C. Fundamentals of AIoT

Chapter #3: Introduction to AIoT and its Components

ET0743
5G and AIoT Applications

Week #4 – #5

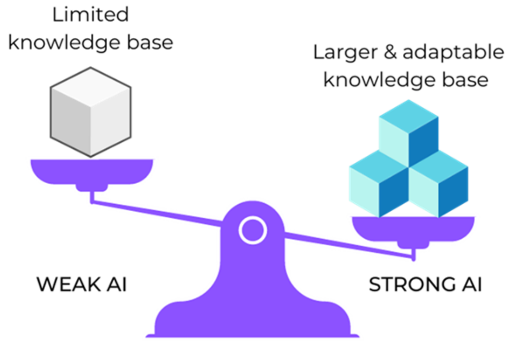
Learning Objectives

At the end of instruction, the learner should be able to:

- Understand the basics of AI and Machine Learning.
- Understand AIoT architectures and integration of AI and IoT.
- Describe examples of AIoT applications.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Road Map of AI



The modern field of artificial intelligence was established in 1956 during a conference held at Dartmouth College in Hanover, New Hampshire. At the meeting, the phrase "artificial intelligence" (AI) was formally established.

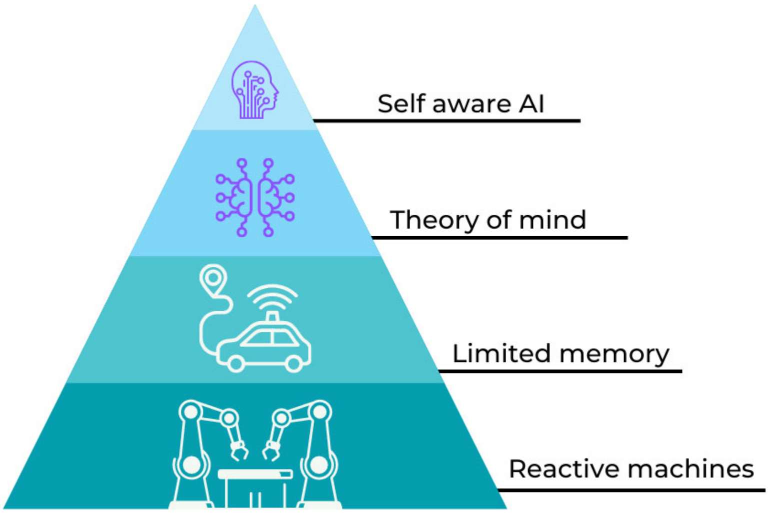
The MIT cognitive scientist Marvin Minsky, who was present at the Dartmouth conference, expressed optimism over the future of this emerging discipline. But at the time, his optimistic assessment was unwarranted because government support for this emerging profession quickly dried up. After a lull in research until 1994, artificial intelligence (AI) gained widespread attention in 1997 when IBM's Deep Blue® defeated the world's greatest chess player, Garry Kasparov.

When Eugene Goostman, the talking computer and "chatbot," tricked judges into believing it was a real person in 2011, artificial intelligence gained attention once more. The British mathematician Alan Turing organised the Turing test, which the chatbot had successfully completed. The purpose of the test is to determine whether a machine qualifies as intelligent. Nowadays, artificial intelligence (AI) is a crucial part of the technology sector, helping to resolve a variety of issues in software engineering, computer science, and operations research.

Today's Artificial Intelligence (AI) is substantially more sophisticated than it was even a few years ago, and it is still developing quickly. Weak AI is what it is known as now. Weak AI, sometimes referred to as narrow AI, is able to carry out a single task for which it was created. Conversely, strong AI has the capacity to learn, think, and adapt just like people. Nevertheless, powerful AI systems are still not realised. Case in point, neural networks are able to recognise images in computer vision, however they are not a perfect match for any human visual system. It has been noted that they have occasionally outperformed the human brain, though. Businesses of all kinds, even startups with only one employee, have already started utilising data and analytics to gain a competitive advantage. AI makes it possible to generate insights, optimise processes, and make centric decisions more quickly.

Various forms of artificial intelligence have been developed to make other AI systems more intelligent.

3.1 Basics of Artificial Intelligence and Machine Learning – Road Map of AI



Artificial Intelligence can be categorised into:

Reactive Machines AI:
Earliest kind of AI, with limited capabilities. These machines are incapable of learning or having a grasp of past experiences. Reactive Machines AI is limited to the core activities for which it was designed. Reactive machines cannot grow; instead, they can only stagnate in repeating behaviours and activities.

Limited Memory AI:
Limited Memory AI: These devices can learn from past data and has features like to those of Reactive machines. As the name suggests, limited memory AI devices have a limited amount of memory. They have a very brief window of time to look back on the past and draw lessons from it. Large amounts of relevant training data are used to train these robots. Data from memory is utilised to build a reference model that is used to solve issues in the future. Limited Memory AI drives contemporary AI applications.

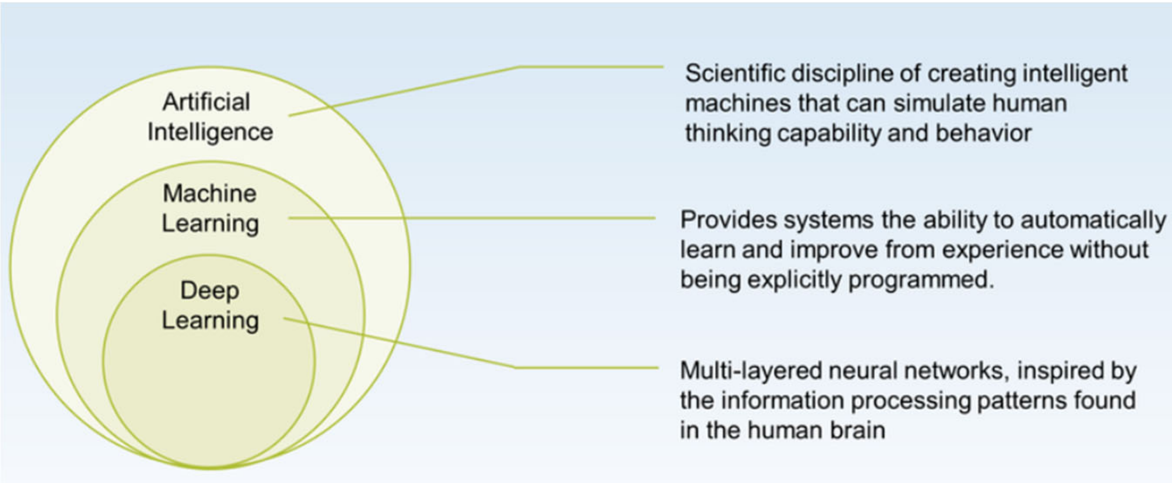
Theory of Mind AI:
This is a sophisticated system that can comprehend the entities it interacts with more

fully. To do this, it must recognise their (i.e. the entities) needs, emotions, beliefs, and thought processes. The development of these systems is still ongoing. The Theory of mind AI is able to quickly identify facial and eye movements and adjust its actions to suit the situation.

Self-aware AI:
This is a theoretical system that is most likely the most sophisticated AI that has ever been imagined by humans. When AI is self-aware, robots or machines understand their own intrinsic characteristics and are completely aware of who they are. AI is also capable of perceiving human emotions and comprehending a variety of situations and states.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Types of AI and relation between AI, ML & DL



By creating intelligent machines that are capable of carrying out tasks that typically require human intelligence, artificial intelligence (AI) has the ability to think like people. These processes include reasoning, self-correction, and learning (the acquisition of knowledge as well as the guidelines for applying it). Machine learning is used by artificial intelligence to imitate human intelligence. The process for reacting to a select few particular actions must be taught to the system. Thus, it creates what is known as a propensity model by combining algorithms with past data. After then, the propensity models start to predict (e.g. scoring leads).

Machine learning (ML) is the automated search for significant patterns in data. AI that enables a system to learn from data instead of explicit programming is called machine learning. Using a variety of methods, machine learning improves, describes, and predicts results by iteratively learning from available data. Accurate models based on that particular data can be generated as the algorithms process the training data.

One may think of deep learning (DL) as a subset of machine learning. It is a part of machine learning that concentrates on how "abstractions and concepts" are formed. Massive amounts of data are usually ingested by deep learning algorithms, which then use supervised or even unsupervised learning to generalise the characteristics/features and categories associated with that data. Neural networks are the foundation of deep learning systems.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Key AI Terms and Definitions

Term	Explanation	Example
Instance	The object about which the AI should make a prediction	An image from a vehicle front camera, which needs to be classified as „contains obstacle" or not
Inference	The process of using a trained model to make predictions about an instance based on new data	Apply an ML model to a new image from the vehicle camera to identify potential obstacles
Label	The outcome of a prediction task (either supplied by the training data or by the AI)	Parts of an image labeled as „traffic light", „pedestrian", „speed limit"
Labeling	The process of manually (at least initially) detecting and tagging data samples as input for model training	Manually identify and tag potential obstacles on a large set of images
Training	Machine learning models are trained by using large, representative sets of data, e.g. labeled training data	Manually identify and tag potential obstacles on a large set of images
Feature	An measurable property or characteristic of an instance	Edges and objects in an image
Model	A statistical representation of a prediction task	A model to predict road traffic
Pipeline	The IT infrastructure for an AI/ML algorithm, including data and model management.	The pipline to manage data flows and prediction model definitions

aiotplaybook.org

Machine Learning (ML) is probably the AI technique that is most significant right now. Machine learning (ML) is the name given to a class of algorithms that automatically get better with experience and data usage. Deep Learning (DL) is a significant subfield in machine learning (ML) that makes use of so-called multi-layered neural networks. Convolutional neural networks (CNNs) and deep neural networks (DNNs) are two examples of deep learning models.

Over the last few decades, machine learning has evolved into a common tool for nearly any activity involving the extraction of information from big data sets. We are surrounded by machine learning-based technology: search engines learn to provide us with the most relevant results, anti-spam software learns to filter our email messages, and software that detects fraud learns to safeguard credit card transactions. Machine learning algorithms are also used by cars to prevent accidents, digital cameras to detect faces, and phone personal assistants to recognise voices. Applications of machine learning are also frequently employed in scientific fields like astronomy, medicine, and bioinformatics.

Using a variety of methods/algorithms, machine learning improves, describes, and predicts results by iteratively learning from available data. By feeding the algorithms with training data, exact models based on that particular data can be generated. When data is used to train the machine learning algorithm, the resulting output is actually the machine learning model. When a model receives input after training, the output will inevitably follow. A predictive algorithm, for instance, will provide a predictive model. A prediction based on the data will be sent when the predictive model receives it. The model is trained using these data. Developing analytics models requires machine learning. ML is a combination of the following:

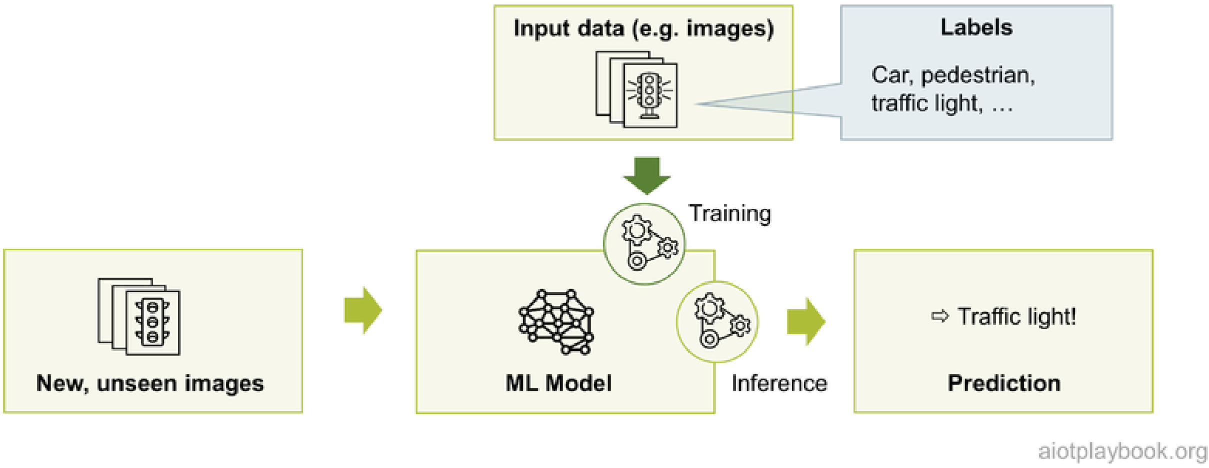
Data + Model → Compute Prediction

Our observations are the data. The model comprises of our expectations based on prior information, which can come from transfer learning or can just be our convictions/beliefs "regarding the universe's symmetries". Humans have inductive biases "built into" our "models". The prediction can be a classification, a course of action, or a quality rating. The significance of guesses in AI is the reason machine learning (ML) has emerged as its foundation.

When is machine learning necessary, then? Programs that learn and subsequently get better based on experience are required because of the intricacy/complexity of a given situation/problem and the necessity for adaptation. Programming tasks that are as sophisticated as driving, picture recognition, and speech recognition is not easy. The programming tool's limiting characteristic/feature is rigidity. Once the programme is built and deployed, it stays the same even if the tasks vary over time or between different users. Machine learning trains programs to change how they behave in response to incoming data, which makes them more adaptable to changes in their surroundings/environment.

Supervised Learning, Unsupervised Learning, and Reinforcement Learning, are the three most popular ML techniques/methods. In order to train a model for use with similar but unlabeled data that is later encountered, the Supervised Learning method uses manually labelled sample data. The Unsupervised
Learning approach/method looks for patterns and structures in data automatically. Reinforcement Learning combines rewards, or penalties, with a trial-and-error methodology. The ensuing sections go into further detail about each technique. The table presented here summarises a few of the key concepts shared by different machine learning methods.

3.1 Basics of Artificial Intelligence and Machine Learning – Supervised Learning



We want to start by looking at Supervised Learning as an AI/ML method. A data collection/set containing certain observations (such as photos) and the labels of the observations (such as classes of items on these photographs, e.g. "traffic light," "pedestrian," "speed limit," etc.) is necessary for supervised learning.

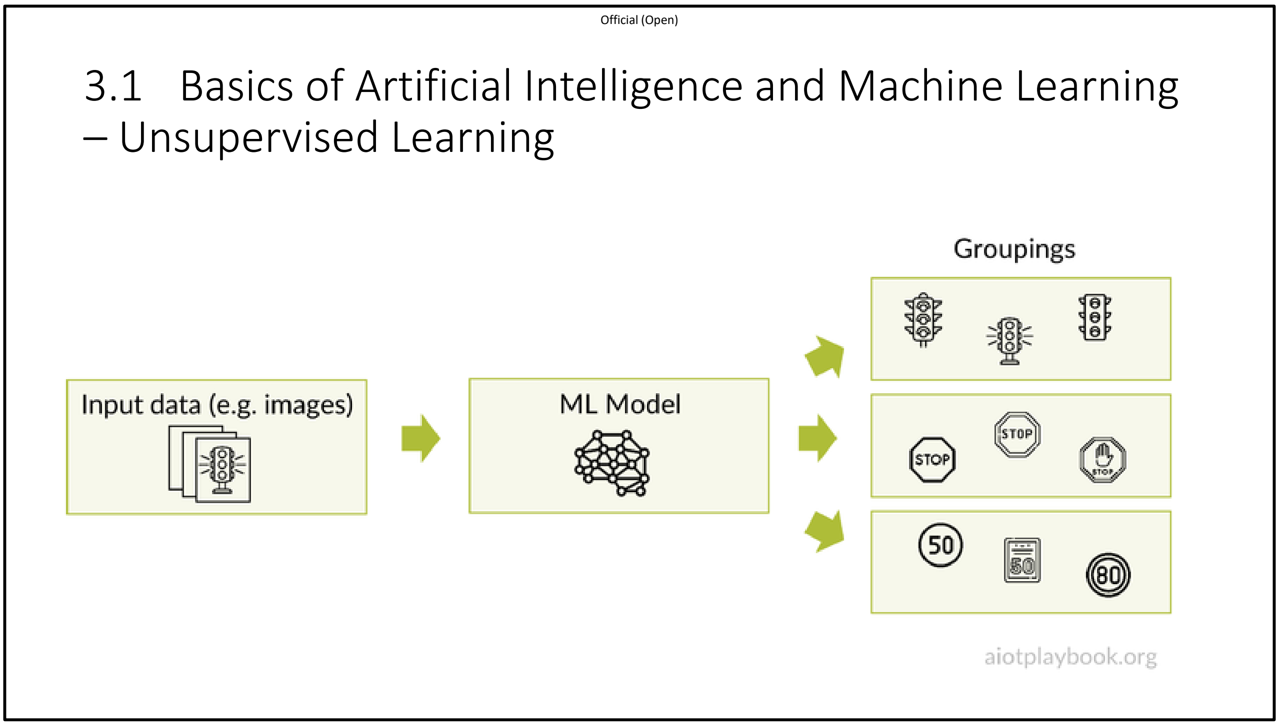
These labelled data sets are used to train the models, which can subsequently be used with observations that were not previously known. When new, unseen data are supplied as input, the supervised learning algorithm generates an inference function to make predictions. By comparing the actual output with the intended output, the model can be further refined; also known as "backward propagation" of errors.

The two main types of supervised models are regression and classification:

- Classification: The output variable is a category (e.g. "stop sign", "traffic light", etc.)
- Regression: The output variable is a real continuous value (e.g. electricity demand prediction)

Some widely used examples of supervised machine learning algorithms are:

- Linear regression – mainly used for regression problems
- Random forest – mainly used for classification and regression problems
- Support vector machines – mainly used for classification problems



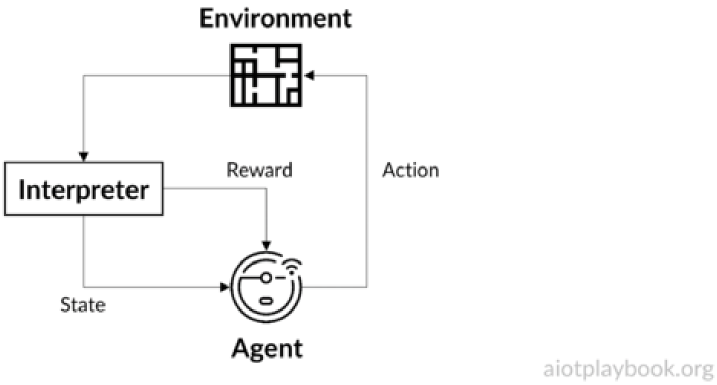
Unsupervised Learning, a class of algorithm that learns patterns from unlabeled data, is the next machine learning technique. Finding patterns in data that were previously unknown is the primary objective. When one lacks data on intended outcomes, unsupervised machine learning is utilised (see Figure).

Typical applications of Unsupervised machine learning include the following:

- Clustering: automatically split the data set into groups according to similarity (but not always easy)
- Anomaly detection: used to automatically discover unusual data points in a data set (e.g. to identify a problem with a physical asset or equipment)
- Association mining: used to identify sets of items that frequently occur together in a data set (e.g. "people who buy X also tend to buy Y“)
- Latent variable models: commonly used for data pre-processing (e.g. reducing the number of features in a data set – dimensionality reduction)

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Reinforcement Learning

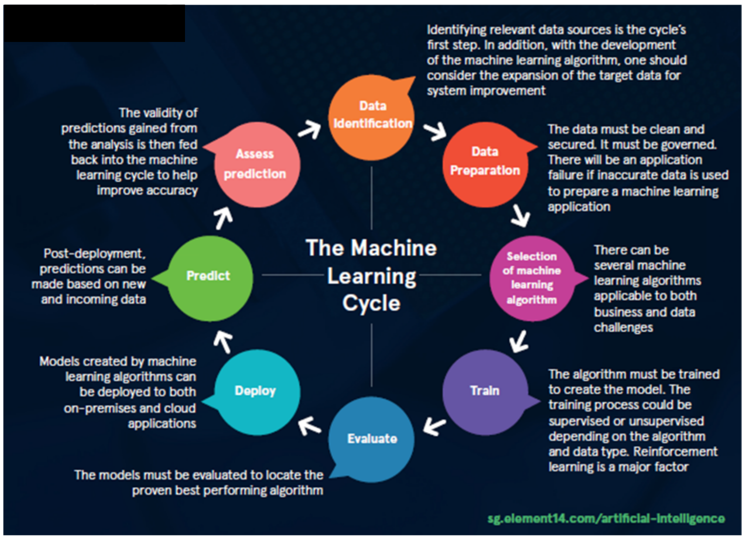


Reinforcement Learning (RL) is the third common machine learning technique. An "Agent" in reinforcement learning (RL) learns how to accomplish its objectives/goals in an uncertain and sometimes complex environment. For instance, the agent may be placed in a simulation where it is rewarded or penalised for the actions that it executes; much like in a gaming situation. Maximising the overall reward is the agent's objective.

The creation of an appropriate simulation environment is one of the primary challenges in Reinforcement Learning. To train autonomous driving algorithms, for instance, the RL environment needs to accurately mimic scenarios like brakes and crashes. The advantage is that training the model in a simulated environment is typically far less expensive than utilising immature models, which run the risk of damaging actual physical things. The next step is to move the model from the training environment to the real world, which presents a challenge.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – The Machine Learning Cycle



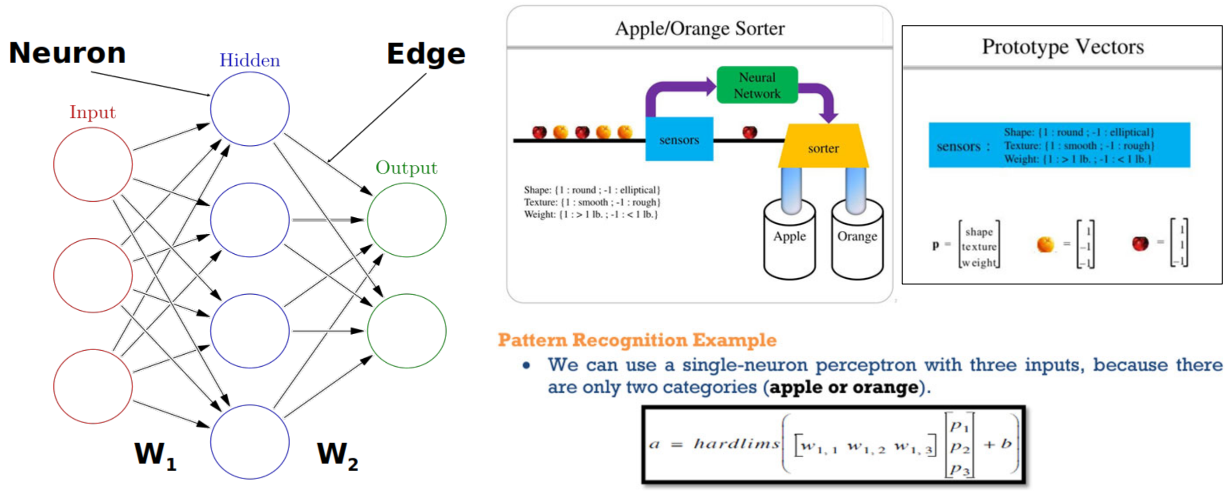
Choosing the right machine learning algorithm is just one of several processes. The steps in the machine learning cycle are shown in the Figure here.

Example: Lab #6 Use Case:

- To be added soon.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Artificial Neural Networks and Deep Learning



Pattern Recognition Example

- We can use a single-neuron perceptron with three inputs, because there are only two categories (**apple or orange**).

$$a = hardlims\left(\begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} + b \right)$$

Artificial Neutral Networks, or ANNs, are a specialised field within machine learning (sometimes simply called Neural Networks). A loose inspiration for ANNs comes from the neural networks that make up biological brains. A group of interconnected nodes called neurons serves as the representation of an ANN. The terms "edges" allude to the connections. Like synapses in the human brain, each edge can send messages to other neurons. After processing the incoming signal, the receiving neuron notifies its associated neurons. Signals are numbers, and statistical functions are used to compute them.

Typically, there is a weighted relationship between neurons and edges that either increases or decreases the signal strength. You can change the weights as you continue to learn. Neurons are typically grouped into layers, with each layer altering the input signals in a different way. These layers allow signals to pass through possibly several times. The employment of numerous layers in these networks is indicated by the adjective "deep" in the context of deep learning.

Example (see Figure):
Assume for the moment that a trader has a warehouse full of several fruit varieties. When the fruit gets to the warehouse, it could become mixed up. The vendor might

need to use a sorting machine to separate the fruits. The machine contains a conveyer belt that is used to load the goods. As the items move through the belt, a system of sensors measures their texture, shape, and weight. If the fruit is oval/elliptical, the shape sensor may report -1; otherwise, it may output 1. If the surface is smooth, the texture sensor may return 1, otherwise -1. In the event that the object weighs one pound or less, the weight sensor may output -1. The neural network receives this sensor data and uses them to determine what kinds of fruit should go on the conveyer belt. Assume for the moment that the fruits on the conveyor belt are apples and oranges.

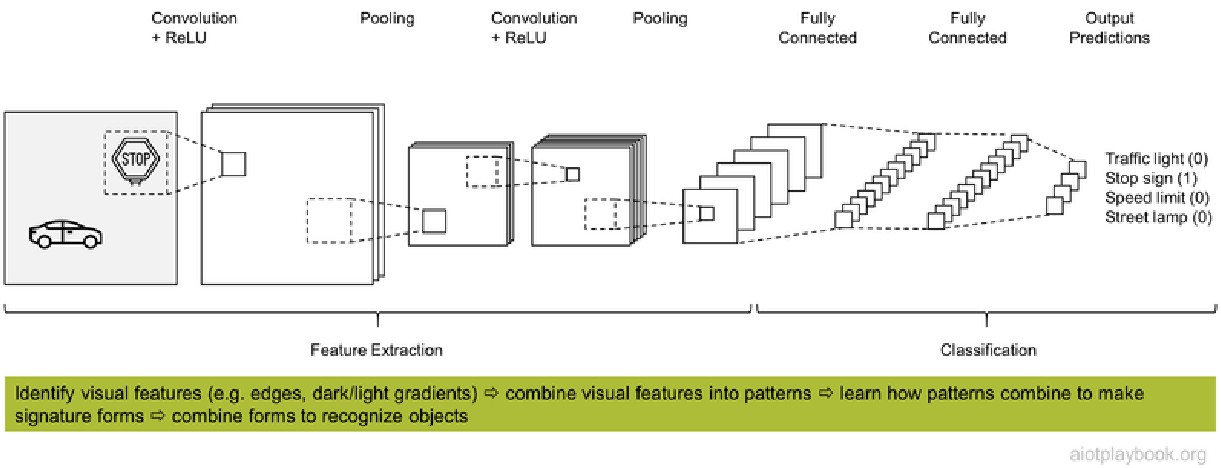
Shape, texture, and weight can all be represented by a three-dimensional vector that represents the value obtained when the fruit passes through the sensor.

As a result, the apple and orange can both be represented as follows. This implies that for each fruit discovered on the conveyer, the neural network will get a single, three-dimensional input vector. To recognise if the object is an orange or an apple, a decision must be made.

N.B. #1 A perceptron is an algorithm for supervised learning of binary classifiers of possible outcomes.
N.B. #2 Neural networks, also called artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are the backbone of deep learning algorithms.

Official (Open)

3.1 Basics of Artificial Intelligence and Machine Learning – Artificial Neural Networks and Deep Learning



Convolutional Neural Networks (CNNs) are a popular way that ANNs are implemented. A variation to multilayer perceptrons is used by convolutional neural networks (CNNs). There are one or more convolutional layers in a CNN. It is possible for these layers to be pooled or fully interlocked. The integral that quantifies the amount that two functions overlap as one passes over the other is called a convolution. Convolution is the process of multiplying two functions to combine them. In the convolutional layer, an input convolutional operation is applied before the result is passed to the next layer. This convolutional procedure allows the network to use less parameters but still be substantially deeper. Convolutional neural networks perform incredibly well in systems for natural language processing and image and video recognition because of this ability.

Example (see Figure):
This example displays a CNN along with its multiple layers. An input image's areas can be categorised into groups like "traffic light" or "stop sign" using this technique. This CNN has four primary functions:

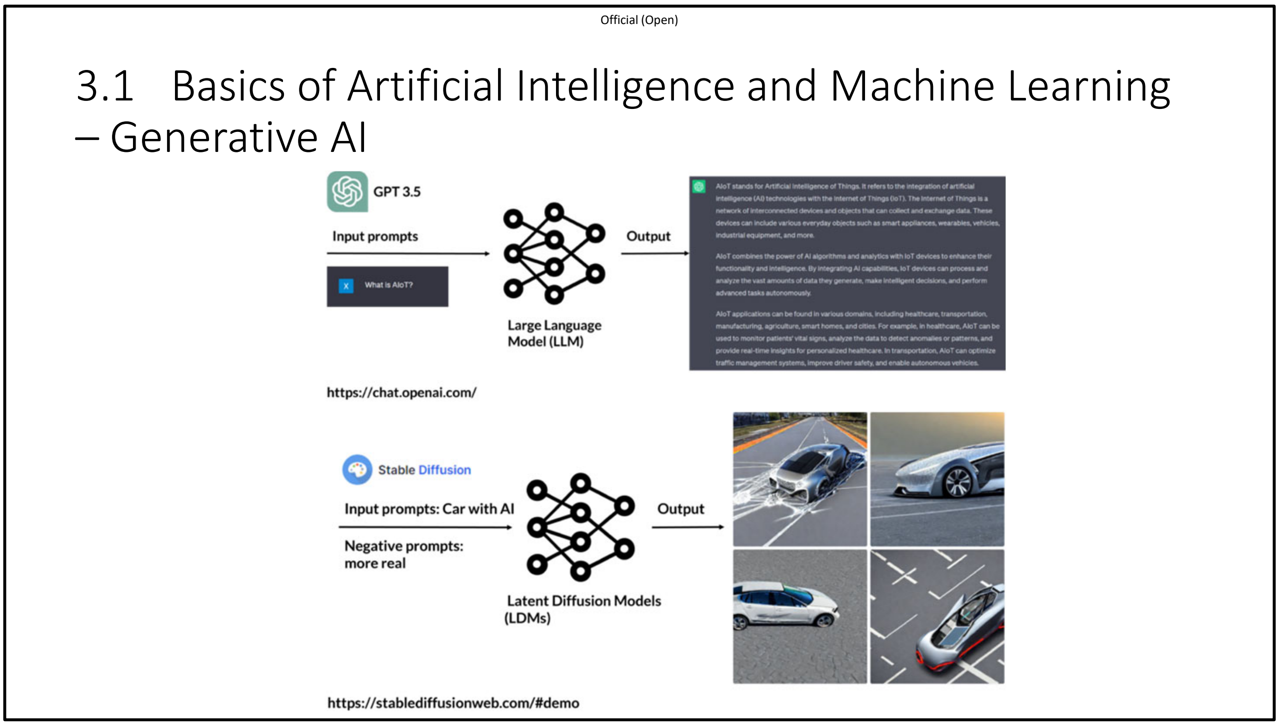
- **Convolution**: Extract features from the input image, preserving the spatial relationship between pixels by using small squares of input data. Convolution is a linear operation: it performs elementwise matrix multiplication and addition.

- **Non Linearity**: ReLU (Rectified Linear Unit) is an operation applied after the convolution operations. ReLU introduces non-linearity in the CNN, which is important because most real-world data are non-linear.

- **Spatial Pooling/down-sampling**: This step reduces the dimensionality of each feature map, while retaining the most important information.

- **Classification (Fully Connected Layer)**: The outputs from the first three layers are high-level features of the input image. The Fully Connected Layer uses these features to classify the input image into various classes based on the training dataset.

N.B. A Convolutional Neural Network (CNN) is a type of deep learning algorithm.



Generative AI is a subset of Deep Learning. It is a type of Artificial intelligence that creates new content based on what it has learned from existing content. The learning process is abstracting the data probability distributions by training large-scale datasets, and after that, it produces a statistical model. Users are usually interacting with a Generative AI via a so-call prompt, for example a question asked to the system. The prompt can include specifics like information on how the answer should be structured, e.g. number of words to be generated.

The generative AI creates new content by using the statistical model to predict the expected response to the provided prompts.

Generative models can be divided into two types: generative language models and generative image/video models.

Natural Language Processing (NLP) techniques are the foundation of generative language models, which use language laws and patterns to generate new text. Often referred to as Large Language Models (LLMs), they typically have billions or higher orders of magnitude of parameters due to training on large-scale textual data such as news, articles, books, or web material. LLaMA, LaMDA (Google Bard), and GPT-3.5

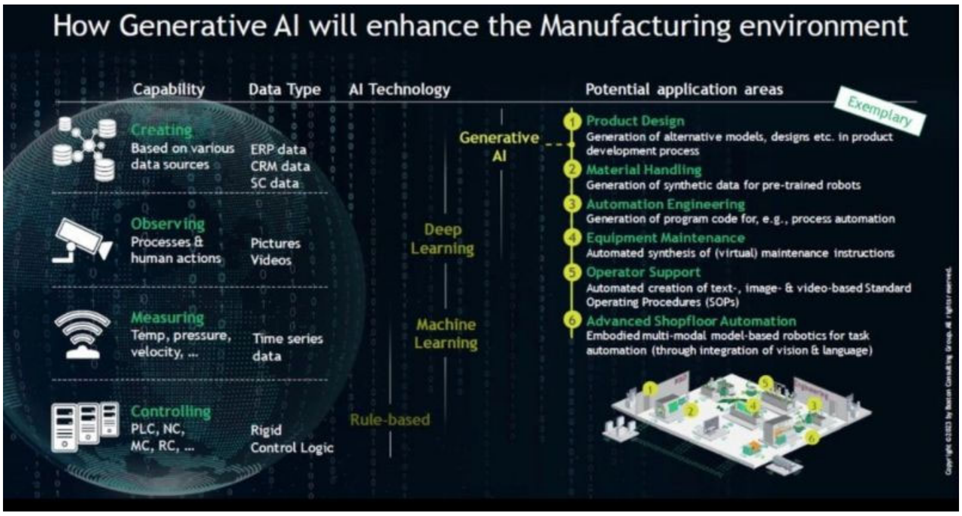
and GPT-4 (OpenAI / ChatGPT) are a few examples of popular LLMs at the moment.

These LLMs' core idea is based on a deep learning model called the "Transformer" architecture. Each word in the input text is encoded by LLMs during model training, and the result is a vector representation known as Word Embedding. The "Transformer" architecture can handle long text dependencies well and makes use of an attention mechanism to better comprehend the correlation between various Word Embeddings. By using this technique/mechanism, LLMs are able to precisely predict the probability distribution of the next word by inferring the correct context from the training task. In a practical application, like a chatbot, the user's input is first transformed into a coherent textual sequence and fed to LLMs, which utilise the preceding/previous words to predict the subsequent ones until a comprehensive response is generated.

The foundation of generative image models is the Computer Vision (CV) technology, which creates new images by learning the structure and features of the original/existing images. The generative adversarial network (GAN) is a more traditional approach. It consists of a discriminator that separates real images from fake and a generator that creates fake images. Both networks compete to create realistic images through repeated iterations. In 2023, Diffusion Models attracted a lot of attention, mainly due to its remarkable results on the text-to-Image task. The model is fed randomly sampled Gaussian noise, and it learns how to denoise data in order to produce output. The DALL-E 2 from OpenAI, Imagen from Google Brain, and Stable Diffusion from StabilityAI are performing in a way that is comparable to that of genuine photos and hand-drawn artwork.

Text (e.g. general writing, note-taking, marketing, sales, support, etc.), code (e.g. code generation, code documentation, text to SQL, web application builders, etc.), graphics (image generation, media advertising, design, etc.), speech, video, 3D, and more are among the many fields in which generative AI finds widespread use.

3.1 Basics of Artificial Intelligence and Machine Learning – Generative AI



It is also predicted that generative AI will transform the manufacturing landscape by providing cutting-edge solutions for all of the industry's application domains.

The ability of generative AI to quickly produce alternative ideas based on pre-established criteria offers up new possibilities for creativity and optimisation in product design. This makes it possible for engineers to effectively investigate a wide range of design options, resulting in the creation of goods that are both highly inventive and functionally superior.

Furthermore, generative AI is essential to the material handling application domain since it generates synthetic data that can be used to pre-train robots to handle a variety of tasks. Generative artificial intelligence (AI) makes it easier to train robots to swiftly adapt to dynamic environments by simulating a variety of real-life scenarios. This increases the robots' dependability and efficiency when handling materials.

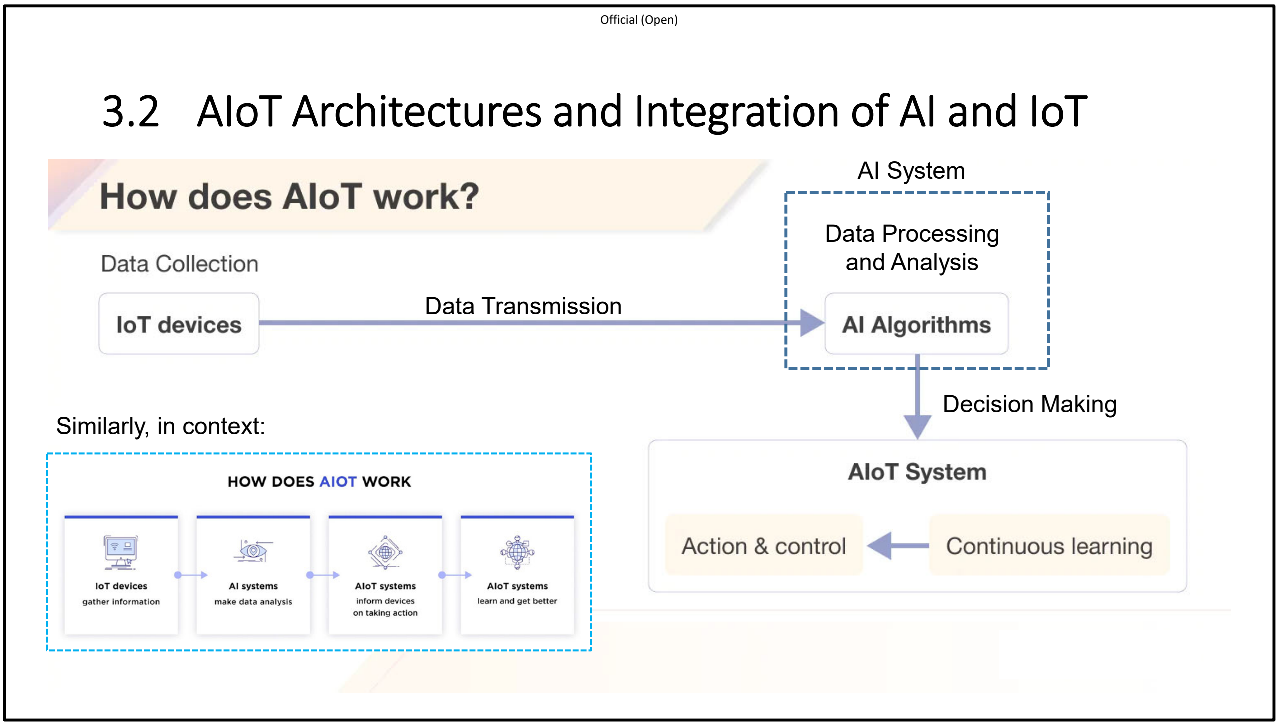
The ability of generative AI to produce program code for process automation greatly expedites the creation of automated systems in automation engineering. Generative AI can reduce the time and effort necessary for programming by producing optimised code that is suited to specific automation tasks by analysing input parameters and intended outcomes.

Furthermore, by examining equipment data and previous maintenance records, generative AI automates the synthesis of maintenance instructions in the context of equipment maintenance. This reduces downtime and improves equipment reliability by facilitating the early detection of possible problems and the creation of detailed maintenance instructions.

Additionally, the automated development of standard operating procedures (SOPs) for operator support is greatly aided by generative AI. Generative AI creates thorough SOPs that lead operators through challenging tasks and guarantee consistency and efficiency in operations by analysing operational data and best practices.

Lastly, generative AI integrates robotics for task automation in sophisticated shopfloor automation, allowing robots to carry out complex jobs on their own. Robots can learn and adapt to new tasks and environments with generative AI, which unlocks new levels of efficiency and flexibility in shopfloor automation. This is made possible by the use of machine learning algorithms.

To summarise, generative AI provides a broad strategy for improving the production environment, encompassing everything from material handling and product design to automation engineering and equipment upkeep. Generative artificial intelligence (AI) uses sophisticated algorithms and machine learning approaches to drive innovation, efficiency, and agility. This puts manufacturing operations in a competitive and future-ready position.



An advanced framework that combines artificial intelligence (AI) and the Internet of Things (IoT) is known as the AIoT architecture. This combination produces a potent ecosystem in which gadgets are intelligent, able to make decisions and learn from their experiences, in addition to being connected.

The seamless integration of AI technology into IoT systems is the core aspect of AIoT. AI gives Internet of Things (IoT) devices the ability to think, allowing them to analyse data, spot trends, and decide for themselves. Usually, the architecture goes through these stages:

1. **Data Collection**: Massive volumes of data are gathered by Internet of Things devices like wearables, cameras, and sensors from their environment.

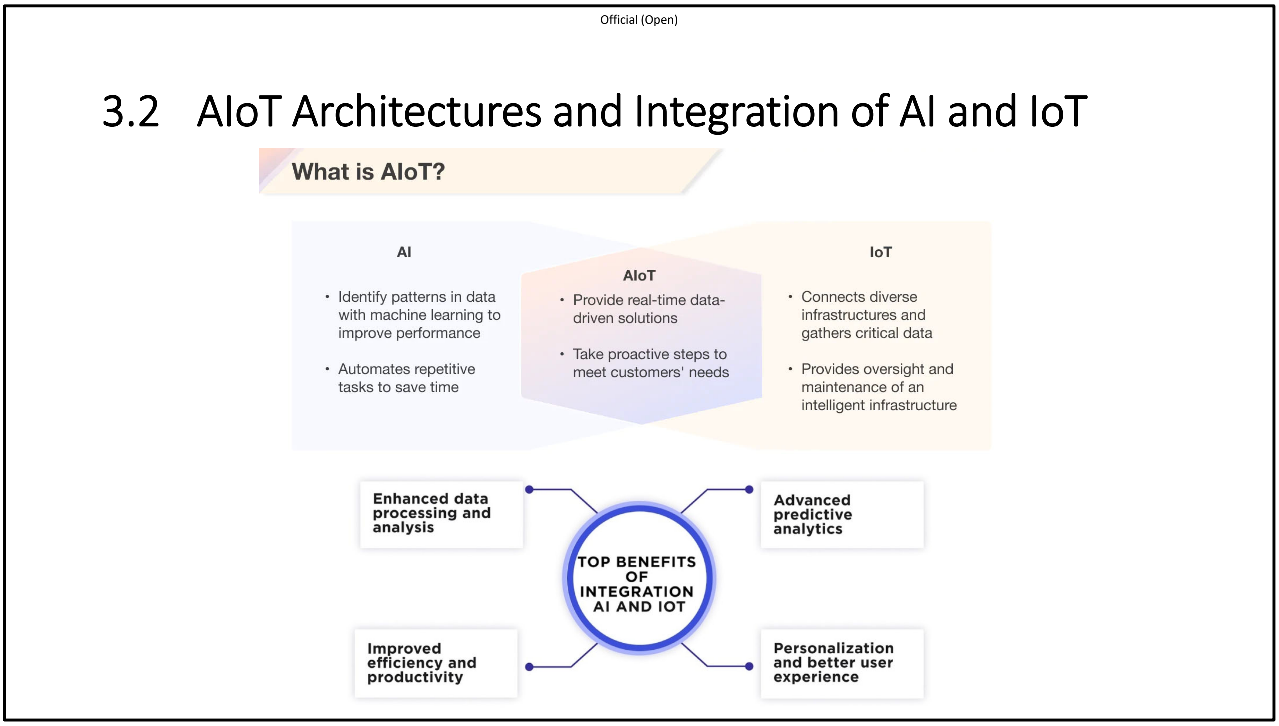
2. **Data Transmission**: For additional processing, the gathered data is sent to a cloud platform or centralised server.

3. **Data Processing and Analysis**: In order to extract insights, identify anomalies, and forecast trends, artificial intelligence (AI) algorithms, including machine learning and deep learning, process and analyse the data.

4. **Decision Making**: Decisions are made with the help of data analysis insights. AI gives the system the ability to make these decisions on its own, without assistance from a person.

5. **Action and Control**: Based on the decisions made, the AIoT system can set off actions or control other devices, resulting in automated and optimised operations.

6. **Continuous Learning**: AIoT systems are made to continuously learn from new data and user interactions, which increases their efficiency and accuracy over time.



Thus, it is evident that IoT and AI technologies can be combined to build intelligent, networked systems, in which AI serves as the IoT's brain. IoT devices gather and send data from many sources to assist AI's learning process and enable automation. Intelligent robots that mimic intelligent behaviour and assist in decision-making with little to no human intervention are produced via AI-enabled IoT. AI aids in the interpretation/making sense of the data that is sent between the IoT devices.

The integration of AI and IoT offers numerous benefits:

Top benefits are:

Real-Time Data Processing and Analysis: Real-time data analysis of information produced by internet-connected devices is made possible by AIoT. This makes it possible to respond and make decisions quickly.

Predictive Maintenance: Based on sensor data, AI models predict equipment breakdowns or the need for repair. It is possible to take preventative actions to avert expensive downtime.

Enhanced User Experience: Personalisation powered by AI enhances user experiences with IoT devices. Consider customised advice from your, for example, smart home setup.

Energy Efficiency: AIoT optimises energy consumption by modifying configurations in response to real-time data. Smart thermostats, for example, adjust based on occupancy patterns.

Other benefits include also:

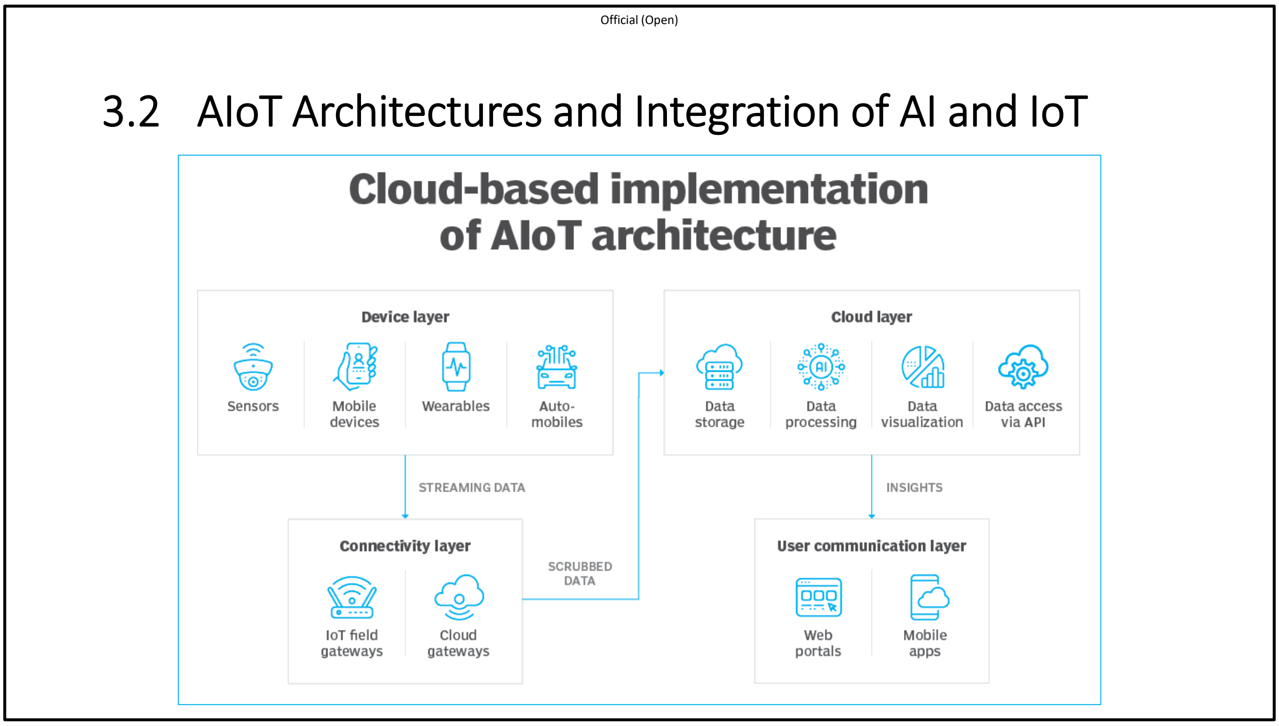
Security and Anomaly Detection: In IoT networks, AI algorithms find anomalies or security flaws. This improves system security as a whole.

New Business Models and Opportunities: AIoT is enabling smarter ecosystems and improving decision-making processes, which are revolutionising business models.

AIoT, for example, can reduce downtime in a smart factory by predicting equipment breakdowns and optimising production processes accordingly. AIoT can more effectively regulate traffic flows in smart cities, lowering pollution and congestion.

The system architecture must be carefully considered when implementing AIoT, the right technologies must be chosen, and seamless platform and device interoperability must be guaranteed. It is a challenging task that calls for both technical know-how and a calculated approach to meet corporate goals.

In conclusion, the AIoT architecture is the result of the fusion of the network of interconnected devices seen in IoT with the analytical and predictive powers of AI. Because of this integration, industries are changing as a result of smarter, more responsive, and more efficient systems. AIoT aims to create systems that are capable of thought, learning, and adaptation by integrating intelligence into the very fabric of these connections, rather than merely linking objects.



Primarily, AIoT systems are set up either as cloud-based or edge-based.

Cloud-based AIoT

Cloud-based IoT, often known as IoT cloud, is the use of cloud computing platforms to manage and process data from IoT devices. Since the cloud is where data is processed, stored, and accessible by numerous apps and services, connecting IoT devices to it is essential.

Cloud-based AIoT is composed of the following four layers:

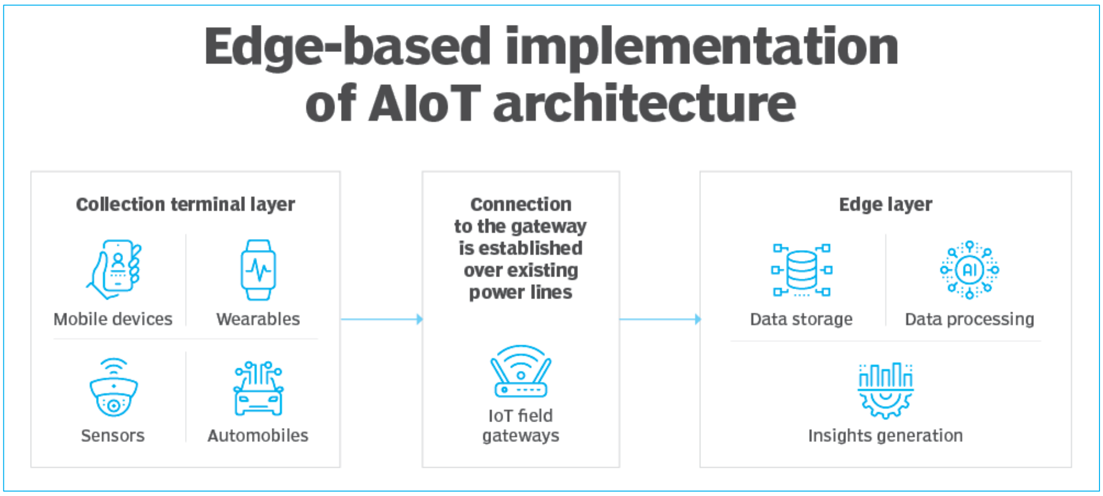
1. **Device layer** : This covers a wide range of hardware, such as automobiles, production equipment, embedded devices, tags, beacons, sensors, and health and fitness gear.

2. **Connectivity layer** : This layer connects cloud storage to controllers, sensors, and other intelligent devices through fields and cloud gateways, which are made up of a hardware or software element.

3. **Cloud layer** : This includes data storage, data visualisation, analytics, data processing using an AI engine, and data access through an API.

4. **User communication layer** : Mobile applications and web portals make up this layer.

3.2 AIoT Architectures and Integration of AI and IoT



Edge-based AIoT

AIoT data can also be processed "at the edge", which refers to processing IoT device data as near to the devices as feasible to reduce bandwidth requirements and prevent potential delays in data analysis.

Edge-based AIoT consists of the following three layers:

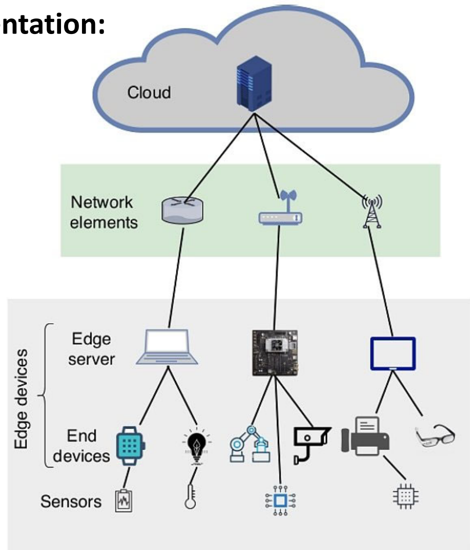
1. **Collection terminal layer** : This includes a variety of hardware items that are connected to the gateway over existing power lines, including embedded devices, vehicles, manufacturing equipment, tags, beacons, sensors, mobility devices, and health and fitness equipment.

2. **Connectivity layer** : This includes the field gateways to which the collection terminal layer is connected via the power lines that are currently in place..

3. **Edge layer** : Facilities for data processing, storage, and insight generation are included in this layer.

3.2 AIoT Architectures and Integration of AI and IoT

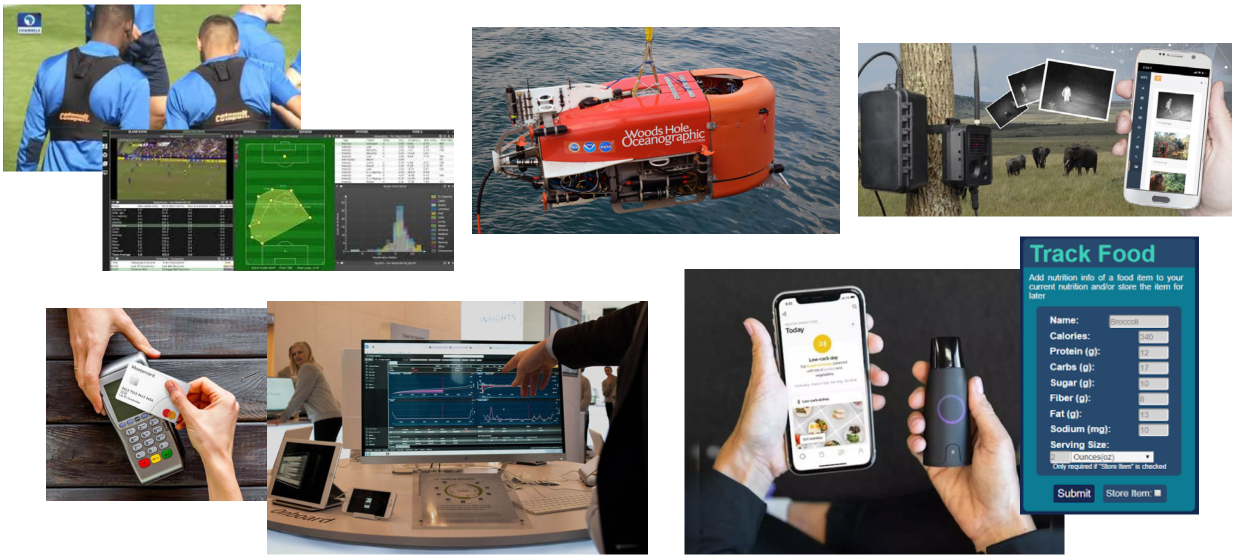
Edge Device Implementation:



The near future will see a lot of AI applications on end devices, often referred to as edge devices, in the upcoming years. These edge devices include the Ultra96 V2 board, Beagle Bone AI, Raspberry Pi, and Smart-edge Agile, among others. The Figure simply illustrates the vast range of applications that edge devices, or computing systems that process data closer to the data source at the network's edge, can offer across various sectors/industries.

Official (Open)

3.3 (Real-world) Examples of AIoT Applications



Here are real-world examples of AIoT applications found outside of smart cities, Industry 4.0, healthcare, and smart homes or buildings (as they will be discussed in later Chapters):

1. **Sports Performance Monitoring** : AIoT technologies are employed in professional sports to optimise training regimens and track players' performance. During training and competition, biometric data including heart rate, movement patterns, and hydration levels are collected via wearable devices with sensors. This data is analysed by AI algorithms to reveal information about the physical state of athletes, pinpoint areas for development, and prevent injuries. Examples include the Catapult Sports system used by professional sports teams like the NBA's Golden State Warriors and the English Premier League's Liverpool FC.

2. **Oceanography and Marine Biology** : Through the gathering and processing of data from remote ocean environments, AIoT is a key component of research in the fields of oceanography and marine biology. Data on ocean temperature, salinity, and marine life distribution are gathered using autonomous underwater vehicles (AUVs) outfitted with Internet of Things sensors. This data is processed by AI algorithms that track the migration patterns of marine species, assess the health of the ocean, and

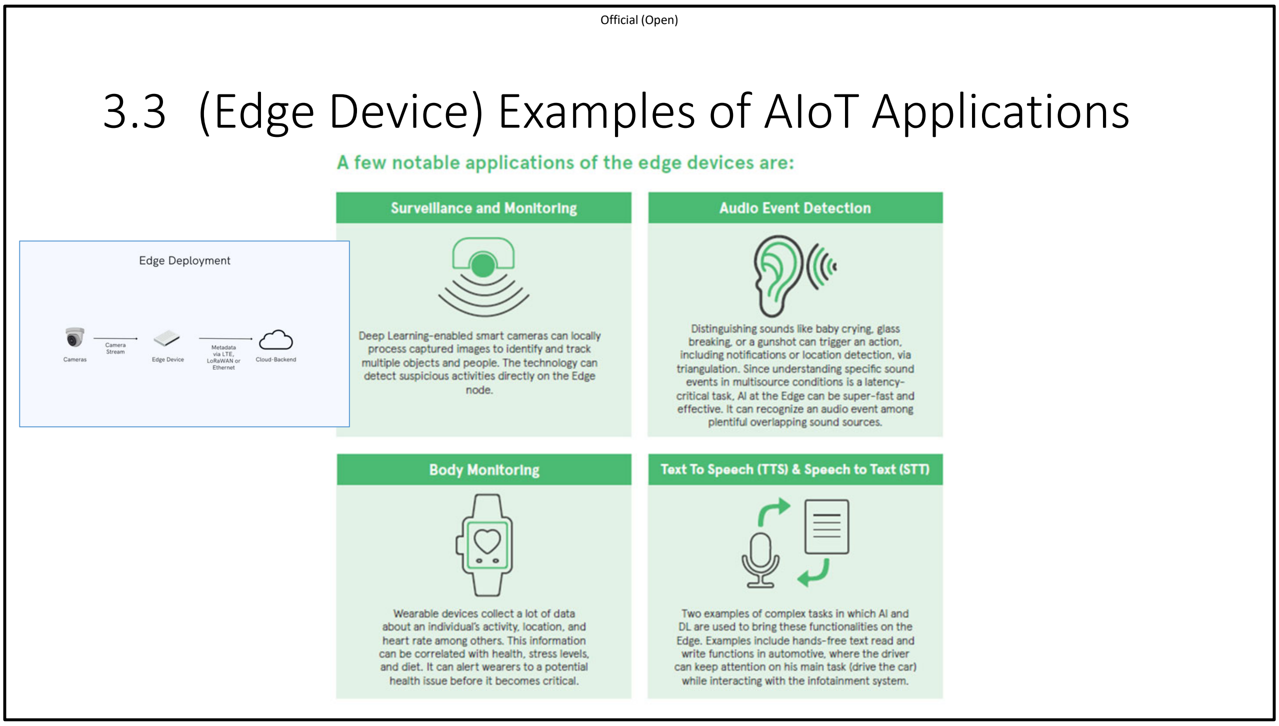
predict environmental changes. Examples include research initiatives led by organizations like the Woods Hole Oceanographic Institution and the Monterey Bay Aquarium Research Institute.

3. **Wildlife Conservation** : In an effort to monitor endangered animals and preserve their habitats, wildlife conservation organisations employ AIoT technologies. Data on the movements of wildlife, population dynamics, and environmental conditions are gathered using Internet of Things devices including acoustic sensors, GPS trackers, and camera traps. AI systems examine this information to find evidence of poaching, monitor animal movements, and evaluate the success of conservation efforts. Examples include projects led by conservation organizations like the World Wildlife Fund (WWF) and the Wildlife Conservation Society (WCS).

4. **Financial Services** : AIoT is used in the financial services sector for risk management, fraud detection, and customised banking services. Real-time tracking of market trends, customer interactions, and financial transactions is facilitated by IoT devices. AI systems examine this information to spot questionable activity, determine a person's creditworthiness, and provide them with tailored financial guidance. Examples include fraud detection systems developed by financial institutions like JPMorgan Chase and Mastercard.

5. **Wearable Technology for Personalised Nutrition** : Users' eating habits are tracked by wearable devices that are coupled with AIoT technology, which then generates individualised nutrition recommendations. Wearable technology, such fitness trackers and smartwatches, gathers information on a user's caloric intake, food intake, and nutritional objectives. With the assistance of AI algorithms, this data is analysed to create customised meal plans, recommend dietary changes, and monitor users' progress towards their health objectives. Examples include wearable nutrition trackers like Lumen and Nutralyz.

These examples show how the integration of AI and IoT technology is generating innovation and impact in multiple areas, illustrating the wide range of AIoT applications in several fields beyond the frequently recognised sectors.



Edge devices play a crucial role in various applications, offering real-time processing capabilities and reducing latency by handling data locally. Here are some notable applications:

1. **Surveillance and Monitoring**:
- Surveillance systems often use edge devices to monitor environments in real time. Edge processing-capable cameras can examine video streams locally to find abnormalities like unauthorised intrusions, suspicious behaviour, or object recognition (such recognising people or cars).
- Additionally, edge devices have the ability to trigger actions or alerts based on preset rules. For example, they can initiate alarms or send notifications to security staff.

<u>Benefits</u>:
Real-time response: Edge devices can trigger alarms or alerts immediately upon detecting suspicious activity.
Reduced network load: Local processing minimises bandwidth requirements.
Privacy: Sensitive data remains on-premises, enhancing privacy.

2. **Audio Event Detection**:

- When an edge device has a microphone, it can analyse sound patterns in real time to perform audio event detection. Urban locations, industrial settings, and smart homes can all benefit from this capabilities.
- Examples include listening for strange noises that indicate a failure in machinery, identifying specific sounds like alarms or screaming, detecting glass shattering, and identifying sounds from the environment like sirens or traffic.

Benefits:
Timely response: Immediate alerts for emergency situations.
Privacy preservation: Audio data stays local, ensuring privacy.

3. **Body Monitoring**:
- Real-time monitoring of many physiological parameters is possible with edge devices that are fitted with sensors such as gyroscopes, accelerometers, or biosensors.
- Wearable medical technology can monitor vital signs like blood pressure, heart rate, respiration rate, and activity level. Edge processing makes it possible to analyse this data instantly, allowing for the prompt notification of anomalies or medical emergencies.
- Wearable technology can measure performance indicators in sports and fitness, such as steps walked, calories burned, and workout intensity, and give users useful information for enhancing their fitness and overall health.

Benefits:
Continuous monitoring: Real-time health insights.
Early intervention: Detecting health issues promptly.

4. **Text-to-Speech and Speech-to-Text**:
- Edge devices are capable of doing local text-to-speech and speech-to-text conversions, as well as other speech processing activities including turning spoken words into text.
- Voice-activated gadgets, accessibility tools, and smart assistants are just a few of the industries in which this capacity finds use.
- By processing sensitive data locally, edge processing-capable smart speakers, for instance, can interpret voice instructions without requiring constant internet connectivity, resulting in faster response times and protecting user privacy.

Benefits:
Low latency: Immediate responses without internet delays.
Offline functionality: Works even when connectivity is limited.

Because they handle data locally, edge devices provide advantages including lower latency, better privacy, and higher reliability in each of these applications. They facilitate prompt response to occurrences and more effective use of available resources by facilitating real-time decision-making. Furthermore, edge computing lightens the load on cloud infrastructure and central servers, increasing the scalability and resilience of systems.