

Microsoft Azure AI Fundamentals

What is AI?

AI stands for Artificial Intelligence, which is a technology that allows computers to perform tasks that typically require human intelligence.

We can use AI to develop incredible software that can:

- enhance healthcare
- help individuals overcome physical limitations
- support advanced infrastructure
- provide exceptional entertainment experiences
- contribute to environmental conservation efforts

Capabilities of Azure AI / Key Workloads of AI

Machine learning → Azure Machine Learning services → Azure Machine Learning studio

Computer Vision → Azure AI Vision service → Azure Vision studio

Natural Language Processing → Azure AI Language services → Azure Language studio

Document Intelligence → Azure AI Document Intelligent Service → Document Intelligence studio

Knowledge Mining → Azure AI Search Service

Generative AI → Azure Open AI service → Azure AI Foundry

Understand Machine Learning

How machine learning works:

assume that our environmental conservation organization wants volunteers to identify & catalog different species of wildflowers using a phone app

1. A team of botanists and scientists collect data on wildflower samples
2. The team labels the samples with the correct species
3. The labelled data is processed using an algorithm that finds relationships between the features of the samples and the labelled species
4. The results of the algorithm are encapsulated in a model
5. When new samples are found by volunteer, the model can identify the correct species label

Introduction to Microsoft AI

Machine learning in Microsoft Azure

Azure Machine Learning services :- A cloud based platform for creating, managing & publishing machine learning models.

Azure Machine Learning studio :- Offers multiple authoring experiences such as :

- Automated machine learning
- Azure Machine Learning designer
- Data metric visualization
- Notebooks

Understand computer vision

Computer vision is an area of AI that deals with visual processing.

Computer vision models and capabilities :-

- Image classification
- Object detection
- Semantic Segmentation
- Image Analysis
- Face detection, analysis, and recognition
- Optical character recognition (OCR)

Computer vision services in Microsoft Azure

You can use Microsoft's Azure AI Vision to develop computer vision solutions.

The service features are available for use & testing in the Azure vision studio and other programming languages. Some features of Azure AI vision include:

- Image Analysis
- Face
- Optical character recognition (OCR)

understand NLP

Natural language processing (NLP) is the area of AI that deals with creating software that understands written & spoken language.

NLP enables you to create software that can:

- Analyze & interpret text in documents, email messages, and other sources.
- Interpret spoken language and synthesize speech responses.
- Automatically translate spoken or written phrases between languages.
- Interpret commands & determine appropriate actions.

NLP in Microsoft Azure

Azure AI Language:- Understanding and analyzing text, training conversational language models that can understand spoken or text-based commands, and building intelligent applications.

Azure AI Speech:- Build natural language processing solutions, with features including speech recognition and synthesis, real-time translations, conversation transcriptions, and more.

understand Document intelligence

- Focuses on handling, analyzing and utilizing large quantities of diverse data commonly found in forms & documents.
- This capability allows for the development of software that can automate the processing of contracts, health records, financial forms, & other similar documents.

Document intelligence in Microsoft Azure

Azure AI Document Intelligence:- Build solutions that manage & accelerate data collection from scanned documents, helps automate document processing in applications and workflows, enhance data-driven strategies and enrich document search capabilities.

Understand Knowledge Mining

The process of knowledge mining involves extracting information from large amounts of typically unorganized data in order to establish a searchable knowledge repository.

Knowledge Mining in Microsoft Azure (Azure AI search)

- An enterprise search solution that has tools for building indexes. The indexes can be used for internal use only, or to enable searchable content on public facing internet assets.
- The built-in AI capabilities of Azure AI services, like image processing, document intelligence, and natural language processing, can be used by Azure AI search to extract data.
- Due to the product's AI capabilities, it is possible to index documents that were previously unsearchable and to swiftly extract and surface insights from massive volumes of data.

Understanding Generative AI

- A subset of AI capabilities known as generative AI is responsible for producing original material.
- When generative AI is included into chat programs, people usually engage with it.
- Natural language input is fed into generative AI applications, which then output suitable replies in a range of formats, such as audio, images, code, & natural language.

Generative AI in Microsoft Azure (Azure OpenAI service)

- Microsoft's cloud solution for deploying, customizing, and hosting generative AI models
- Use Azure AI studio to create generative AI solutions, such as custom copilot chat-based assistants that use Azure OpenAI service models.

Challenges & Risks with AI

- 1) Bias can affect results
- 2) Errors may cause harm
- 3) Data could be exposed
- 4) Solutions may not work for everyone
- 5) users must trust a complex system
- 6) who's liable for AI-driven decisions?

V.SMP

Responsible AI Principles

Fairness → Equality for everyone, no discrimination

Reliability & Safety → Safe to use, no harm to user

Privacy & security → Protect Data.

Inclusiveness → Empower & engage everyone; Accessible to all.

Transparency → User must be aware of what & how AI system works.

Accountability → Legal process, if something goes wrong, the entire system is legally responsible.

2 kinds of questions:-

Scenario/use-case →

Q1) Which responsible AI Principle is being violated here?

Q2) Which responsible AI Principle is being followed/implemented here?

Fundamentals of Machine Learning

Machine Learning intersects data science and software engineering.

The goal is to create a predictive model using data for software applications or services.

Collaboration is needed between:

- Data scientists who prepare and train the model.
- Software developer who integrate the model into applications for making predictions (inference).

What is Machine Learning?

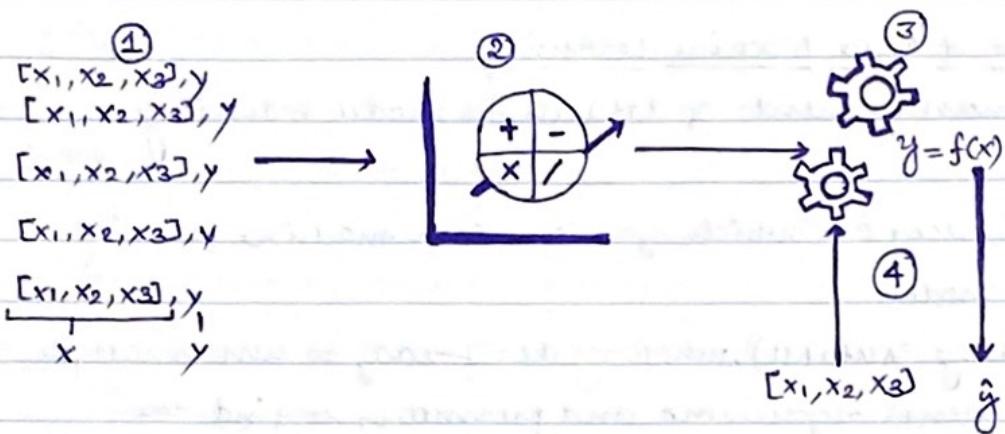
The fundamental idea of machine learning is to use data from past observations to predict unknown outcomes or values.

- Based on the weather forecast, an ice-cream shop owner may utilize an app that integrates past sales data & meteorological information to estimate how many cones of ice-cream they will sell on a given day.
- An automated test to determine if a new patient is at risk of diabetes based on measurements such as blood glucose level, weight, and other parameters may be done by a physician using clinical data from previous patients.
- The identification of several penguin species (such as Adelie, Gentoo, or Chinstrap) based on measurements of a bird's flippers, bill, and other physical characteristics might be automated by an Antarctic researcher using historical observations.

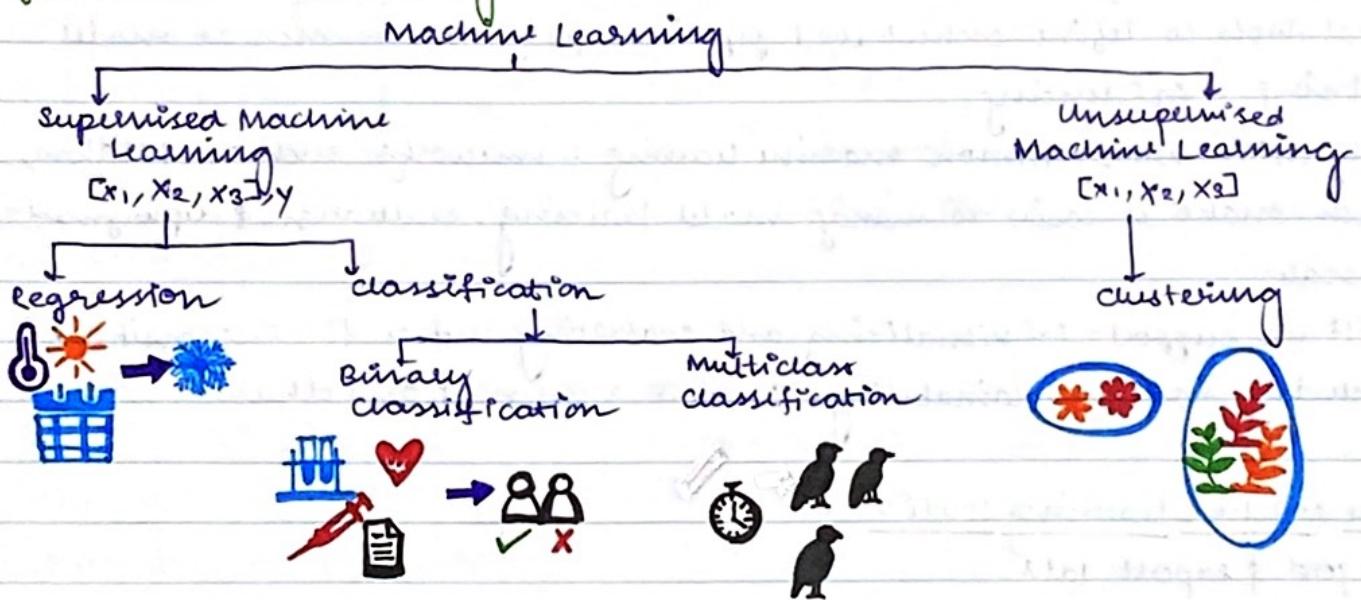
Machine learning as a function

- Because machine learning is based on mathematics and statistics, it's common to think about machine learning models in mathematical terms.
- Fundamentally, a machine learning model is a software application that encapsulates a function to calculate an output value based on one or more input values.
- The process of defining that function is known as training.
- After that function has been defined, it can be used to predict new values in a process called inferencing.

steps involved in training & inferencing



Types of Machine Learning



Deep learning → It is an advanced form of ML that tries to emulate the way the human brain works/learns.

Azure Machine Learning

The key to deep learning is the creation of an artificial neural network that simulates electrochemical activity in biological neurons by using mathematical functions.

Microsoft Azure Machine Learning is a cloud service for training, deploying, and managing machine learning models.

It is designed to be used by data scientists, software engineers, devops professionals, and others to manage the end-to-end lifecycle of machine learning projects.

Features and capabilities of Azure Machine Learning

- centralized storage and management of datasets for model training and evaluation.
- on-demand compute resources on which you can run machine learning jobs, such as training a model.
- automated machine learning (AutoML), which makes it easy to run multiple training jobs with different algorithms and parameters to find the best model for your data.
- visual tools to define orchestrated pipelines for processes such as model training or inferencing.
- integration with common machine learning frameworks such as MLflow, which make it easier to manage model training, evaluation, & deployment at scale.
- Built-in support for visualizing and evaluating metrics for responsible AI, including model explainability, fairness assessment, and others.

Azure Machine Learning Studio

- import & export data
- create & use remote resources
- run code in notebooks
- use visual tools to create jobs & pipelines
- use automated machine learning to train models.
- view details of trained models, including evaluation metrics, responsible AI information, and training parameters
- deploy trained models for on-request & batch inferencing
- import & manage models from a comprehensive model catalog.

A) Services on the Azure Platform

Azure AI services are based on 3 principles that dramatically improves speed-to-market:

- Prebuilt and ready to use
- Accessed through APIs
- Available on Azure

→ Create Azure AI service Resources



→ use the Azure portal to create an Azure AI service resource

→ use Azure AI services

→ Associate the AI service resource

→ Understand authentication for Azure AI services

(Authentication, the process of confirming that a user or service is who they are and that they are permitted to access the service, is how this is accomplished.)

→ Keys & Endpoint

Computer Vision

Computer vision focuses on developing tools that let AI programs "see" and interpret their environment.

The ability to process images is the key to creating software that can emulate human visual perception.

Image and image processing

The true nature of a picture in relation to program data

- images as pixel arrays
- Using filters to process images

Machine learning for computer vision

When using image processing software for activities like picture editing, the ability to apply effects to photos using filters is helpful.
→ But the main objective of computer vision is frequently to derive meaning - or at the very least, useful insights - from images.
→ To do this, machine learning models that have been trained to identify characteristics using a vast amount of preexisting image data must be developed.

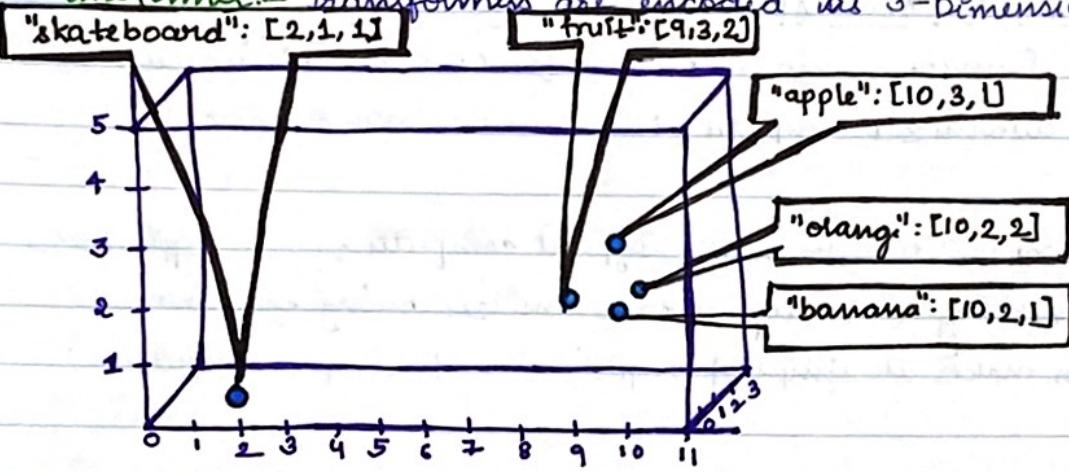
Convolutional Neural Networks (CNNs)

CNNs use filters to extract numeric feature map from images, and then feed the feature values into a deep learning model to generate a label prediction.

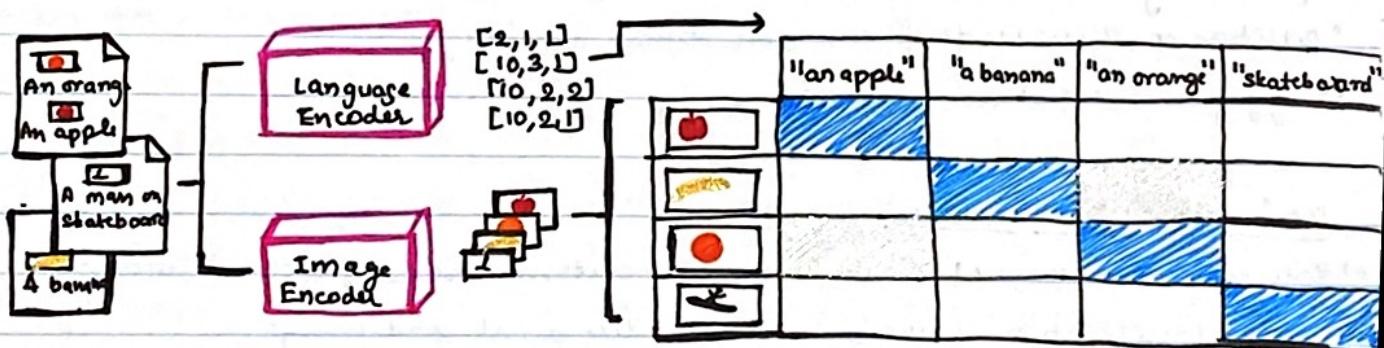
Transformers and Multi-Modal Models

To find different classes of objects in a single image, object detection models integrate CNN feature extraction layers with the recognition of image regions of interest.

Transformer:- Transformers are encoded as 3-Dimensional vectors.

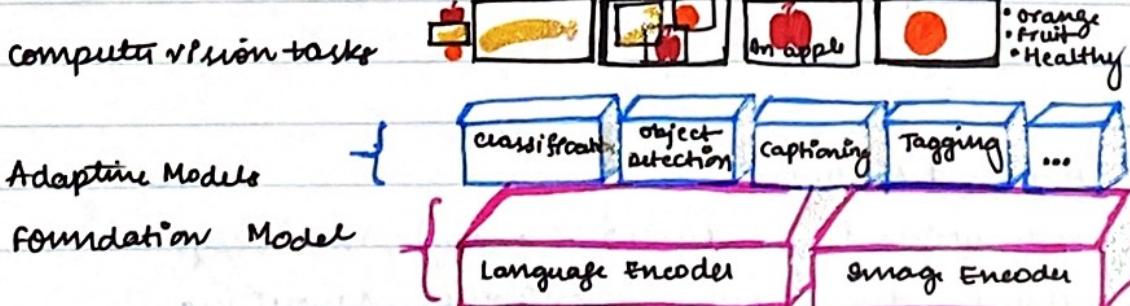


Multi-modal models: In multi-modal models, the model is trained using a large volume of captioned images, with no fixed labels.



Microsoft Florence Model

- **Image classification**:- Identifying to which category an image belongs.
- **Object detection**:- Locating individual objects within an image.
- **Captioning**:- generating appropriate descriptions of images.
- **Tagging** :- Compiling a list of relevant text tags for an image



Azure AI vision

Analyzing image with the Azure AI vision service

- Based on the Florence foundation model, Microsoft's Azure AI Vision service offers prebuilt and customized computer vision models with a range of potent features.
- Using "off-the-shelf" capabilities for many typical computer vision applications, together with the flexibility to develop custom models using your own photos, Azure AI Vision makes it simple & rapid to create complex computer vision solutions.
- Azure AI Vision supports multiple image analysis capabilities, including:
 - Optical Character Recognition (OCR) - Extracting text from images.
 - Generating captions & descriptions of images
 - Detection of thousands of common objects in images
 - Tagging visual features in images

Training custom models

- You can use Azure AI Vision to train a custom model for object detection or image classification if the built-in models aren't good enough.
- With Azure AI Vision, you can train complex models with very few training photos because bespoke models are built upon the pre-trained foundation model.

Fundamentals of Facial Recognition

Face detection and analysis is an area of artificial intelligence (AI) which uses algorithms to locate & analyze human faces in images or video content.

use of Face Detection & Analysis

There are many applications for face detection, analysis, and recognition. For eg:- Security, social media, intelligent monitoring, advertising, Missing persons, Identity validation

- Face detection involves identifying regions of an image that contain a human face, typically by returning bounding box coordinates that form a rectangle around the face.
- Facial features such as nose, eyes, eyebrows, lips, and others, can be utilized to train machine learning models for face analysis, which then returns more information.

Facial Recognition

- Facial analysis can also be used to train a machine learning model to recognize well-known people based only on their facial traits.
- This is called face recognition, and it trains the model by using several pictures of the same person.
- By doing this, the model is trained to recognize those people in newly-taken photographs that it hasn't been trained on.

Face analysis on Azure:- Microsoft Azure provides multiple Azure AI services that you can use to detect and analyze faces, including:

- Azure AI Vision → Basic face analysis.
- Azure AI Video Indexer → Analyze face in videos.
- Azure AI Face → in depth facial analysis
- Azure AI Service

Face service:- The Azure Face service can return the rectangle coordinates for any human faces that are found in an image, as well as a series of attributes related to those faces such as:- Accessories, Blur, Exposure, Glasses, Headpose, Mask, Noise, Occlusion.

Responsible AI use

Anyone can use the Face service to:-

- Detect the location of faces in an image
- Determine if a person is wearing glasses

- Detection determine if there's occlusion, blur, noise, or over/under exposure for any of the faces
 - Return the head pose coordinates for each face in an image
- * To use the Face service, you must create one of the following types of resource in your Azure subscription: Face & Azure AI Service.

Optical Character Recognition (OCR)

Optical character recognition (OCR), a technology that allows artificial intelligence (AI) to convert words in photographs into machine-readable text, allowing machine to interpret the text included in the images.

uses of OCR

- Text processing that is automated can increase productivity by doing away with the requirement for human data entry.
- Recognizing handwritten and printed text in photographs is useful for a variety of tasks, including taking notes, scanning cheques for bank deposits, digitizing historical or medical data, and more.

Azure AI Vision

- The ability for computer systems to process written and printed text is an area of AI where computer vision intersects with natural language processing.
- Vision capabilities are needed to "read" the text, and then natural language processing capabilities make sense of it.

Azure AI Vision's OCR Engine

- Images can have text that can be read by machines extracted using the Azure AI Vision service.

- The OCR engines that facilitates text extraction from pictures, PDFs, and TIFF files is called Azure AI Vision's Read API.
- Because OCR for pictures is designed for common, non-document images, integrating it into your user experience scenario is made simpler.

Azure AI Vision's OCR Engine

Calling the Read API returns results arranged into the following hierarchy:-

- Pages → one for each page of text, including information about the page size and orientation.
- Lines → The lines of text on a page.
- Words → The words in a line of text, including the bounding box coordinates and text itself.

Vision studio on Azure

To use the Azure AI Vision service you must first create a resource for it in your Azure subscription. You can use either of the following resource types:-

- Azure AI Vision

- Azure AI services

There are several ways to use Azure AI Vision's Read API:-

- Vision studio

- REST API

- Software development kits (SDKs) : Python, C++, Javascript

Points to be Remembered

* There are some considerations that can help improve the accuracy of the detection in the images:-

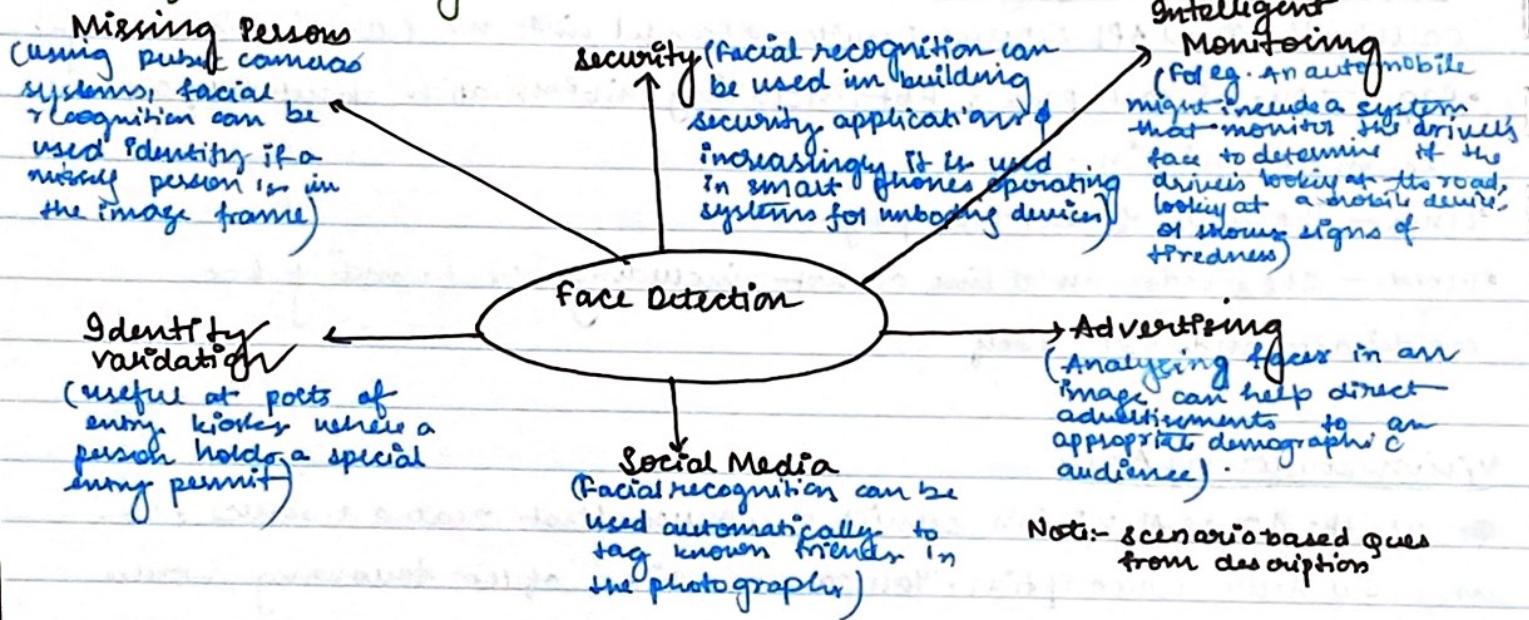
- Image Format → JPEG, PNG, GIF & BMP

- File Size :- 6 MB or less

- Face Size Range :- From 36x36 pixels upto 4096x4096 pixels. smaller or larger face will not be detected.

• Other issues → face detection can be impaired by extreme face angles, extreme lighting & occlusion (objects blocking the face such as hand).

Note :- convolutional filtering (Also known as Laplace filter) :- Highlights the edges of the image.



Natural Language Processing (NLP)

It is the area of AI that deals with creating software that understands written and spoken language.

NLP enables you to create software that can:-

- Analyze & interpret text in documents, emails, messages, & other sources.
- Interpret spoken language, & synthesize speech responses.
- Automatically translate spoken or written phrases between languages.
- Interpret commands & determine appropriate actions.

Text Analytics

- Text analysis describes NLP processes that extract information from unstructured text.
- Azure AI Language is a cloud-based service that includes features for understanding and analyzing text. Azure AI Language includes various features that support sentiment analysis, key phrase identification, text summarization, and conversational language understanding.
- Statistical analysis of a corpus of text (a collection of texts) is one of the early methods used in computer-assisted text analysis, and it is used to infer semantic meaning.
- Steps in analyzing a corpus:
 - Tokenization
 - Frequency analysis

Features of Text Analytics

Azure AI Language is a part of the Azure AI service offerings that can perform advanced natural language processing on unstructured text. Azure AI Language's text analysis features include:

- **Named Entity Recognition**: - Identifies people, places, events, and more. This feature can also be customized to extract custom categories.
- **Entity Linking** → Identifies known entities together with a link to Wikipedia.
- **Personal Identifying Information (PII) detection**: - Identifies personally sensitive information, including personal health information (PHI).
- **Language Detection**: - Identifies the language of the text & return a language code such as "en" for English.
- **Sentiment Analysis and Opinion Mining**: - Identifies whether text is positive or negative.
- **Summarization**: - Summarizes text by identifying the most important information.
- **Key Phrase extraction**: - Lists the main concepts from unstructured text.

* Question Answering

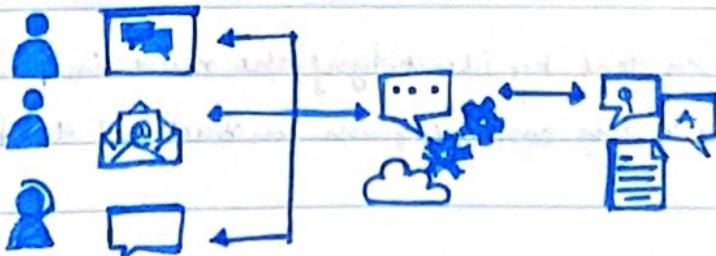
- solutions that facilitate communication between an AI agent & a person are referred to as conversational AI.
- Bots are general term for conversational AI Agents.
- Bots can interact with user via email, social media, web chat interfaces, and other methods.
- command Based Bot → Like Alexa, Siri, Google.
- Question answering supports natural language AI workloads that require an automated conversational element.
- Typically, question answering is used to build bot applications that respond to customer queries.
- You can create a user support bot solution on Microsoft Azure using a combination of 2 core services:-
 - Azure AI Language
 - Azure AI Bot Service

Creating a custom Question Answering Knowledge Base - we can use Azure AI Language studio to create, train, publish, and manage question answering projects.

- Define questions and answers
- Test the project

Build a bot with Azure AI Bot Service - we can create a custom bot by using the Microsoft Bot Framework SDK to write code that controls conversation flow and integrates with your knowledge base.

connect channels :- when the bot is ready to be delivered to users, you can connect it to multiple channels; making it possible for users to interact with it through web chat, email, Microsoft Teams, and other common communication media.



- Microsoft Azure supports conversational language understanding through its AI Language service.

* Conversational language Understanding (CLU)

- This feature enables you to author a language model & use it for predictions.
- Authoring a model involves the steps to define entities, intents, and utterances.

Utterance :- Words said by the user.

"Turn off the light"

Entity :- Thing in reference being spoken about.
"light"

Intent :- what does the user want to do?

"The light must be turned off."

- Generating predictions involves publishing a model so that client applications can take user input and return responses.

Azure resources for CLU :- To use conversational language capabilities in Azure, you need a resource in the Azure subscription. We can use the following types of resource:

- Azure AI Language
- Azure AI Services

The steps involved are : Authoring, Training and Prediction.

* Azure AI Speech (Speech Studio)

- With the help of AI speech capabilities, we can create captions from audio, answer questions aloud to computers, operate home and car systems with voice commands, and much more.
- An acoustic model that converts the audio signal into phonemes (representation of specific sounds). Phonemes include intent.

* Speech Recognition understanding what is spoken & who is speaking
speech to text

* Speech Synthesis : speech synthesis produces speech. Text to speech.

→ Microsoft Azure offers both speech recognition and speech synthesis capabilities through Azure AI speech service, which includes the following application programming interfaces (APIs):

- The speech to text API
- The text to speech API

→ To use Azure AI Speech in an application, you must create an appropriate resource in your Azure subscription. You can choose to create either of the following types of resources:-

- A speech resource
- An Azure AI service resource.

Speech-to-text API :- Perform real-time or batch transcription of audio into a text format.

The source for transcription can be a real-time audio stream from a microphone or an audio file.

- Real-time transcription
- Batch transcription.

Text-to-speech API :- Allows you to convert text input to audible speech, which can either be played directly through a computer's speaker or written to an audio file.

- speech synthesis voices.

language Translation

- Automated translation capabilities in an AI solution enable closer collaboration by removing language barriers.
 - One of the many challenges of translation between languages is that words don't have a one to one replacement between languages. Machine translation advancements are needed to improve the communication of meaning & tone between languages.
- Literal & semantic translation
 - Text & speech translation

Translation in Azure

- The Azure AI Translator service, which supports text-to-text translation.
 - The Azure AI Speech service, which enables speech-to-text & text-to-speech translation.
- We can use Azure AI Translator with a programming language of your choice or the REST API. You can use some of its features with Language Studio.

Azure resources for Azure AI Translator & Azure AI Speech

→ There are dedicated Translator and Speech resource types for these services which you can use if you want to manage access & billing for each service individually.

Azure AI Translator

→ Azure AI Translator includes following capabilities:

Text translation:- Used for quick & accurate text translation in real time across all supported languages.

Document translation:- Used to translate multiple documents across all supported languages while preserving original document structure.

Custom translation:- Used to enable enterprises, app developers, and language service providers to build customized neural machine translation (NMT) systems.

- speech translation with Azure AI speech - it includes the following capabilities
 - speech to text → used to transcribe speech from an audio source to text format.
 - text to speech → used to generate spoken audio from a text source.
 - speech translation → used to translate speech in one language to text or speech in another language.

Document Intelligence and Knowledge Mining

Fundamentals of Azure AI Document Intelligence

- The term "document intelligence" refers to AI skills that enable text processing and information extraction from text.
- As an extension of optical character recognition (OCR), document intelligence takes the next step a person might after reading a form or document.
- It automates the text data extraction, comprehension, and saving processes.
- Azure AI Document Intelligence supports features that can analyze documents and forms with pre-built and custom models.

Capabilities of Document Intelligence

- Document intelligence relies on machine learning, models that are trained to recognize data in text.
- The ability to extract text, layout, and key-value pairs are known as document analysis.
- Document analysis provides locations of text on a page identified by bounding box coordinates.

Azure AI Document Intelligence consists of features grouped by model type:

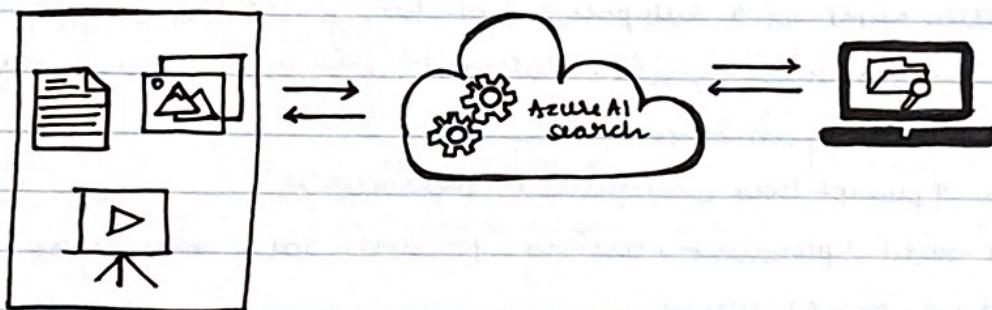
- Prebuilt models
- Custom models
- Document analysis

Fundamentals of Knowledge Mining and Azure AI Search

- Knowledge mining solutions provide automated information extraction from large volumes of often unstructured data.
- One of these knowledge mining solutions is Azure AI search, a cloud search service that has tools for building & managing indexes.
- Azure AI Search can index unstructured, typed, image-based or hand-written media.
- The indexes can be used to make content on publicly accessible internet assets searchable or for internal usage only.

What is Azure AI Search?

- Azure AI Search provides the infrastructure & tools to create search solutions that extract data from various structured, semi-structured, and unstructured documents.



Azure AI Search Features

- Azure AI Search provides the infrastructure & tools to create search solutions that extract data from various structured, semi-structured & non-structured documents.

Azure AI Search → PaaS

(Microsoft manages the infrastructure & availability, allowing your organization to benefit without the need to purchase or manage dedicated hardware resources)

Azure AI search comes with the following features:

- Data from any source
- Multiple options for search and analysis
- AI enrichment
- Linguistic analysis
- Configurable user experience
- Azure scale, security, and integration

→ Create search solution:-

- Azure portal's Import data wizard
- With the REST API
- With a software development kit (SDK)

Azure AI Foundry

The Azure AI Foundry portal is a web portal that brings together multiple Azure AI-related services into a single, unified development environment. Specifically, Azure AI Foundry combines:-

- The model catalog & prompt tool development capabilities
- The generative AI model deployment, testing, & custom data integration capabilities of Azure OpenAI service
- Integration with Azure AI services for speech, vision language, document intelligence & content safety

Identify elements of a search solution

A search index contains your searchable content. In an Azure AI search solution, you create a search index by moving data through the following indexing pipeline:

- 1) Start with a data source
- 2) Indexer
- 3) Document cracking

→ Enrichment

→ Push to index

→ Result populated as search index.

Generative AI

- The term "generative AI" refers to a class of AI capabilities that produce unique material. Usually, generative AI that is integrated into chat programs is used for human interaction.
- By utilizing machine learning to interact with the environment and carry out activities without clear instructions of what to produce, artificial intelligence (AI) mimics human behavior.

Notes:

- GPT-3.5-Turbo, GPT-4 and GPT-4o: Conversation-in & message-out language models.
- GPT-4-Turbo with Vision: A language model developed by OpenAI that can analyze images & provide textual responses to questions about them. It incorporates both natural language processing & visual understanding.
- DALL-E: A language mode that generates original images, variations of images, & can edit images.

Generative AI Capabilities

Generative AI applications take in natural language input, and return appropriate responses in a variety of formats such as natural language, images, or code.

- Natural language generation
- Image Generation
- Code Generation

Language Models

Gen AI applications are powered by language models.

They are a specialized type of machine learning model that can be used to perform natural language processing (NLP) tasks, like:

- Determining sentiment or otherwise classifying natural language text

- summarizing text

- comparing multiple text sources for semantic similarity

- generating new natural language

→ On Microsoft Azure, you can find foundation models in the Azure OpenAI service and in the Model Catalog

→ In addition to the Azure OpenAI models, the Model Catalog includes the latest open-source models from Microsoft and multiple partners, including:

- OpenAI

- Hugging Face

- Mistral

- Meta and others

Copilots

Copilots are generative AI assistants that are integrated into applications often as chat interfaces. They provide contextualized support for common tasks in those applications.

Copilot → ChatGPT + GenAI (similar to ChatGPT)

Microsoft Copilot:- Microsoft Copilot features can be found throughout all different Microsoft applications. They unlock productivity across your organization, safeguard your business, and build and extend your AI capabilities.

- Web browsing with AI

- AI assistance for information workers

- Use AI to support business processes

- AI assisted data analytics
- Manage IT infrastructure and security
- AI powered software development.

Extending & Developing copilots

If your organization makes the decision to customize Microsoft copilot or develop custom copilots. Microsoft provides 2 tools that you can use, Copilot Studio & Azure AI Studio.

Introduction to Azure AI Studio

- In order to create a single, centralized workspace where developers can work with data scientists and other experts to create AI solutions.
 - Azure AI Studio combines features from Azure Machine Learning, Azure OpenAI service, and other Azure AI services.
 - Azure AI Studio is a web portal that brings together multiple Azure AI related services into a single, unified development environment.
- It also includes support for:
- Creating Azure AI hubs that provide a collaborative development workspace for data scientists, developers, and others to build AI solutions.
 - Creating projects in which assets & code for a specific solution are managed.
 - Scalable, on-demand compute.
 - Integration with data sources & other cloud services.
 - Web-based code development environments.
 - SDKs & CLI libraries for automation.

Azure AI studio (Azure AI Foundry) functionalities

- Deploying models from the model catalog to real-time inference endpoints for client applications to consume.
- Deploying & testing generative AI models via an Azure OpenAI service.

- integrating data from custom data sources to support a RAG (retrieval augmented generation) approach to prompt engineering for generative AI models.
- using prompt flow to define workflows that integrate models, prompts, and custom processing
- integrating content safety filters into a generative AI solution to mitigate potential harms.
- extending a generative AI solution with multiple AI capabilities using Azure AI services
- RAG (Retrieval Augmented Generation) :- Azure AI Search + Azure OpenAI Azure AI Studio/Azure AI Foundry :- <https://ai.azure.com/>
- Request that you ask to Copilot (ChatGPT) → Prompt
- Response that you receive from Copilot (ChatGPT) → Completion
- Azure AI Foundry → Previously known as Azure AI Studio → one stop destination for working with Azure AI services, especially generative AI (OpenAI)
- Generative AI → Build your own ChatGPT which will only answer questions about any topic (Harry Potter Universe) choose an open AI Model.
- Azure OpenAI Service → Azure OpenAI Studio → Azure AI Foundry Copilot → ChatGPT → Generative AI (similar to ChatGPT)
- Language, Vision, DocIntel → Deployed as container → Any app can access them via APIs → You request a URL & action that you want to perform - You get a response.

When to use Azure AI Studio

It is the go-to platform when you need to:

- Create and manage AI projects
- Develop generative AI applications
- Explore available AI models
- Leverage Retrieval-Augmented Generation (RAG)

- Monitor & evaluate AI models
- Integrate with Azure services
- Build responsibly

Responsible generative AI in AI Studio

Given its potent potential, generative AI carries some risk, therefore data scientists, developer, and other stakeholders in generative AI solution creation must take a responsible approach that recognizes, quantifies, & reduce risks.

Responsible generative AI solution

The 4 stages in the process are:

- Identify potential harms that are relevant to your planned solution.
- Measure the presence of these harms in the outputs generated by your solution.
- Mitigate the harms at multiple layers in your solution to minimize their presence & impact, and ensure transparent communication about potential risks to users.
- Operate the solution responsibly by defining & following a deployment & operational readiness plan.

Identify Potential Harms

There are 4 steps in this stage:

1. Identify potential harms
2. Prioritize identified harms
3. Test & verify the prioritized harms
4. Document & share the verified harms.

Measure Potential Harms

After compiling a prioritized list of potential harmful output, you can test the solution to measure the presence & impact of harms. There are 3 steps involved:

- 1) Prepare a diverse selection of input prompts that are likely to result in each potential harm that you have documented for the system.
- 2) Submit the prompts to the system and retrieve the generated output.
- 3) Apply pre-defined criteria to evaluate the output and categorize it according to the level of potential harm it contains.

Mitigate potential harms

After determining a baseline and way to measure the harmful output generated by a solution, you can take steps to mitigate the potential harms, and when appropriate retest the modified system & compare harm levels against the baseline.

Mitigation techniques can be applied at each of four layers:

- Model
- Safety system
- Metaprompt & grounding
- User Experience

Confusion Matrix (Machine Learning)

- A confusion matrix, also known as an error matrix, is a table that helps evaluate the performance of machine learning classification model by comparing predicted values against the actual values for a dataset. It visualizes the performance of an algorithm, displaying the counts of true true, false -ve, false +ve and true -ne.
- It shows where a model gets confused by mapping the predictions to the original class, which is useful in supervised learning where the output distribution is known.

- True +ve (TP) - correctly predicted +ve values
- True -ve (TN) - correctly predicted -ve values
- False +ve (FP) → incorrectly predicted +ve values (Type I error)
- False -ve (FN) - incorrectly predicted -ve values (Type II error)
- Usefulness - the confusion matrix is useful because it helps identify where the model is making the most mistakes. It is useful when there is a significant class imbalance in the data.
- Metrics - from a confusion matrix, one can calculate various metrics such as accuracy, precision, recall, specificity, and the F1 score.
- Applicability - confusion matrices can be used with various classifier algorithms like Naive Bayes, logistic regression, and decision trees. They can be extended to multi-class classifiers.

Accuracy :-

$$(TN + TP) \div (TN + FN + TP + FP)$$

TN → True Negative

TP → True Positive

FN → False -ve

FP → False +ve

Recall :-

$$TP \div (TP + FN)$$

	+ve	-ve
True	TP	FP
False	FN	TN
Predicted	+ve	-ve

Precision :-

$$TP \div (TP + FP)$$

F1 score :- $(2 \times \text{precision} \times \text{recall}) \div (\text{precision} + \text{recall})$

Recall → True +ve Rate (TPR)

→ False +ve Rate (FPR) = $FP \div (FP + TN)$

ROC (Received Operator characteristic)

AUC (Area under the curve)

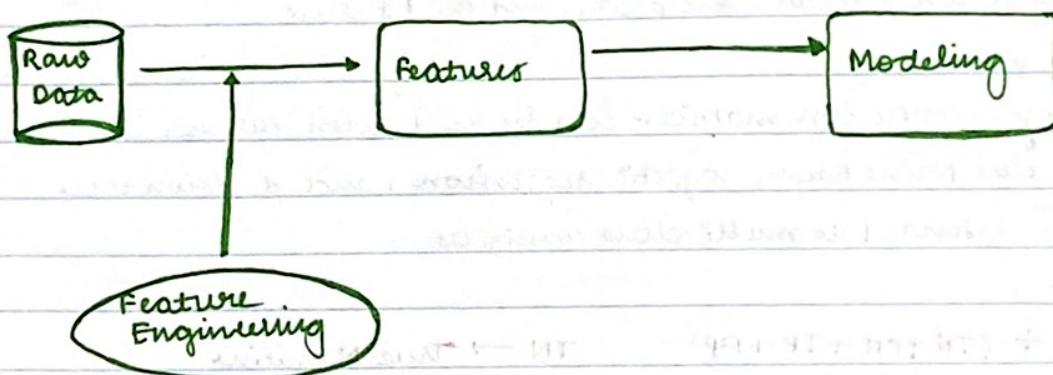
Model Evaluation

Examining the values of a confusion matrix.

Feature engineering

Feature engineering refers to the systematic approach of selecting, creating, transforming, and extracting features (input variables) from raw data to improve the accuracy and efficiency of machine learning models.

A feature can be understood as any measurable property or characteristic used by the model to make predictions.



Feature selection

Feature selection is a fundamental process in machine learning that involves identifying and selecting a subset of relevant input features from a larger dataset. This technique aims to improve model performance by reducing noise & complexity, thereby enhancing the model's interpretability & computational efficiency.

639-1 language code.
en - English
fr - French
zh - Chinese
3166 - 1

DOMS	Page No.
Date	/ /

Additional Points to Remember

- Tagging involves associating an image with metadata that summarizes the attributes of the image.
- Detecting image types involves identifying clip art images or line drawings.
- Content organization involves identifying people or objects in photos & organizing them based on the identification.
- Categorizing involves associating the contents of an image with a limited set of categories.
- Detecting objects identifies common objects and, for each, returns bounding box coordinates.
- Image categorization assigns a category to an image, but it does not return bounding box coordinates.
- Tagging involves associating an image with metadata that summarizes the attributes of the image, but it does not return bounding box coordinates.
- OCR detects printed & handwritten text in images, but it does not return bounding box coordinates.
- Each phrase returned by an image description task of the Azure AI Vision includes the confidence score.
- An endpoint & a key must be provided to access the Azure AI Vision service.
- Bounding box coordinates are returned by services such as object detection, but not ^{by} image description.
- Face attributes are a set of features that can be detected by the Face Detect API.
- Attributes such as accessories (glasses, mask, headwear) etc can be detected.
- For example, face ID, and face landmarks do not allow you to determine whether a person is wearing glasses or headwear.

- Azure AI Custom Vision is an image recognition service that allows you to build & deploy your own image models. The Azure AI vision service, Azure AI Face service, & Azure AI language service do not provide the capability to train your own image model.
- Azure OpenAI is the only service capable of generating text that can be used in chat applications to create conversational experiences. The other workloads (Azure AI Language, Azure AI Metrics Advisor, Azure AI Vision) are Azure Cognitive Services used for different purposes, but not for generating texts used in chat applications.
- Knowledge mining is an artificial intelligence (AI) workload that has the purpose of making large amounts of data searchable. While other workloads (image analysis, object detection, semantic segmentation) that also have the purpose of making large amounts of data searchable, leveraging indexing for faster access to large amounts of data, this is not their primary purpose.
- Translating text between different languages from product services is an NLP workload that uses the Azure AI Translator service & is part of Azure AI services. It can provide text translation of supported languages in real-time.
- Performing sentiment analysis feature of the Azure AI service for language. It can provide sentiment labels, such as -ne, neutral, & positive for text-based sentences & documents.
- OCR & spatial Analysis are part of the Azure AI Vision service.
- Sentiment analysis, entity recognition, & key phrase extraction are not part of the computer vision service. These are associated with Azure AI language service.
- Transparency provides clarity regarding the purpose of AI solutions, the way they work, as well as their limitations.
- The privacy & security, reliability & safety, & accountability principles focus on the capabilities of AI, rather than raising awareness about its limitations.

- the inclusiveness principle is meant to ensure that AI solutions empower & engage everyone regardless of criteria such as physical ability, gender, sexual orientation, or ethnicity.
- Privacy & security, reliability & safety, & accountability do not discriminate based on these criteria, but also do not emphasize the significance of bringing benefits to all parts of the society.
- Fairness involves evaluating & mitigating the bias introduced by the features of a model. Privacy is meant to ensure that privacy provisions are included in AI solutions.
- Transparency provides clarity regarding the purpose of AI solutions, the way they work, as well as their limitations.
- Accountability is focused on ensuring that AI solutions meet ethical & legal standards that are clearly defined.
- Accountability defines the framework of governance & organizational principles, which are meant to ensure that AI solutions meet ethical & legal standards that are clearly defined.
- Fairness, inclusiveness & transparency do not define the framework of governance & organization principles, but provide guidance regarding the ethical & legal aspects of the corresponding standards.
- The transparency principle states that AI systems must be designed in such a way that users are made fully aware of the purpose of the systems, how they work, & which limitations can be expected during use.
- The inclusiveness principle states that AI systems must empower people in a true & engaging way.
- Fairness is applied to AI systems to ensure that users of the systems are treated fairly.
- the privacy & security principle are applied to the design of AI systems to ensure that the systems are secure & to

Respect user privacy

- The accountability principle ensure that AI systems are designed to protect and meet any ethical & legal standards that are applicable.
- The privacy & security principle states that AI systems must be designed to protect any personal and / or sensitive data.
- The inclusiveness principle state that AI systems must empower people in a true & engaging way.
- The fairness principle is applied to AI system to ensure that users of the systems are treated fairly.
- Stemming normalizes words before counting them.
- Frequency analysis counts how often a word appears in a text.
- N-grams extend frequency analysis to include multi-term phrases.
- Vectorization captures semantic relationships between words by assigning them to locations in n-dimensional space.
- Summarization, also known as stemming, normalizes words before counting them.
- Removing stop words is the 1st step in the statistical analysis of terms used in a text in the context of NLP.
- Counting the occurrences of each word takes place after stop words are removed.
- Creating a vectorized mode is not part of statistical analysis. It is used to capture the semantic relationship b/w words.
- Encoding words as numeric features is not part of statistical analysis. It is frequently used in sentiment analysis.
- Language Name, ISO 6391 Code, # score are their values returned by the language service of NLP (Natural Language Processing) in Azure.
- Bounding box coordinates are returned by the Azure AI Vision services in Azure.
- Wikipedia URL is one of the potential values returned by entity linking at entity recognition.

- Entity recognition includes the entity linking functionality that returns links to external websites to disambiguate terms (entities) identified in a text
- Key phrase extraction evaluates the text of a document & identifies its main talking points
- Azure AI language detection identifies the language in which text is written
- Sentiment analysis evaluates text & returns sentiment scores & labels for each sentence
- The universal language Model used by the speech-to-text API is optimized for conversational & dictation scenarios. The acoustic, language & pronunciation scenarios require developing ^{your} own model.
- The Azure AI translation service supports text-to-text translation only, but does not support speech-to-text, text-to-speech & speech-to-speech.
- Entity linking, PII detection, & sentiment analysis are all 3 elements of the Azure AI service for Azure AI language.
- Azure AI Vision deals with image processing
- Azure AI Content Moderator is an Azure AI service that is used to check text, image, & video content for material that is potentially offensive
- Language identification, speaker recognition, & voice assistance are all elements of the Azure AI Speech service.
- Text-translation & document translation are part of the Translator service
- Hierarchical clustering groups data points that have similar characteristics
- Multiple linear regression models a relationship between two or more features & a single label.
- Linear regression uses a single feature
- Logistic regression is a type of classification model, which

return either a Boolean value or a categorical decision

- The regression algorithms predict a numeric label (continuous value) associated with an item based on its features. e.g. linear regression, polynomial regression & Decision tree regression
- clustering algorithms groups data points that have similar characteristics but doesn't assign specific labels.
- classification algorithms are used to predict the category to which an input value belongs & predicts a categorical label.
- unsupervised learning is a category of learning algorithms that includes clustering, but not of regression or classification. It refers to algorithms that learn patterns in data without labeled output.
- clustering algorithms group data points that have similar characteristics or clustering algorithms group observations based on similarity of their features
- Classification assigns labels to data based on features but doesn't group them by similarity.
- Regression predicts a numeric value for a given observation, not grouping
- supervised learning involves learning from labeled data which does not include clustering. It includes regression & classification
- classification is used to predict categories of data. It can predict which category or class an item of data belongs to.
- In this example sentiment analysis can be carried out on the Twitter posts with a numerical value applied to the posts to identify if classify the tweet as -no sentiment
- clustering is a machine learning type that analyzes unlabeled data to find similarities in the data.
- Regression is a machine learning scenario that is used to predict numeric values.
- Data transformation is not a machine learning type

- * A company is using machine learning to predict house price based on appropriate house attributes. For the machine learning model, which attribute is the label?

The price of the house is the label you are attempting to predict through the machine learning model.

This is typically done by using a regression model.

→ Floor space size, number of bedrooms, & age of the house are all input variables for the model to help predict the house price label.

→ Weather temperature & weekday or weekend are features that provide a weather temperature for a given day & a value based on whether the day is an

→ Azure AI speech provides speech-to-text & text-to-speech capabilities

→ Tokenization is part of speech synthesis that involves breaking text into individual words such that each word can be assigned phonetic sounds.

→ Transcribing is part of speech recognition which involves converting speech into a text representation.

→ Key-phrase extraction is part of language processing, not speech synthesis.

→ Lemmatization, also known as stemming, is part of language processing, not speech synthesis.

→ Azure AI services provide direct access to both Azure AI Translate & Azure AI Speech services through a single endpoint & authentication key.

→ Azure AI Language service can be used to access the Azure AI Language service, but not the AI Translator & Azure AI Speech services.

- > The machine learning service is used to design, implement, & deploy Machine learning models.
- > Azure AI Bot service provides a framework for developing, publishing, & managing bots in Azure.
- > Entity linking identifies & disambiguates the identity of entities found in a text.
- > Key phrase extraction is not used to extract entities & is used instead to extract key phrases to identify the main concepts in a text.
- > Named entity recognition cannot provide a link for each entity to view further information.
- > Text translation is part of the Azure AI Translator service.
- > Speech recognition uses audio data to analyze speech & determine recognizable patterns that can be mapped to distinct user voices.
- > Azure AI Speech synthesis is concerned with vocalizing data, usually by converting text to speech.
- > Azure AI speech translation is concerned with multilingual translation of speech.
- > Language identification is used to identify language spoken in audio when compared against a list of supported languages.
- > Azure AI Bot service provides a platform for conversational artificial intelligence (AI), which designates the ability of software agents to participate in a conversation.
- > Azure AI Translator is part of NLP but it does not serve a platform for conversational AI.
- > Azure AI Vision deals with image processing.
- > Azure AI Document Intelligence extracts information from scanned forms & invoices.
- > Extracting key phrases from text to identify the main terms is an NLP workload.
- > Predicting whether customers are likely to buy a product based on previous purchases requires the development of a machine learning model. It is related to Predictive analysis.

- Monitoring for sudden increases in quantity of failed sign-in attempts is a different workload. It is related to cybersecurity.
- Identifying objects in landscape images is a computer vision workload.
- Translating text between different languages from products services is an NLP workload that uses the Azure AI Translator service & is part of Azure AI Services. It can provide text translation of supported languages in real-time.
- Performing sentiment analysis on social media data is an NLP task that uses the sentiment analysis feature of the Azure AI Service for Language. It can provide sentiment labels, such as -ve, neutral, & +ve for text-based sentences & documents.
- OCR is a feature of Azure AI Vision that extracts text from images, scanned documents, or other visual media.
- Spatial Analysis is part of Azure AI Vision, enabling the detection of object locations & relationships in visual data, such as tracking people in a room or identifying object movement.
- Entity Recognition is a NLP task that identifies & classifies entities (like names, dates, locations) in text. It's unrelated to processing images or visual content, which is the focus of Azure AI Vision.
- Key Phrase Extraction: this task identifies important terms or concepts in text data, helping summarize textual content. It's part of text analytics under Azure AI Language.
- Sentiment Analysis: this involves determining the sentiment (+ve, -ve or neutral) of text, typically used for text-based applications like analyzing reviews or social media posts. It doesn't involve visual content.

- Fairness is when implementing an AI solution for business is an approach, fairness ensures that the system provides unbiased decisions & avoids discrimination against individuals or groups based on factors like gender, race, or other sensitive attributes. This principle is critical to maintain trust & equity in AI-driven financial services.
- Features are used to generate predictions for the label, which is compared to the actual label value. In regression machine learning algorithms, features are used to generate predictions for the label, which is compared to the actual label value. There is no direct comparison of features or labels between the validation & training datasets.
- In a regression machine learning algorithms, training set contains unknown feature & label values.
- outcome is the label value that you want to predict, also known as the independent variable.
- Number of people in the house, weather temperature, & time of year are features, & all values that are dependent on the label. No. of people in the house, weather temperature, & time of year can influence the water consumed in a household.
- Splitting data into training & validation datasets leaves you with 2 datasets - the first & largest of which is the training dataset you use to train the model. The second, smaller dataset is the held back data & is called the validation dataset, as it is used to evaluate the trained model. If normalizing or summarizing the data is required, it will be carried out as a part of data transformation.
- Cleaning missing data is part of preparing the data & the data transformation process.
- A dataset is required to create an automated machine using (Automated ML) arm. A workspace must be created before you can access Machine Learning studio. An Azure container instance & an AKS cluster can be deployment target, after training of a model is complete.

- A job must be created in Machine Learning studio to use Machine learning to train a regression model. A workspace must be created before you can access Machine Learning studio.
- An Azure container instance for on the cluster can be created as a deployment target, after training of a model is complete.
- Time-series forecasting, regression, & classification are supervised machine learning models. Automated ML learning can predict categories or classes by using a classification algorithm, as well as numeric values as part of the regression model algorithm & at a forecast future point in time by using time-series data.
- Ensemble pipeline is not a machine learning model.
- Clustering is unsupervised machine learning & automated ML only works with supervised learning algorithms.
- Normalize Data is a data transformation module that is used to change the values of numeric columns in a dataset to a common scale, without distorting differences in the range of values. The clean Missing Data module is part of preparing the data a transformation process.
- Select columns in Dataset is a dataset transformation component that is used to choose a subset of columns of interest from a dataset. The train clustering model is not a part of data transformation. The evaluate model is a component used to measure the accuracy of training models.
- A webpage or an entity document, such as a text file containing question & answer pairs, can be used to generate a knowledge base. You can manually enter the knowledge base question & answer pairs.
- You cannot directly use an image or an audio file to import a knowledge base.
- Image description is not a capability included in the DALL-E model, therefore, it is not a use case that can be implemented by using DALL-E 1 while the other 3 capabilities are offered by DALL-E in Azure OpenAI.

- Copilots are plugins to provide end users with the ability to get help with common tasks from a generative AI model. Copilots are often integrated into applications to provide a way for users to get help with common tasks from a generative AI model. Copilots are based on a common architecture, so developers can build custom copilots for various business-specific applications.
- The safety system layer includes platform-level configurations & capabilities that help mitigate harm. E.g., the True OpenAI service includes support for content filters that apply criteria to suppress prompts & responses based on the classification of content into 4 severity levels (safe, low, medium, & high) for 4 categories of potential harm (hate, sexual, violence, & self-harm).
- Generative AI can return responses, such as natural language, images, or code, based on natural language input.
- Generative AI models offer the capability of generating images based on prompts by using DALL-E-models, such as generating image from natural language.
- System messages should be used to set the context for the model by describing expectations. Based on system messages, the model knows how to respond to prompts. The other techniques are also used in generative AI models, but for other use cases.
- Identifying potential harms is the 1st stage when planning a responsible generative AI solution.
- True OpenAI natural language models can take in natural language & generate responses. GPT models are excellent at both. Models tend to generate natural language (image generation, image editing, image variation).
- Data grounding refers to ensuring that a model's responses are based on reliable, factual data.
- Machine learning is a broader concept that encompasses various techniques including embeddings, but it is not specifically the method for searching or comparing text similarity.

- system messages guide model behaviour & responses but are not used for text comparison or similarity tasks.
- embeddings is an AI model that converts text into numerical vectors for analysis. Embeddings can be used to search, classify, & compare sources of text for similarity.
- Modern image classification solutions are based on deep learning techniques.
- semantic segmentation provides the ability to classify individual pixels in an image depending on the object that they represent.
- Both linear regression & multiple linear regression use training & validate predictions to predict numeric values, so they are not part of image classification solutions.
- OCR can extract printed or handwritten text from images. In this case, it can be used to extract text from scanned medical records to produce a digital archive from paper-based documents.
- Identifying wildlife in an image is an example of a computer vision solution that uses object detection & is not suitable for OCR.
- Identifying a user requesting access to a laptop is done by taking images from the laptop's webcam & using facial detection & recognition to identify the user requesting access.
- Translating speech to text is an example of using speech translation & uses the Azure AI speech service as part of Azure AI services.
- Which 3 parts of the ML process does the Azure AI Vision eliminate? choosing a model, evaluating a model & training a model.

The computer vision service eliminates the needs for choosing, training & evaluating a model by providing pre-trained models to use computer vision yet must create an Azure resource. The use of the computer vision involves inferring

→ Azure AI Custom Vision is an image recognition service that allows you to build & deploy your own image model. The Azure AI vision service, Azure AI Face service, & Azure AI Language service do not provide the capability to train your own image model.

id

seed

#

image

in

l. #

train

use

use #