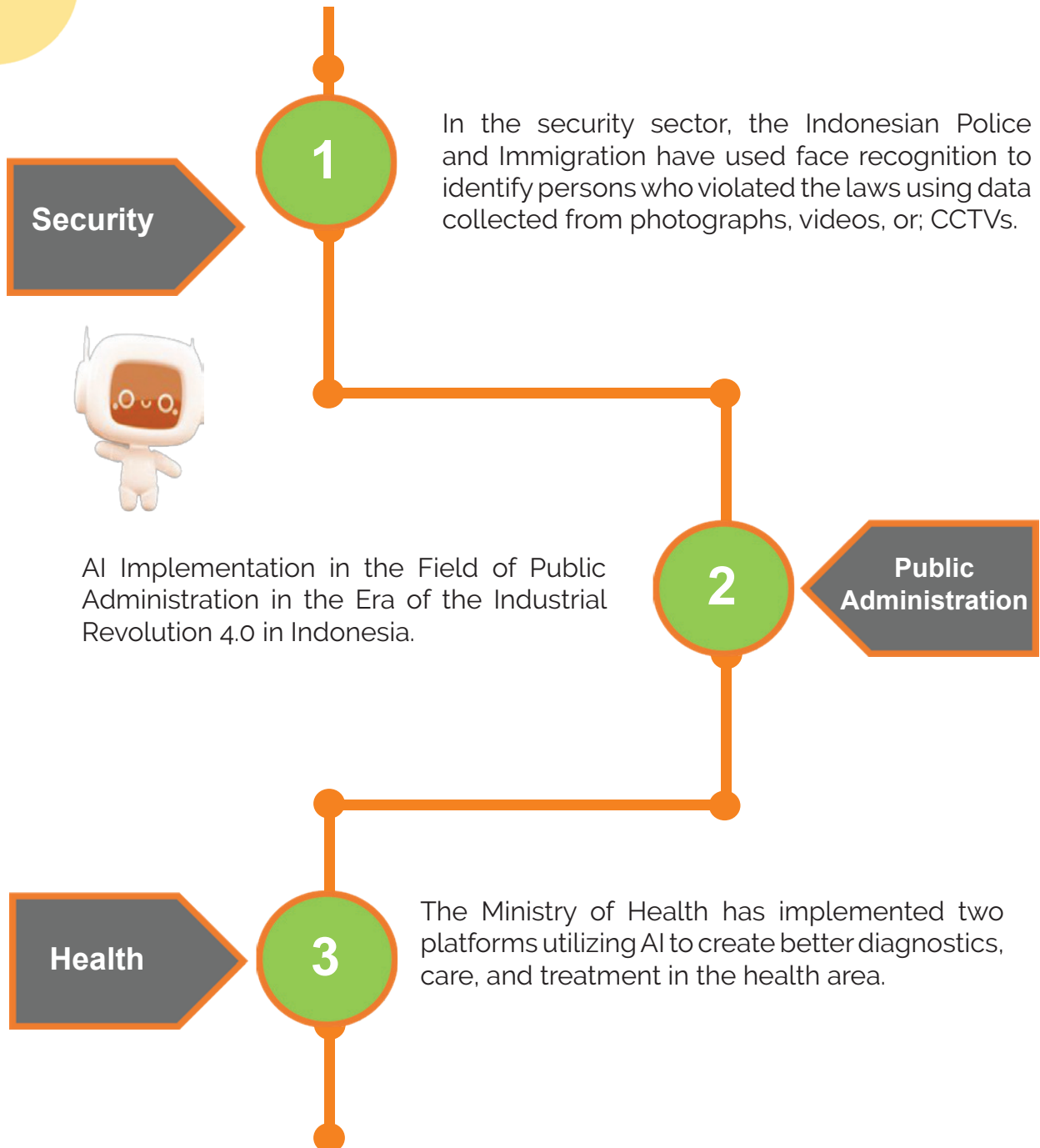
6

Agriculture

INDONESIA



Key Initiatives – Indonesia's National AI Strategy from 2020 to 2045, now known as Stranas KA (Strategi Nasional Kecerdasan Artifisial)



REPUBLIC OF KOREA



Key Initiatives – Artificial Intelligence R&D Strategy

Business

1



e-Policy Fund System for SMEs

The Ministry of SMEs and Startups (MSS) has been pushing ahead with the digital transition of policy financing for SMEs from 2021 to 2023 in a phased manner. This is to provide non-face-to-face service of loaning policy funds throughout all the processes from application, evaluation, and to loan commitment as a part of its efforts to brace for post COVID-19 era. The e-Policy Fund System allows SMEs and venture companies to apply for policy fund and conclude loan agreement online, bringing more convenience in using the policy funds.

The MSS also established the AI Supporter in April 2022, which is an advanced AI and data-based evaluation system, in order to enhance fairness in evaluation of policy fund applicants and reduce the delinquency rate.

UNI-PASS of the Korea Customs Service

Big data platform collects a wide range of internal and external data on customs administration, and the results of analysis on such large data are provided through the big data portal. Based on this, the Korea Customs service (KCS) developed 12 analysis models in 2021 by combining AI algorithm analysis such as forecast, regression, classification, etc. and various visualization methods for processes of each work area.

Thus, the KSC detected an illegal foreign exchange transaction exceeding KRW 200 billion. It also identified 290 false businesses through similarities analysis on unstructured data of each business.

Also, in a response to possible shocks from global supply chains as seen in the urea solution shock, the KCS supported establishing measures to stabilize supply and demand in a timely manner by providing each Ministry with big data analysis results related to customs trade.



Smart KIPOnet of the Korean Intellectual Property Office

Based on the results of accumulated AI empirical studies, the Korean Intellectual Property Office (KIPO) developed models for patent classification earning and recommendation and built up the AI Patent Classification and Recommendation System which automatically recommends the CPC patent classification.

KIPO also established the AI Patent Search System which automatically extracts keyword and sentence from documents subject to evaluation, and then searches and recommends the most similar document. About 5 million patent documents were used as learning data, and the accuracy of search was improved by utilizing not only key words, but also the unique information of patent documents such as key sentence drawn by AI and the CPC.



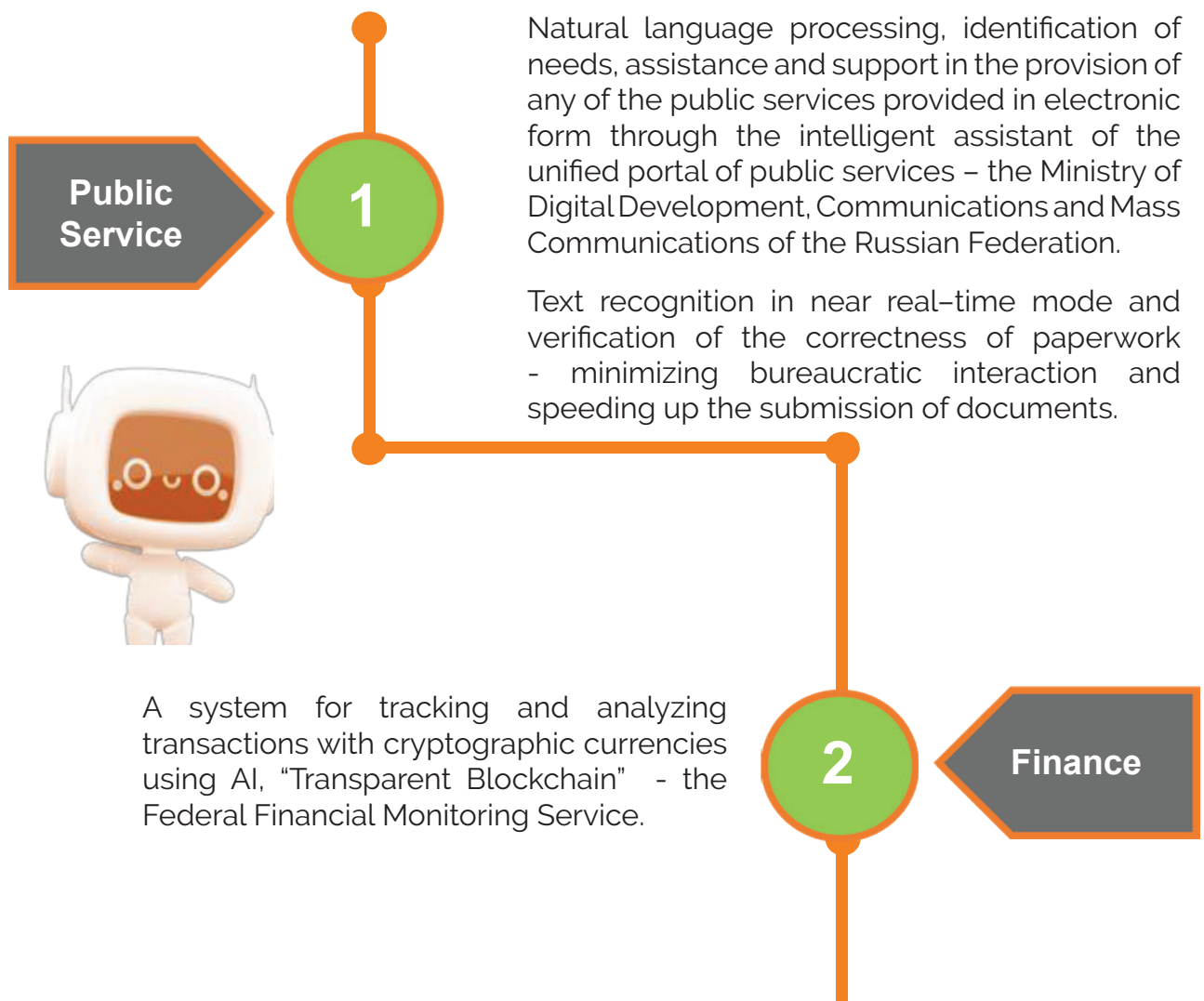
RUSSIA



Key Initiatives – National Strategy for the Development of Artificial Intelligence for period until 2030.

Approved Concept for the development of regulation until 2024.

Legislative Acts that form special legal regimes.





Communication

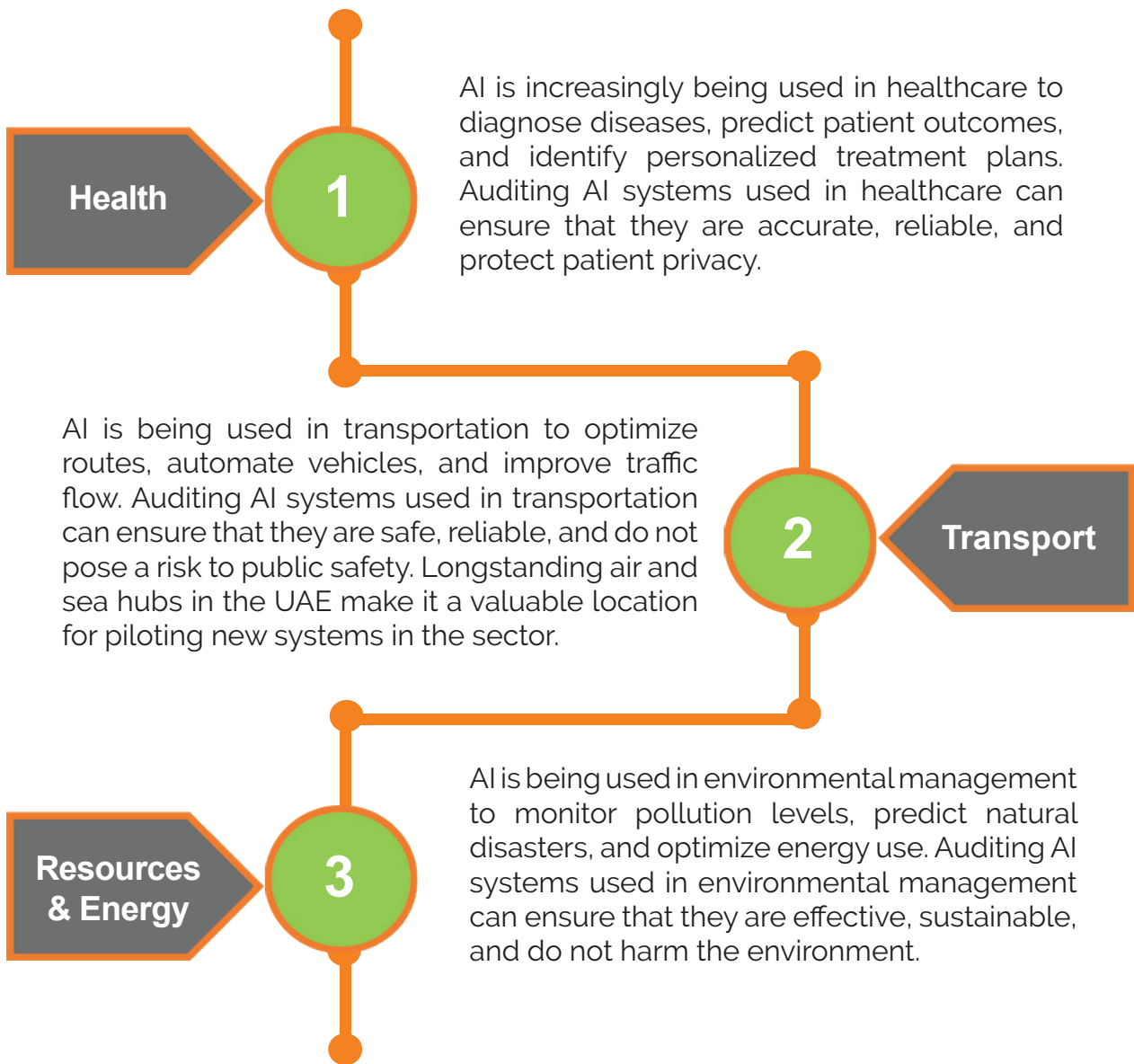
3

Digital assistant “robot Max” is an intelligent chatbot that performs a “smart” search on the portal of State and municipal services, receiving online answers to users’ questions, identifying their needs.

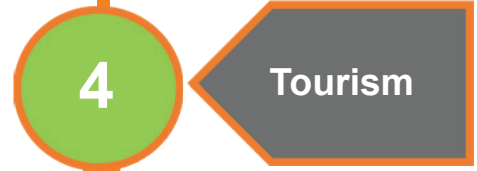
UAE



Key Initiatives - UAE Strategy for Artificial Intelligence



This is another area where AI is increasingly being used to personalize experiences, optimize operations, and improve customer service. Auditing AI systems used in tourism and hospitality can ensure that they are effective, secure, and do not infringe on guest privacy. opportunity for globally becoming first in customer support AI, creating integrated and personalized services for tourists in the UAE.



AI is being used to detect, prevent and respond to cyber threats. Auditing AI systems used in cybersecurity can ensure that they are effective, reliable and secure in protecting critical infrastructure, sensitive data, and intellectual property. a strategic imperative, given the rise of AI, the UAE will also concentrate on building robust systems for protection.





Chapter-3

AI Tools,
Technology
and Framework

AI Tools, Technology and Framework

Introduction

The field of AI is rapidly evolving, and new tools, technologies, and frameworks are continually being developed to advance the capabilities of AI systems and make them more accessible to developers. Some of the AI tools, technologies and framework for auditing AI have been covered in this chapter.

(a) AI Tools in Audit

Some of the common tools which have been used by SAI India to implement AI in Audit process are –

1. Python – Open source and easy-to-use. It offers flexible and extensible large number of libraries for AI and can assist in Audit related tasks.
2. R – Open source and easy to use. Offers many packages that are commonly used for machine learning and can be helpful tool in Audit analysis.
3. Azure Services – It is a proprietary tool from Microsoft which has various inbuilt machine learning and AI algorithms which can be used for advanced audit analysis.
4. KNIME – Open source data analytics platform for data science, ML, AI, Big data etc.

(b) Tools for Auditing AI

Over the past few years, the importance of open source software in the realm of AI has increased significantly. The availability of numerous tools and resources for academicians and developers is another advantage of open source AI software.

Auditors can explore the use of these tools in Audit to verify the AI or ML model. Some of the tools are discussed below :

- (i) **Fortuna¹** : Fortuna, an open source toolbox for ML model uncertainty assessment, has been made broadly available, according to an announcement from AWS. Any trained neural network can generate calibrated uncertainty estimates using the calibration techniques provided by Fortuna, such as conformal prediction.

1 <https://www.opensourceforu.com/2023/01/fortuna-an-open-source-library-for-measuring-model-uncertainty-is-now-available>

2 <https://ai-fairness-360.org/#:~:text=AI%20Fairness%20360%2C%20an%20LF,throughout%20the%20AI%20application%20lifecycle>

An accurate assessment of the anticipated uncertainty is necessary for applications that demand the ability to make key decisions. When there is uncertainty, one might assess the accuracy of model predictions, rely on human judgement, or decide if a model can be used safely.

- (ii) **AI Fairness 360²** : AI Fairness 360, an LF AI incubation project, is an extensible open source toolkit that can help users examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. It is designed to translate algorithmic research from the lab into the actual practice of domains as wide-ranging as finance, human capital management, healthcare, and education. The toolkit is available in both Python and R. IBM moved AI Fairness 360 to LF AI in July 2020.

The main features of this tool kit are 10 state-of-the-art bias mitigation algorithms, 70 Fairness metrics and industry specific applications.

Other toolkits in this family includes : AI Explainability 360, AI Privacy 360, Adversarial Robustness 360, Uncertainty Quantification 360 and AI FactSheets 360.

(c) **Auditing Framework for Auditing AI**

Based on the role of the SAI, the Auditing framework in Audit of AI can be covered under five components viz. Governance, Performance, Regulatory, Societal and Safety as shown below in diagram:



3 <https://indiaai.gov.in/responsible-ai/homepage>

A Responsible AI Resource Kit³ has been designed by National Association of Software and Service Companies (NASSCOM) which is the culmination of a joint collaboration between NASSCOM and leading industry partners to seed the adoption of responsible AI at scale.

The Resource Kit comprises sector-agnostic tools and guidance to enable businesses to leverage AI to grow and scale with confidence by prioritising user trust and safety.

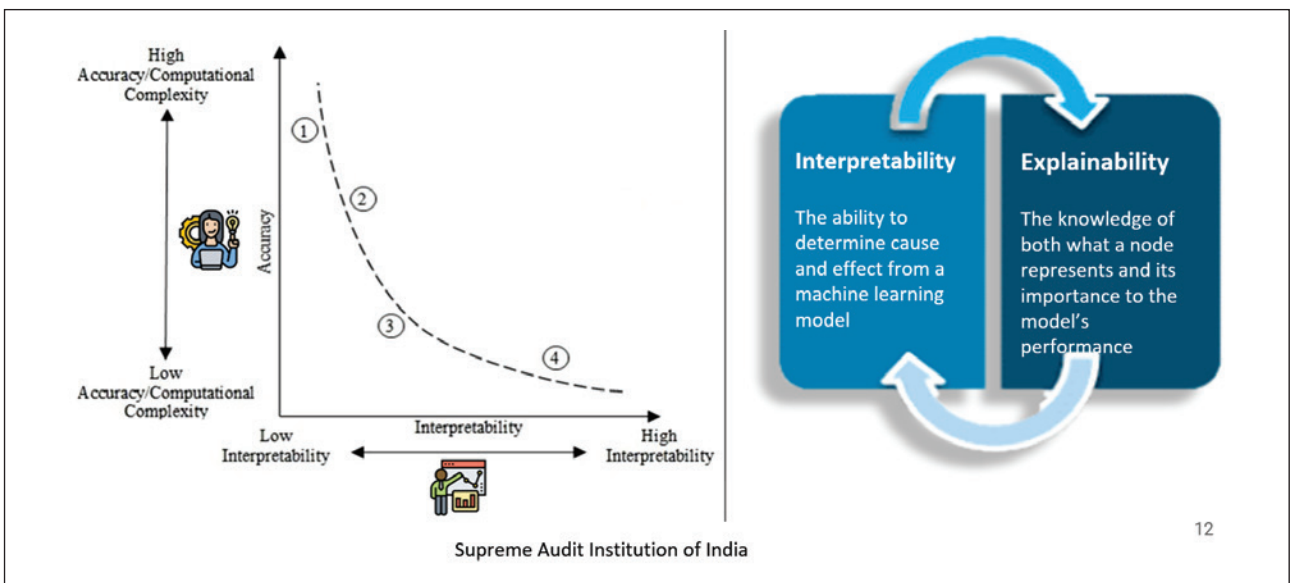


The Audit Toolkit can be prepared covering these areas identified for compliance as below:

Audit Framework	Objectives	Base Questions for Tool Kit
Governance	Explainability — Users should be able to request explanations for significant decisions taken by AI systems. Explanations must be provided free of cost to the user and should contain human understandable summary of how the system arrived at a particular decision	<ul style="list-style-type: none"> • What are the documents available explaining the decision process? • Whether Process flow chart of the AI Algorithm is available? • Whether Explanation on the decision process in simple non-technical and intuitive language available for users? • What type of algorithms are used for AI – Standard/ Customised/ Generic etc?

	<p>Accountability — Organisational structures and policies should be created to clarify who is accountable for the outcomes of AI systems. Human supervisory control of AI systems is recommended</p>	<ul style="list-style-type: none"> • Whether adequate organisational structures and policies are defined? • Whether clear roles and responsibilities are fixed? • Whether adequate user feedback mechanism exists? • Whether human supervisory control exists?
	<p>Transparency- AI systems should be transparent about how they were developed, their processes, capabilities and limitations, to the extent possible.</p>	<ul style="list-style-type: none"> • Whether training data set is shared for ensuring transparency? • Whether the algorithm logic is shared? • Whether proper documentation exists for each lifecycle of the system for audit?
<p>Performance</p>	<p>AI systems must produce consistent and reliable outputs in all scenarios. Appropriate grievance redressal mechanisms should be put in place to address cases of adverse impact</p>	<ul style="list-style-type: none"> • Whether risk based quality assurance checks has been done? • Whether there is proper monitoring on the performance of the output? • Whether grievance addressal system is being used to fine tune the system?
<p>Regulatory</p>	<p>Compliance — Throughout their lifecycle, AI systems should comply with all applicable laws, statutory standards, rules, and regulations – Organizations should be watchful of the evolving AI regulatory landscape and ensure compliance at all times</p>	<ul style="list-style-type: none"> • Whether regulatory framework exists? • Whether compliance to laws, rules and regulations are audited through proper agency? • Whether privacy law has been considered for data collected during implementation?

Societal	Protection and Reinforcement of Positive Human Values — AI systems should be designed and operated such that they align with human values. AI should promote positive human values for the progress of humanity as a whole	<ul style="list-style-type: none"> • Whether critical review of the system to verify anticipated benefits, harm and impact has been done? • Whether the system is developed keeping Human values in centric? • Whether any unethical use-cases has been considered?
	AI systems must be fair and inclusive, not fostering prejudices, discrimination or preference for an individual, a community or a group based on their sensitive attributes (e.g., race, gender, ethnicity)	<ul style="list-style-type: none"> • What type of data has been used for training? • Whether the training data is having any of the bias? • Whether all type of user groups has been covered in the training?
Safety	AI system should be robust and secured against adversarial attacks and malicious use. Identifying and mitigating system vulnerabilities is critical	<ul style="list-style-type: none"> • Whether adequate protocols and measures are in place against unauthorised access? • Whether adequate controls are in place to guard against adversarial attacks? • Whether security audit is done?



(d) Data Standardization for AI

Data standardization is the process of creating standards and transforming data taken from different sources into a consistent format that adheres to the standards. This is a pre-condition to enable implementation of AI in SAI as this will support cross-verification, training, and fraud detection across government organizations.

Data standardization is critical to facilitating and improving the use of data, especially as related to data portability (i.e., the ability to transfer data without affecting its content) and interoperability (i.e., the ability to integrate two or more datasets). One such initiative has being undertaken by ISO-TC-295 which is being adopted by various countries.

AI systems must produce consistent and reliable outputs in all scenarios. Standards for data quality in machine learning and analytics are essential for producing reliable insights and eliminating errors. Governance standards in AI and big data analytics address how these technologies can be managed and overseen. Standards for trustworthiness, ethics, and societal concerns should also be developed from the beginning to ensure responsible deployment.

The value of digital transformation is the ability to provide insights. Technologies such as AI, analytics and big data are enablers for this transformation. The international standards SC 42 is developing will remove adoption barriers while addressing concerns enabling broad responsible adoption.


(e) Global Standards for AI

UNESCO's global agreement : The recommendations on the ethics of AI can guide governments. 193 countries reached a ground breaking agreement at UNESCO in November 2021 on how AI should be designed and used by governments and tech companies. A common rule book has been framed which will serve as a compass to guide government and companies to voluntarily develop and deploy AI technologies that conform with the commonly agreed principles.

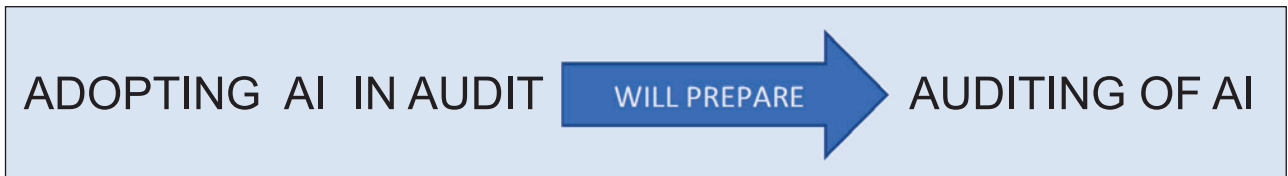
ISO/IEC JTC 1/SC 42 : is an international standards committee responsible for developing artificial intelligence (AI) standards. SC 42 is a joint committee of ISO and IEC. As the focal point of standardization on AI within ISO and IEC, SC 42's program work looks at the entire AI ecosystem. Additionally, SC 42 is scoped to provide guidance to ISO and IEC committees developing Artificial Intelligence applications. Its current program work includes standardization in the areas of foundational AI standards, Big Data, AI trustworthiness, use cases, applications, governance implications of AI, computational approaches of AI, testing, ethical and societal concerns.

(f) Global Footprint of Artificial Intelligence

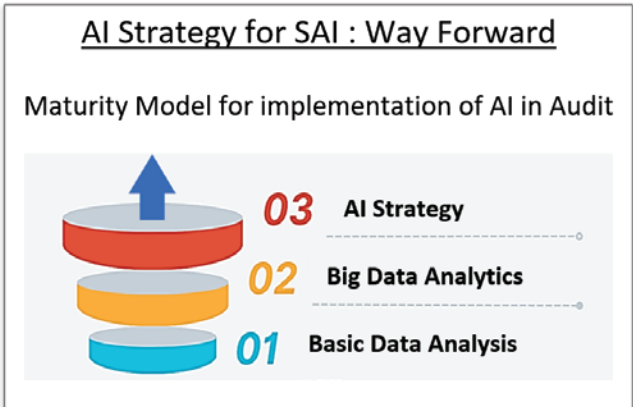
Artificial intelligence (AI) is a rapidly advancing technology that has a global impact. AI has the potential to revolutionize many aspects of society, from business and healthcare to education and the environment. Most of the countries and some unions have adopted specific strategy in AI. The various initiatives adapted globally are shown below:

 Australia AI Action Plan	 Brazil National AI Strategy (EBIA)	 Canada Pan-Canadian AI Strategy	 China Next Generation AI Development Plan
 Denmark Strategy for Digital Growth	 Finland Age of AI	 France Strategy for AI	 Germany Federal Government Strategy on AI
 India National Strategy for AI	 Italy AI at the service of Citizens	 Japan AI Technology Strategy	 Mexico Towards an AI Strategy
 Russia National Strategy for development of AI by 2030	 Saudi Arabia National Strategy for Data and AI	 Singapore Model AI Governance Framework	 Republic of Korea AI R&D Strategy
 Sweden National approach to AI	 UAE Strategy for AI	 UK Industrial Strategy: AI	 USA Algorithmic Accountability Act of 2022
 OECD AI Ethics guidelines, Principles on AI	 UNESCO Global agreement on AI principles and challenges	 European Union General Data Protection Rules (GDPR) 2016	

Auditing cannot be left behind in what may be the next digital frontier – artificial intelligence. To prepare, SAI auditors must understand AI basics, the roles that SAI can and should play, and AI risks and opportunities. To meet these challenges, SAI should leverage the Framework to deliver systematic, disciplined methods to evaluate and improve the effectiveness of risk management, control, and governance processes related to AI.



- Adopting AI capabilities in Audit**
- Opportunities and Risks
 - Mapping Audit processes and Tasks
 - Use of AI in Big Data Analytics
 - Exploring Tools for AI
 - Capacity Building Programme





Chapter-4

Voices

Voices

Introduction

SAI India had organized a Seminar on AI. SAI India would also hold a side event on AI with domain experts during the SAI20 summit. The valuable insights of the speakers in these events would help us to gain valuable understanding in the field of Artificial Intelligence. Their work would also help the SAI20 members to gain a better understanding of the latest advancements in the field of Artificial Intelligence. Some of these work are as below :

(1) Dr. Pavan Duggal

He is an Advocate in Supreme Court of India. He is also the Founder & Chairman of International Commission on Cyber Security Law along with being the President of Cyberlaws. Net and has been working in the pioneering area of Cyber Law, Cyber Security Law & Mobile Law.

He is heading the Artificial Intelligence Law Hub and Blockchain Law Epicentre and Founder cum Honorary Chancellor of Cyberlaw University.



REVOLUTION OF ARTIFICIAL INTELLIGENCE – SOME LEGAL NUANCES

We are actually seeing the world changing.

Today's landscape is a landscape of energy, of opportunity, and also of scepticism and challenges. Everybody today is enamoured by the new flavour of the season, GPT-4 and ChatGPT. ChatGPT and GPT-4, bringing across unprecedented legal challenges for which the world is still not ready.

The news that GPT-4 has already passed the US bar examination assumes importance, since it takes a normal person atleast many years to even write the bar examination after getting a law degree.

ChatGPT amassed a 100 million users in five days. Everybody in the professional world is bothered. Are these AI chatbots or models going to take the jobs away? Looking at the current level of evolution, this scenario may not come true. However, if you are someone like a paralegal, then tighten your belts, because going forward, a lot of your work is going to be done by these AI algorithms.

There are numerous legal, policy and regulatory complexities that ChatGPT and GPT4 have begun to start throwing up. The jury is divided as to whether or not the output from ChatGPT or GPT-4 is legal and whether it can be used in a court of law. The author is very clear today that an output from GPT-4 or ChatGPT is nothing but computer output, in other words, it is an output that is generated by a computer. Hence, it becomes an electronic record, under Section 2 of the Information Technology Act, 2000 and is legal under Section 4 of the IT Act, 2000.

Recently, a lawyer in South India while arguing a matter went about and gave his proposition for research to ChatGPT and ChatGPT came up with a number of authorities and citations. The lawyer argued in the court on the basis of the research note generated by ChatGPT, without cross checking through the references or authorities or citations. It subsequently came to be known that none of these citations ever existed.

Amazon has reported recently that their books segment has suddenly started booming. Everybody is writing books because a lot of persons have started using ChatGPT for writing books. If tomorrow, it is determined that your book was not written by you, but by ChatGPT or some other similar tool, the kind of impact it can have on your reputation, your goodwill, your standing will be phenomenal to say the least. It is true that artificial intelligence algorithms like ChatGPT and GPT-4 give you a direction but if you're going to blindfold rely upon these, without any application of mind, think again- you could be in big trouble.

In fact, AI products can be used for auditing services and can be extremely helpful for auditors for the purposes of conducting audit. The coming up of Artificial Intelligence shows the need for its appropriate auditing and need for appropriately regulating the same. A step in this direction has already begun. Recently, at the time of writing, China has introduced a draft law which is entitled "Administrative Measures for Generative Artificial Intelligence Services", which has been released by the Cyberspace Administration of China to ensure "the healthy development and standardised application of generative AI technology". It implies that all AI products developed in China will be required to undergo "a security assessment", before being released to the public.

From a regulator's standpoint, these Generative AI technologies tantamount to be a big black-box, where it is not exactly known as to what is contained in the said black-box and how does it actually function to come up with its output. There is thus a need for ensuring appropriate regulation of AI through appropriate auditing.

The legal provisions need to evolve which can mandate the developers of AI algorithms to share with the lawmakers and policymakers as to how their cognitive intelligence is actually working and what are the essential ingredients of the black-box that their Generative AI algorithms actually represent. Hence, this is a very interesting area of legal jurisprudence that has to evolve under AI law jurisprudence in the coming times.

Further issues arise pertaining to what happens if you rely on ChatGPT output, or the GPT-4 output, and if it turns out to be false or inaccurate? Can OpenAI be sued? OpenAI has actually made the disclaimers and the terms and conditions deliberately pretty solid, because they want to protect themselves. But that does not mean that they cannot be sued. India is a litigious society and it is only a question of time before they could potentially be sued.

When one reads the privacy policy, Open AI stipulates the position that data shared on their platforms, is going to be stored on servers that are located in the United States of America and will be used for the purposes of further bettering the services of the algorithm.

Here, there is an underlying message for the users so as to be careful. OpenAI is encouraging its users not to share their confidential notes, or confidential information, personal information on their algorithms as these are likely to be used.

Data protection poses huge legal issues for generative AI. Open AI says that it is a California company, and governed by California law. But it has also gone ahead and complied with the GDPR. But what happens in locations where there are no or different data protection?

As a lawyer, I will equate AI language algorithms to balls of fire as you don't exactly know what's going to happen. This is because bias is intrinsically and inherently built in these systems. Consequently, legally dealing with the issue of bias is going to be a huge problem, as we move forward.

Lot of people think AI and cybersecurity are different fields. I, on the other hand, think that these will have to be seen in conjunction with each other. Nobody is talking about the cybersecurity of AI. Some experts are talking about the increasing possibility of potential hacking into ChatGPT or GPT-4. Already ChatGPT has been used for creating a new malware that is incapable of being detected by current anti-malware products.

One survey tells us that by 2062, human intelligence is going to be superseded by artificial intelligence. I disagree because I believe that the tipping point is going to come much earlier. So we have to be prepared on how things will be while going forward, in terms of potentially protecting AI against sophisticated cyber attacks.

Recently, an institution called Future Of Life Institute has actually issued an open letter that has been signed by various people. It's an open letter to all giant AI companies, asking them to pause all giant AI experiments and development for a period of six months, so as to enable the regulators of the world to have a review and in order to enable those regulators to have legal frameworks in place. However, the ferocious development of Artificial Intelligence continues unabated.

There are numerous questions being raised, regarding these AI algorithms and models, involving legal, ethical and moral issues. Do these AI or these algorithms recognize human values? Are they ethical? Do they recognize and analyze what's going to be ethical? A holistic approach and collaboration will have to be the mantra, as humans train artificial intelligence to be more ethical. At Artificial Intelligence Law Hub, we are also now collating what is happening in terms of evolving legal jurisprudence, on AI.

To conclude, it can be stated that new emerging technologies like generative AI are empowering us in a manner that has never been seen before. While being grateful, we need to be careful of the legal, technical and other pitfalls connected therewith. So while the law catches up with newly emerging technologies, please protect yourself. Please ensure that you do not unnecessarily become victims of cybercrimes or misuse of newly emerging technologies. Due diligence, caution and care will have to be the mantra for all digital ecosystem stakeholders to follow, as they deal with the unfolding of the AI algorithm revolution and its connected issues and challenges.

(2) Ms. Debjani Ghosh

Debjani Ghosh is the President of the National Association of Software & Services Companies (NASSCOM) since April 2018. As the NASSCOM President, she actively work with the industry and Government to accelerate the growth of the Industry and strengthen India's position as hub for Innovation and Digital Talent.

With multiple feathers in her cap, before joining nasscom, she was the first woman to lead Intel India and the Manufacturers' Association for Information Technology, She believes that learning ever stops, and a trainable mindset is the biggest strength one can have.



The Artificial Intelligentsia

It is essential for us to move beyond the binary of AI regulatory alignment versus innovation

It was almost unthinkable to have a tech keynote in the last few years without the potential of AI being called out at great length. However, it is only in the last few months that the world has truly woken up to what it really means. The exponential pace at which generative AI is advancing and the speed at which companies are releasing products and its compounding effect is something not seen before.

To put it in perspective, ChatGPT (generative pre-trained transformer) gained its first 1 million users just five days after launching in November 2022. It reached the 100-million-user mark two months later, the fastest ramp in a consumer internet app to date. By comparison, it took TikTok about nine months after its global launch to reach 100 million users. Instagram took more than two years. And just last week, a new experimental open-source product, AutoGPT, took the internet by storm. AutoGPT gets GPT-4 to act autonomously enabling it to perform a task with little human intervention. And it can self-prompt.

But as AI systems become more sophisticated, concerns about their potential risks are also growing. Which is why we must urgently engage in critical discussions about the alignment and risks of AI. Alignment is a must to ensure that AI innovation is in line with human well-being and to

minimise potential risks associated with its development and deployment. And we know that it cannot happen at the cost of innovation. The challenge ahead is to achieve the goal of AI innovation with alignment.

Many experts feel that this is when government needs to step in with regulation. For starters, the industry should come out with a well-thought-through and transparent governance framework that will act as guard rails to define how their tools can be used, and not. This will give some time to track progress and key risks, and shape a more effective plan for regulation.

At a domestic level, India should set up an interdisciplinary task force of policymakers, researchers, technologists and ethicists, to study and design guidelines, ethical standards and regulatory frameworks that align with the values of society. This will need to be an ongoing process as the technology evolves. We have to give ourselves room to unlearn and adapt to these changes. Some focus areas:

- Conducting comprehensive risk assessments throughout the lifecycle of the solution, with the aid of critical guiding frameworks, can help evaluate potential harm and ensure responsible innovation.
- Development of best practices for developing, deploying and using generative AI in a safe and trustworthy manner through strict adherence to applicable data protection and intellectual property rules, as well as deployment of best-in-class cybersecurity infrastructure.
- Rethinking the talent agenda to not just minimise the risk of job displacements but also leverage the productivity multiplier of AI across sectors.
- Committing to steer and support universal AI literacy and awareness programmes and policy projects focused on designing guard rails for trustworthy adoption of generative AI technologies, in collaboration with partners from industry government and international public Institutions.

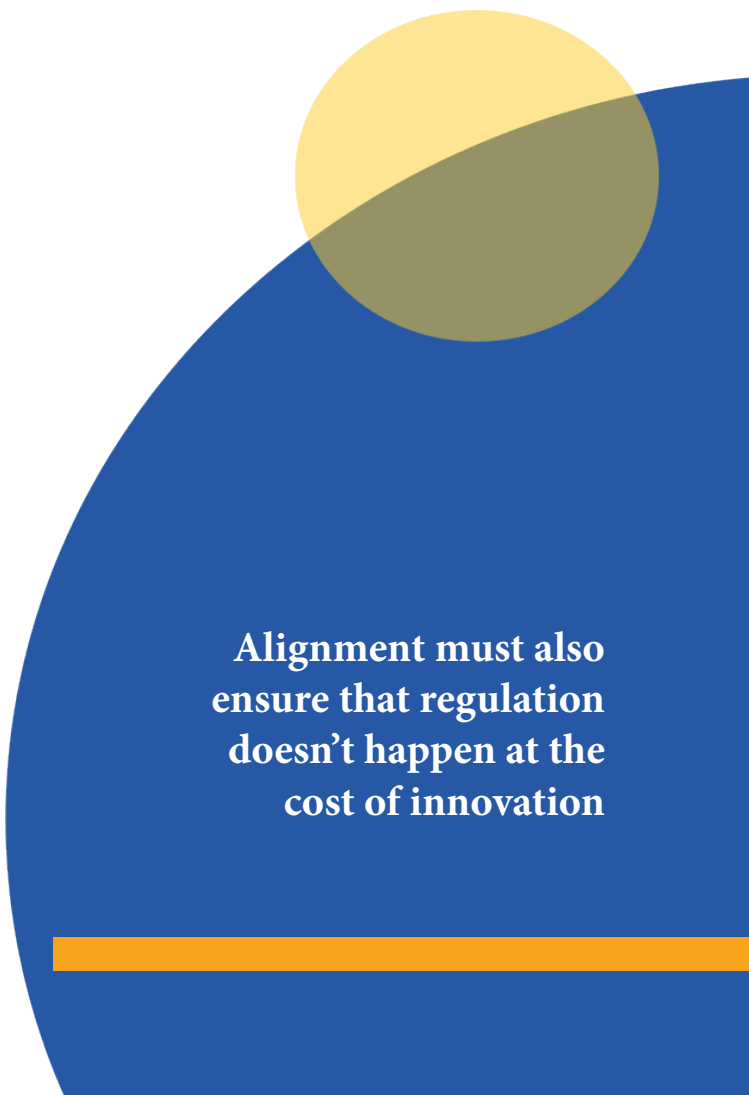
At an international level, India should take the lead to work with G20 countries to build a global framework for design principles that can be applied to the regulatory thinking of countries. This is going to be extremely important to build trust and to avoid the risk of nefarious activities by any nation. For example, instead of imposing broad regulations on the training of large AI language models in all areas, it may be more effective to prioritise the development of ethical guidelines and regulations for specific applications that pose significant threats to humanity

We need a risk-based usage-case approach. This approach would require global cooperation, and trust among countries and companies to slow down the development of potentially harmful technologies until adequate safeguards are in place.

It is also important to separate the regulation of research and science from that of products and product development. Technological advancements are often the result of continuous research and experimentation. Regulating products and services that use AI technology can ensure they meet safety standards, are ethically produced, and do not harm consumers or the environment.

The debate between regulation and innovation in AI is complex and multifaceted. It is essential to move beyond the binary view of alignment versus innovation.

By finding a balance between innovation and responsible regulation, the potential of AI can be harnessed for the betterment of society while minimizing risks.



Alignment must also ensure that regulation doesn't happen at the cost of innovation

(3) Sanjeev Singh

Sanjeev Singh is an experienced digital transformation and risk management expert with nearly 20 years' experience in designing and implementing complex, transformational changes by leveraging data analytics, AI and ML. He is currently posted as the Principal Commissioner of Income Tax at Bengaluru with responsibilities for implementation of the new Faceless Assessment Scheme and Faceless Penalty Scheme. He has made significant contributions to the Income Tax Department.



Leveraging Artificial Intelligence and Machine Learning for Effective Tax Administration

Background

With the advent of information technology and analytics, tax administrations can leverage Artificial Intelligence (AI) and Machine Learning (ML) to provide quality taxpayer service and implement risk-based approach for compliance management. This can help tax administrations to incorporate appropriate interventions and develop capabilities to address tax risks.

AI Implementation Framework

This article presents an AI implementation framework to identify 6 focus areas for effective tax administration. It also presents some common practices and key functionalities under each focus area.

Service Delivery

Information technology, AI and ML can be used to provide efficient and quality taxpayer services. Tax administrations use electronic filing and centralised processing systems for achieving efficiency and accuracy in filing and processing of returns and forms. Use of Chatbot with Natural Language processing (NLP) and ML provides context specific response to the queries of taxpayers with higher level of engagement. Tools for analysis and monitoring of social media posts and sentiments helps in getting instant feedback and identification of problems faced by taxpayers. Integration of social media analysis tool with ticketing system helps in seamless closure of issues.

Information Lifecycle Management

Information lifecycle management builds a robust and dynamic foundation for leveraging AI and ML in tax administration. Tax administrations generally collect third-party information (e.g. cash transactions, property transactions, foreign remittances, contract payments, fees, interest, dividend, securities transactions etc.) for effective risk assessment. Use of data quality rules and automated identification and communication of defects and exceptions enables reporting entities to rectify the defect in reported information.

Data enrichment involving standardization and clustering of key information fields (e.g. name, address, email, mobile, account etc.) reported in various information sources enables identification of best contact details (e.g. email, mobile, address) for communication. It also helps in identification of hidden relationships between persons. Many tax administrations implement entity resolution systems for deduplication of Tax Identification number (TIN) and population of TIN in information not having valid TIN for correct linking and utilisation of information.

In addition to data exchange with domestic and international data exchange partners, tax administrations, implement consent-based exchange with financial institutions (e.g. sharing of key information of IT return to pre-fill loan application) which reduces compliance cost for business and promotes voluntary compliance.

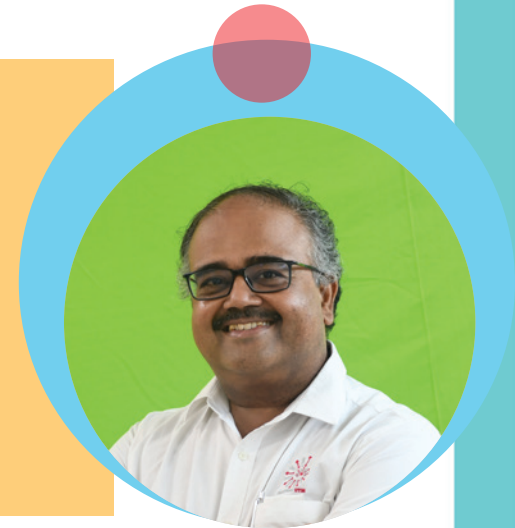
Many tax administrations display information collected from third party reporting and data exchange to the taxpayer. This helps in promoting voluntary compliance and validation of information. If the taxpayer is provided with the functionality to provide feedback on displayed information (e.g. information is duplicate, relates to other person/year, is partly incorrect etc.), it enhances the maturity of information reporting and processing systems.

Alert Generation

Tax administration leverage hybrid approach (consisting of rule engine, data analytics, text analytics, network analytics, geospatial analytics, machine learning etc.) for generation of wide range of alerts related to tax obligations (e.g. non-filing of return, inaccurate particulars in return, non-payment of taxes, etc.). Time critical alerts (e.g. high risk outward remittance) needs to be generated in nearreal time.

(4) Prof. B. Ravindran

He is the head of the Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI) at IIT Madras and a professor in the Department of Computer Science and Engineering. He is also the co-director of the reconfigurable and intelligent systems engineering (RISE) group at IIT Madras, which has nearly 80 members associated with it currently. He also serves on the advisory boards of several startups in the data analytics and AI space.



Managing Human-Allied Artificial Intelligence

We are fortunate to be living in interesting times, much like our ancestors who lived through the industrial revolution or the Renaissance! We are probably seeing the beginnings of the remaking of our society through the transformative power of artificial intelligence (AI). The techniques and paradigms collectively known as AI are not entirely new. The AI revolution traces its roots back to the Fifties and even earlier in certain cases. The hype around AI is also not new in 1958¹ people thought that AI systems that can walk, talk, see, write, reproduce and be conscious of their existence were around the corner! Again, real AI success is also not new the first self-learning checkers program debuted on live television in 1956². What makes the current scenario different is that AI is beginning to affect everyday activities and social interactions in an intimate fashion. Be it connecting with friends, reading the news, looking for jobs, shopping, or even ordering dinner.

With the recent advent of Midjourney, ChatGPT, Bard, and their ilk, the hysteria about AI has reached a crescendo. Generative AI is being seen as a technology that will disrupt the social order and cause large-scale job displacement. Leaders in the community as well as rabble-rousers are demanding a ban on the technology, claiming that this will even be the end of humanity. This is reminiscent of the public debate at the start of the industrial revolution or, more recently, computer automation of workplaces. The Time Magazine, in an article in 1966, envisaged that by the 1980s, most of the human population would be paid to be idle due to the increasing automation of



¹ July 8, 1958, The New York Times “NEW NAVY DEVICE LEARNS BY DOING”.

² Arthur Samuel Biography (<https://history-computer.com/arthur-samuel-biography-history-and-inventions/>)

jobs. Obviously, such dire predictions have not come to pass, and computers have had a significant impact on the quality of life all around. Generative AI and related advances are here to stay. The genie is out of the bottle, and we need to discuss how to shepherd the applications of the technology. Many documents on principles of responsible AI or trustworthy AI have been published³⁴⁵⁶. Here are the key ideas behind these principles.

- 1. Explaining AI decisions:** For us to trust the outcomes from AI models, we need to understand the reasons for the decisions/recommendations made by the model. Many successful AI applications are achieved by what are known as black box models. While we know the mechanistic process of computation, we do not fully understand the reason for the outcomes. Even models that are not black boxes can only be explained in terms of the statistical properties of the model. A responsible AI system will offer explanations so that a layperson affected by the decisions can understand them. This would require aligning the AI model to the accepted modality of explanations in the application domain.
- 2. Data Discipline:** This is one aspect of responsible AI that has received the most attention. The General Data Protection Regulation⁷ (GDPR) in the European Union was the first comprehensive framework for controlling collection, access, and usage of data. The rights of the end user whose data is used for building AI systems are made paramount in these frameworks. India is working on the draft Digital Personal Data Protection Bill⁸ and will be soon adopting the fairly rigorous framework.
- 3. Performance Guarantees:** One aspect of responsible deployment of AI models that are not discussed extensively is the performance expectations of such systems. Most modern AI systems report performance in statistical terms, and it is often not clear what that translates to in terms of accepted norms of performance in the intended application. While AI systems might have a performance guarantee of, say, no more than 7 out of every 100 positive predictions are wrong, we do not have an easy way to measure what this means in terms of other metrics such as loss of productivity, lost opportunity cost, etc. Further, AI systems are tested on fixed data distributions and could prove to be arbitrarily brittle on data outside this distribution. We need to develop regulations and, consequently, the technology to mandate performance guarantees for each application domain of interest. At the same time, one needs to understand that we

³ NITI Aayog's Principles for Responsible AI (<https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>)

⁴ NASSCOM's Responsible AI Architect's Guide (<https://indiaai.gov.in/responsible-ai/pdf/architect-guide.pdf>)

⁵ UNESCO Ethics of AI Agreement (<https://www.unesco.org/en/articles/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>)

⁶ EU Ethical Guidelines for Trustworthy AI (<https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>)

⁷ EU's GDPR Regulation (<https://gdpr-info.eu/>)

⁸ India's Draft Personal Data Protection Bill (https://www.meity.gov.in/writereaddata/files/The%20Digital%20Personal%20Data%20Protection%20Bill%2C%202022_0.pdf)

cannot guard against all possible eventualities and hence would need to build appropriate insurance models for AI systems, and actuarial scientists need to develop appropriate risk/underwriting models for AI systems in various applications.

- 4. Fairness and ethics:** These again are aspects that have received much attention in both research and popular media. AI systems that routinely identify people of a certain race as more likely to be criminals or that women are more likely to be nurses than doctors have been reported often. AI chatbots have been known to become toxic in their language with suitable encouragement from the other participant. While we have made much progress in addressing such issues, techniques and policies must be developed for the local social context. India has her own set of discriminative axes, and we cannot blindly adopt the US-Euro centric views on discrimination. Prejudices exist against certain castes or people from certain regions. And these change from region to region and from group to group. Some of these biases will be reflected in AI systems trained on this data and we need to identify and guard against these biases. This requires us to first understand the human biases and the ways in which they manifest and calls for closer collaboration between AI researchers and social scientists.
- 5. Human-AI partnerships:** The full impact of AI will be available when we develop Human-allied AI workflows. Responsible deployment of AI systems will require us to understand how work will be disrupted with AI and how we will develop new workflows with AI-in-the-loop. Companies will have to re-train/re-skill workers to operate effectively in a hybrid environment.

The Centre for Responsible AI (CeRAI) has been set up at IIT Madras to study these and other themes under the three themes of Making AI Understandable, AI & Safety, and AI & Society. The centre is set up in a multi-stakeholder consortium model, with participation from industry, government, legal experts, social scientists, and industry bodies apart from different academic institutions. There is an exciting journey ahead of all of us in shaping the AI-enabled transformation of the world.

(5) Dr. Rohini Srivathsa

She is the National Technology Officer at Microsoft India. In her role, she is responsible for driving innovation and growth through 'tech-intensity' – across industry and the government. She is a thought-leader, advisor and contributor to several industry bodies, government, and academia forums and publications.



Artificial intelligence may well represent the most consequential technological development of our lifetime. Today's cutting-edge AI is a powerful tool for advancing critical thinking and stimulating creative expression, making it possible not only to search for information but to seek answers to questions, help people uncover insights amid complex data and processes, and speed up our ability to express our ideas.

New technologies have the potential to bring out both the best and worst in people. In the days ahead, AI will be celebrated for its benefits but also scrutinized and, to some extent, even feared. For AI to benefit everyone, it must be developed and used in ways which secure people's trust. The opportunities for society are significant, and the responsibilities for every stakeholder in society are heightened.

At Microsoft, we've been working to build a responsible AI approach since 2017 when we launched the AI and Ethics in Engineering and Research (AETHER) Committee with researchers, engineers and policy experts to focus on responsible AI issues and help craft the AI principles that we then adopted in 2018. Microsoft's approach follows a risk-management philosophy and is focused on proactively establishing guardrails for AI systems so that we can anticipate and mitigate risks, while maximizing the benefits.

Over the past few years, principles around developing AI responsibly have proliferated and, most stakeholders tend to agree on the need to prioritize issues like transparency, fairness, accountability, privacy, and security. While principles

are necessary, having them alone is not enough. The hard work begins when we need to turn those principles into practice. Microsoft's Responsible AI Standard, a framework to guide how we build AI systems, is a significant step for our practice of responsible AI. It sets out practical approaches for identifying, measuring, and mitigating harm ahead of time, and requires teams to adopt controls that can secure beneficial use and guard against misuse.

We understand it is critically important to think holistically about the AI systems we choose to build. We need to think deeply about sociotechnical impacts. Teams need to build the muscle of asking ground-zero questions, such as, "Why are we building this AI system?" and, "Is the AI technology at the core of this system ready for this application?" In the fast-evolving and nuanced practice of responsible AI, it is impossible to reduce all the complex sociotechnical considerations into an exhaustive set of pre-defined rules. And this has led to a process for ongoing review and oversight of high-risk cases and close examination of issues and implications.

At the foundation of all the efforts towards responsible AI, we should have three key goals:

1. AI should be built and used responsibly and ethically. Countries and communities should use democratic law-making processes to engage in whole-of-society conversations about where the lines should be drawn to ensure that people have protection under the law. And just like AI itself, regulatory approaches around the world will need to be interoperable and adaptive.
2. AI should advance international competitiveness and national security. Technological prowess is core to international competitiveness and national security in today's dynamic environment. AI is the next frontier of that competition. Democratic societies will need strong technology leaders to help advance AI, with broader public policy leadership from governments, industry and academia on topics including data, AI computing infrastructure and talent.
3. AI should serve society broadly, not narrowly. We need tech advances to keep pace with the ability of people and institutions to adapt, so that workers can be empowered by AI, students can achieve better educational outcomes and

every person and every organization can enjoy inclusive and sustainable economic growth. AI must serve people and the planet and can play a pivotal role in helping address the climate crisis.

In the year of India's G20 presidency, its role as a digital leader is creating global influence, especially on leveraging tech to drive impact at scale. The Indian tech ecosystem is proactively engaged in conversations related to Responsible AI to help evolve the right frameworks. Building on the idea of 'AI for All', the Government of India has outlined the considerations and approaches for managing AI systems and has documented the principles for operationalizing the responsible deployment of AI. NASSCOM has launched a Responsible AI Hub and Toolkit to bring together these efforts in India with strong participation from industry, government, and civil society stakeholders.

As AI models continue to advance, we will need to address new and open research questions, close measurement gaps and design new practices, patterns, and tools. This transformative moment for AI calls for a wider lens on both the positive and negative impacts of the technology and a broader dialogue among stakeholders. We need to have wide-ranging conversations and commit to joint action to ensure that AI is built responsibly for the future.



सत्यमेव जयते

Comptroller and Auditor General of India
<https://www.cag.gov.in>