# Part 1 : Introduction to ZSL

## Definition and Motivation , Challenges in conventional supervised learning

**Slide 1:**

Intro 🡪

Welcome to the Generative AI and Cybersecurity FDP. I’m Sudhanva & I will be giving a presentation on ZSL

Tell about ZSL 🡪

In this section, we will introduce the fundamental concept of ZSL and its significance in the realm of machine learning.

**Slide 2 :**

Zero-shot learning allows a model to recognize what it **hasn’t seen before.**

1. Conventional supervised learning methods require large amounts of labeled data for each category, making it impractical for scenarios with a vast number of possible categories.
2. ZSL overcomes this limitation by enabling models to learn to generalize from known categories to the unknown ones through semantic embedding’s and attribute-based reasoning.
3. Let's dive deeper into the motivation and key components of ZSL on the next slide.

**Slide 3:**

* In conventional supervised learning, models can only classify objects into categories they have seen during training.
* However, this becomes impractical when we have a large number of possible categories, and obtaining labeled data for all is infeasible.
* Imagine you’re tasked with designing the latest and greatest machine learning model that can classify all animals. Yes, all animals.
* Using your machine learning knowledge, you immediately understand that we need a labeled dataset with at least one example for every single animal. There’s 1,899,587 described species in the world, so you’re gonna need a dataset with roughly 2 million different classes.

**Slide 4:**

1. The motivation behind ZSL is to enable machines to understand and categorize unseen objects by leveraging the knowledge acquired from known classes.
2. This is particularly useful in scenarios where new categories emerge regularly or where labeling data for unseen categories is prohibitively expensive or time-consuming.
3. ZSL opens up possibilities for more flexible and adaptable machine learning systems.

## **Outline of the presentation**

**Slide 5 :**

1. ZSL in NLP
2. ZSL in CVPR

Zero-shot learning is a subfield of transfer learning. The general idea of zero-shot learning is to transfer the knowledge already contained in the training instances to the task of testing instance classification. It relies on no labeled examples to learn a task . Zero-Shot learning is proposed to learn intermediate semantic layers and properties, then apply it to predict a new class of unseen data .

# ZSL in NLP

## Text Classification

**Slide 6:**

1. NLP is a field that focuses on enabling computers to understand, interpret, and generate human language, making it an essential part of many applications such as chatbots, sentiment analysis, and machine translation.
2. However, traditional NLP methods often require extensive labeled data for every possible category, making them impractical for scenarios with numerous unseen categories. This is where Zero-Shot Learning comes to the rescue, allowing NLP models to generalize to new, previously unseen categories without explicit training.

One more thing that comes to mind when we talk about the zero-shot text classification zero-shot classification is the few-shot classification which is also similar to the zero-shot classification but this type of modelling possesses the usage of very few labelled samples in the time of training.

We find the implementation of the few-shot classification methods in OpenAI where GPT-3 is a well-known few-shot classifier.

((( [Zero-shot text classification](https://huggingface.co/tasks/zero-shot-classification) is a task in natural language processing where a model is trained on a set of labeled examples but is then able to classify new examples from previously unseen classes. )))

**Slide 7:**

* One of the key components enabling ZSL in NLP is the use of semantic embeddings.
* Semantic embeddings are dense vector representations of words or phrases, capturing their semantic meanings in a continuous space.
* Popular examples of semantic embeddings include Word2Vec, GloVe, and FastText, which have proven to be effective in various NLP tasks.
* These embeddings enable knowledge transfer from seen to unseen categories, making ZSL feasible in NLP.

**Slide 9:**

* Zero Shot Classification is the task of predicting a class that wasn't seen by the model during training.
* This method, which leverages a pre-trained language model, can be thought of as an instance of [transfer learning](https://www.youtube.com/watch?v=BqqfQnyjmgg) which generally refers to using a model trained for one task in a different application than what it was originally trained for.
* This is particularly useful for situations where the amount of labeled data is small.
* Instead of training the model on each individual class, we leverage semantic embeddings to infer relationships between seen and unseen categories.
* For example, if our model has seen examples of movie genres like "Action," "Romance," and "Comedy," it can generalize to classify texts into new genres like "Sci-Fi" or "Thriller" without explicit training.

**Slide 10:**

1. As we know, text classification is a task of natural language processing where the model needs to predict the classes of the text documents.
2. In the traditional process, we are required to use a huge amount of labelled data to train the model, and also they can’t predict using the unseen data.
3. Adding zero-shot learning with text classification has taken natural language processing to the extreme.
4. The main goal of any model related to the zero-shot text classification technique is to classify the text documents without using any single labelled data or without having seen any labelled text.
5. We mainly find the implementations of zero-shot classification in the transformers.
6. In the hugging face transformers, we can find that there are more than 60  transformers that work based on zero-shot classification.

**Slide 11: Applications in Sentiment Analysis**

* Sentiment analysis is another area where Zero-Shot Learning can be immensely valuable.
* Traditional sentiment analysis models are usually trained on labeled data for specific sentiments (e.g., positive, negative, neutral).
* With ZSL, we can go beyond these predefined sentiments and classify text into more nuanced and specific emotional categories.
* For instance, using semantic embeddings, a model trained on basic emotions like "happy," "sad," and "angry" can recognize emotions like "excited," "fearful," and "relieved."

## ZSL in CVPR

**Slide 12 :**

[Zero shot image classification](https://huggingface.co/tasks/zero-shot-image-classification)

In this section, we'll explore how ZSL is applied in Computer Vision and Pattern Recognition (CVPR) tasks. CVPR tasks involve image classification, object recognition, object tracking and more.

ZSL in CVPR allows us to recognize and categorize objects from unseen classes, extending the capabilities of traditional supervised learning. Let's dive in!

**Slide 13 & 14 : Image Representation**

**Zero-shot image classification is a computer vision task to classify images into one of several classes, without any prior training or knowledge of the classes.**

Zero-shot or open vocabulary image classification models are typically multi-modal models that have been trained on a large dataset of images and associated descriptions. These models learn aligned vision-language representations that can be used for many downstream tasks including zero-shot image classification.

This is a more flexible approach to image classification that allows models to generalize to new and unseen categories without the need for additional training data and enables users to query images with free-form text descriptions of their target objects .

**Slide 15:** CLIP (Multimodal model)

* **CLIP** is trained using a staggering amount of **400 million image-text pairs.** For comparison, the ImageNet dataset contains 1.2 million images.
* CLIP (Contrastive Language-Image Pre-Training) is a neural network trained on a variety of (image, text) pairs. It can be instructed in natural language to predict the most relevant text snippet, given an image, without directly optimizing for the task, similarly to the zero-shot capabilities of GPT-2 and 3.
* CLIP takes an Image, text pairing as input to learn a multi-modal embedding space. CLIP jointly trains an image encoder and text encoder to maximize the cosine similarity of the image and text embedding of the correct pair and minimize the cosine similarity of the image and text embeddings of the incorrect pairings.

**Slide 17:** Object Detection

Zero-shot object detection is supported by the OWL-ViT model which uses a different approach. OWL-ViT is an open-vocabulary object detector. It means that it can detect objects in images based on free-text queries without the need to fine-tune the model on labeled datasets.

OWL-ViT leverages multi-modal representations to perform open-vocabulary detection. It combines [CLIP](https://huggingface.co/docs/transformers/model_doc/clip) with lightweight object classification and localization heads.

Open-vocabulary detection is achieved by embedding free-text queries with the text encoder of CLIP and using them as input to the object classification and localization heads. associate images and their corresponding textual descriptions, and ViT processes image patches as inputs.

## Challenges and Future Directions

**Slide 18 :**

Like every concept, Zero-Shot Learning has its limitations. Here are some of the most common challenges you'll face when applying Zero-Shot Learning in practice.

**Bias -** During the training phase, the model has access only to the data and labels of the seen classes. This biases the model towards predicting unseen data samples during the test time as one of the seen classes. In cases where, during the test time, the model is evaluated on samples from both seen and unseen classes, the bias problem becomes more prominent.

**Domain shift -** Domain shift results when the statistical distribution of the data in the training set (seen classes) and the testing set (which may be samples from the seen or unseen classes) are significantly different.

**Hubness -** Hubness is a phenomenon that occurs when high-dimensional data is projected into a lower-dimensional space. [It can cause problems in zero-shot learning because it can lead to an over-representation of certain classes](https://www.v7labs.com/blog/zero-shot-learning-guide)

**Semantic loss -** While training on the seen classes, the model learns only the important attributes for distinguishing between these seen classes. But, some latent information may be present in the seen classes, which are not learned if they don’t contribute significantly to the decision-making process. However, this information might be important in the testing phase on the unseen classes. This is what we call semantic loss.

**Final Slide**

1. In conclusion, Zero-Shot Learning represents a groundbreaking paradigm in machine learning that offers unprecedented flexibility, cost-effectiveness, and the ability to tackle real-world challenges across diverse domains.
2. Its capability to recognize and classify unseen categories has the potential to revolutionize industries and enhance the capabilities of AI systems, opening up new opportunities for innovation and problem-solving.
3. As researchers continue to explore and refine ZSL techniques, we can look forward to a future where machines learn to adapt and understand the world more like humans do, paving the way for exciting advancements in AI and its practical applications.

# Reference Links

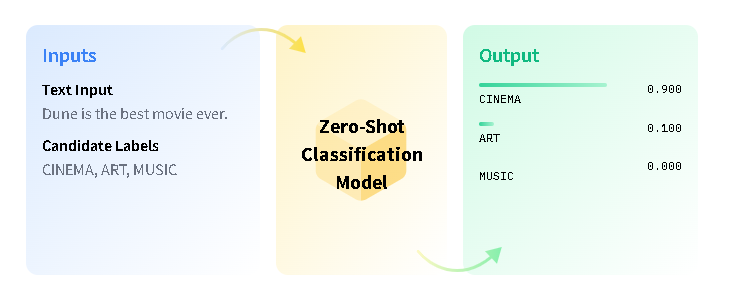
[Zero shot image classification](https://huggingface.co/tasks/zero-shot-image-classification) is the task of classifying previously unseen classes during training of a model.

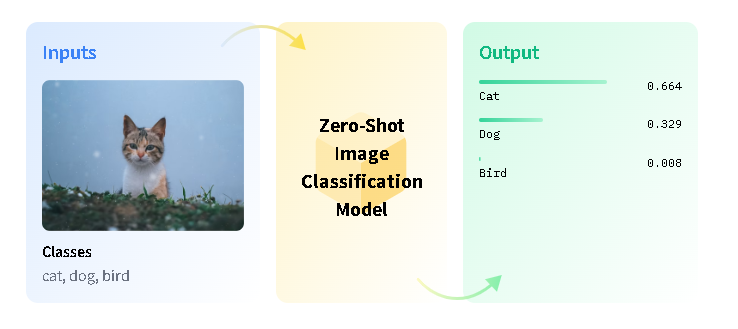
https://bpiyush.github.io/projects/1\_zero\_shot\_learning/

ZSL has a range of applications, including [image classification](https://blog.roboflow.com/how-to-use-openai-clip/), object detection, [object tracking](https://blog.roboflow.com/zero-shot-object-tracking/), semantic segmentation, style transfer and natural language processing.

<https://github.com/joeddav/zero-shot-demo>

https://blog.roboflow.com/zero-shot-learning-computer-vision/





Ext Notes

Natural language processing is a very exciting field right now. In recent years, the community has begun to figure out some pretty effective methods of learning from the enormous amounts of unlabeled data available on the internet. The success of transfer learning from unsupervised models has allowed us to surpass virtually all existing benchmarks on downstream supervised learning tasks. As we continue to develop new model architectures and unsupervised learning objectives, "state of the art" continues to be a rapidly moving target for many tasks where large amounts of labeled data are available.

One major advantage as models continue to grow is that we see a very slow decrease in the reliance on large amounts of annotated data for downstream tasks. This week the team at Open AI released a preprint describing their largest model yet, GPT-3, with 175 billion parameters. The paper is entitled, ["Language Models are Few-Shot Learners"](https://arxiv.org/abs/2005.14165), and shows that extremely large language models can perform competitively on downstream tasks with far less task-specific data than would be required by smaller models.

Old Links :   
https://towardsdatascience.com/understanding-zero-shot-learning-making-ml-more-human-4653ac35ccab

https://blog.roboflow.com/zero-shot-learning-computer-vision/

**It is important to notice that ZSL is not a form of unsupervised learning**

<https://jrodthoughts.medium.com/understanding-zero-shot-learning-8da323a5b588>

<https://towardsdatascience.com/zero-shot-vs-similarity-based-text-classification-83115d9879f5>

<https://joeddav.github.io/blog/2020/05/29/ZSL.html>

### huggingface :

https://huggingface.co/tasks/zero-shot-classification

https://huggingface.co/tasks/zero-shot-image-classification

https://huggingface.co/blog/clipseg-zero-shot

low-priority :

https://analyticsindiamag.com/a-complete-tutorial-on-zero-shot-text-classification/

https://jedleee.medium.com/zero-shot-topic-classification-fb92d1b33cfb

# Old Notes :

Zero Shot Classification is the task of predicting a class that wasn't seen by the model during training.

# **Background : Zero-Shot Learning**

Zero-shot learning (ZSL) is a model's ability to detect classes never seen during training. The condition is that the classes are not known during supervised learning.

Earlier work in zero-shot learning used attributes in a two-step approach to infer unknown classes. In the computer vision context, more recent advances learn mappings from image feature space to semantic space. Other approaches learn non-linear multimodal embeddings. In the modern NLP context, language models can be evaluated on downstream tasks without fine tuning.

## About the Task

Zero-shot image classification is a computer vision task to classify images into one of several classes, without any prior training or knowledge of the classes.

Zero shot image classification works by transferring knowledge learnt during training of one model, to classify novel classes that were not present in the training data. So this is a variation of [transfer learning](https://www.youtube.com/watch?v=BqqfQnyjmgg). For instance, a model trained to differentiate cars from airplanes can be used to classify images of ships. This is particularly useful for situations where the amount of labeled data is small.

The data in this learning paradigm consists of

* Seen data - images and their corresponding labels
* Unseen data - only labels and no images
* Auxiliary information - additional information given to the model during training connecting the unseen and seen data. This can be in the form of textual description or word embeddings.