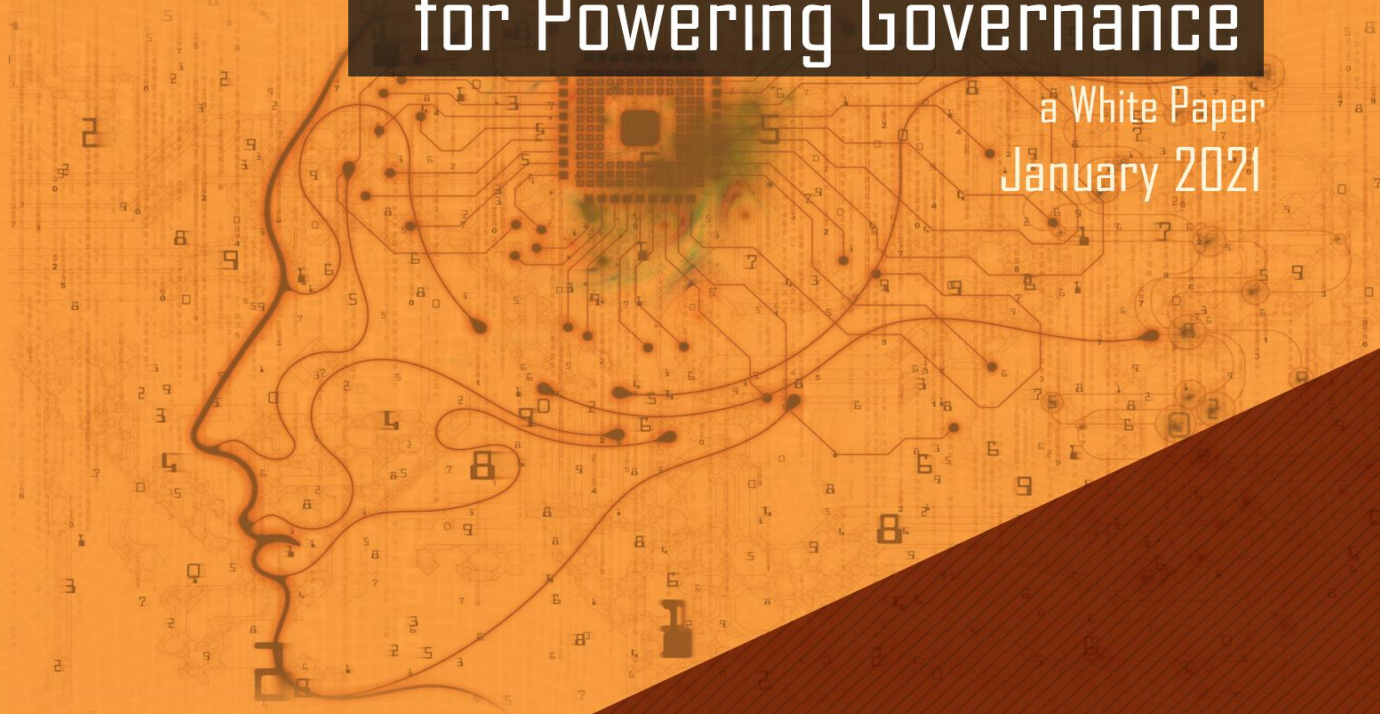
एनआईसी
National Informatics Centre

Digital India
Power To Empower

सत्यमेव जयते

ARTIFICIAL INTELLIGENCE
for Powering Governance

a White Paper
January 2021



Centre of Excellence in Artificial Intelligence

Ministry of Electronics and Information Technology
Government of India

Artificial Intelligence

for

Powering Governance

(a White Paper)

January 2021

Centre of Excellence in Artificial Intelligence

National Informatics Centre

Ministry of Electronics and Information Technology

Government of India

Executive Summary

Government has already come a long way in digital transformation and is actively capturing data for assessing effectiveness of implementation of various Government Schemes. Now is the time to explore new ways to analyze, integrate and share data to fuel program innovations and service redesign for better execution of schemes at ground level.

The Government and the governed are intertwined delicately in the fabric of society. The data collected by government is increasingly being analysed to see effectiveness of implementation of government schemes. However, it is reactive in nature. To make decisions capable of being fluid and effective at the execution level, it is needed for Artificial Intelligence to be able to learn from the data and simulate the results for the diverse ecosystem in a diverse country like ours.

It has become imperative to explore Emerging Technologies like Artificial Intelligence to build better models of Governance, stop wastage & improve service delivery. If we have to realize the dream of a connected government, it is imperative that transformative solutions that are projected by disruptive technologies like AI, are imbued by the government and society at large.

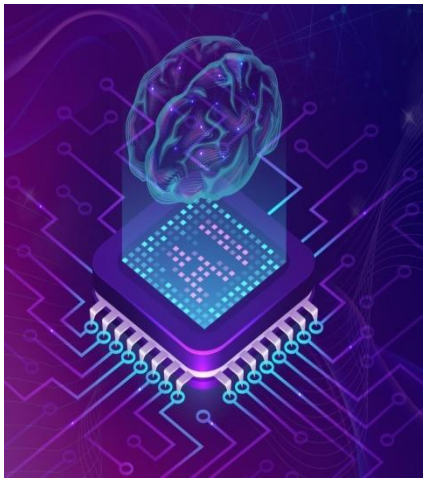
With this in mind, India's premier Government IT Organisation NIC, setup Centre of Excellence in Artificial Intelligence in January 2019. The objectives of CoE are to take eGovernance to the next level, by making use of AI Technologies like Computer vision, Natural Language processing and Conversational AI like chatbots and voicebots, facilitating convenience and ease of use of facilities by Citizens and empowering Government to understand citizens' requirements by harnessing their feedback in social media and other forms of unstructured data like images, videos and documents.

This page has been intentionally left blank.

Contents

This page has been intentionally left blank.

1. Basics of Artificial Intelligence (AI)



India, being the fastest growing economy with the second largest population in the world, has a significant stake in the Artificial Intelligence (AI) revolution. Niti Aayog in its National Strategy for Artificial Intelligence has identified '#AI for All' as the theme for leveraging the full potential of AI to meet country's unique needs and aspirations. It has identified five critical sectors for AI intervention namely Healthcare, Agriculture, Education, Smart Cities & Infrastructure and Smart Mobility & Transportation[1].

I. What is AI –

Artificial Intelligence is simulation of human cognitive processes by machines. This signifies a change from normal setup where the input and the process are fixed and the system outputs results on expected lines.

Two very distinct approaches are available for reasoning: inductive and deductive. Both approaches have their advantages. In a traditional classroom settings, a maths teacher may conduct lessons by introducing and explaining concepts to students, and then expecting students to practice the concepts; this is deductive reasoning approach and in respect to systems it is the approach used by traditional programming applying a rule to the data.

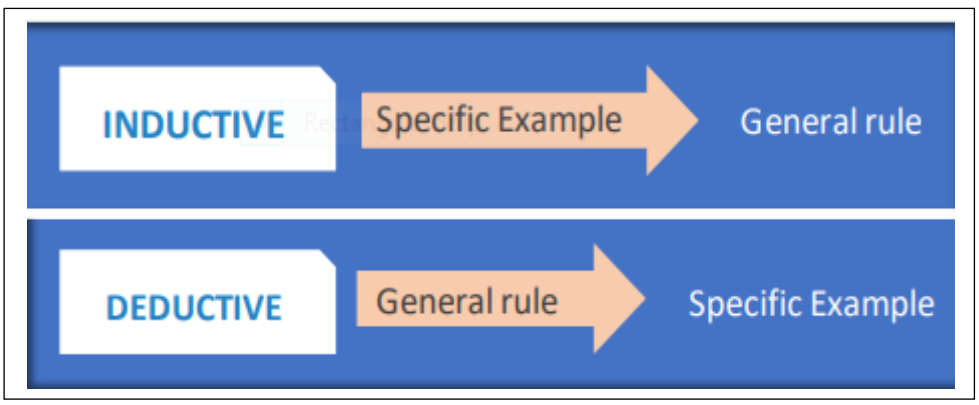


Figure 1 : Deductive vs Inductive Reasoning[2]

Conversely, inductive learning is used in tutorials to help students learn more effectively by observing examples. This way we can learn the rules by observation and check to see if there is a pattern. We can then try applying the rule in different situations to see if it works. This approach of learning from data, is used by Machine Learning algorithms and Deep Learning models in AI.

All such smart technologies, which can take input from the environment, learn and respond probabilistically in new and changed environment rather than deterministically (If *condition* then *some response* rule may fail for cases it has not observed before), with a degree of confidence that exceeds a threshold level set up by humans, qualify for being called as Intelligent Agents.

Where AI is involved, intelligent agents can take signals from environment and respond automatically, learning patterns from historical data, that programmers cannot always anticipate.

Figure 2 : AI based Headcount of Children in ICDS Scheme by NIC - DGRC Patna

II. What is Machine Learning –

Machine Learning is an integral part of Artificial Intelligence. There are various aspects to this learning. In machine learning we (i) take some data, (ii) train a model on that data, and (iii) use the trained model to make predictions on new data. The process of training a model can be seen as a learning process where the model is exposed to new and unfamiliar data, step by step.

At each step, the model makes predictions and gets feedback about how accurate its generated predictions were. This feedback, which is provided in terms of an error according to some measure (for example distance from the correct solution), is used to correct the errors made in prediction.

Machine Learning (ML) algorithms usually work on structured data and can mainly be categorized as of two types, Unsupervised learning where system tries to cluster data on its own, and Supervised Learning where labelled or annotated historical data is provided to the model, and the system gets trained on this data. Commonly used for forecasting and trend prediction like in agriculture/weather forecasting, Financial cash flow/ credit risk management etc.

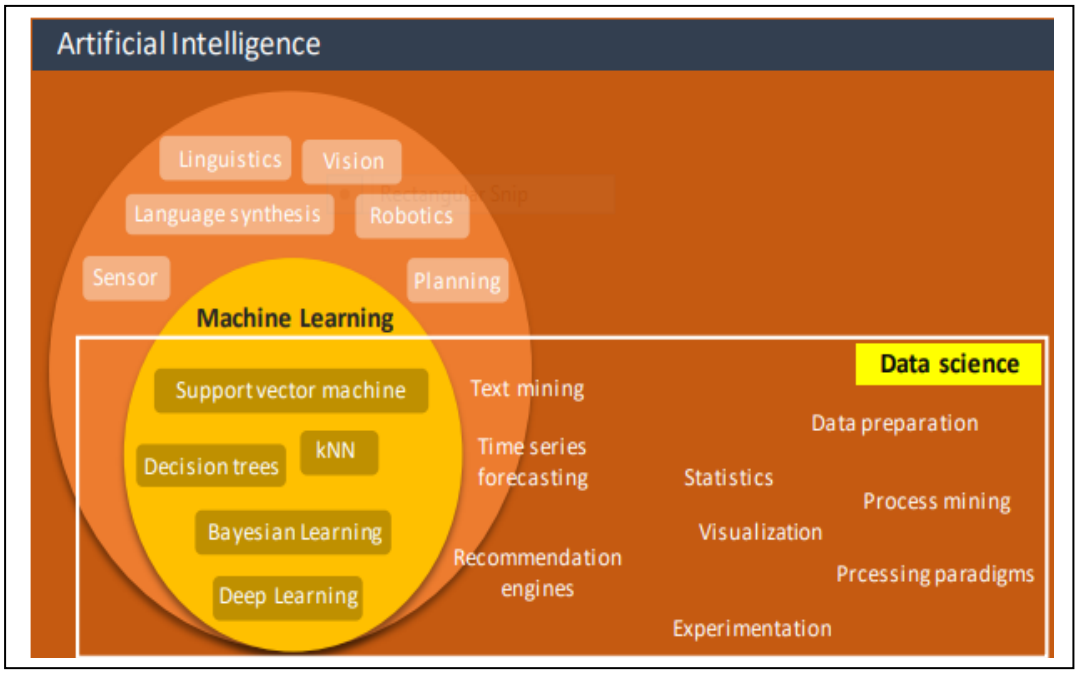


Figure 3 : AI, Machine Learning & Deep Learning in Perspective [3]

III. What is Deep Learning (DL) –

Deep Learning is a subset of Machine Learning and is the state of art technology. Deep Learning involves Neural Network algorithms and it has been developed on the concept of neurons in the brain.

There are many layers of neurons that abstracts information and achieves complex nonlinear function on similar lines to working of brain. It is not always possible for human to decipher the complex logic developed by training these neural network models.

Here also the ground truth or annotated/labeled data need to be given for the predicted variable(s) so that system learns from the annotated data. Advantage here is that the system in this case does the feature extraction by itself. However, wherever trace back is required like in case of banks where trust needs to be built for the model being used, machine learning is better suited as all the features are directly engineered by domain experts.

IV. AI Disciplines –

Image and Video Analytics - Computer Vision systems can be trained to detect objects in images, do human face detection, one to one face verification, face recognition from multiple images and in videos. Image and Video analytics can also be used for surveillance, emotion detection, trajectory analysis, scene reconstruction and much more.

Figure 4 : Helmetless Driving by NIC Andhra Pradesh State Unit

Text Analytics - Second is Cognitive science Communication, and it is the domain of AI that deals with Natural Language Processing (NLP) and Text analysis. This is a comparatively harder problem then vision, as the data is less structured here, text may be mixture of two or more languages, many ambiguities exist even in one language like the word 'going to bank' as in bank of a river or bank for encashing a cheque.

Sentiment analysis based on social media feedback is a popular exercise in text analysis. Text Summarisation and Topic Modeling for news etc. is another area where AI is gaining popularity.

Conversational AI – Another area of AI usage is building Chatbots and Voicebots, to assist human by answering repetitive questions. It uses many cognitive aspects like Natural Language Understanding (NLU) component and a Dialogue component to manage the conversation session. Speech to Text (Automatic Speech Recognizer or ASR) to understand human speech and Text to Speech (Speech Synthesizer) for giving response to user.

Figure 5 : NICCI Chatbot developed by NIC, Rajasthan State Unit

V. Maturity of AI Technologies –

Usage of Artificial Intelligence Technologies, has become mainstream in the last couple of years primarily due to exponential growth of data and hardware having been able to match up with Machine Learning algorithms.

Computer Vision technologies have matured, and many models exist for transfer learning, instead of trying to develop one from scratch. These models have been trained on millions of images worldwide. NLP and machine translations are harder as annotated corpus and trained models in Indian languages are scarce and considerable work needs to go in, to develop any text based models.

In addition to identifying typical historical trends, AI is excellent at identifying transactional patterns in structured data, which is useful for isolating transactions that do not align with those patterns. In addition to pattern recognition, AI and machine learning are powerful tools for identifying relationships between data sets [4].

2. Some AI Usecases –

AI work in NIC being pursued can be categorized under 4 group heads.

Figure 6 : AI Service Disciplines Work Categories by NIC

Conversational AI in form of virtual assistants, chatbots and voicebots have gained popularity nowadays, as it is used as a means of augmenting the system to automate the task of answering user queries repetitive in nature and provides administration support to speed up task completion by relieving them of such tasks which can be answered by the system.

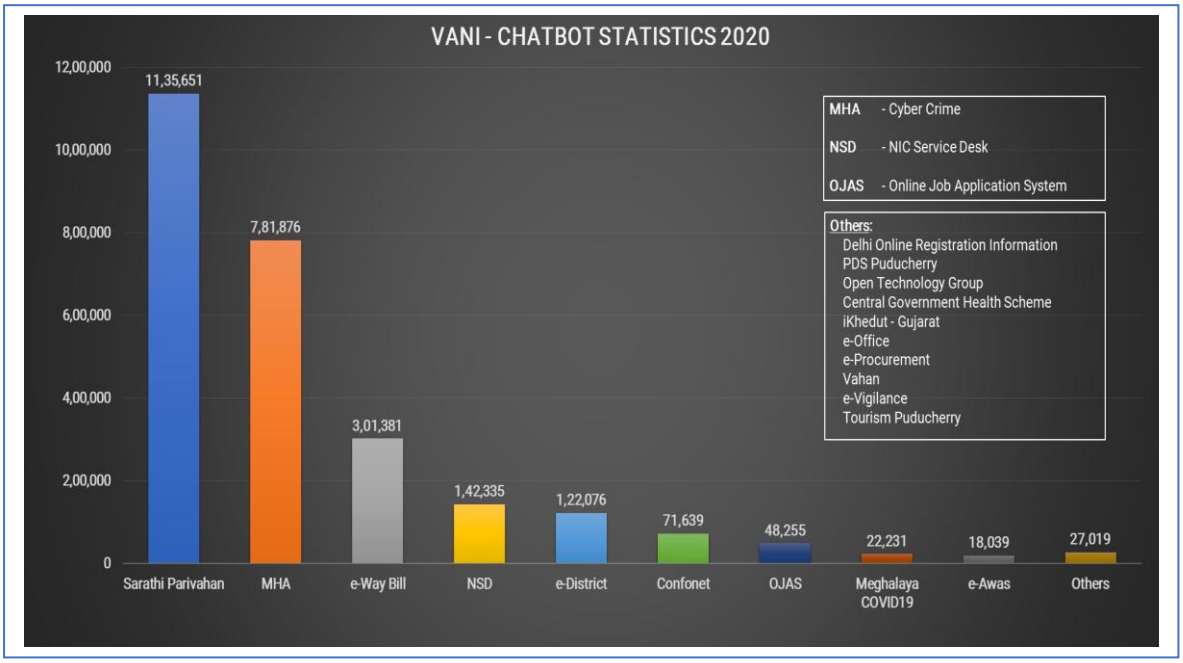


Figure 7 : Virtual Assistance by NIC - Chatbot Usage

20 Chatbots and 8 Voice based Bilingual Support Services released by NIC for different departments/ sections of citizens, are listed below and more are in the offing :

- NICCI chatbot for citizen on Rajasthan Govt. portal Pehchan
- CONFONET chatbot for National Consumer Disputes Redressal Commission website
- Chatbot for RTO on license related queries
- Chatbot for OJAS (Gujarat Public Service Commission)
- Chatbot iKhedut chatbot for farmers in Gujarati & English
- Chatbot for eAWAS and vigilance – Chandigarh
- Chatbot for PDS – DBT Puducherry
- eWayBill chatbot
- Covid19 Chatbot for Meghalaya in English, Garo & Khasi languages

One Important Usecase of Voice Based Bilingual support in English & Hindi over basic telephony is showcased here :

Vaidya Vani – vOPD Teleconsultancy Services – This is a Tele-medicine Application where the patients are attended virtually, just like they are attended to, while on a physical visit to the OPD department of the hospital. The application is designed keeping in view the lowest rung of the society that they can use a common telephone to do a virtual visit to the OPD.

The application is designed and developed by the Artificial Intelligence and Resources division of National Informatics Centre in collaboration with Lady Hardinge Medical college & Hospitals, New Delhi.

There are facilities for 3 levels of doctors to be logged in at the same time for a department, namely: Junior, Senior & Specialist Doctors.

The patient calls a standard 8 digit telephone number which has facilities for catering to multiple calls at the same time. The patient selects the conversational language, then the name of the department like ENT, eye, ortho, medicine, etc. by speaking. Then the patient has to register by providing the mobile number for receiving the prescription via SMS.

After this, the patient is connected to the Junior Resident doctor. If there are many patients, they would be waiting in the queue and the patient would also be suitably notified.

After connecting to the Jr. resident doctor, the doctor views the patient details already filled along with the phone number and the selected conversational language. This enables the doctor to continue the conversation in that language. The doctor would then query the patient, fill in the remaining demographic details and fill the patients health issues including allergies, associated ailments and addictions for the preliminary diagnosis.

Once completed, the Jr resident doctor would transfer the patient to the senior doctor for further diagnosis and prescription.

Figure 8 : virtual OPD for Patients using Automatic Speech Recognition

On the screen at the top right, there is the return caller button which lights up if the caller had a call drop while waiting in the queue or while conversing with the doctor. If the caller dropped off at any levels of the queue, they would be returned to the respective queue with priority of getting connected to the doctor at the next turn.

If the patient is on a revisit, the doctor can view the details using the search button and also the history of the patient, the doctor can peruse each previous details from various departments

If the doctor finds the current case requires emergency attention, he can direct the patient appropriately by clicking the emergency button. The patient hears the necessary information over phone and also gets a SMS and an email with emergency situation handling information as perceived by the doctor.

The Sr. resident and Specialist doctors also have similar screens and facilities. The sr. resident doctor or specialist doctor only can create and send prescriptions. The patient moves from one queue to the other much like on a physical visit.

This entire process has been facilitated using Automatic Speech Recognition for converting Speech to Text in Hindi & English.

Some Other Important Voice based Support in English & Hindi on NIC's VANI are as follows :

- PM Kisan Samman Nidhi Yojana Bilingual Voice Support in English & Hindi
- Voice based support for IVFRT, Kailash Mansarovar Yatra etc.

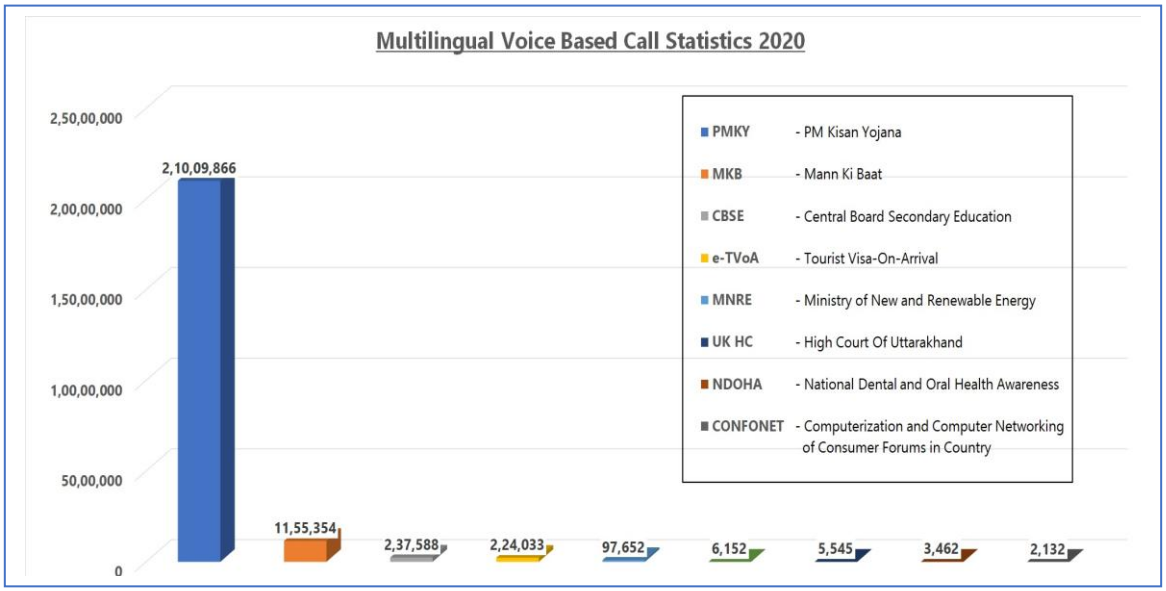


Figure 9: VANI - Bilingual Voice Support in Hindi & English

II. IVANI - Image & Video Analytics Assistance by NIC - Important Usecases

Swachh Bharat Mission Urban - Government of India has announced a number of schemes that are focused on improving the material condition of below poverty line citizens both in urban and rural areas. Most of these schemes are implemented as well as managed through a Digital Platform.

The citizens apply for these schemes online directly through the platform or through Citizen Service Centres and also submit documents such as Identity Credentials, bank account details, applicant photo etc. which help in validating their identity and establish their eligibility for the scheme.

These applicants upload photos of progress of work to get instalments from these portals through Direct Benefit Transfer (DBT). For establishing right utilisation of funds under these schemes, checks are usually conducted by officials before transfer of funds to beneficiary's bank account. However delays are usually inevitable in such cases .

A Typical Use case was Swachh Bharat Mission where Image Analytics was used for active intervention in facilitating citizens get their final instalment with ease.

As citizens could see automatically whether they have uploaded correct geo-tagged constructed toilet photo or not, through this intelligence augmentation through SwachhAI Mobileapp and citizens were spared from waiting in queue for verification of uploaded photo and could reload correct photo instantly if required.

Figure 10: SwachhAI in aid of the Citizens

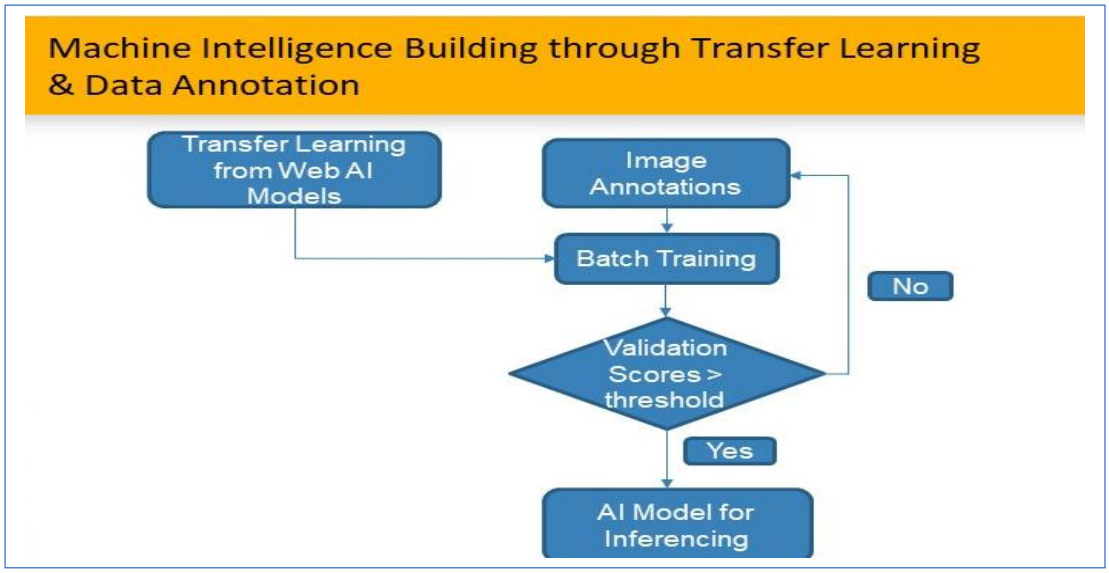


Figure 11 : AI Life Cycle for Model Building

SwachhAI, an AI enabled mobile app was **launched by Hon'ble MOS(I/C) of Housing and Urban Affairs Sh. Hardeep Singh Puri** on 13th August 2019.

Figure 12: Image Classification with 99% accuracy

Image analytics can also be used for checking progress in work undertaken by Govt. departments like construction of water tanks, road construction etc. over time, crop pest detection in agriculture, radiology for human health care etc.

Facial Recognition/ Face Verification Services - Another area where work is progressing is the area of facial recognition. Facial recognition has been attempted by NIC West Bengal Unit to track missing child from missing children database. Facial recognition works by identifying distinct points on an individual's face and creating a unique map of it. It is therefore more akin to a fingerprint rather than a photograph. NIC has successfully facilitated facial verification services in the following applications :

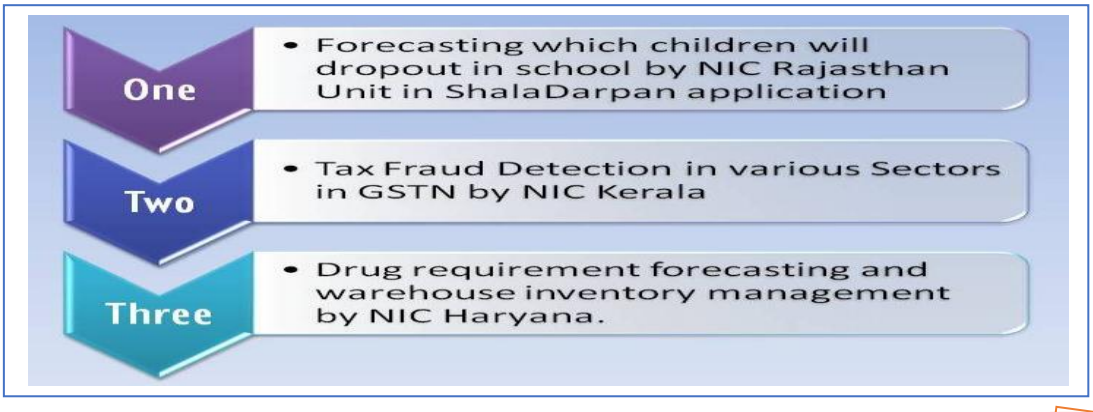
Figure 13 - Hon'ble Chief Minister launching mPension Manipur

Video analytics and AI can be used in managing traffic, managing crowd, checking attendance in video conferencing, parking lot management, looking for suspicious behavior at airports and other security sensitive installations etc. Feeds from web cameras, mobiles, CCTV cameras etc. can be used to improve the quality of governance.

III. AI in Trend Forecasting –

Machine Intelligence using structured data is an area which can yield rich dividends for the society. Attribute Data can be either Categorical such as Male/Female, Malignant/Benign or Numerical such as age, tumor size etc. These attribute data is called feature set or predictor variables and we need to predict values for or classify the dependent variable based on this feature set.

Some use cases attempted by NIC are as follows :



GSTN Tax Fraud Detection in Various Sector – Salient Features of the Machine Learning Life Cycle process adopted by NIC Kerala for the process of analysing sectorial data in GSTN filing is as follows :

- Prepossessing was done on the GST tax return data.
- Heat map analysis was done to identify parameter pairs with high overall correlation (for each sector).
- Pearson's correlation coefficient between the pairs was calculated and setup for clustering.
- Cluster analysis was carried out followed by Benford analysis on all the 5 sectors namely gold, steel, vehicle, textile, timber.
- The suspicious list of dealers were generated for all the sectors.
- he list of suspicious dealers in timber sector was shared with the department for examination.
- Provided for timber sector was found to contain most number of dealers to whom notices were served.

Tax return data for the period of 2017-2019 was considered. A sample for unsupervised learning for the patterns in the tax filing data for the timber sector is given below as an illustration.

.

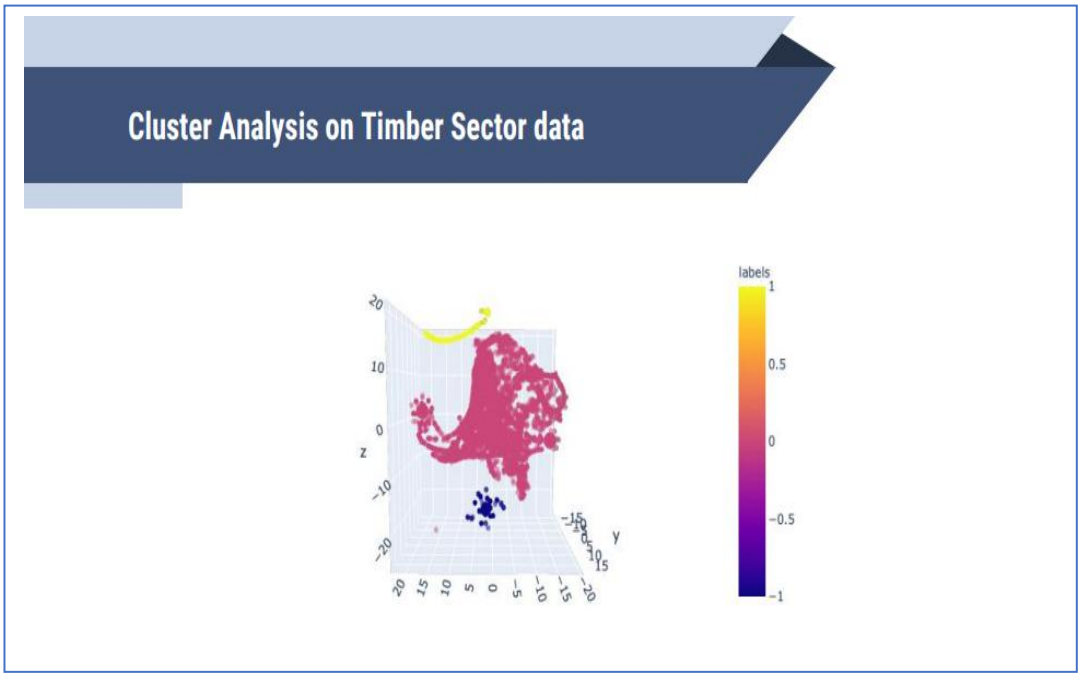


Figure 14: GSTN Sector wise Analysis for Tax Fraud Detection

IV. TAANI - Text Analytics Assistance by NIC –

Text Mining or Text analytics is that branch of AI that delves into natural language processing and can transform unstructured text into meaningful insight from large text corpus. Natural Language has an extremely rich form and structure. It is also very ambiguous.

A lot of work goes into extracting the textual content that may be from web, email, news, social media etc. and pre-processing it and making it possible to derive the semantics of the text.

Cognitive Search for MACP cases - COE-AI has developed a model for Cognitive Search for assisting lower Judiciary on Motor Accident Claim Petitions (MACP Case orders) case precedents.

It is extracting text corpus from case order pdfs, customised Named Entity Recognition(NER) to identify Victim's age, income, dependents and status like deceased or grievously injured or just injured from the case orders.

It has applied unsupervised learning to build language model for legal text and uses the results and annotated text for classifying case order outcomes which is a supervised learning model, so that it can predict for an incoming petition what will be the case outcome along with a confidence score for the predicted outcome. It informs of the Citations used and compensation calculated for similar cases. It is currently being integrated with eCourts.

MinT tool - One Sentiment Analysis Tool has also been developed by NIC's Karnataka State Unit. Usually there is a need for a mining and classification of feedback/ suggestions/ complaints/ comments of citizens in audio/ text format on different government schemes/services received from Emails, Web Applications, Social Media sites like Twitter, Facebook.

It generates a dashboard which can help senior officers to understand citizens' needs and priorities. The tool does intent analysis, sentiment analysis and trend prediction. It is already implemented for a government department and is able to treat feeds from multiple Indian languages. It has been implemented with a Government Department in Karnataka.

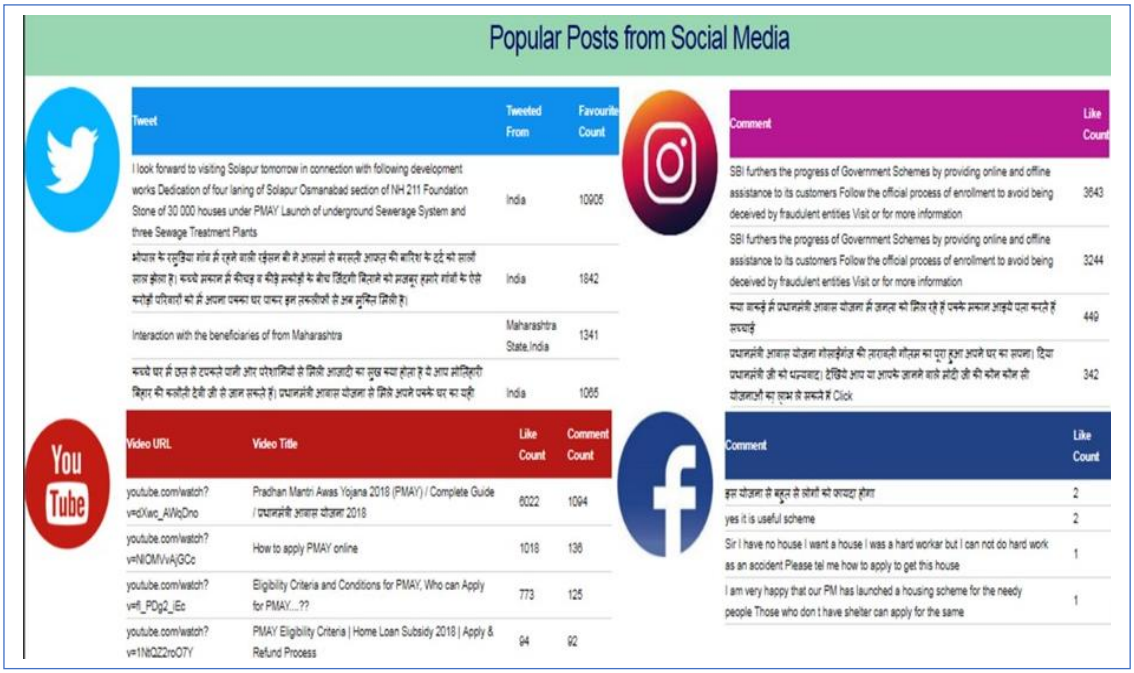


Figure 15 : Sentiment Analysis

3. Road to adapting AI –

There are also some specific challenges to knowledge discovery from the data and the road to adapting AI in the organisations. Some of them are as follows :

I. Technology Adoption –

Organisations that move to integrate AI into their flow of work may need to make their executives learn more about AI; deepen their perspective on how to organize their business around AI; and rethink and develop a more expansive view of the landscape in which their business operates.

Have the budget, time, and resources to bring on consultants and new hires to flesh out AI applications, without expecting any kind of near-term return on investment.

Have a leadership team which believes in AI, and is willing to learn new terminology and methods to foster the growth of AI in their organization.

Have a clear understanding of the problem, and robust understanding of the domain to validate why AI is being used as the solution.

II. Data Readiness for Machine Learning –

Foremost to AI adoption is data. Nicholas Piette who is the Chief Evangelist at Talend, said the following words in this regard [5]:

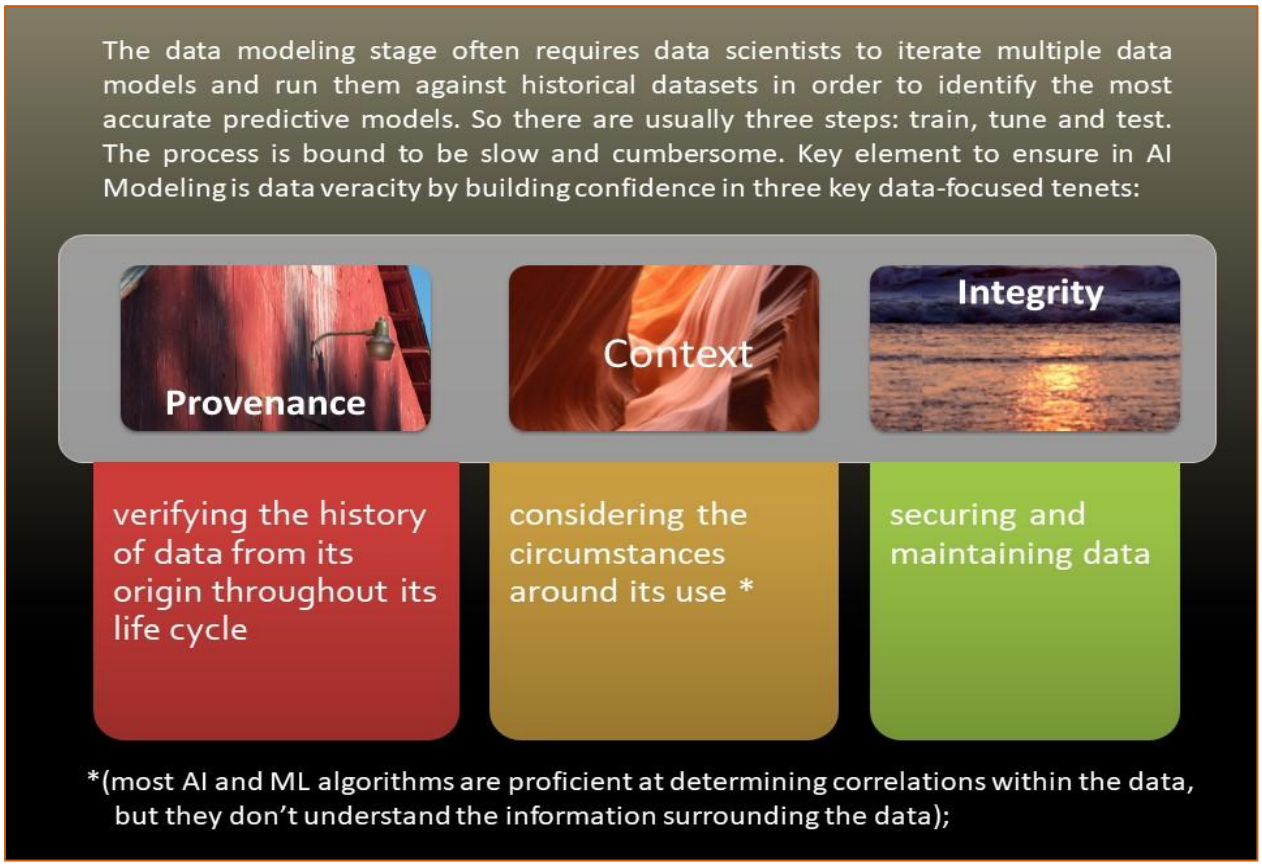
"100% of AI projects are subject to fail if there are no solid efforts beforehand to improve the quality of the data being used to fuel the applications. Making no effort to ensure the data you are using, is absolutely accurate and trusted—in my opinion—is indicative of unclear objectives regarding what AI is expected to answer or do. I understand it can be difficult to acknowledge, but if data quality mandates aren't addressed up front, by the time the mistake is realized, a lot of damage has already been done. So make sure it's forefront."

Figure 16 : The 5 Rs for Data Quality – Nicholas Piette

Exploratory Data analysis is foremost to applying ML to data, this may lead to data cleaning. At the end of the day, users will need to have data scientists on their team to make the most of the AI tools. Whoever is feeding the data into these tools, need to have the confidence that the data is clean, free of biases and free of anomalies.

As we adopt multiple machine learning tools to assess data at various stages of a process or for a particular task, they may need to restructure the data into the format suited to that machine learning tool. A data quality assessment matrix is provided in Annexure – II for ready reference.

III. Data Modeling –



IV. Data Security, Privacy and Ethics –

As AI/ML algorithms are part of open source frameworks, they are as easily available to hackers who can also start leveraging the benefits of AI/ML as much as a genuine user, which is in a way inevitable, with so many countries pouring in funds on advancing military AI and cyber warfare, not to mention dark web. So it is important to understand the context of privacy & security of the data.

With Big data, this challenge had surfaced and exacerbated with AI. Even with levels of anonymization of data, it is possible to add up the chunks of information to construct the whole and pose a challenge to the privacy. However, it is possible to introduce noise in the data to introduce enough ambiguity that it may not be possible to reconstruct personal information from inferences made from different anonymized datasets. There are many other techniques being used nowadays to obfuscate sensitive information in the data when outsourcing model building exercise to third parties[6].

In context of ethics, it is useful to mention that AI models are dependent on data to get trained and data can introduce bias into the models unintentionally like model being biased towards majority of training data and less towards data which are more infrequent. More so, in public sector use of AI, accountability has to be built in use of the AI model otherwise hidden social and ecological cost can be seen as an after effect in such a discourse.

4. COE-AI @ NIC

Centre of Excellence in Artificial Intelligence (COE-AI), was inaugurated by Honorable Minister of Electronics & Information Technology, Shri Ravi Shankar Prasad in National Informatics Centre (NIC) Headquarters, on 10th January 2019.

Objective is 'Inclusive AI'. The primary goal of setting up COE-AI is to give impetus to the use of Artificial Intelligence Technologies to improve service delivery to citizens by enhancing eGovernance in various key sectors of Government Initiatives.

Key role of the COE-AI@NIC is educating the government departments/ agencies about the potential use of AI and plan following activities:

- Recommend technologies and architectures
- Disseminating knowledge and best practices
- Maintain website https://ai.nic.in with different use case studies
- Creating technology demonstrators

NIC is skilling its officers in AI in batches, through in-house programs and residential trainings in institutes of repute.

Infrastructure Facilities – Keeping in view the high performance compute requirement for computer vision applications, as well as text analytics models using deep learning, NIC established supercomputing facilities at NIC Data

Centre to supplement and complement the future AI requirements of national level eGovernance projects running on Meghraj cloud.

Deep learning frameworks come with pre-built components that are easy to understand and code. A good framework reduces the complexity of work and helps smoothen model building exercise. The supercomputers have these open source frameworks and docker engine built in it which are attuned to parallel computation. Dockers allow for light weight images to be built with associated libraries and AI models developed with transfer learning sourced from the web.

NIC is building a Development platform that will help AI users to train datasets and further build an Inference Platform for deployment. With this vision, NIC has already procured 2 PFlops of Supercompute.

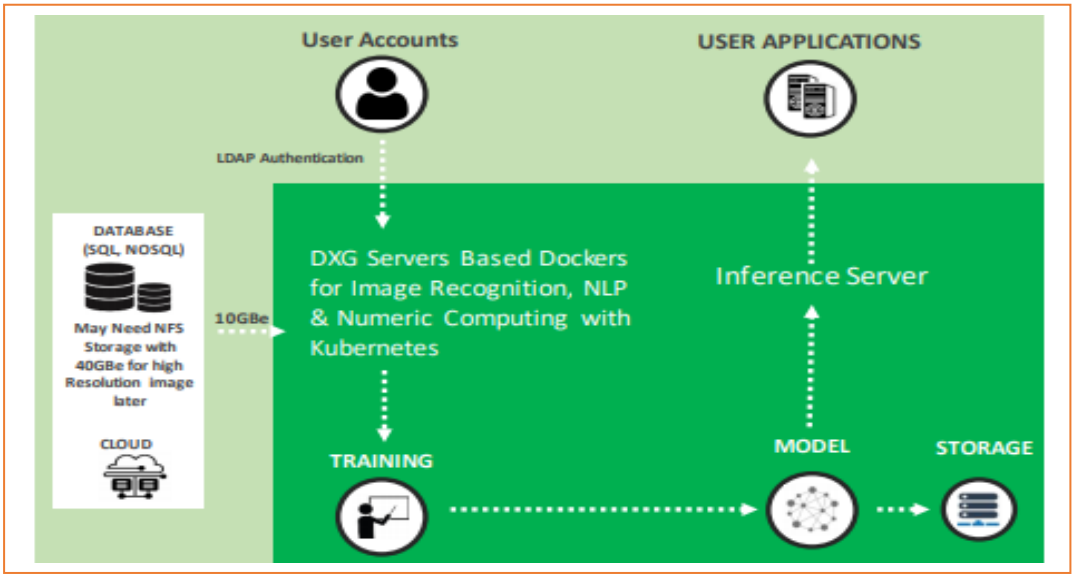


Figure 17 : Conceptual Architecture of AI Platform as a Service

AI Development Platform as a Service - AI models mostly use supervised learning, or semi-supervised learning, where the system can be used to search for patterns in the data and cluster them, and in next stage use such classes for further model training. Data Annotation Tools access may be made available to users of data annotation.

At present the AI Model Development platform has been provided by NIC to its officers in 25 Central projects and state units. For training, users are given accounts and docker image of choice depending on the type of AI application framework required at the backend where they can train an AI model from scratch or with transfer learning from AI models on web.

AI Inferencing Platform Testbed - The models that are built can be transferred to inferencing servers for deployment which will be less costly than development platform. With this, NIC plans to keep the data within its data centres while providing security & privacy to the users

AI Services Offerings - First AI Service category that is being made available for AI as a Service is that of Conversational AI, which will be facilitated as a paid service under NICSI. Second AI Service Category that is being made available is Face Verification Services. Also Personally Identifiable Information from different ID cards will be made available as a service for matching with the online form contents.

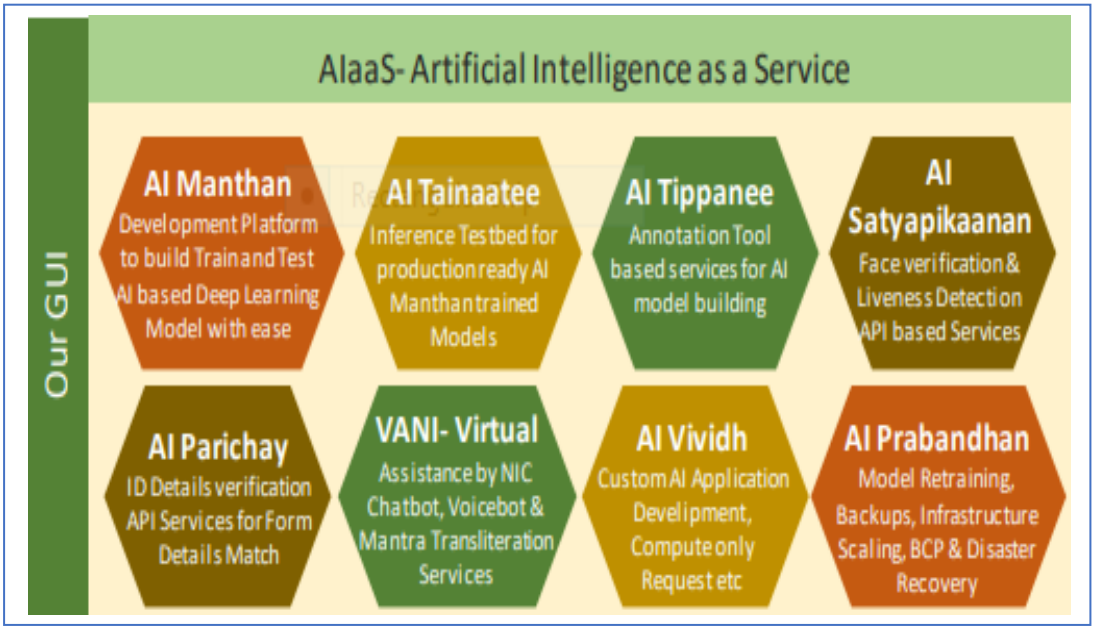


Figure 18 : AIaaS on NAIC - AI as a Service on NIC AI Cloud

AI Platform Services Offerings Support – NIC plans to build a Services Support model to offer services to more and more Government Departments & organizations. It is planned to have two types of Managed Services :

- **Standard Support Services –** This will be a fully managed service where AI Development Platform or API based services is made available to user 24 X 7.
  The user will be provided with one time training for using the services. The user will also be provided with six months professional consultancy support as and when required to get the AI models off the ground.

- **Premium Support Services –** Under Premium support, the department will be on an annual subscription charge and support will be provided anytime for any retraining of the AI model required once it has been deployed.

References :

1. http://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf
2. https://www.netlanguages.com/blog/index.php/2017/06/28/what-is-inductive-learning/
3. https://www.sciencedirect.com/topics/computer-science/artificial-intelligence
4. https://www.afponline.org/ideas-inspiration/topics/articles/Details/why-hasn-t-finance-embraced-ai-for-forecasting-yet/
5. https://medium.com/@Talend/for-ai-to-change-business-it-needs-to-be-fueled-with-quality-data-b03cc508cd85
6. https://towardsdatascience.com/security-privacy-in-artificial-intelligence-machine-learning-part-6-up-close-with-privacy-3ae5334d4d4b
7. https://tdwi.org/articles/2017/06/29/ai-as-platform-as-a-service.aspx

Data Quality Matrix :

Define Data Quality Requirements
Is Data Captured Online at Source
Is the Data Capture Distributed across cross cultural Domains
Is Data Granularity maintained across Distributed Data Capture
Is there possibility of same data capture through multiple entry points like CSCs, Direct Entry by Citizens, Surveys by LSG…
Is the Data across the organization captured through Predefined Templates
Has any Data Codification Standardisation practices been adopted
Does any time barred Bulk entry get posted into database without data validations
Is data captured though different systems in different locations merged
Has Deduplication process been run & are the records uniquely identifiable
Is Data Range check done on each category of data entered in the system
Is Text data entered recoded to eliminate spelling differences
Is Data Source Captured along with timestamp of entry/updation
Is Original record archived on Updation
Is Data captured at source stored on different servers in different file formats
Is same Data stored on different servers in different file formats consistently changed on Updation
Is Data backup done at scheduled intervals
Is Data restore checked for consistency at scheduled intervals

Profile, Analyse & Assess Data Quality
Are Free Text Entries like issues or remarks Grouped & analysed for frequency, for arriving at a reduced Category set for that column
Is Mean & Variance & Range calculated & distribution observed for Numerical data
Are Outliers and their frequency checked and documented
Is it documented how Mean, Variance & Range distribution change after removing outliers
Has it been brought to the notice of user department & reasons assessed
Has it been documented from user experience whether to use that variable with or without outliers
Are Zero or null values documented. Can they be replaced by mean or modal value or a randomized value within the data range
Is the Data a sparse matrix
Is it possible that the data is sparse or dense depending on target groups
Is there reasonable assurance that the data collection methods being used donot produce systematically biased data
Are Data Collection and analysis method documented in writing and being used to ensure the same procedures are followed each time?
Are mechanisms in place to prevent unauthorized changes to the data?
Is data required from any other sector to complete the use case for AI application in domain under consideration?

Define Data Quality Metrics & Thresholds
Are the entities modeled within the enterprise captured and represented uniquely within the relevant application architecture
Does the data correctly represents the “real-life” objects they are intended to model
Is data consistency maintained across datasets such that two data values drawn from separate data sets do not conflict with each other
Is data completeness ensured, it may be seen as encompassing usability and appropriateness of data values
Is Timeliness ensured, it can be measured as the time between when information is expected and when it is readily available for use
Is Data currency ensured, it may be measured as a function of the expected frequency rate at which different data elements are expected to be refreshed
Is data conforming i.e. whether instances of data are stored, exchanged, or presented in a format that is consistent with the domain of values, as well as consistent with other similar attribute values
Is threshold for data conformance defined for ‘Acceptable’, ‘Questionable but usable’ and ‘Unusable’

Know Our Team

CONTACT US

Centre of Excellence in Artificial Intelligence

nic-aird@nic.in

ai.nic.in

011-24305211/24305747

This page has been intentionally left blank.

Centre of Excellence in Artificial Intelligence

Email: nic-aird@nic.in

Website: ai.nic.in

Phone: 011-24305211/24305747