

Advanced Certification in

# Applied Data Science, Machine Learning and AI



**Electronics & ICT Academy**  
**Indian Institute of Technology Guwahati**

An Initiative of Ministry of Electronics & Information Technology(Meity)



इंटरकॉम की एवं  
सूचना प्रौद्योगिकी मंत्रालय  
MINISTRY OF  
ELECTRONICS AND  
INFORMATION TECHNOLOGY

## Projects

Duration: 9 Months (240Hrs)

DS | ML | DL | AI | Generative AI

Advanced Generative AI & LLMs Are Covered



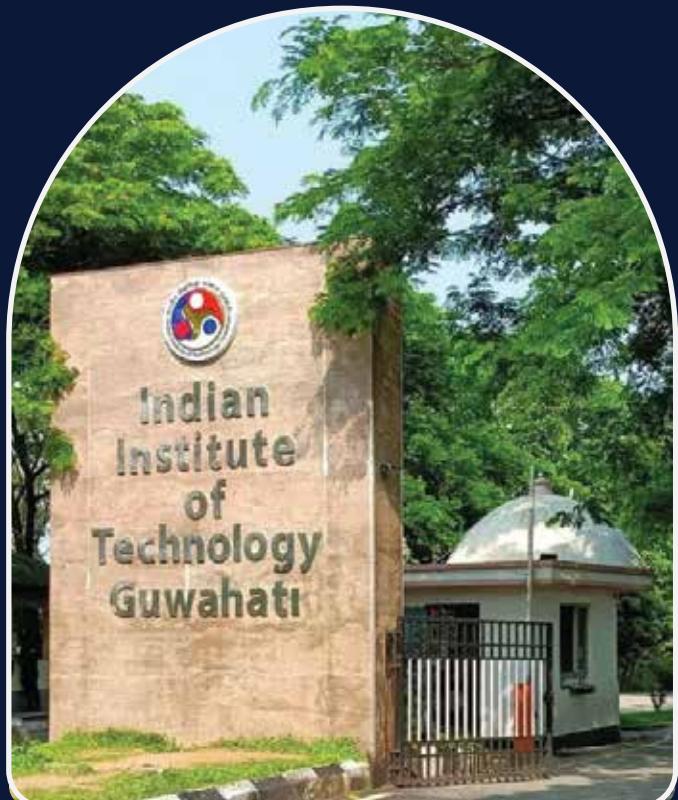
In Collaboration With



## About The IIT Guwahati

Indian Institute of Technology Guwahati (IIT Guwahati), the sixth member of the IIT fraternity, was established in 1994. The institute began its academic programmes in 1995. Currently, it comprises eleven departments and three interdisciplinary academic centres, covering major disciplines in engineering, science, and humanities. IIT Guwahati offers BTech, BDes, MA, MDes, MTech, MSc, and PhD programmes.

Within a short span of time, IIT Guwahati has established world-class infrastructure and earned a strong academic reputation.



## About The E&ICT Academy IIT Guwahati

Electronics and ICT Academy aims to provide specialized training to the faculties and working professionals, Arts, Commerce, Science colleges and polytechnic institutes by developing short term training programmes on fundamental and advanced topics in IT, Electronics & Communication, Product Design, Manufacturing. In addition, the Academy conducts specialized customized training programmes and research promotion workshops for corporate sector & educational





## About The Program

Most of the technological advancements in the recent times you see around have an element of Data Science, Machine Learning, AI & Generative AI in it. This 9-month Applied Data Science, Machine Learning & AI course is uniquely designed by E&ICT Academy, IIT Guwahati, and industry leaders, to make you master exactly what the Data Science, Machine Learning & AI jobs market demands. The objective of this certification program is to perfectly prepare you for the Data Science, Machine Learning & AI job roles you aspire for. You will learn Artificial Intelligence & Data Technologies, AI & Advanced Analytics, Intelligent Systems & Data Science, Machine Intelligence Technologies, AI/ML & Data-Driven Solutions skills through multiple ML, Deep Learning and AI business projects, transforming you into a sought-after New Age DS, ML & AI Specialist.

### Integrated with Generative AI & LLMs



Generative AI Modules &  
Architecture



Building LLM Applications



Generative AI Tools & Case  
Studies

# Capstone Projects

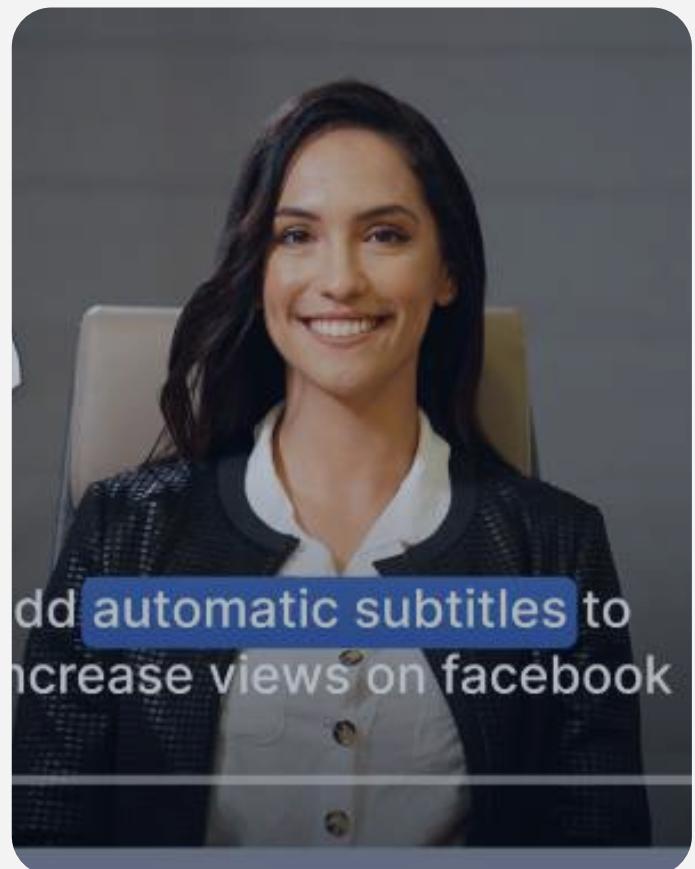
Industry projects will be a part of your Advanced Certification In Applied Data Science, Machine Learning & AI to consolidate your learning. Industry projects will ensure you have the real-world experience to start your career in DS, ML-AI.

- **10+ Essential Tools**
- **Designed by Industry Experts**
- **Get Real-world Experience**

## Auto Image Captioning

Automatic image captioning is widely used by search engines to retrieve and show relevant search results to the users. For example - to categorize personal multimedia collections, for automatic product tagging in online catalogs, and other areas of business and researches. Use CNN and LSTM to create a model that can automatically add captions to the image.

**Tools you will be using:** OpenCV, Tensorflow2/ Keras, Numpy, Pandas, Matplotlib



## Twitter Sentiment Analysis using Tweepy

Analyzing the tweets helps us in understanding the thoughts and sentiments of people over any popular topic. It helps us to understand what people are thinking about the trend. Here, as a part of this project, you will use Tweepy, Textblob, nltk, and other NLP libraries to analyze the sentiments from the Trending Twitter's tags.

**Tools you will be using:** Tweepy, Textblob, NLTK



## Building a RASA Based Chatbot

Rasa is a framework for developing AI-powered, industrial grade, powerful chatbots. The developers use it to create intent-based chatbots. In this project, we are going to understand some of the most important basic aspects of the Rasa framework and use RASA NLU and RASA CORE to build a conversational chatbot.

**Tools you will be using:** RASA NLU, RASA Core



## Real-Time Age, Emotion and Gender Detection using CNN

Use CNN and OpenCV to create a model which would detect the person's age, emotion, and gender in real-time.

**Tools you will be using:** OpenCV, Tensorflow2/ Keras, Numpy, Pandas, Matplotlib



## Resort Menu Prediction

Emerald Oyster is affected by the high cost of meals at multiple restaurants within the resort. To resolve this issue, they need a new menu and choices based on the economical condition of the customer. Your task will be to identify the economic class of the customer based on the data collected so that the board can resolve this with proper remodeling.

**Tools you will be using:** Pandas, SKLearn, Matplotlib, Seaborn, Numpy



## Telecom Churn Prediction

A telecom company wants you to analyze its data, to keep its customers. You will be provided with the 'Telecom Churn' dataset. Use it to create a model to predict which customer will switch to other telecom service providers, based on the relevant customer data.

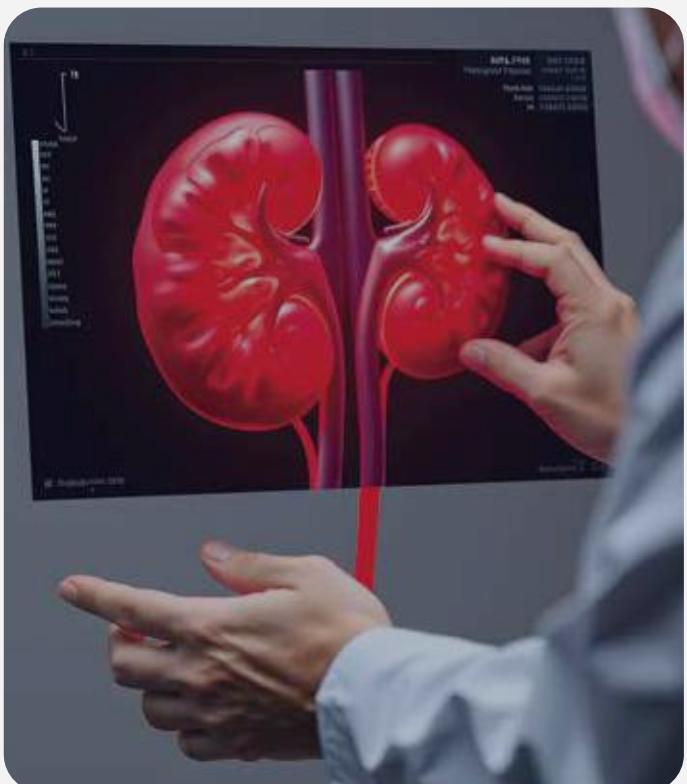
**Tools you will be using:** Pandas, Numpy, Seaborn, Matplotlib, SKLearn



## Chronic Kidney Disease Prediction

Chronic kidney disease (CKD) is a covert disease. Accurate prediction of CKD progression overtime is necessary for reducing its costs and mortality rates. The dataset is taken over a 2-month period in India. It has 400 rows with 26 features like red blood cells, pedal edema, sugar, etc. Use this to classify whether a patient has chronic kidney disease.

**Tools you will be using:** Pandas, SKLearn



## Gnar Automobiles

Gnar Automobiles engages in the distribution and sale of automobiles and light commercial vehicles. The owner of the Gnar Automobiles deals with a number of distributors across countries in different origins. As every origin sends cars with various specifications. The owner wants to determine the origin of the cars based on the specifications of the cars to further increase business opportunities.

**Tools you will be using:** Pandas, Matplotlib, Numpy, mlxtend



## Big Mart - Customer Segmentation

The data scientists at Big Mart have collected 2013 sales data of 1559 products across 10 stores in different cities. Big Mart CEO wants to understand the customer demographics and customer retention (Customer who can converge easily) so that the marketing team can market their products and services by conducting various strategies accordingly.

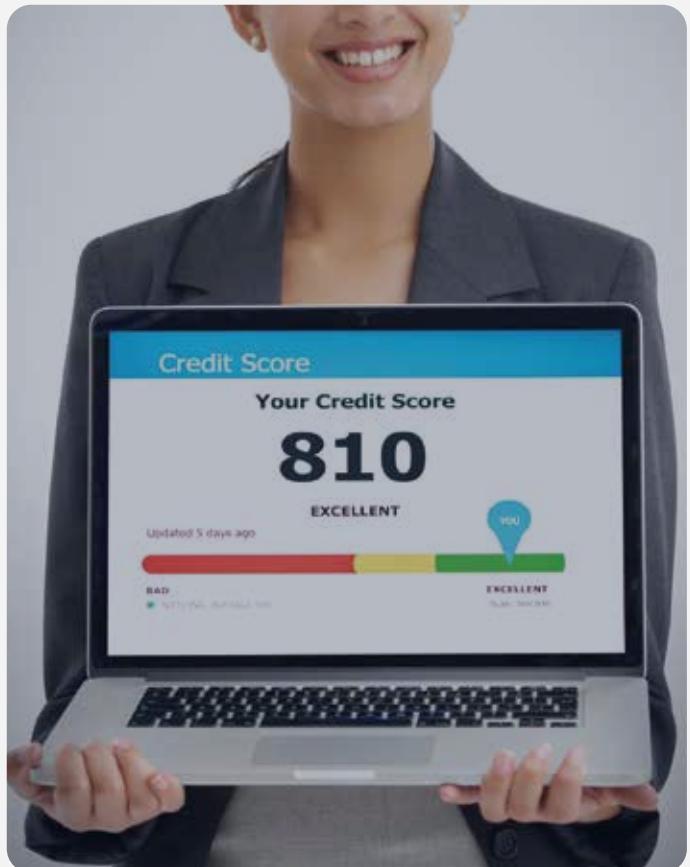
**Tools you will be using:** Pandas, scipy, Numpy



## Credit Score Prediction

Sydney based Caltech bank plans a new loan scheme for its customers and wants to analyze its customer data to find out how the customer's earning is associated with their credit score. Use clustering methods to find the high credit score clusters of customers. It will summarize the existing loan scheme and help Caltech bank to decide about the new loan scheme.

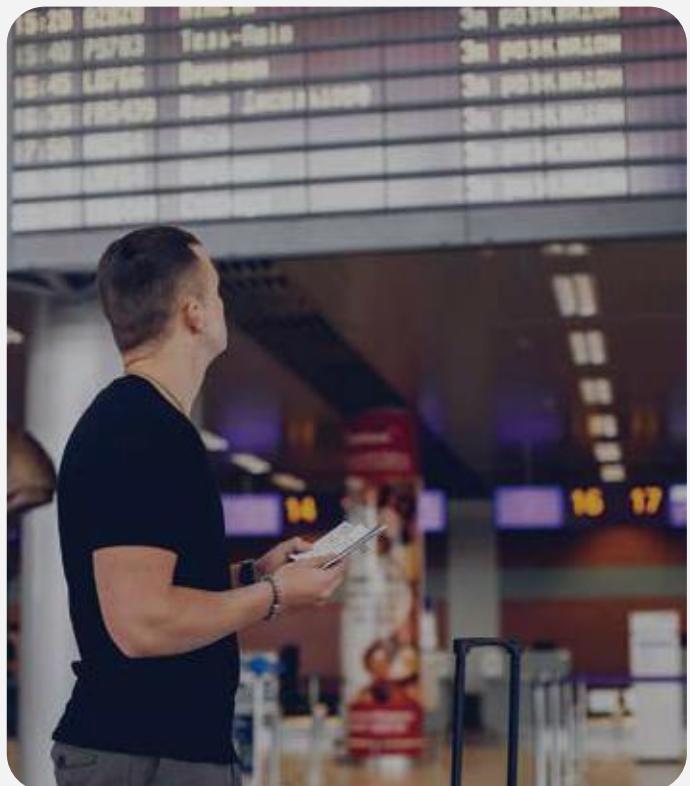
**Tools you will be using:** Pandas, scipy, Numpy



## Forecast Air Passenger Traffic

An Airline called Star Air has the data of its passengers across months. The data is classified in date/time and the passengers travelling per month. Build a model to forecast the demand (passenger traffic) in Airplanes. You will learn to use Pandas, Scipy, Numpy with hands-on experience of other tools, features and libraries.

**Tools you will be using:** Pandas, scipy, Numpy



## Housing Price Prediction

The dataset is collected from the 1990 California census containing data of one row per census group. The dataset has various demographics and details captured. Based on this data, we have to create a model using Pandas, Scipy, Numpy that can determine the housing price of the house based on the details provided.

**Tools you will be using:** Pandas, scipy, Numpy



## Traffic Sign Classification using CNN in Tensorflow 2.0

Detection and recognition of traffic signs are crucial for the development of self- driving cars, which have a direct impact on driving behaviors. You will learn to build a CNN model using OpenCV, TensorFlow2, Keras, Numpy, Pandas, Matplotlib to detect and classify the traffic signals for new self-driving cars.

**Tools you will be using:** OpenCV, Tensorflow2/ Keras, Numpy, Pandas, Matplotlib



## Smart Surveillance System

A shopping centre needs a surveillance system to detect persons and other items. As a machine learning engineer, you will create a model to detect objects using a pre-trained Mask RCNN model. You will be using OpenCV, Tensorflow2, Keras, Numpy, Pandas, Matplotlib and others.

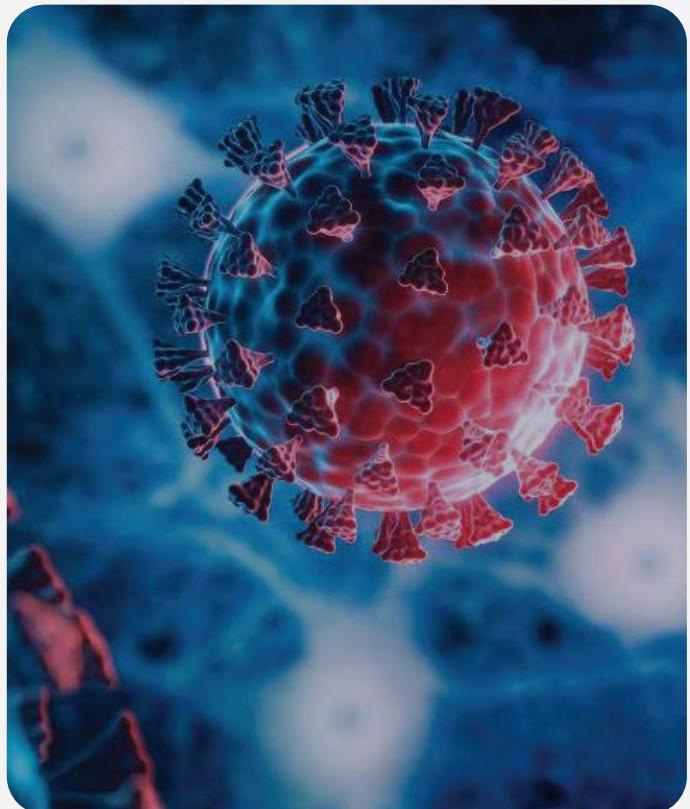
**Tools you will be using:** OpenCV, Tensorflow2/ Keras, Numpy, Pandas, Matplotlib



## COVID Analysis in India

COVID-19 pandemic is the greatest global humanitarian challenge the world has faced since World War II. The pandemic has spread widely, and the number of cases is rising daily. The government is working to slow down its spread. Web Scrape the data from the official government website and find various insights by comparing the trend of COVID in India as compared to the world.

**Tools you will be using:** Prophet, ARIMA, Pandas, SKLearn, pyplot, Seaborn, Matplotlib, beautiful soup



## Analyze & Visualize Employee Attrition

Employee retention is one of the biggest metrics that a company should have in mind when thinking of growth. Employee attrition is caused when the total strength of the company is greatly reduced as more employees leave the company than expected. Uncover the factors that lead to employee attrition and explore the reasons as to why people are leaving the organization and predict whether an employee will leave the organization or not by creating a Web App using Streamlit that takes inputs from user's online.

**Tools you will be using:** Sklearn, PyCaret, Streamlit, SHAP, Pandas-Profilng



## Analyze & Visualize Video Games Sales

SRS Ltd. is a Korean and video game company headquartered in Seoul. You as a Data Scientist is required to analyze the trend in Global Sales according to the Genres on the Video Games Sales Dataset from 1980 & visualize the change in Net Sales of different publishers from the year 2005 to 2015.

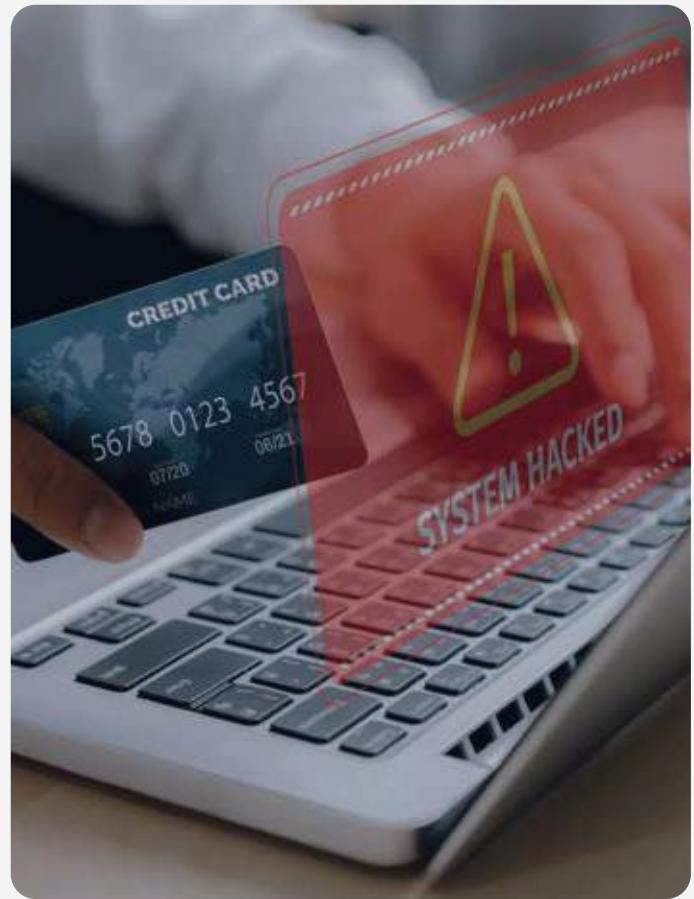
**Tools you will be using:** Plotly Express, Streamlit



## Identify fraudulent credit card transactions

Credit card fraud happens when a fraudster or a thief steals your credit card or the information from that card to make unauthorized purchases in your name or take out cash advances using your account. Credit card companies such as Citibank, HSBC, and American Express need to recognize fraudulent credit card transactionso that customersare not charged for items that they did not purchase.

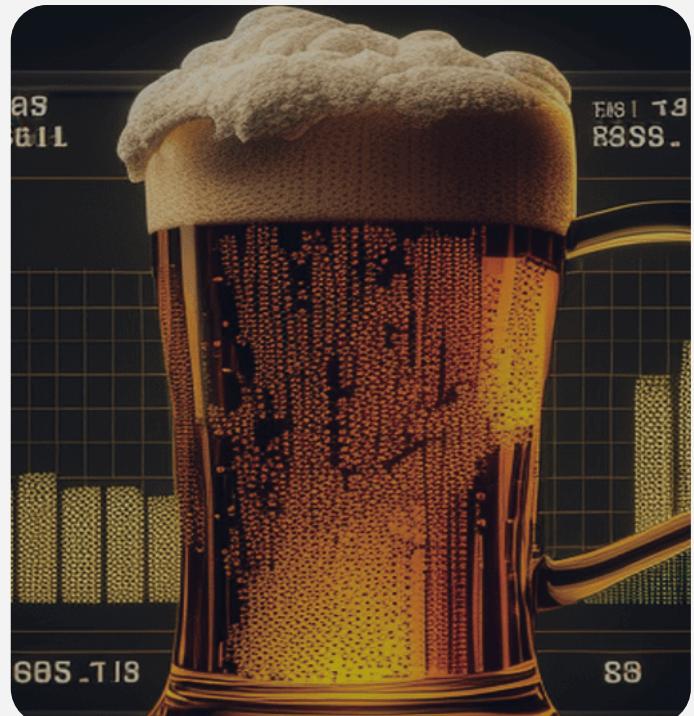
**Tools you will be using:** Sklearn, PyCaret, Streamlit, ELI5, Pandas-Profilng



## Beer Consumption Prediction

Beer is the most consumed drink in the world. Not without reason, it is perfect for almost every situation,from happy hour to large wedding parties. You will be given a data sample collected in São Paulo, Brazil. Use this to predict the quantity of beer consumption based on the features that contain climate conditions of a given day.

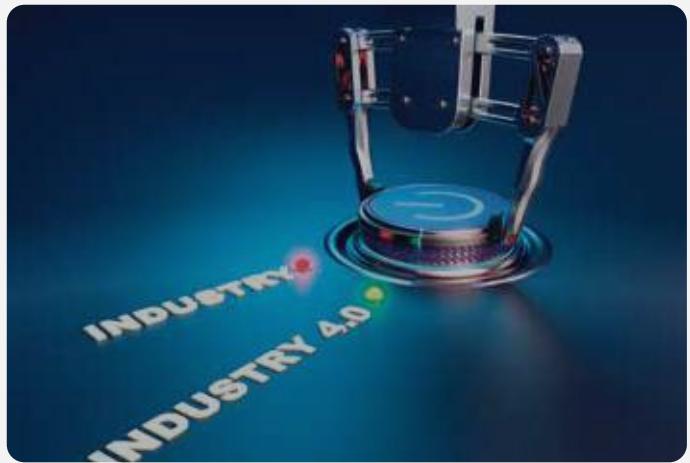
**Tools you will be using:** Pandas, Seaborn, SKLearn



## Industry 4.0

Automationsuch as split decisions based on real time data

**Tools you will be using:** ETHERNET, Modbus, wifi, ble, arm, nodemcu, rs485, aws, nodered.



## Smart Healthcare

Use of technologyto device-to-data centric solutionsin healthcare.

**Tools you will be using:** ARM, ARDUINO, Bluetooth, Sensors, wifi, Node Red



## Smart Farming

Monitoringcrop filed through sensors and automation

**Tools you will be using:** LoraWan, ARM, Soil sensor, Moisture sensor, Cellular AI, Temperature and humidity sensor



## Smart Grids

Use of smart grids for energy efficiency on a real-timebasis.

**Tools you will be using:** DLMS, LORAWan, Bluetooth, 4g, Arm, node mcu, google cloud, aws



## Predictive Maintenance

To monitor, optimize & maintenance of assets on a real-time basis

**Tools you will be using:** Accelerometer, arm, nodemcu, Edge impulse, Neural network, IBM Watson, wifi



## Condition Monitoring

Help in reducing damage and maintenance costs with AI solutions

**Tools you will be using:** Current sensing, wifi, ble, xbee, lorawan, nodered, aws



## Smart Building Automation

Simplifying tasks such as control building, security, temperature etc via devices.

**Tools you will be using:** Bacnet, Modbus, Ethernet, arm, nodemcu, nodered, ibm bluemix



## Environment Monitoring

Remote environment monitoring connected virtually via different devices.

**Tools you will be using:** PM2.5 sensor, air quality sensor, ARM, ble mesh, lorawan, nbiot



## Waste management

Reduces fuel consumption while dumping waste in the city.

**Tools you will be using:** wifi, nbiot, phsensor, moisture sensor, gas sensor.



## Connected Supply Chain

Easier to track where goods are stored

**Tools you will be using:** temperature and humidity sensor, accelero meter, lte, nbiot, gps, gas sensor, pressure sensor



## Smart Logistics

Increase in the real-time decision-making process in supply chain management.

**Tools you will be using:** Nbiot, GPS, gas sensor, vibration sensor, ARM, Arduino



## Text-to-Image Generation

**Description:** Combine the power of NLP and CV to generate realistic images from textual descriptions, such as creating visual prototypes for design concepts or generating fictional landscapes.

**Tools and Technologies:** PyTorch, TensorFlow, OpenCV, Diffusion Models, GANs, Transformers.

**Outcomes:** Learn to preprocess text and image data, design transformer-based models, and implement diffusion models for image synthesis.



## Interactive Chatbot for Healthcare

**Description:** Build a conversational agent using GPT-based LLMs that assists users with healthcare queries, such as symptom checks or medication reminders.

**Tools and Technologies:** Rasa, PyTorch, Hugging Face Transformers, Streamlit, NLP Libraries (spaCy, NLTK).

**Outcomes:** Gain experience fine-tuning LLMs for domain-specific tasks, handling conversational context, and deploying chatbots.

## Voice Cloning and Audio Synthesis

**Description:** Develop a model to synthesize human-like speech or clone specific voices using generative models like WaveNet or Tacotron.

**Tools and Technologies:** TensorFlow, PyTorch, Librosa, WaveNet.

**Outcomes:** Learn to preprocess audio data, generate waveforms, and build real-time voice applications.



## Smart Surveillance System Using Generative Models

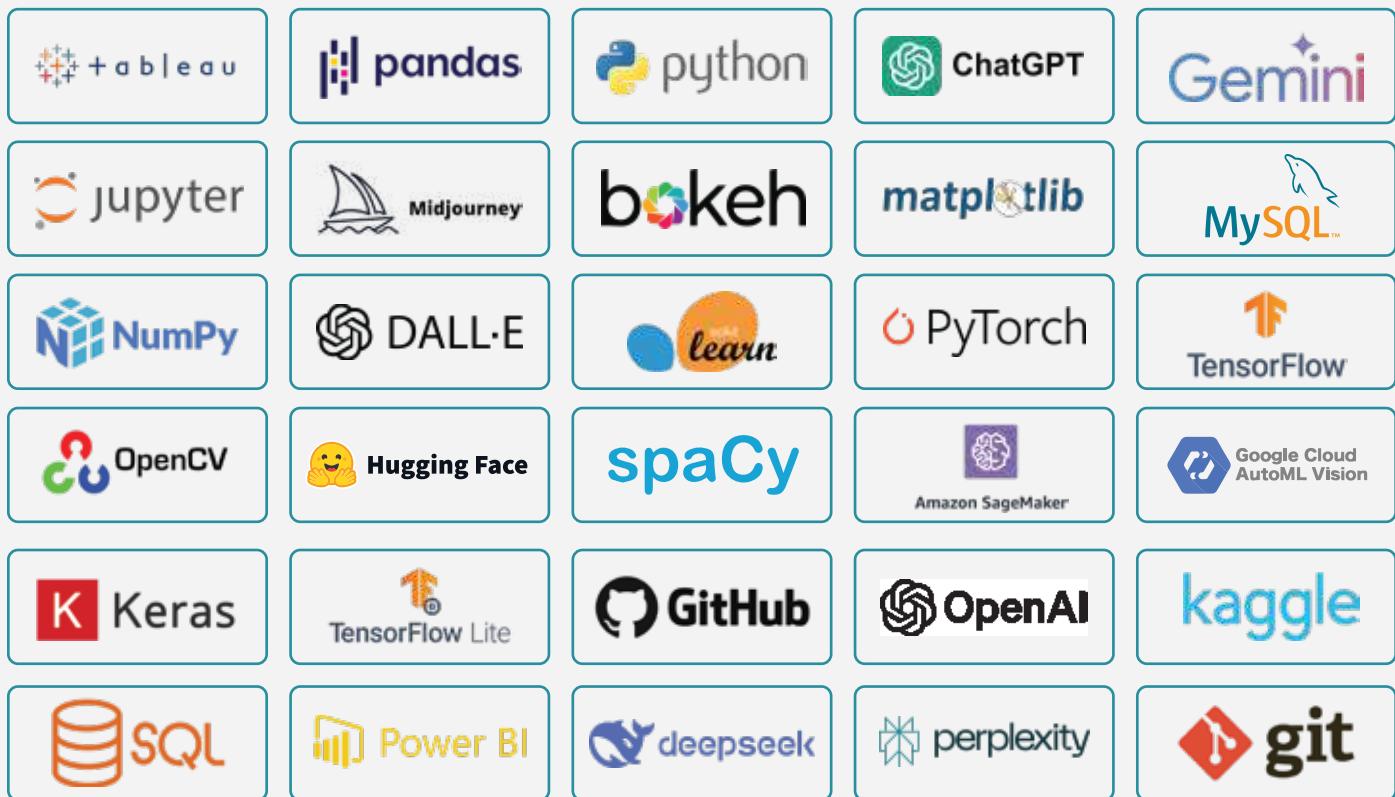
**Description:** Detect and track anomalies or objects in video footage using generative models for data augmentation and anomaly detection.

**Tools and Technologies:** TensorFlow, PyTorch, OpenCV, GANs.

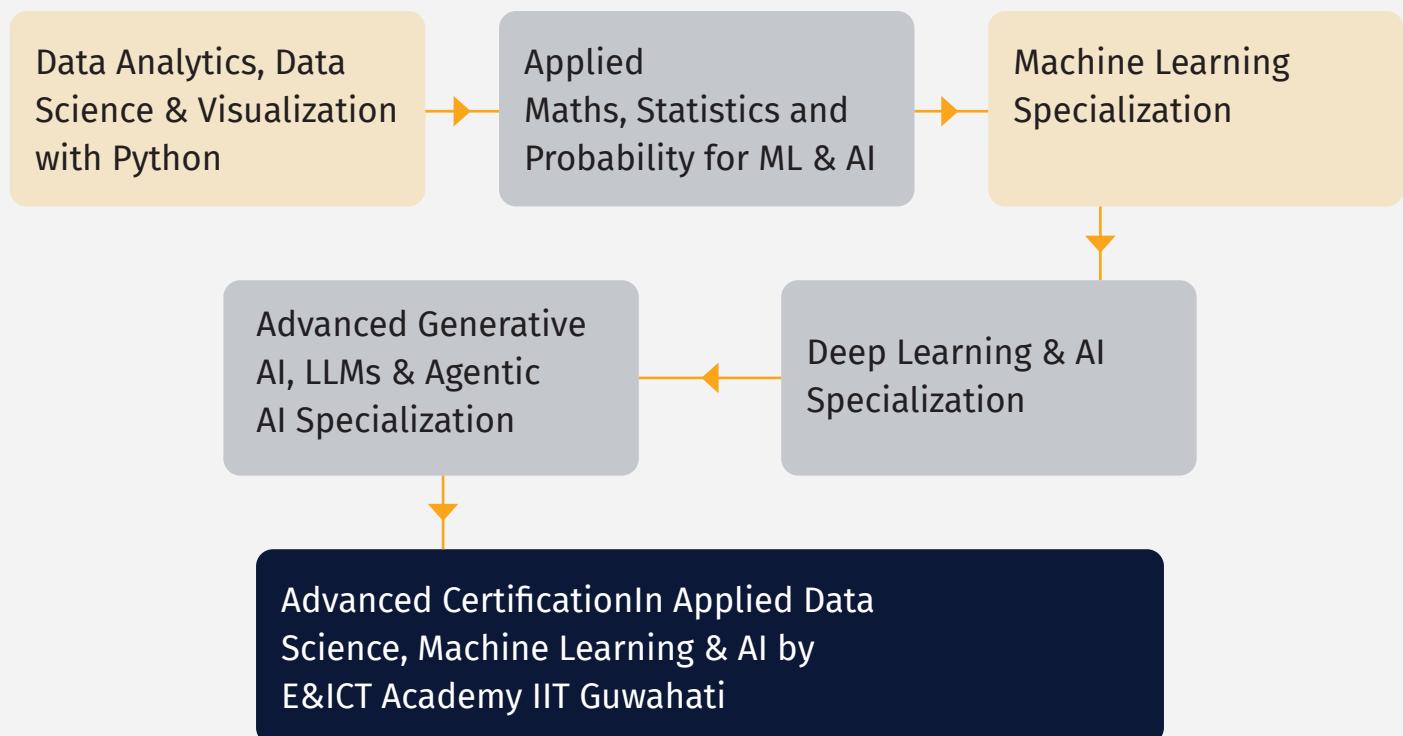
**Outcomes:** Explore applications in computer vision, object detection, and real-time surveillance.



# Tools, Languages & Frameworks



## Learning Path



# CONTACT

## THANK YOU!

*Contact Details*



**Electronics & ICT Academy**  
**Indian Institute of Technology Guwahati**

An Initiative of Ministry of Electronics & Information Technology(Meity)



इलेक्ट्रॉनिकी एवं  
सूचना प्रौद्योगिकी मंत्रालय  
MINISTRY OF  
ELECTRONICS AND  
INFORMATION TECHNOLOGY

### Program Information

Advanced Certification In Applied Data Science,  
Machine Learning & AI

By E&ICT Academy, IIT Guwahati

+91 8595486148

admissions.eictiitg@theiotacademy.co

[eict.iitg.ac.in](http://eict.iitg.ac.in)

[theiotacademy.co](http://theiotacademy.co)