

**INTERNSHIP REPORT**

**On**

**“Short-Term Internship on Artificial Intelligence”**

**At**

**“Artificial Intelligence Medical & Engineering Researchers Society”**

**(AIMER Society)**

**Submitted By**

**TADI MOHANAPAVANI (22KP1A1250)**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**CERTIFICATE**

This is to certify that the internship work embodied in this report entitled “Short-Term Internship on Artificial Intelligence” during June & July-2024 was carried out by student name 22KP1A1250 of Information Technology, Guntur for partial fulfillment of B.Tech degree to be awarded by JNTU Kakinada. This Internship work has been carried out is to the satisfaction of department.

Place:

Date:

Signature of the Internal Guide

Signature of the HOD

Signature of the External

## **References and Acknowledgements**

References:

- 1.ChatGPT-4o
- 2.Google
- 3.Youtube
- 4.Hugging Face
- 5.Tensorflow
- 6.google teachable machine

## **ACKNOWLEDGEMENT**

An industrial attachment cannot be completed without significant help from others. First we gratefully acknowledge the help and support from our parents, teachers, employers, friends and others, whose support has been invaluable for me. I would like to thank the following people for their contribution in this industrial attachment

We express our gratitude to my honorable SOWJANYA BATTULA, HOD, Information Technology, who has monitored and directed us for this report.

We would like to express my gratitude to Artificial Intelligence Medical And Engineering Researchers Society for providing me with this opportunity. Special thanks to SAI SATISH for their guidance and support throughout the internship. It has been enriching experience that has been significantly contributed to my professional and personal growth

## **ABSTRACT**

**Computer Vision** capabilities include extensive experience with image processing, object detection, classification, and human pose estimation using tools like Google Teachable Machine and Mediapipe Studio. The proficiency extends to implementing robust algorithms such as YOLO and Faster R-CNN for real-time applications across various domains, including medical imaging, traffic monitoring, and drone technology. Expertise in image labeling tools like Roboflow highlights a strong understanding of annotation's importance in AI pipelines.

Additionally, foundational knowledge of OpenCV underpins practical problem-solving across vision tasks.

In **Artificial Intelligence**, the user demonstrates skills in developing and training advanced models, including Convolutional Neural Networks (CNNs), Large Language Models (LLMs), and lightweight models using TensorFlow Lite. Key areas of focus include **Generative AI** for content creation, **visual question answering**, and sentiment analysis. The integration of models like BERT, GPT, and Google's Vision API further underscores the user's ability to bridge the gap between visual and textual data. The deployment and management of AI applications using platforms like Ultralytics Hub and Dialogflow illustrate an understanding of creating scalable and interactive solutions.

**Cybersecurity** expertise encompasses fundamental principles such as the CIA Triad (Confidentiality, Integrity, Availability) and AAA Framework (Authentication, Authorization, Accounting). Proficiency includes identifying and mitigating web vulnerabilities like SQL Injection and Cross-Site Scripting (XSS) and implementing network protection measures using firewalls. Familiarity with the OWASP framework and tools like Acunetix demonstrates hands-on experience in securing systems against cyber threats.

## Table of Contents

Certificate	
Acknowledgement	
Abstract	
Chapter 1 Industry/company profile.....	5
1.1 Firm History .....	5
1.2 Management Board of the firm.....	6
1.3 Vision and Mission.....	6
1.4 Services.....	6
1.5 Employees and their experience.....	6
Chapter 2 Internship Experience & Job Description (your contribution in)	
2.1 Name and short Description of the project .....	7
2.2 Involvement in Software/Hardware development process....	8

2.3 Involvement in Planning.....	8
2.4 Images/Screen shots.....	9-25
Chapter 3 Conclusion.....	104-106

## **Chapter 1 Industry/company profile :**

### **1.1 Firm History, AIMER SOCIETY**

Name: Artificial Intelligence Medical and Engineering Researchers Society (AIMER Society)

Overview:

The Artificial Intelligence Medical and Engineering Researchers Society (AIMER Society) stands as a premier professional organization at the forefront of the advancement of Artificial Intelligence (AI) within the realms of medical and engineering research. This esteemed society is committed to driving innovation and excellence in AI by fostering a collaborative environment among researchers, practitioners, and students from diverse backgrounds and disciplines.

The AIMER Society's mission is to serve as a catalyst for the development and application of cutting-edge AI technologies that can address complex challenges in healthcare and engineering. By creating a vibrant and inclusive platform, the society facilitates the exchange of knowledge, ideas, and best practices among its members. This collaborative approach ensures that AI research is not only innovative but also practically applicable, leading to real-world solutions that can significantly improve medical outcomes and engineering processes.

In pursuit of its mission, the AIMER Society organizes a wide array of activities and initiatives designed to promote AI research and development. These include annual conferences, symposiums, and workshops that bring together leading AI experts to discuss the latest advancements and trends. Such events provide invaluable opportunities for networking, collaboration, and professional growth.

Mission:

The mission of the AIMER Society is to promote the development and application of AI technologies to solve complex medical and engineering problems, improve healthcare outcomes, and enhance engineering solutions. The society aims to bridge the gap between theoretical research and practical implementation, encouraging interdisciplinary collaboration and real-world impact.

### **Impact and Achievements:**

- Developed AI models for early diagnosis and treatment of medical conditions.
- Contributed to significant advancements in engineering solutions through AI technologies.
- Fostered a global community of AI researchers and practitioners.

- Organized successful conferences and workshops with high participation and impactful outcomes.
- Published influential research papers and articles in reputed journals.

Future Goals:

- Expand the scope of research and applications in AI to cover emerging fields and technologies.
- Increase collaboration with international AI societies and organizations.
- Enhance training and certification programs to meet the evolving needs of AI professionals.
- Promote ethical AI practices and address challenges related to AI governance and societal impact.

## **1.2 Management Board of the firm:**

- Website: AIMER Society Website <http://www.aimersociety.com>
- Email: info@aimersociety.org
- Phone: +91 9618222220
- Address: Sriram Chandranagar, Vijayawada

## **1.3 Vision and Mission:**

- To advance research in AI and its applications in medical and engineering fields.
- To provide a platform for researchers, practitioners, and students to share knowledge and collaborate on AI projects.
- To organize conferences, workshops, and seminars for the dissemination of AI research and knowledge.
- To support the professional development of AI researchers and practitioners through training programs, certifications, and networking opportunities.
- To foster ethical AI practices and address societal challenges related to AI deployment.

## **1.4 Services:**

- Conferences and Workshops: Organizing annual conferences, symposiums, and workshops that bring together leading AI experts, researchers, and practitioners to discuss the latest advancements and trends in AI.
- Research Publications: Publishing high-quality research papers, journals, and articles on AI technologies and their applications in medical and engineering fields.
- Competitions and Contests: Hosting AI model development and chatbot contests to encourage innovation and practical applications of AI among students and professionals.
- Training Programs: Offering training and certification programs in AI and related technologies to enhance the skills and knowledge of members.
- Collaboration Projects: Facilitating collaborative projects between academia, industry, and healthcare institutions to drive AI innovation and practical solutions.

Membership:

The AIMER Society offers various membership categories, including individual, student, and corporate memberships. Members gain access to exclusive resources, networking opportunities, and discounts on events and publications. The society encourages participation from AI enthusiasts, researchers, practitioners, and organizations interested in the advancement of AI technologies.

## **1.5 Employees and their experience:**

The AIMER Society is led by a team of experienced professionals and experts in the fields of AI, medical research, and engineering. The leadership team is responsible for strategic planning, organizing events, and guiding the society towards achieving its mission and objectives.

## **Chapter 2 Internship Experience & Job Description**

### **2.1 Name and short Description of the project**

S.NO	TOPICS	LINKS
1	OBJECT DETECTION	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_dronetechnology-computervision-opencv-activity-7201969090268155905-ieAZ?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/pavani-pavani-b1225928b_dronetechnology-computervision-opencv-activity-7201969090268155905-ieAZ?utm_source=share&amp;utm_medium=member_android</a>
2	IMAGE CLASSIFICATION	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_object-detection-using-yolov8-aimer-society-activity-7201617563095621633-81C3?utm_source=share&amp;utm_medium=member_androidhttps://www.linkedin.com/posts/pavani-pavani-b1225928b_object-detection-using-yolov8-aimer-society-activity-7201617563095621633-81C3?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/pavani-pavani-b1225928b_object-detection-using-yolov8-aimer-society-activity-7201617563095621633-81C3?utm_source=share&amp;utm_medium=member_androidhttps://www.linkedin.com/posts/pavani-pavani-b1225928b_object-detection-using-yolov8-aimer-society-activity-7201617563095621633-81C3?utm_source=share&amp;utm_medium=member_android</a>
3	MEDICAL IMAGE ANALYSIS AND LABELING :	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210203136534224896-pROT?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210203136534224896-pROT?utm_source=share&amp;utm_medium=member_android</a>
4	POSE ESTIMATION	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210206196463464449-Lo49?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210206196463464449-Lo49?utm_source=share&amp;utm_medium=member_android</a>
5	MEDIAPIPE STUDIO	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210204238277197824-608N?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-deeplearning-activity-7210204238277197824-608N?utm_source=share&amp;utm_medium=member_android</a>
6	GOOGLE DIALOGFLOW	<a href="https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-conversa">https://www.linkedin.com/posts/pavani-pavani-b1225928b_ai-machinelearning-conversa</a>

		<a href="https://www.linkedin.com/posts/mohana-pavan-b1225928b_machinelearning-ai-healthcare-reinnovation-activity-7212362385234251776-SXjb?utm_source=share&amp;utm_medium=member_android">tionalai-activity-7210211913606082560-yPu8?utm_source=share&amp;utm_medium=member_android</a>
7	VISUAL QUESTIONS AND ANSWERING	<a href="https://www.linkedin.com/posts/mohana-pavan-b1225928b_machinelearning-ai-healthcare-reinnovation-activity-7212362385234251776-SXjb?utm_source=share&amp;utm_medium=member_android">https://www.linkedin.com/posts/mohana-pavan-b1225928b_machinelearning-ai-healthcare-reinnovation-activity-7212362385234251776-SXjb?utm_source=share&amp;utm_medium=member_android</a>

## 2.2 Involvement in Software/Hardware development process

Here we have to used the software application to do this project is

- > hugging face website
- >google colab
- >pycharm

## 2.3 Involvement in Planning

### COMPUTER VISION :

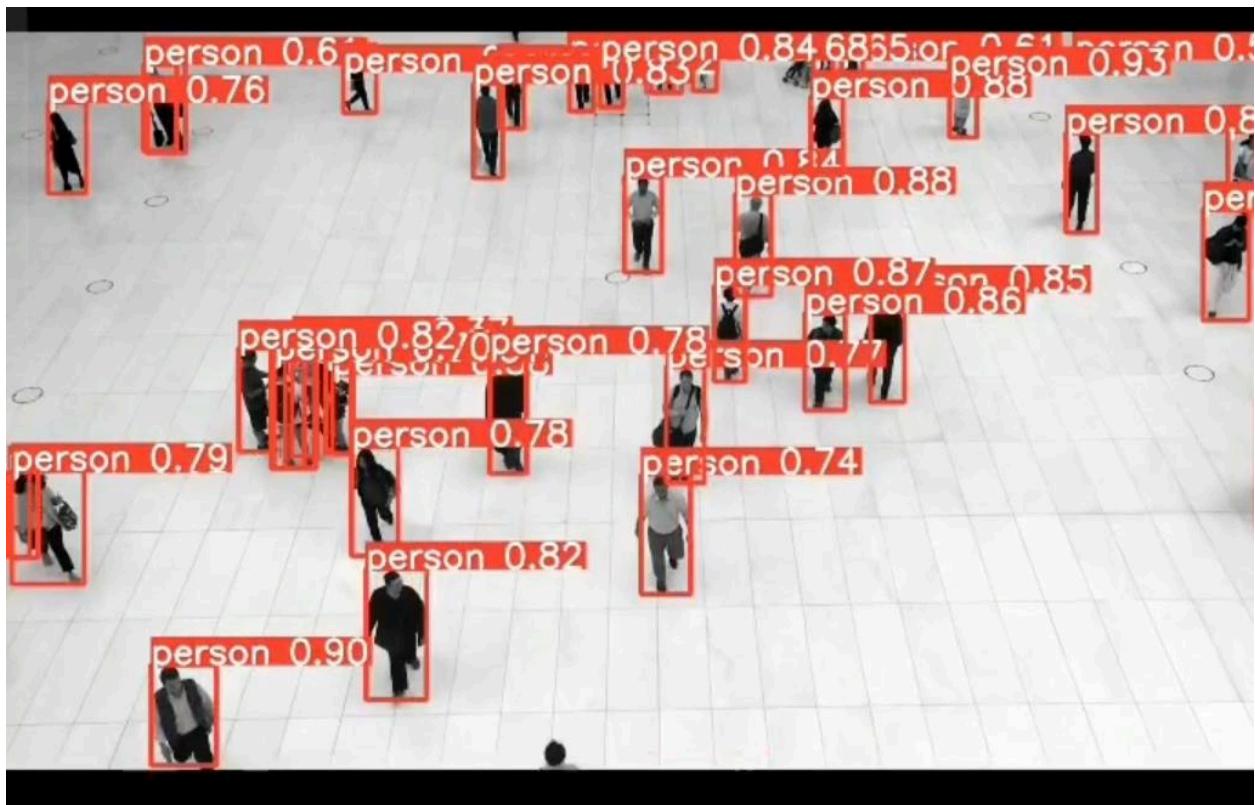
Computer vision is nothing but identifying objects and people in images and videos .It is a type of artificial intelligence, that analyze the visual world and understand their environment. It applies machine learning models to identify and classify objects in digital images and videos. Let us see some example of computer vision , public security -facial recognition , facial detection and recognition are some of the most prominent computer vision technology examples. We come across this AI application in a lot of different shapes and forms. But the public security sector is the most significant driver of the pervasive use of facial detection. Computer vision is used in a lot applications from retail to security, healthcare, construction, automotive, manufacturing, logistics,

### OBJECT DETECTION :

Object detection is the automatic inference of what an object is in a given image or video frame. It's used in self-driving cars, tracking, face detection, pose detection, and a lot more. There are 3 major types of object detection – using OpenCV, a machine learning-based approach, and a deep learning-based approach  
Object detection is a technique that uses neural networks to localize and classify objects in images. This computer vision task has a wide range of applications, from medical imaging to self-driving cars.

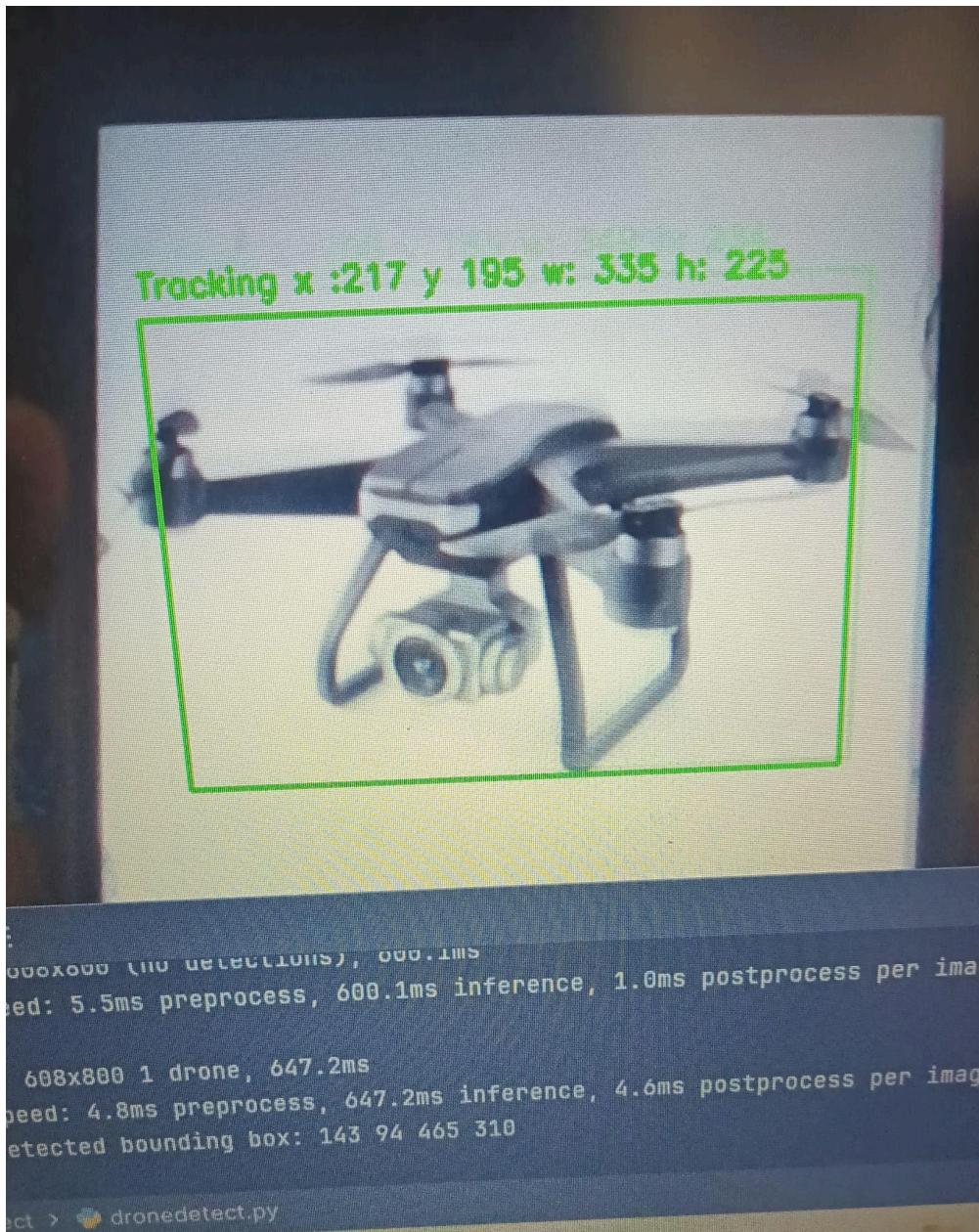
Object detection is a [computer vision](#) task that aims to locate objects in digital images. As such, it is an instance of [artificial intelligence](#) that consists of training computers to see as humans do,

specifically by recognizing and classifying objects according to semantic categories. Object localization is a technique for determining the location specific objects in an image by demarcating the object through a bounding box. Object classification is another technique that determines to which category a detected object belongs. The object detection task combines subtasks of object localization and classification to simultaneously estimate the location and type of object instances in one or more images.



In the above image i have trained the model to detect the people and applied a image to the model and it detected the people in the image we can use tensorflow or yolo to train the model first we need to datasets to the model then we can train the model here i applied people dataset to model so that it can detect people in the image

I also did another project detecting x/y-coordinates of an object i trained the model using yolov8 and write a code in python to open web cam and track the x/y-coordinates of the object



## IMAGE CLASSIFICATION :

### **Image Classification using Google's Teachable Machine**

Machine learning is a scientific field that allows computers to learn without being programmed

directly. students, engineers, and data scientists use machine learning to create projects and

goods, the application of machine learning is trendy. However, the development of machine

learning models involves high parameters, also the model training process can be from hours to days. Hence, low-end systems can not train successful machine learning models, or crucial system problems are likely to arise.

However, several Machine Learning environments are easily available on the internet that do not need any system specification or framework specifications and use cloud technology to train the model in the best possible time. Some of these open-source machine learning environments are Google Colaboratory, and Kaggle Kernel, these are an excellent platform for deep learning and machine learning applications in the cloud. Both of them are Google products and require the knowledge of data science to develop and train models using them. However, Google introduced a new open-source platform for training machine learning models that developers can code i.e. Google's Teachable Machine.

The Google Teachable Machine is an online open-source environment that is used to develop and train machine learning and deep learning supervised models without using any programming language.

**Below is the step-by-step approach on how to use the Teachable Machine to develop and train machine learning models:**

- Go to <https://teachablemachine.withgoogle.com/>

≡

About FAQ Get Started

# Teachable Machine

**Train a computer to recognize your own images, sounds, & poses.**

A fast, easy way to create machine learning models for your sites, apps, and more – no expertise or coding required.

[Get Started](#)

- Click on *Get Started* and choose whether to open an existing project or create a new project. In-order to create a new project we have three options i.e. Image Project, Audio Project, or Pose Project. Click on the Image project.

≡ Teachable Machine

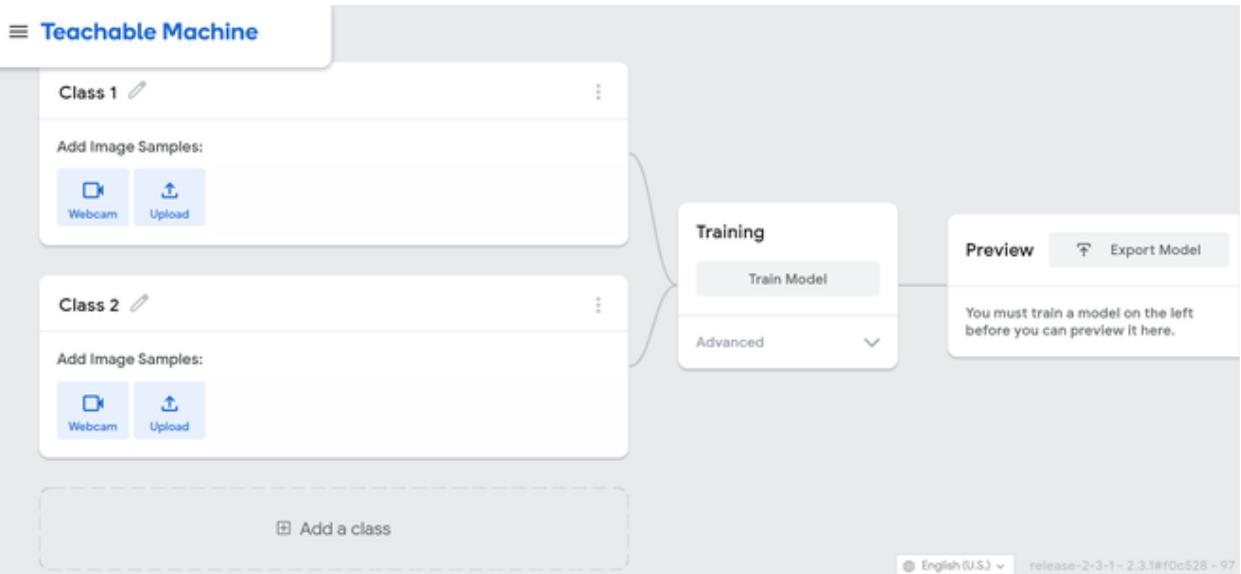
Open an existing project from Drive. Open an existing project from a file.

**Image Project**  
Teach based on images, from files or your webcam.

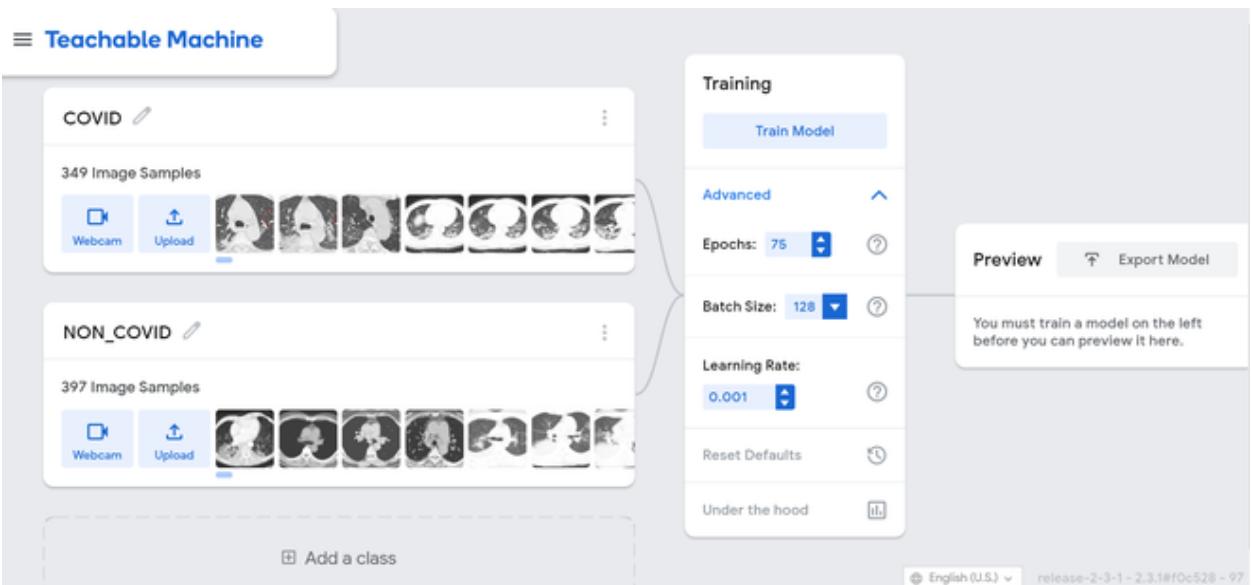
**Audio Project**  
Teach based on one-second-long sounds, from files or your microphone.

**Pose Project**  
Teach based on images, from files or your webcam.

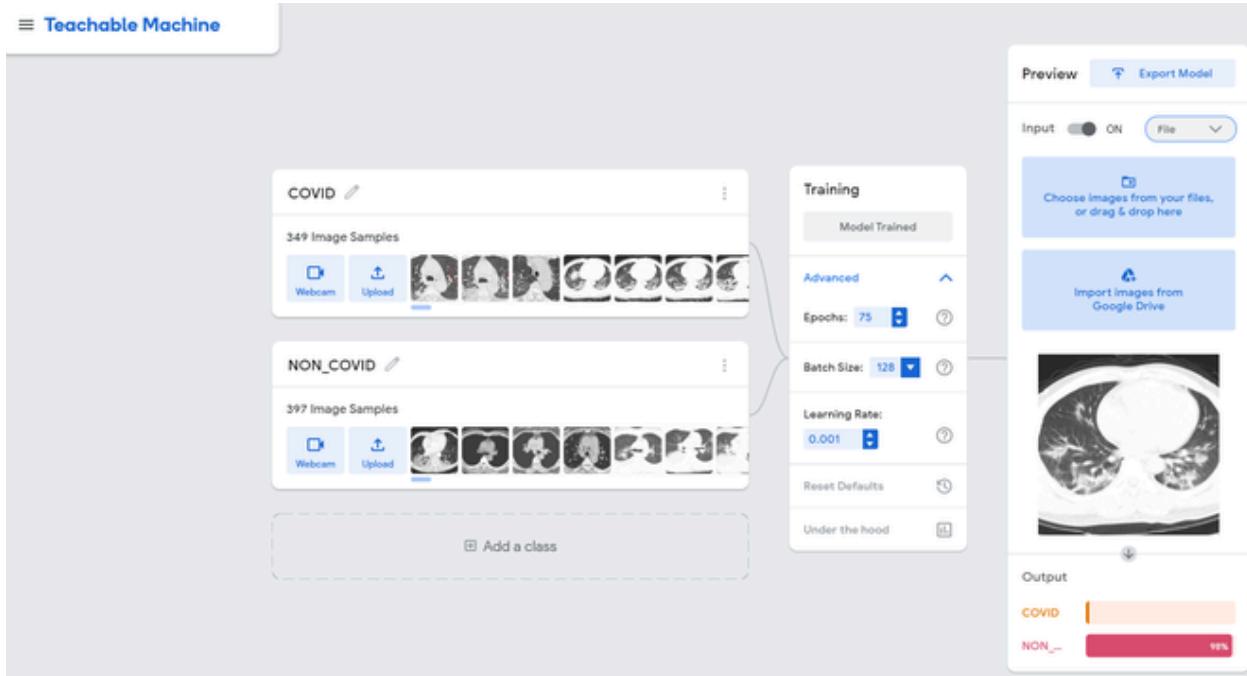
- After clicking on Image Project, the below web page will be displayed.



- Add a number of classes, rename them, and upload sample images for each class. The dataset we are going to use is COVID 19-Lung CT Scans.
- Then click on Advanced and adjust *Epochs*, *Batch Size*, and *Learning Rate*.



- Now click on Train Model, it will require some time to process. After the model is trained, click on under the hood to get accuracy and other details.



- Click on *Export Model* to download the model or generate a shareable public link for the model.

```

import tensorflow.keras
from PIL import Image, ImageOps
import numpy as np

# Disable scientific notation for clarity
np.set_printoptions(suppress=True)

# Load the model
model = tensorflow.keras.models.load_model('keras_model.h5')

# Create the array of the right shape to feed into the keras model
# The 'length' or number of images you can put into the array is
# determined by the first position in the shape tuple, in this case 1.

```

In this way, we can easily develop machine learning models using Google's Teachable Machine.

## YOLO :

You Only Look Once (YOLO) is a state-of-the-art, real-time object detection algorithm introduced in 2015 by **Joseph Redmon**, **Santosh Divvala**, **Ross Girshick**, and **Ali Farhadi** in their famous research paper “**You Only Look Once: Unified, Real-Time Object Detection**”.

The authors frame the object detection problem as a regression problem instead of a classification task by spatially separating bounding boxes and associating probabilities to each of the detected images using a single convolutional neural network (CNN).

By taking the **Image Processing with Keras in Python** course, you will be able to build Keras based deep neural networks for image classification tasks.

If you are more interested in Pytorch, **Deep Learning with Pytorch** will teach you about convolutional neural networks and how to use them to build much more powerful models



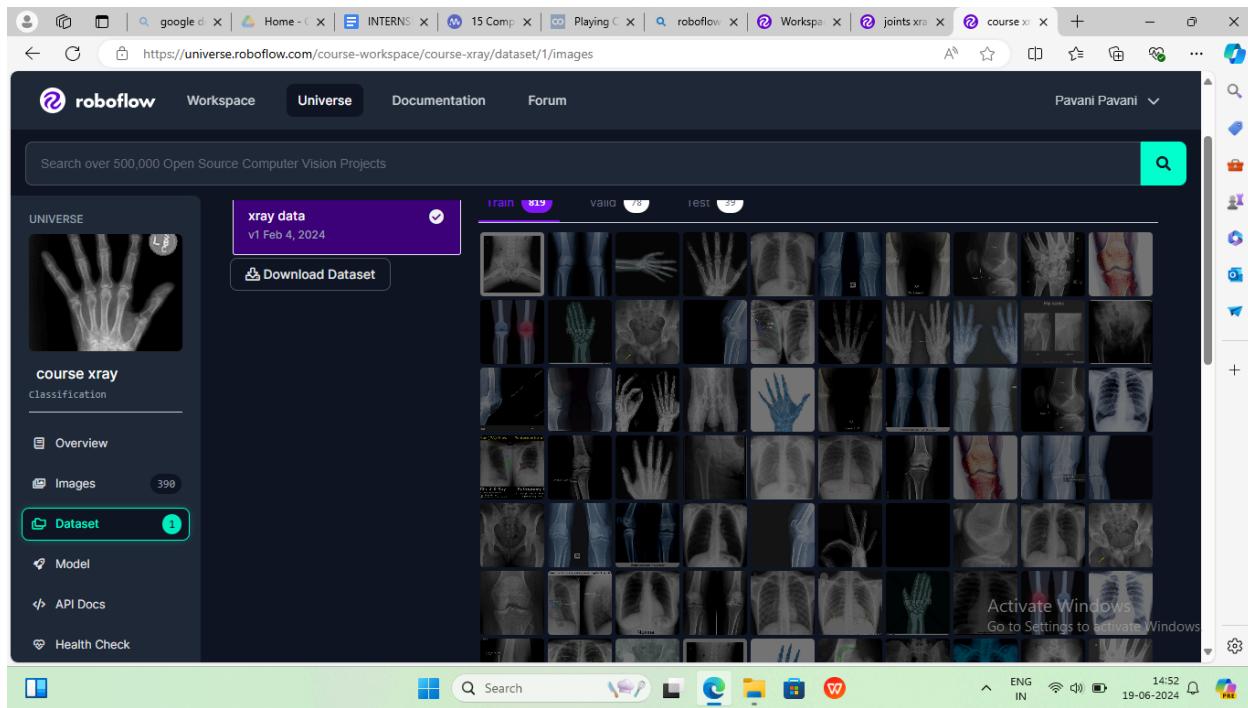
Here i used yolov8 for detecting objects of a class

## MEDICAL IMAGE ANALYSIS AND LABELLING :

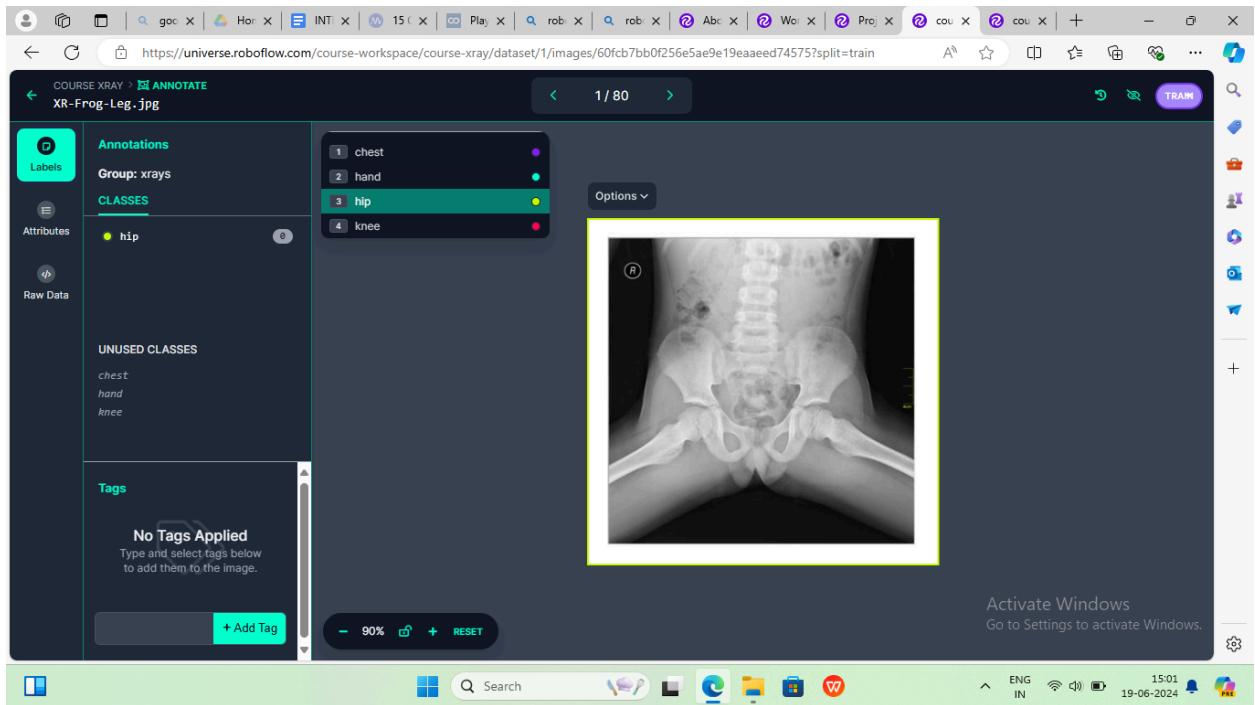
Roboflow empowers developers to build their own computer vision applications, no matter their skillset or experience. They streamline the process between labeling your data and training your model.

After building **their own applications**, they learned how difficult it can be to **train and deploy a computer vision model**. writing excess code to **format our data**. It was difficult to **collaborate**, and **benchmarking machine learning tools** was a lot of work.

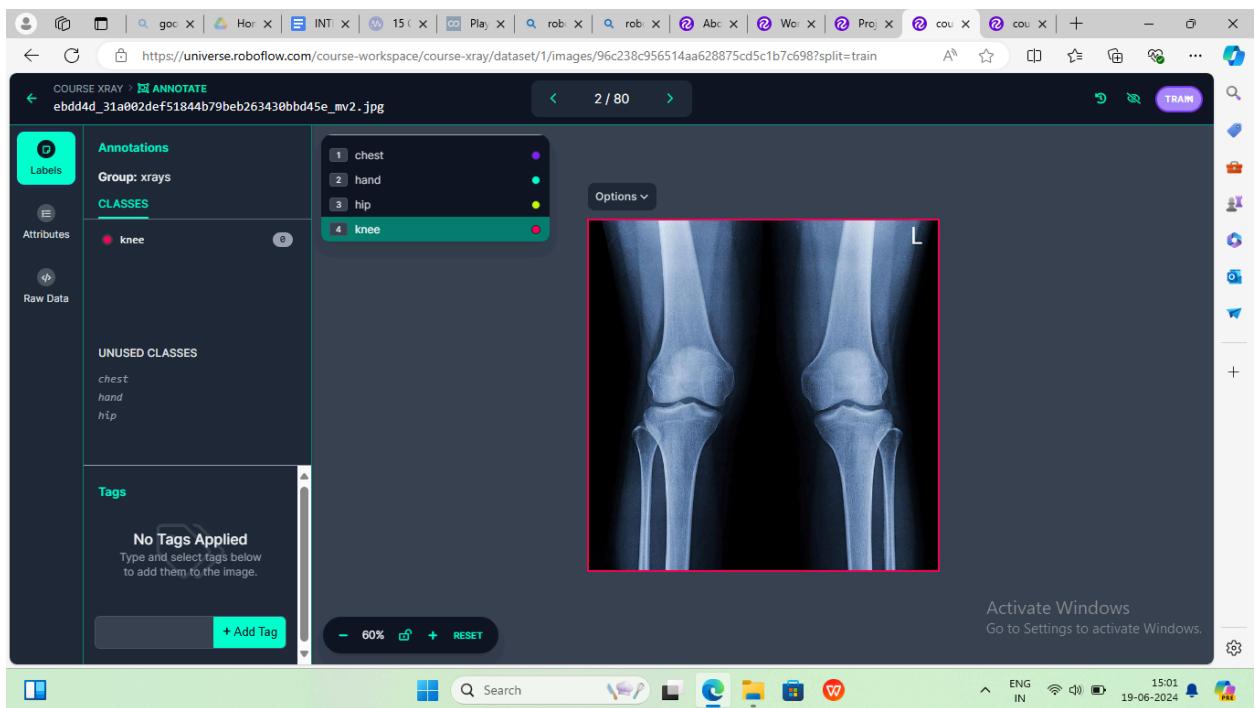
That's why they launched Roboflow in January 2020. believe every developer should have computer vision available in their toolkit. Roboflow mission is to remove any barriers that might prevent them from succeeding.



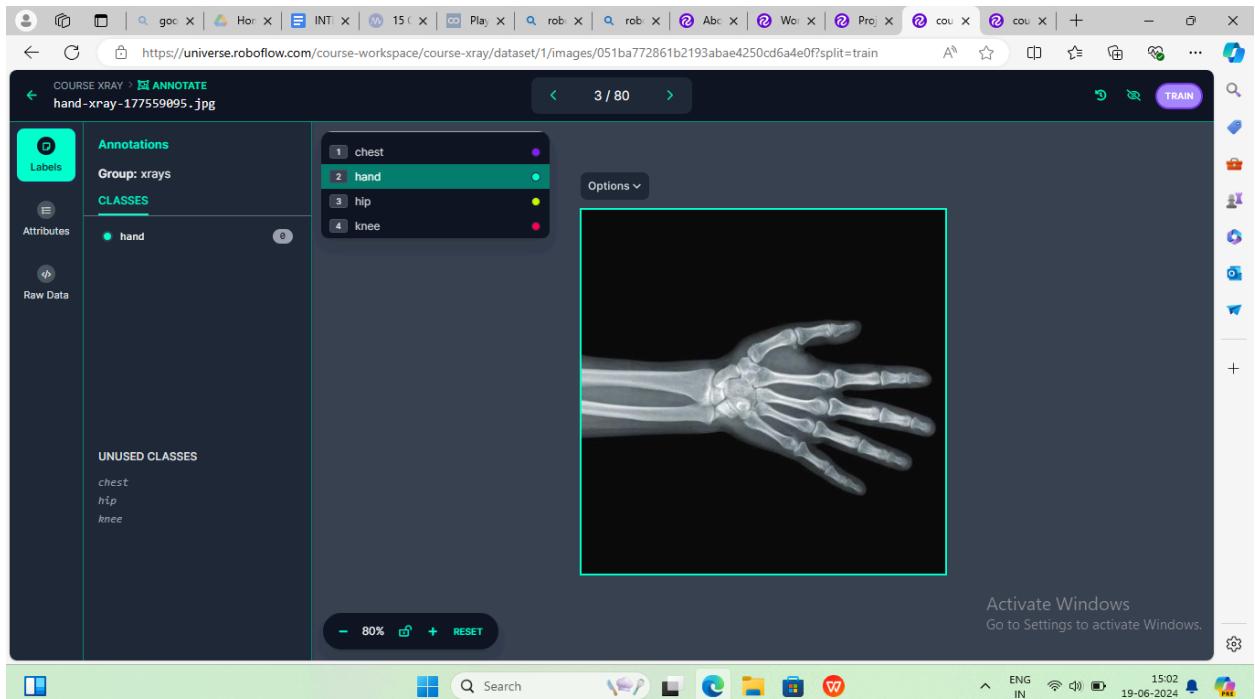
Here we took x-ray images of the hands legs chest hip of human next step is to **Annotations** Of images



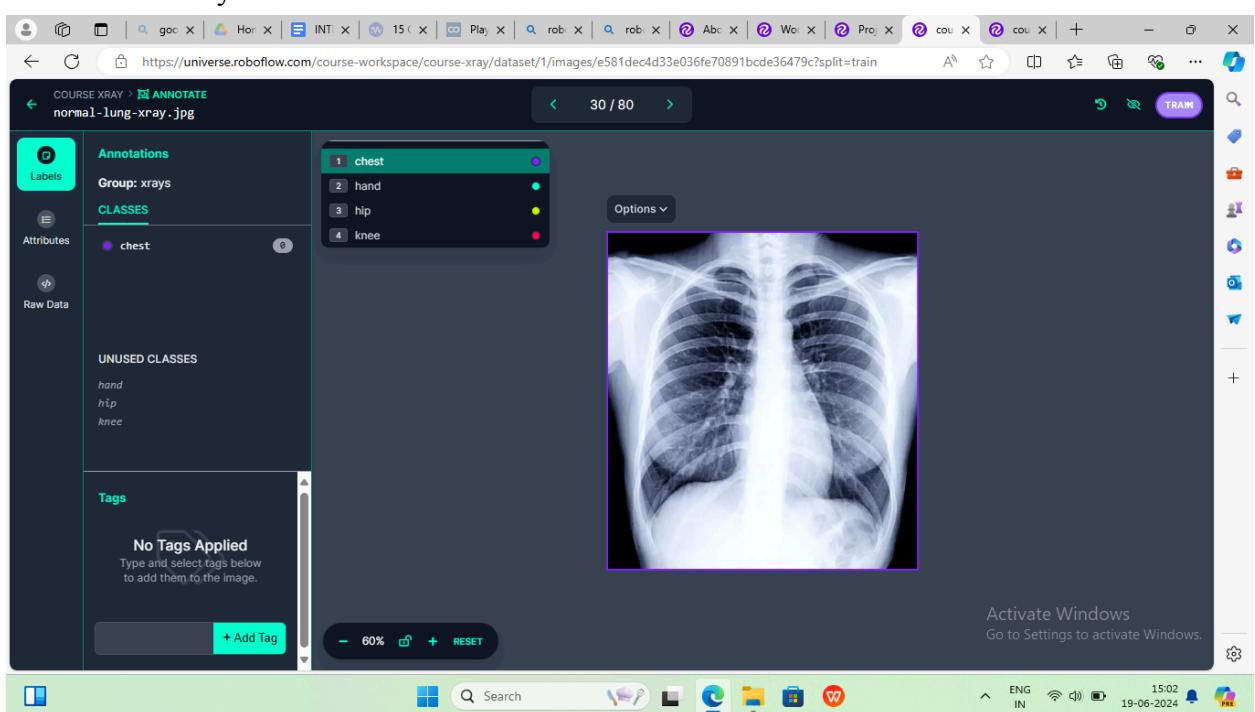
This is hip we labeled it as hip



Here it is x-ray of legs so we labeled it with leg



Here it is an x-ray of hand so we labeled it as hand

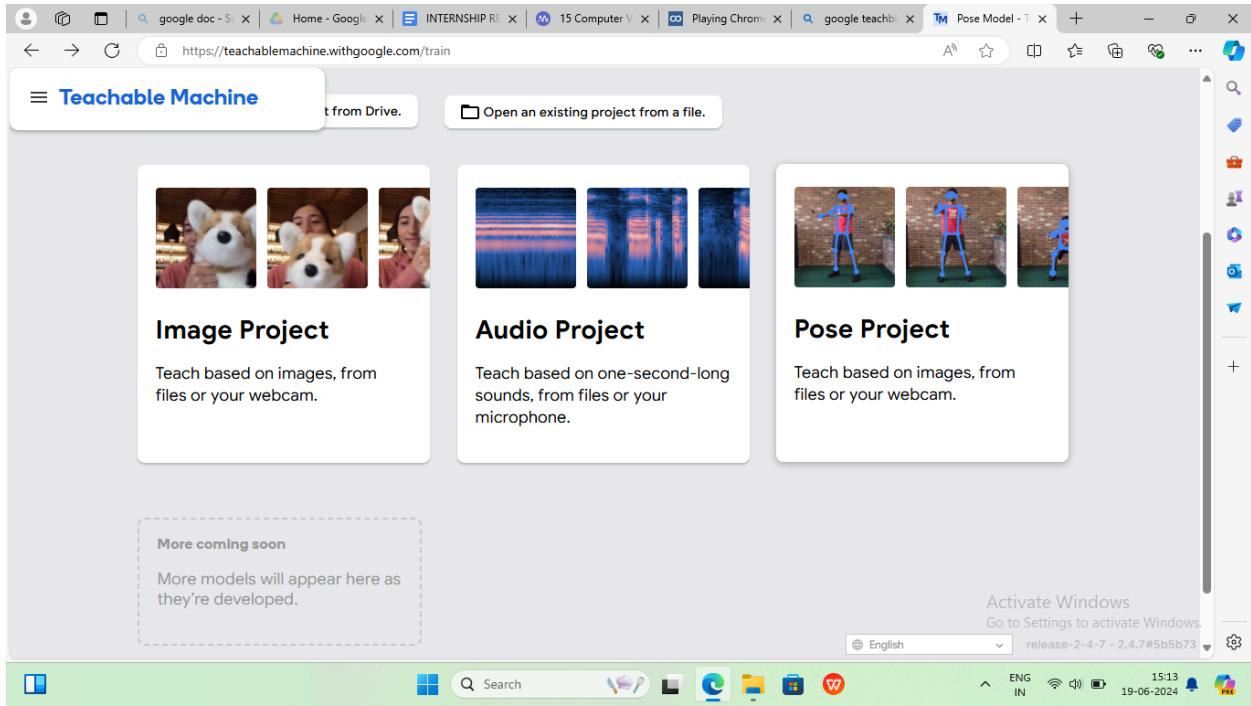


Here it is x-ray of chest so labeled it as chest

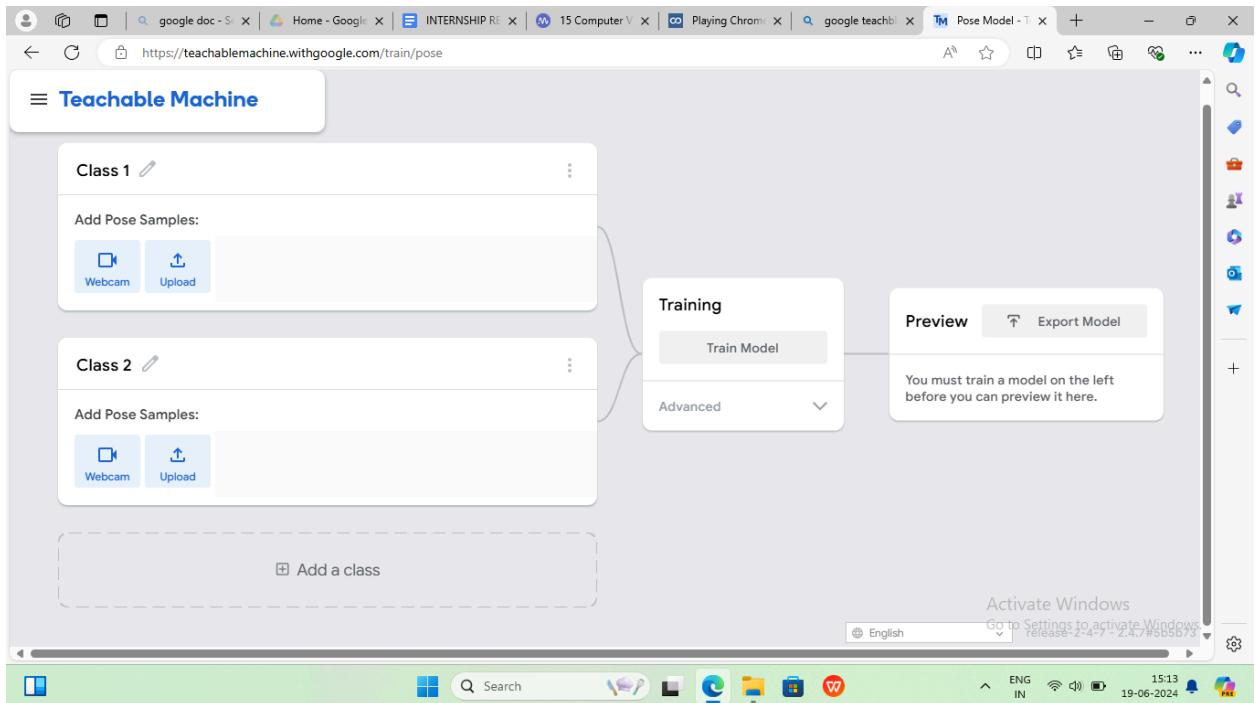
So this is we label the images in roboflow

HUMAN POSE ESTIMATION :

We can use google teachable machine for human pose estimation actually we use this for voice estimation human pose estimation object pose estimation as well



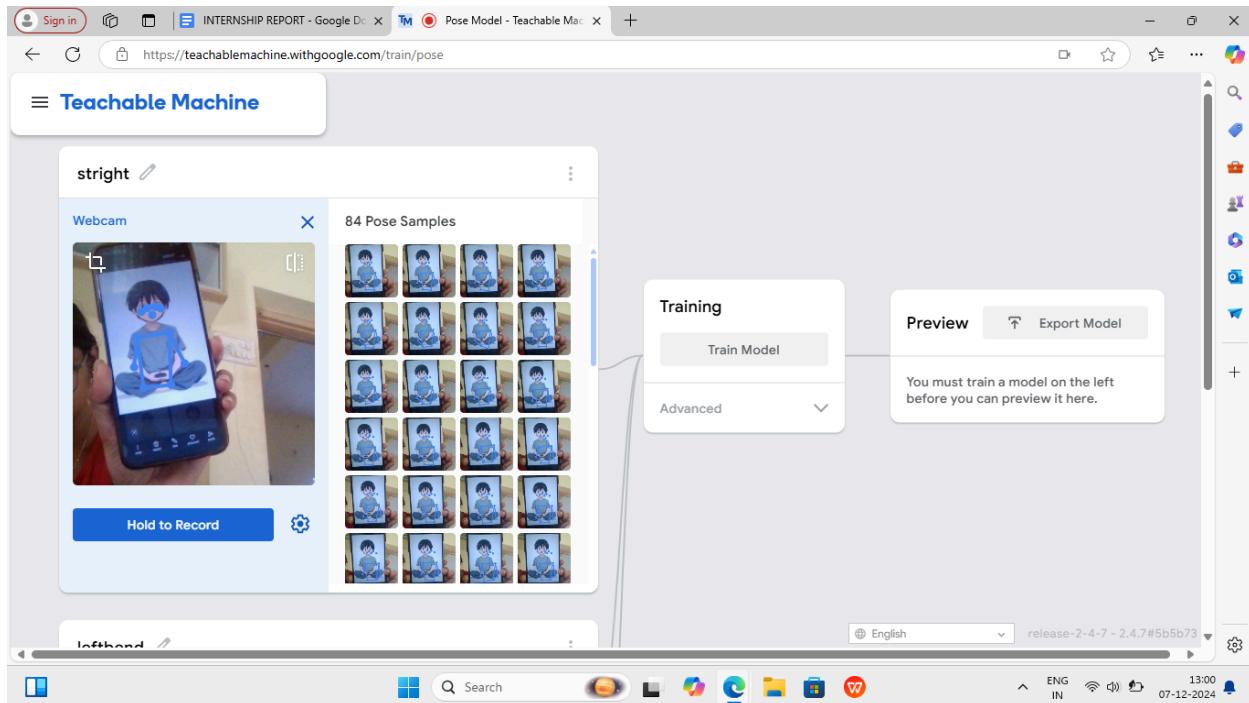
Here let us choose pose project for human pose estimation then we get train the model by labeling the poses we can just insert images or use webcam to show poses



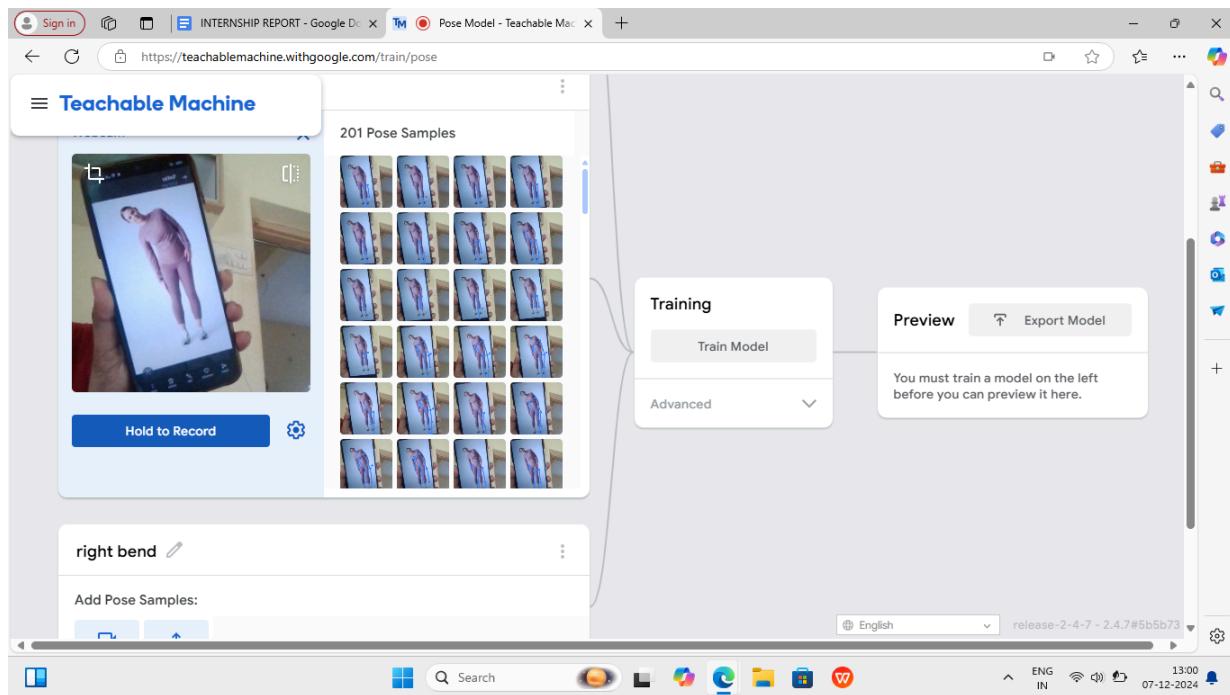
Here we name classes with pose name you add as many classes as you can then you train the model preview the model or test the model

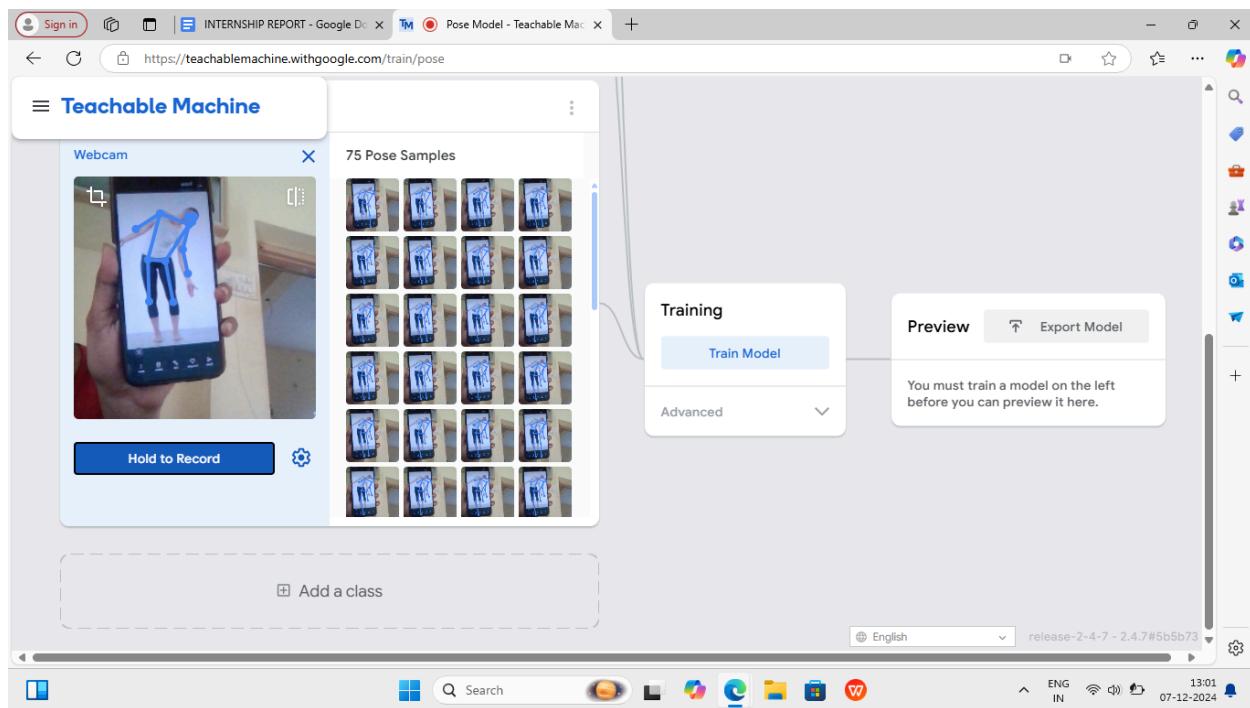
Sit straight, and 'Hold to Record' for about 5 seconds — around 30–50 pose samples

would be populated by Teachable Machine.



Do the same for the remaining two classes, tilting your head to the left for one class and tilting to the right for another.

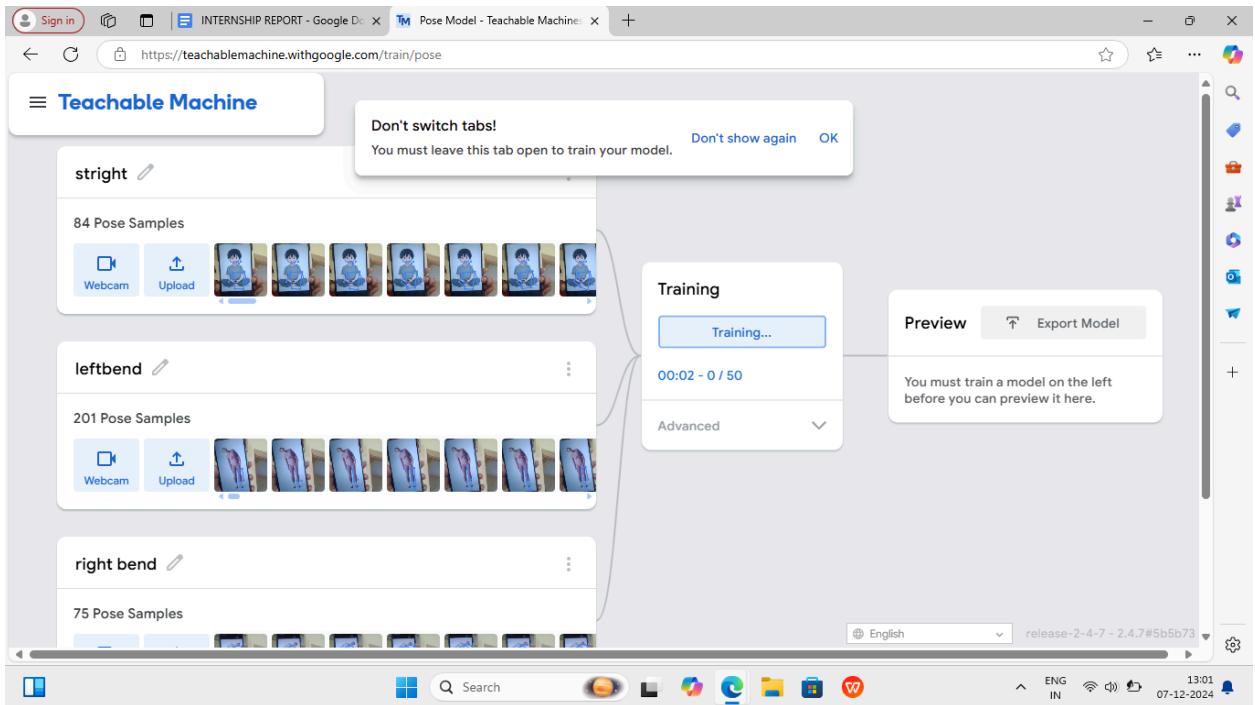




## Step 4

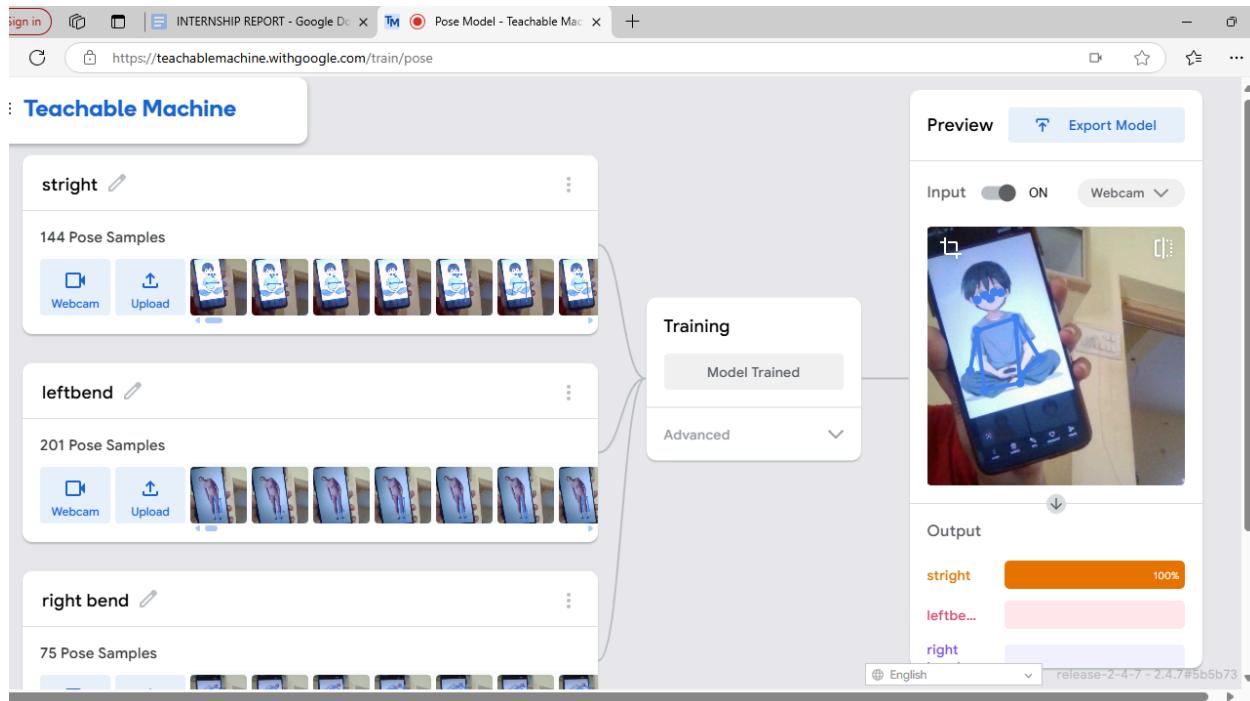
Click on ‘Train Model’. We’ll not bother with any of the advanced settings for this project. It might take a couple of minutes to train the model.

You should be able to see the preview on the right when the model training is done. You can check by tilting your head if the model is able to classify your poses correctly now.

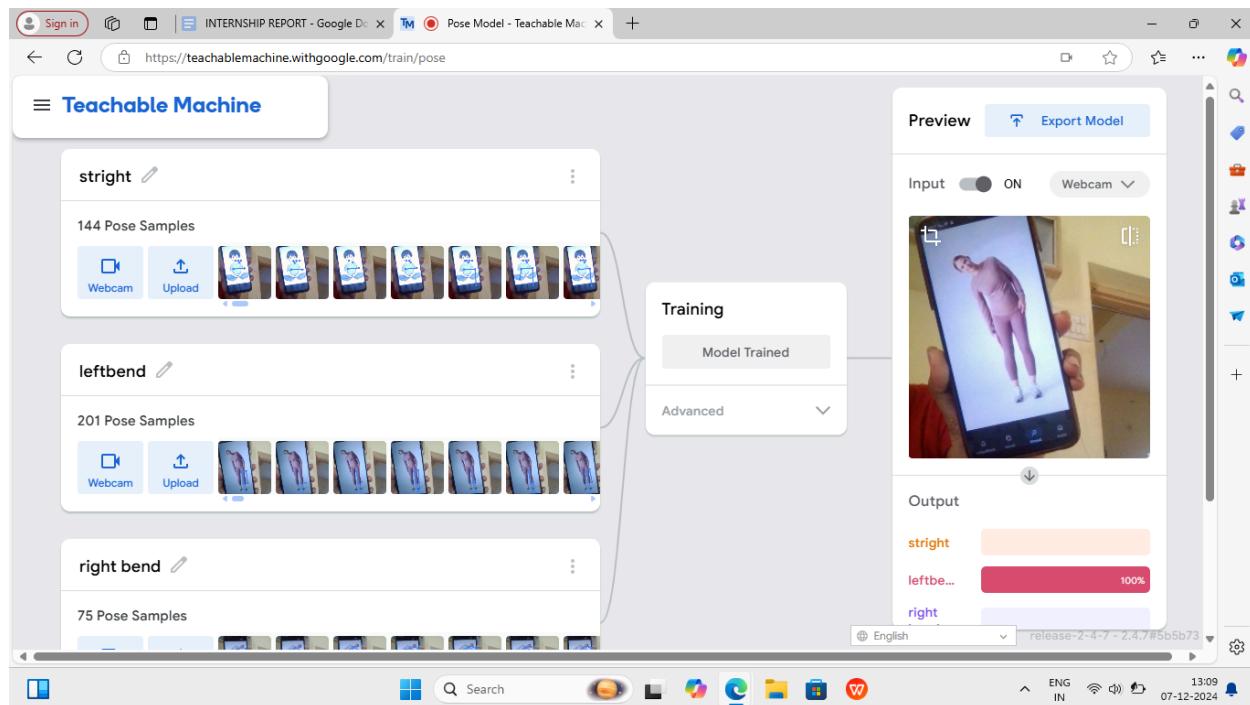


## Step 5

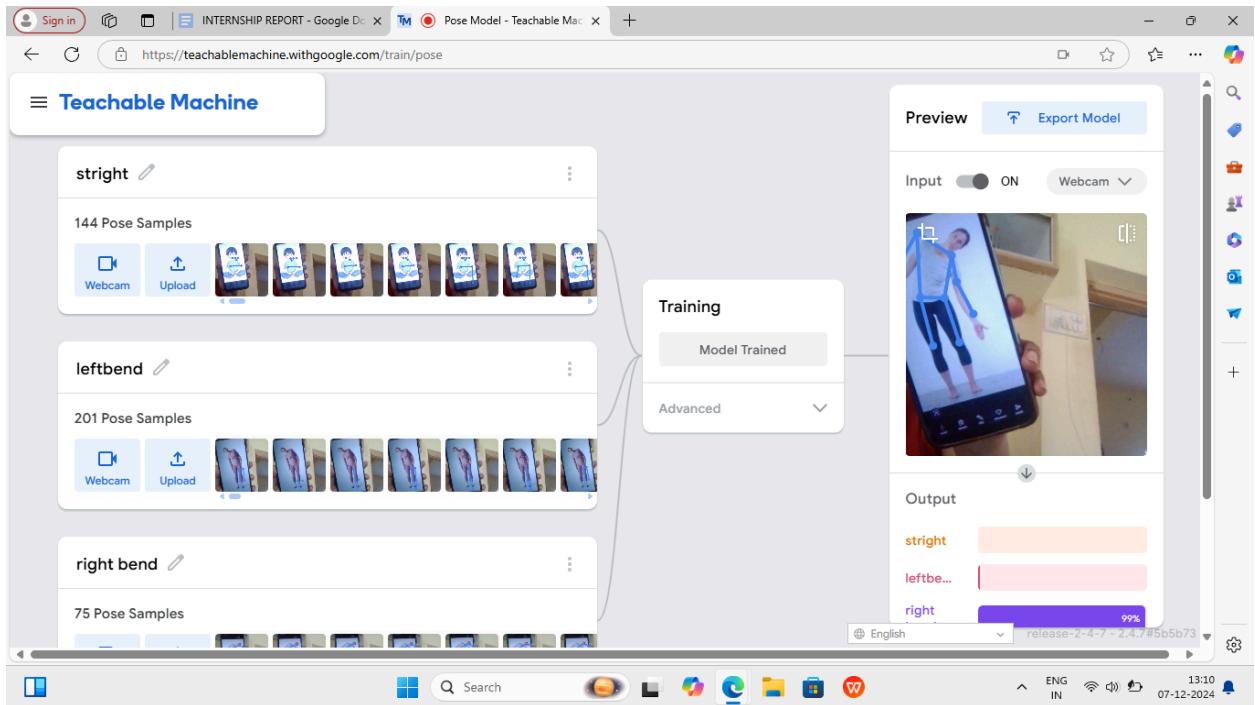
Click on the preview model now try the poses your going get poses name  
First let us try the straight pose



So here we got straight with 100 %  
Let us try other two poses as well



So see we got right bend as right bend pose with 100%

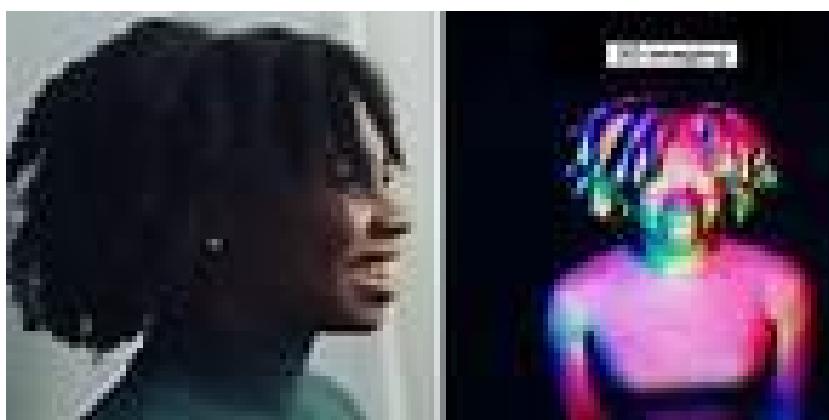


So we got left bend also right

### AI MODELS :

A.I has taken over the world. Let us look at the top A.I models currently being developed.

### PROJECT STARLINE :



This is a google project and this project is called starline and only a handful of people have experienced it. project starline is an experimental video communication method that allows the user to see a 3D model of the

person they are communicating with. google announced the project at it's 2021 I/O developer conference, saying that it will allow the users to talk naturally, gesture and make eye contact. The system uses AI to create a depth map of the user and their room creating an incredibly realistic 3D effect also the display the person on the call into a lifelike 3D model with head tracking and a parallax effect making it feel like your looking into their space.

### **DREAM MACHINE :**

You can now turn any images into a smooth realistic video so this luma ai new text to video model called dream machine and its available to the public now we can go try it out for ourself no beta invite is needed and now lets take some of the top memes and animate them



### **Voice engines :**

Open ai introduced voice engines which can clone someones voice with just 15 seconds of audio its an absolutely wild now they actually developed this way back in late 2022 but instead of releasing it to the public they have been quietly testing it with a small group of trusted partners and businesses. Why? Because they know that this kind of powerful ai voice cloning from very very small samples could be seriously misused especially with the cyber crimes noam brown an employee tweeted right after the announcement saying “if you havent disabled voice authentication for your ban account and had a conversation with your family about ai voice impersonation yet now would be a good time



## DEVIKA :

There is now an open source alternative to devin called devika that can understand high level human instructions , break them down into steps , research relevant information, and write code to achieve the given objective. Devika utilizes large language models, planning and reasoning algorithms, and web browsing abilities to intelligently develop software now I'm yet to try it out myself but you can simply set up devika right now by just below following instructions

**Step 1:** Clone the Devika GitHub repository by executing the following command:

```
bash
```

```
git clone https://github.com/stitionai/devika.git
```

**Step 2:** Install all required dependencies by entering the Devika directory and running the pip command:

```
bash
```

```
cd devika pip install -r requirements.txt
```

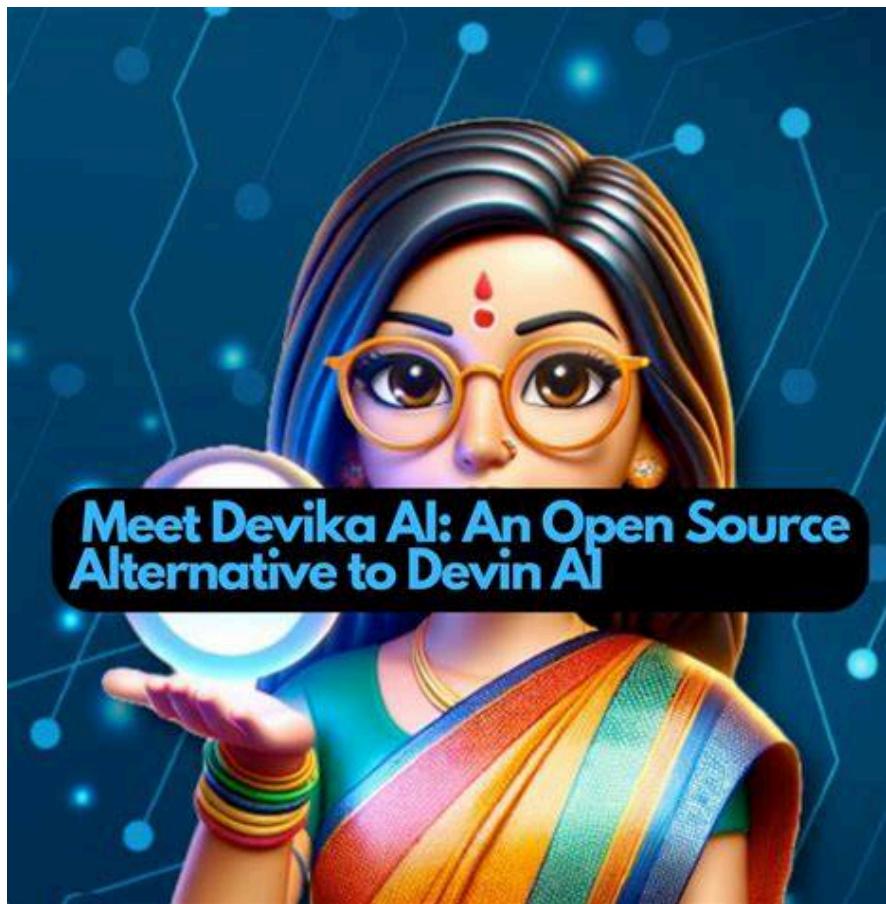
**Step 3:** Configure the necessary API keys and settings within the config.toml file according to your needs.

**Step 4:** Start the Devika server with this command:

```
python devika.py
```

**Step 5:** Access Devika's web interface by opening <http://localhost:3000> in your web browser.

According to github repo, devika was able to complete the task like building a full functional game in python and under the hood devika utilizes llms planning and reasoning algorithms and web browsing abilities to intelligently develop a software now the goal of the project is to be an open-source alternative to devin with an overly ambitious aim to meet the same score as devin in the SWE-bench benchmark and eventually beat it now



GENERATIVE AI :

AI has been a hot technology topic for the past decade, but generative AI, and specifically the arrival of ChatGPT in 2022, has thrust AI into worldwide headlines and launched an unprecedented surge of AI innovation and adoption. Generative AI offers enormous productivity benefits for individuals and organizations, and while it also presents very real challenges and risks, businesses are forging ahead, exploring how the technology can improve their internal workflows and enrich their products and services. According to research by the management consulting firm McKinsey, one third of organizations are already using generative AI regularly in at least one business function.<sup>1</sup> Industry analyst Gartner projects more than 80% of organizations will have deployed generative AI applications or used generative AI application programming interfaces (APIs) by 2026

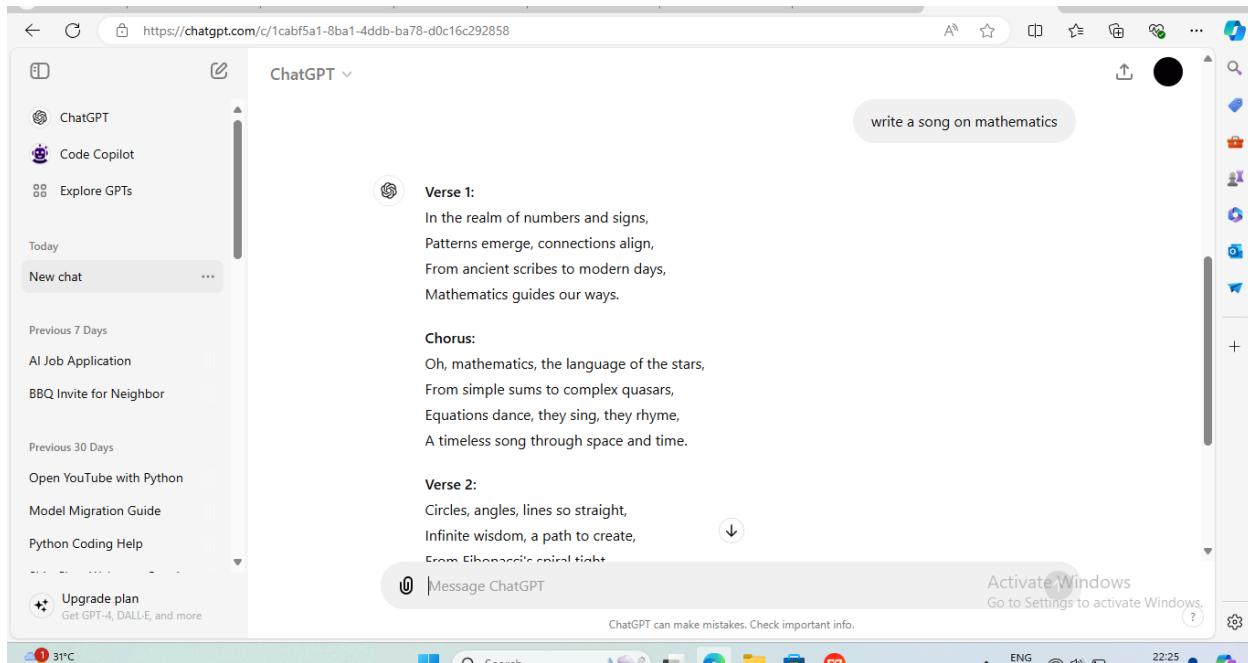
## ChatGPT

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms"	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?"	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?"	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

»

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them even more useful.

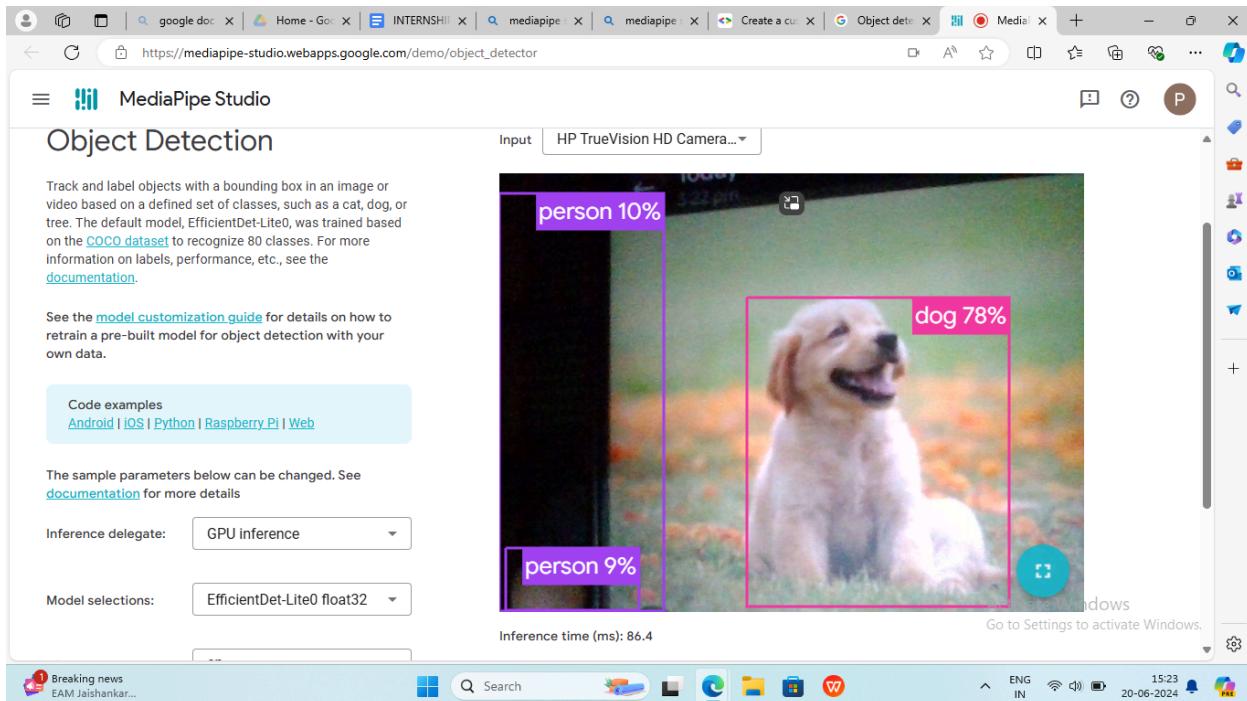
Let us try chatgpt for creating a song on mathematics



So here it has given song on mathematics within seconds i would say in coming days these are likely to steal our jobs

## MEDIPIPE STUDIO :

MediaPipe Solutions provides a suite of libraries and tools for you to quickly apply artificial intelligence (AI) and machine learning (ML) techniques in your applications. You can plug these solutions into your applications immediately, customize them to your needs, and use them across multiple development platforms. MediaPipe Solutions is part of the MediaPipe [open source project](#), so you can further customize the solutions code to meet your application needs.



Here it has detected dog and also person behind the photo of dog well it is incredible as it has been detecting the person behind the photo as well and it is an open source

## OPENCV BASICS :

OpenCV, or the Open Source Computer Vision Library, is a open-source library widely used in computer vision projects. OpenCV is designed to offer for real-time computer vision and serves as a platform that facilitates many image and video analysis applications.

The OpenCV library plays an important role in enabling developers and researchers to use the advantages of visual data processing. Its capabilities of various applications, from basic image processing to more complex tasks such as object identification and **facial recognition**. By offering a many collection of algorithms, methods and image data processing operations, OpenCV facilitates the development of intelligent systems capable of recognizing and classifying visual content.

One notable application of OpenCV is in object recognition, where its algorithms facilitate identifying and localizing objects within images or videos. This capability also can extend to other computer vision tasks, such as facial recognition, movement tracking, and support for augmented reality technologies.

## Installation steps

- Python installation:** Make sure to have Python installed on the system. Python can download the latest version from the official **Python website**.
- Environment and Package Manager (optional):** You may use package managers like Anaconda to ease library handling. Download and install Anaconda from their **official website**.
- OpenCV Installation:** Open a command prompt or terminal window and use the following command to install OpenCV using pip (Python's package installer):

```
!pip install opencv-python
```

## GOOGLE DIALOGFLOW :

### Generative AI agent

Spin up agents with just a few clicks using Vertex AI Agent Builder and Dialogflow. Connect your webpage or documents to your Dialogflow CX agent and leverage foundation models for generating responses from the content, out of the box. You can also call a foundation model to perform specific tasks during a virtual agent conversation or respond to a query contextually, significantly reducing development effort and making virtual agents more conversational

The screenshot shows the Google Dialogflow interface for creating a welcome intent. On the left, there's a sidebar with navigation links: Dialogflow Essentials, Global, PizzaBot, en, Intents (selected), Entities, Knowledge [beta], Fulfillment, Integrations, Training, Validation, History, Analytics, Prebuilt Agents, and Small Talk. The main area is titled 'Default Welcome Intent'. It has sections for 'Contexts', 'Events' (with 'Welcome' selected), and 'Training phrases'. A warning message states: 'Template phrases are deprecated and will be ignored in training time. More details [here](#)'. Below this, it says: 'When a user says something similar to a training phrase, Dialogflow matches it to the intent. You don't have to create an exhaustive list. Dialogflow will fill out the list with similar expressions. To extract parameter values, use annotations with available system or custom entity types.' Under 'Training phrases', there are three examples: 'just going to say hi' and 'heya'. On the right, there's a 'Try it now' button, an 'Agent' section with 'USER SAYS' (hi) and 'DEFAULT RESPONSE' (Hi, My name is Pizza Bot, I am here to help you with your pizza order.), and a 'DIAGNOSTIC INFO' section with 'Activate Windows' and 'Go to Settings to activate Windows.'

So here i trained it to respond as a pizza bot it can only helps with pizza orders

## **VISUAL QUESTIONS AND ANSWERING :**

Visual Question Answering (VQA) is the task of answering open-ended questions based on an image. The input to models supporting this task is typically a combination of an image and a question, and the output is an answer expressed in natural language.

Some noteworthy use case examples for VQA include:

- Accessibility applications for visually impaired individuals.
- Education: posing questions about visual materials presented in lectures or textbooks. VQA can also be utilized in interactive museum exhibits or historical sites.
- Customer service and e-commerce: VQA can enhance user experience by letting users ask questions about products.
- Image retrieval: VQA models can be used to retrieve images with specific characteristics. For example, the user can ask “Is there a dog?” to find all images with dogs from a set of images.

**Visual Question Answering is the task of answering open-ended questions based on an image. They output natural language responses to natural language questions.**

### **Inputs**



**Question**

What is in this image?

**Visual Question Answering Model**

**Output**

elephant

0.970

elephants

0.060

animal

0.003

**About Visual Question Answering**

## Use Cases

### Aid the Visually Impaired Persons

VQA models can be used to reduce visual barriers for visually impaired individuals by allowing them to get information about images from the web and the real world.

### Education

VQA models can be used to improve experiences at museums by allowing observers to directly ask questions they interested in.

### Improved Image Retrieval

Visual question answering models can be used to retrieve images with specific characteristics. For example, the user can ask "Is there a dog?" to find all images with dogs from a set of images.

### Video Search

Specific snippets/timestamps of a video can be retrieved based on search queries. For example, the user can ask "At which part of the video does the guitar appear?" and get a specific timestamp range from the whole video.

### Task Variants

#### Video Question Answering

Video Question Answering aims to answer questions asked about the content of a video.

### Inference

You can infer with Visual Question Answering models using the vqa (or visual-question-answering) pipeline. This pipeline requires [the Python Image Library \(PIL\)](#) to process images. You can install it with (`pip install pillow`).

```
from PIL import Image

from transformers import pipeline

vqa_pipeline = pipeline("visual-question-answering")

image = Image.open("elephant.jpeg")

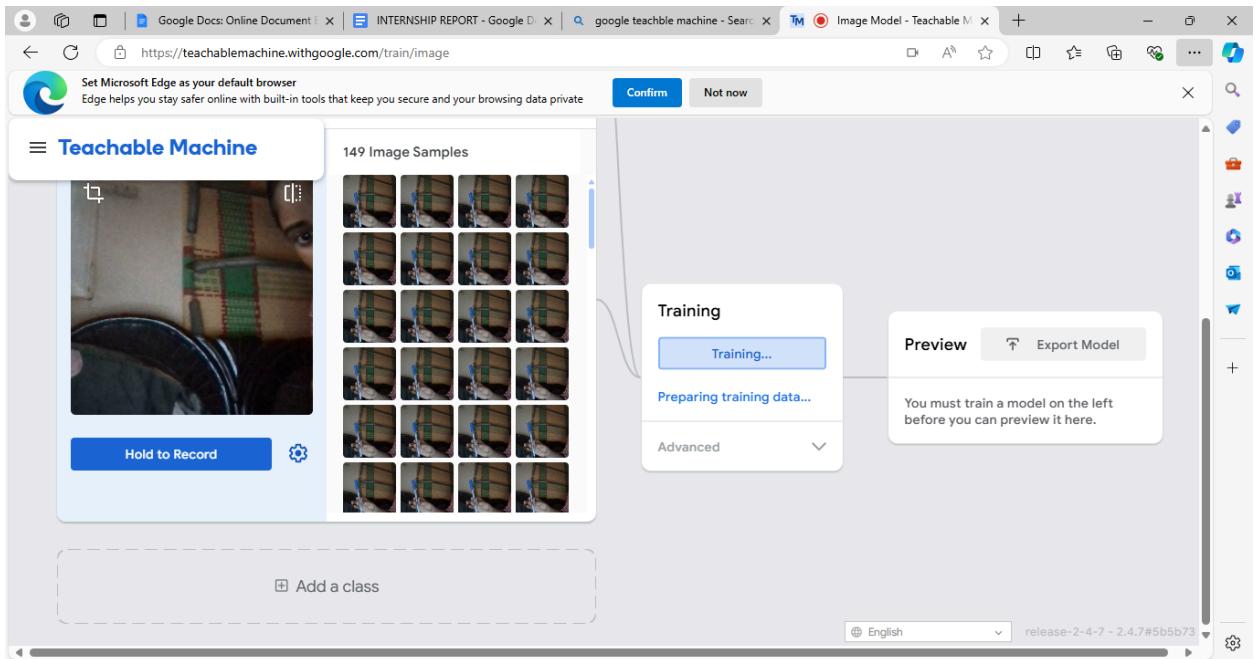
question = "Is there an elephant?"

vqa_pipeline(image, question, top_k=1)

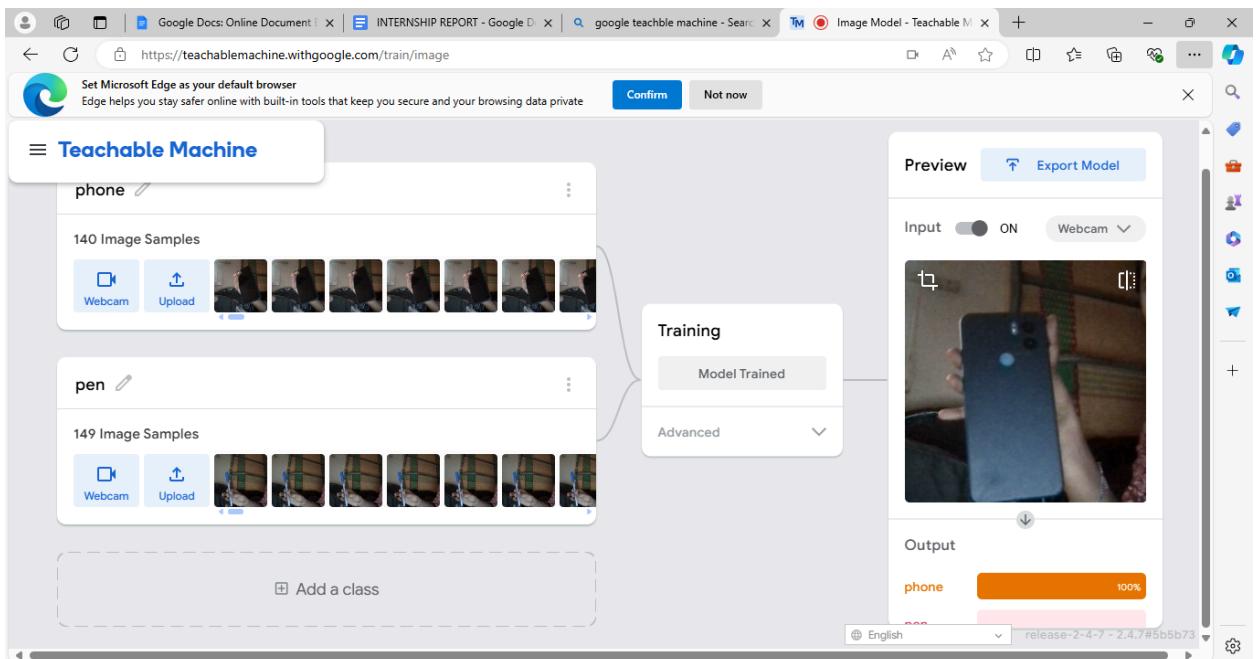
#[{'score': 0.9998154044151306, 'answer': 'yes'}]
```

## Useful Resources

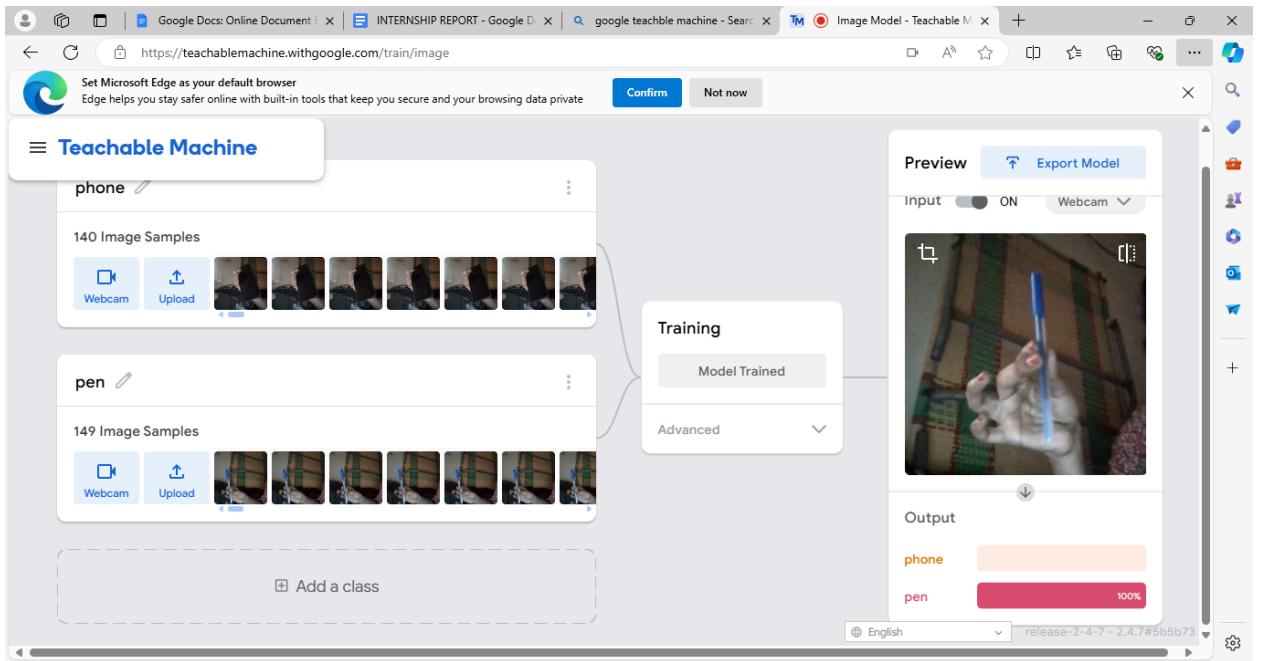
- [An introduction to Visual Question Answering - AllenAI](#)
- [Multi Modal Framework \(MMF\) - Meta Research](#)



Here i trained the model for detecting the phone and lets us check weather it detects phone and pen or not



Here i showed phone and it detected as phone



Here i showed pen and it detected it as pen this is visual question and answering

**OpenCV Tutorial:** utilizing the OpenCV library for image and video processing within a Python environment. We dive into the wide range of image processing functionalities OpenCV offers, from basic techniques to more advanced application

OpenCV, or the Open Source Computer Vision Library, is a robust open-source library widely adopted in computer vision projects. OpenCV is designed to offer an integrated framework for real-time computer vision and serves as a platform that facilitates numerous image and video analysis applications.

The OpenCV library plays a pivotal role in enabling developers and researchers to harness the advantages of visual data processing. Its capabilities span various applications, from basic image processing to more complex tasks such as object identification and **facial recognition**. By offering an extensive collection of algorithms, methods and image data processing operations, OpenCV

facilitates the development of intelligent systems capable of recognizing and classifying visual content.

One notable application of OpenCV is in object recognition, where its algorithms facilitate identifying and localizing objects within images or video streams. This capability also extends to other computer vision tasks, such as facial recognition, movement tracking, and support for augmented reality technologies.

In this article, we explore the following:

- *Image processing functionalities enabled by the OpenCV library*
- *How to install the OpenCV library within a Python environment*
- *An overview of image processing techniques and how to implement them within OpenCV*
- *Video Processing techniques with OpenCV*

## OpenCV Installation

### Prerequisites

To go through this OpenCV tutorial, you'll need an intermediate knowledge of Python and a basic understanding of computer vision. Familiarity with image processing principles will also improve the learning experience throughout the tutorial, although key concepts are defined when mentioned.

### Installation steps

1. **Python installation:** Make sure to have Python installed on the system. Python can download the latest version from the official [Python website](#).

2. **Environment and Package Manager (optional):** You may use package managers like Anaconda to ease library handling. Download and install Anaconda from their [official website](#).
3. **OpenCV Installation:** Open a command prompt or terminal window and use the following command to install OpenCV using pip (Python's package installer):

```
!pip install opencv-python
```

OpenAI

You may use the following command in case you are using Anaconda:

```
!conda install -c conda-forge opencv
```

OpenAI

Verification

To confirm if the installation is successful, run the following code on a Python Interpreter or Jupyter NoteBook:

```
import cv2
```

```
print(cv2.__version__)
```

OpenAI

**On successful installation, this should print the version number of the installed OpenCV library. Failure of the version number not printed on the terminal or console implies that the installation process wasn't successful.**

**Important Note if using Google Colab:** For adapting the code to display images within Google Colab notebooks, it's necessary to modify the traditional `cv2.imshow` function to `cv2_imshow` provided by Google Colab's specific libraries. This adjustment ensures compatibility with the Colab environment, which does not support `cv2.imshow`. To implement this change, import the `cv2_imshow` function from Google Colab's patches as follows:

```
from google.colab.patches import cv2_imshow
```

```
# Displaying an image in a window
```

```
cv2_imshow(image)
```

OpenAI

Introduction to OpenCV features

## Loading and Displaying an Image in OpenCV

The code snippet below introduces a fundamental entry point to utilizing the OpenCV library for image processing. Loading an image into a development environment is conducted in various areas, from simple computer vision projects to complex real-time object detection applications in

**production. Displaying an image is also crucial for observing the content of an image or processing result. In the code below, we read a JPG from a file and display it using the OpenCV library.**

The following code snippet could be used to load and display images in OpenCV:

```
import cv2

# Loading image from a file

image = cv2.imread('your/image/path.jpg')

# Displaying an image in a window

cv2.imshow('Image', image)

cv2.waitKey(0)

cv2.destroyAllWindows()
```

OpenAI



*Office workers surrounding a laptop. Source: Canva*

In this code, the image is read from a file using ‘cv2.imread’ and then displayed in a window with the help of ‘imshow’. The ‘cv2.waitKey(0)’ function ensures the window remains open until the user closes it by pressing a key.

Image processing is manipulating pixel data in an image in such a manner that it modifies the visual appearance of the initial image, either for feature extraction, image analysis, or other computer vision-related tasks. Examples of image processing techniques are color processing, filtering, segmentation, edge detection, etc.

The code snippet below demonstrates a simple image processing technique on the image loaded in the previous code snippet. The image processing technique shown here is a simple color conversion, specifically, a grayscale conversion. Grayscale conversion is a fundamental technique in image processing, reducing the image to a single channel of intensity values.

```
# Converting an image to grayscale form  
  
gray_image = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)  
  
# Displaying the grayscale image  
  
cv2.imshow('Gray Image', gray_image)  
  
cv2.waitKey(0)  
  
cv2.destroyAllWindows()
```

OpenAI

Here, `cv2.cvtColor()` converts the loaded image into a grayscale.



*Office workers surrounding a laptop(Grayscale). Source: Canva*

## Edge Detection

Edge detection is a fundamental technique in image processing and computer vision. It identifies points in an image where brightness changes sharply or has discontinuities. Detecting edges is essential in computer vision, especially when detecting features within image data. Edge detection generally functions by detecting sharp changes in pixel intensities. By joining these points of sharp changes in image brightness, lines and edges that outline objects in an image are formed.

With edge detection algorithms, minuscule textural and structural features of an object on an image can be detected.

Several edge detection algorithms, such as [Canny Edge Detector](#), [Sobel Method](#), and [Fuzzy Logic method](#), are utilized in object recognition, image segmentation, and feature extraction. The following code snippet demonstrates basic edge detection using OpenCV:

```
# Applying Canny edge detection (a simple edge detection method)
```

```
edges = cv2.Canny(gray_image, 50, 150)
```

```
# Displaying the edge-detected image
```

```
cv2.imshow('Edges', edges)
```

```
cv2.waitKey(0)
```

```
cv2.destroyAllWindows()
```

OpenAI

The Canny edge detection algorithm is made readily available in the OpenCV library by the underlying function 'cv2.Canny()' that applies to detect the edges of an object for the grayscale image.



*Office workers surrounding a laptop(Grayscale). Source: Canva*

### Resizing and Rotating Images

Resizing and rotating images are fundamental requirements for image processing since they facilitate the adaptation and transformation of visual content. Resizing an image involves changing its dimensions, whether enlarging or reducing, and this function is essential for scaling and placing images in specific environments.

Rotating is an adjustment that alters the angle at which a specified rotation angle views an image. Both operations are commonly applied as image pre-processing steps for various purposes, including computer vision, machine learning, and graphics. For example, in training deep neural networks, rotating and resizing images are data augmentation techniques used to increase the dataset variability of the training data fed into neural networks during training. Data augmentation aims to train the network with a dataset with enough variance to enable the trained network to generalize unseen data well.

To resize and rotate images using OpenCV, you can use the following code:

```
# Resizing the image
```

```
resized_image = cv2.resize(image, (width, height))
```

```
# Rotating the image by 45 degrees
```

```
rotation_matrix = cv2.getRotationMatrix2D((width / 2, height / 2), 45, 1)
```

```
rotated_image = cv2.warpAffine(resized_image, rotation_matrix, (width, height))
```

```
# Displaying the resized and rotated images
```

```
cv2.imshow('Resized Image', resized_image)
```

```
cv2.imshow('Rotated Image', rotated_image)
```

```
cv2.waitKey(0)
```

```
cv2.destroyAllWindows()
```

OpenAI

While `'cv2.resize'` is used to resize the image, and `'cv2.getRotationMatrix2D'` along with `'cv2.warpAffine'` are both used for rotating the image.



*Office workers surrounding a laptop(Rotated). Source: Canva*

### Loading and Processing Videos

Videos are a series of images presented sequentially. In computer vision, video processing capabilities are crucial to developing and building applications that process real-time and recorded data. In contrast to the static nature of images, video processing deals with temporal sequences of images, allowing for the analysis of motion, changes over time, and the extraction of temporal patterns.

The dynamic nature of content within video data requires video processing techniques to be efficient, especially in critical systems, for example, in applications such as surveillance, autonomous driving, activity recognition, and many more, where understanding movement and change is essential.

**A significant advantage of using OpenCV for video processing tasks is the provisioning of a suite of comprehensive features designed to handle video streams efficiently. OpenCV provides video capture, processing, and analysis tools as it does with images.**

**The following code snippet can be used to load and process a video located locally:**

```
import cv2

# Open a video capture object

cap = cv2.VideoCapture("Your video's path.mp4")

# To check if the video has opened successfully or not

if not cap.isOpened():

    print("Error: Couldnot open the video.")

    exit()

# Loop through the video frames

while True:

    # Reading a frame from the video
```

```
ret, frame = cap.read()
```

```
# Break the loop if the video has ended
```

```
if not ret:
```

```
    break
```

```
# Display the original frame
```

```
cv2.imshow('Original Video', frame)
```

```
gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
```

```
cv2.imshow('Black and White Video', gray_frame)
```

```
# Perform edge detection on the video
```

```
edges_frame = cv2.Canny(gray_frame, 50, 150)
```

```
cv2.imshow('Edges Video', edges_frame)
```

```
# Break the loop if 'q' key is pressed

if cv2.waitKey(25) & 0xFF == ord('q'):

    break

# Release the video capture object and close all windows

cap.release()

cv2.destroyAllWindows()
```

## OpenAI

In this code, the ‘cv2.VideoCapture’ class is used to access a video file and read frames through its method called ‘cap.read’. It then shows first the original video, then converts to black and white, and finally does edge detection for each frame. The loop runs until the video gets empty or before taking input from the keyboard to quit. The ‘cv2.waitKey(25)’ allows a short time delay between the frames to display them, thus ensuring an even playback.

## People Detection With OpenCV and Yolo

### Object Detection and People Detection

Object detection is a fundamental computer vision task that aims to detect and localize objects in an image or video and provide recognition results – a list of detected class labels and corresponding

**bounding boxes. People detection, a variant of object recognition, focuses on identifying and localizing people within visual information.**

### **Applications of People Detection**

**Applications of people detection include surveillance, crowd management, people counting, tracking, human-computer interaction, and smart environments. In surveillance, this technology aids in detecting suspicious activities, while in crowd management, it supports gesture recognition and tracking. People detection plays a crucial role in human-computer interaction and is also employed in smart environments for resource optimization and enhanced security.**

### **YOLO Object Detection Model**

**YOLO, an acronym for "You Only Look Once," has emerged as a commonly used system that efficiently achieves real-time object recognition. Contrary to classical object detection methods, YOLO divides the image into a grid of cells, simultaneously predicting bounding boxes and class probabilities for each cell.**

You can read more about [YOLO Object Detection](#) in our comprehensive guide.

### **Person Detection on an Image**

**Below is a code snippet using OpenCV and Yolo for person detection in an image:**

```
import cv2

net = cv2.dnn.readNet('yolov3.weights', 'yolov3.cfg')
```

```
layer_names = net.getUnconnectedOutLayersNames()

image = cv2.imread('image path.jpg')

height, width = image.shape[:2]

blob = cv2.dnn.blobFromImage(image, 1/255.0, (416, 416), swapRB=True, crop=False)

net.setInput(blob)

outs = net.forward(layer_names)

for out in outs:

    for detection in out:

        scores = detection[5:]

        class_id = np.argmax(scores)

        confidence = scores[class_id]

        if confidence > 0.5 and class_id == 0: # Class ID for person
```

```
box = detection[:4] * np.array([width, height, width, height])
```

```
(x, y, w, h) = box.astype(int)
```

```
cv2.rectangle(image, (x, y), (x + w, y + h), (0, 255, 0), 2)
```

```
cv2.imshow('Person Detection', image)
```

```
cv2.waitKey(0)
```

```
cv2.destroyAllWindows()
```

OpenAI

## Person Detection in a Video

Similarly, the code structure follows similar principles for person detection in a video. Instead of processing one static image, the script or application continuously processes video frames.

The same principle underlies the identification of people in both images and video clips, with the main difference being the factor of time. With images, the task involves recognizing and positioning people in single static frames.

This application is most suited for surveillance snapshots and other non-overlapping photographs. In contrast, with videos, the goal is to identify and track individuals as they move onto subsequent frames. This is sometimes referred to as object tracking in videos. The temporal aspect is crucial when considering human movements and interactions over time, making video-based person

**detection** a vital tool for real-time applications, including security surveillance, crowd monitoring, and behavior analysis.

## Real-life Applications and Ethical Considerations

Some other real-world use cases of people detection include smart cities, retail analytics, and public safety. Nevertheless, there are ethical considerations related to privacy issues, consent, and misuse of the technology. Responsible deployment across domains is achieved by balancing the benefits and ethical concerns.

## Conclusion

This tutorial demonstrated the foundational aspects of working with the OpenCV library for various computer vision tasks. Starting with the basics, it showed how to install OpenCV in a Python environment. OpenCV operations for loading and displaying images were showcased along with image processing techniques, such as converting images to grayscale and applying edge detection. Overall, it is clear that OpenCV can perform sophisticated image transformations and analysis, including applying image processing techniques to video frames.

Notably, OpenCV plays a crucial role in applied computer vision as It offers versatility and robust capabilities, enabling computer vision engineers and machine learning practitioners to address many challenges related to visual data processing. The key advantage of utilizing the OpenCV library is that it streamlines the process of applying computer vision algorithms and efficiently moves application concepts into practical implementations.

Enhance your understanding of the concepts discussed in this article by delving deeper into [image processing in Python](#) through a well-curated course. This course broadens your toolkit for image processing and computer vision with additional libraries like Sci-kit Image and Numpy

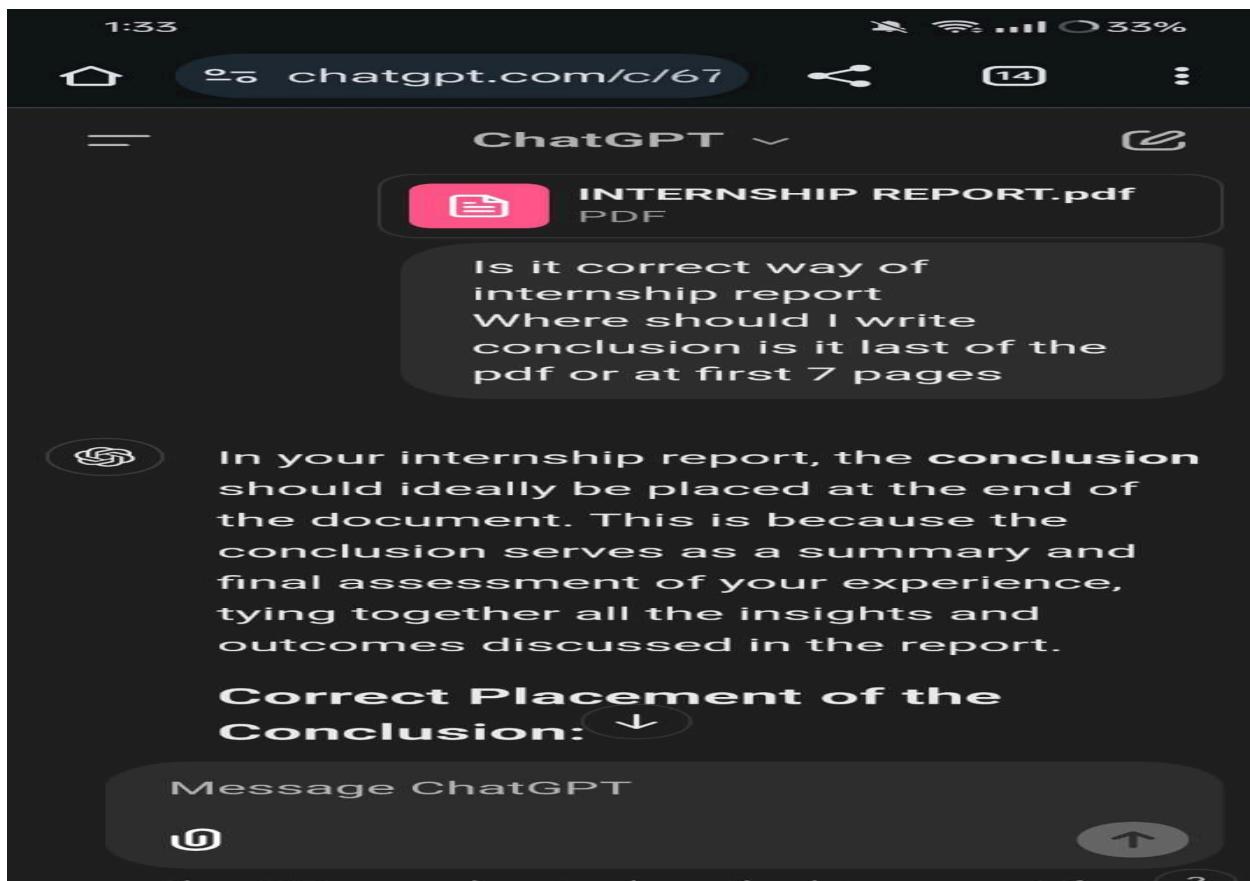
## Document Question Answering

**Document Question Answering** (also known as **Document Visual Question Answering**) is the task of answering questions on document images. Document question answering models take a (document, question) pair as input and return an answer in natural language. Models usually rely on multi-modal features, combining text, position of words (bounding-boxes) and image.

## Inputs

### Question

IS the report okey conclusion should be placed at last or at first 7 pages?



### Document Question Answering Model

## Output

### Answer

## ChatGPT

In your internship report, the conclusion should ideally be placed at the end of the document. This is because the conclusion serves as a summary and final assessment of your experience, tying together all the insights and outcomes discussed in the report.

### Correct Placement of the Conclusion:

The Table of Contents in your report already lists the conclusion in Chapter 3, which is at the end of the report. This is appropriate and follows the standard format.

If the conclusion is within the first seven pages, it might seem misplaced and disrupt the logical flow of the report.

### General Structure of an Internship Report:

Title Page

Certificate

Acknowledgements

Abstract

Introduction/Industry Profile

Details of Internship Work

Conclusion (at the end)

Keeping the conclusion as the last chapter ensures that readers can review all the necessary details before reaching the final summary. If you have further questions or need clarification, feel free to ask!

## LLM

Large language models (LLMs) are a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks.

LLMs have become a household name thanks to the role they have played in bringing generative AI to the forefront of the public interest, as well as the point on which organizations are focusing to adopt artificial intelligence across numerous business functions and use cases.

Outside of the enterprise context, it may seem like LLMs have arrived out of the blue along with new developments in [generative AI](#). However, many companies, including IBM, have spent years implementing LLMs at different levels to enhance their natural [language understanding \(NLU\)](#) and

natural language processing (NLP) capabilities. This has occurred alongside advances in machine learning, machine learning models, algorithms, neural networks and the transformer models that provide the architecture for these AI systems.

LLMs are a class of [foundation models](#), which are trained on enormous amounts of data to provide the foundational capabilities needed to drive multiple use cases and applications, as well as resolve a multitude of tasks. This is in stark contrast to the idea of building and training domain specific models for each of these use cases individually, which is prohibitive under many criteria (most importantly cost and infrastructure), stifles synergies and can even lead to inferior performance.

LLMs represent a significant breakthrough in NLP and [artificial intelligence](#), and are easily accessible to the public through interfaces like Open AI's Chat GPT-3 and GPT-4, which have garnered the support of Microsoft. Other examples include Meta's Llama models and Google's bidirectional encoder representations from transformers (BERT/RoBERTa) and PaLM models. IBM has also recently launched its [Granite model series](#) on [watsonx.ai](#), which has become the generative AI backbone for other IBM products like watsonx Assistant and watsonx Orchestrate.

In a nutshell, LLMs are designed to understand and generate text like a human, in addition to other forms of content, based on the vast amount of data used to train them. They have the ability to infer from context, generate coherent and contextually relevant responses, translate to languages other than English, summarize text, answer questions (general conversation and FAQs) and even assist in creative writing or [code generation tasks](#).

They are able to do this thanks to billions of parameters that enable them to capture intricate patterns in language and perform a wide array of language-related tasks. LLMs are revolutionizing applications in various fields, from chatbots and virtual assistants to content generation, research assistance and language translation.

As they continue to evolve and improve, LLMs are poised to reshape the way we interact with technology and access information, making them a pivotal part of the modern digital landscape.

EbookGenerative AI + ML for the enterprise

How large language models work ?

LLMs operate by leveraging deep learning techniques and vast amounts of textual data. These models are typically based on a transformer architecture, like the generative pre-trained transformer, which excels at

handling sequential data like text input. LLMs consist of multiple layers of neural networks, each with parameters that can be fine-tuned during training, which are enhanced further by a numerous layer known as the attention mechanism, which dials in on specific parts of data sets.

During the training process, these models learn to predict the next word in a sentence based on the context provided by the preceding words. The model does this through attributing a probability score to the recurrence of words that have been tokenized—broken down into smaller sequences of characters. These tokens are then transformed into embeddings, which are numeric representations of this context.

To ensure accuracy, this process involves training the LLM on a massive corpora of text (in the billions of pages), allowing it to learn grammar, semantics and conceptual relationships through zero-shot and self-supervised learning. Once trained on this training data, LLMs can generate text by autonomously predicting the next word based on the input they receive, and drawing on the patterns and knowledge they've acquired. The result is coherent and contextually relevant language generation that can be harnessed for a wide range of NLU and content generation tasks.

Model performance can also be increased through prompt engineering, [prompt-tuning](#), fine-tuning and other tactics like reinforcement learning with human feedback (RLHF) to remove the biases, hateful speech and factually incorrect answers known as “[hallucinations](#)” that are often unwanted byproducts of training on so much unstructured data. This is one of the most important aspects of ensuring [enterprise-grade LLMs](#) are ready for use and do not expose organizations to unwanted liability, or cause damage to their reputation.

#### LLM use cases

LLMs are redefining an increasing number of business processes and have proven their versatility across a myriad of use cases and tasks in various industries. They augment conversational AI in chatbots and virtual assistants (like IBM Watson Assistant and Google's BARD) to enhance the interactions that underpin excellence in customer care, providing context-aware responses that mimic interactions with human agents.

LLMs also excel in content generation, automating content creation for blog articles, marketing or sales materials and other writing tasks. In research and academia, they aid in summarizing and extracting information from vast datasets, accelerating knowledge discovery. LLMs also play a vital role in language translation, breaking down language barriers by providing accurate and contextually relevant translations. They can even be used to write code, or “translate” between programming languages.

Moreover, they contribute to accessibility by assisting individuals with disabilities, including text-to-speech applications and generating content in accessible formats. From healthcare to finance, LLMs are [transforming industries](#) by streamlining processes, improving customer experiences and enabling more efficient and data-driven decision making.

Most excitingly, all of these capabilities are easy to access, in some cases literally an API integration away.

Here is a list of some of the most important areas where LLMs benefit organizations:

Text generation: language generation abilities, such as writing emails, blog posts or other mid-to-long form content in response to prompts that can be refined and polished. An excellent example is retrieval-augmented generation ([RAG](#)).

Content summarization: summarize long articles, news stories, research reports, corporate documentation and even customer history into thorough texts tailored in length to the output format.

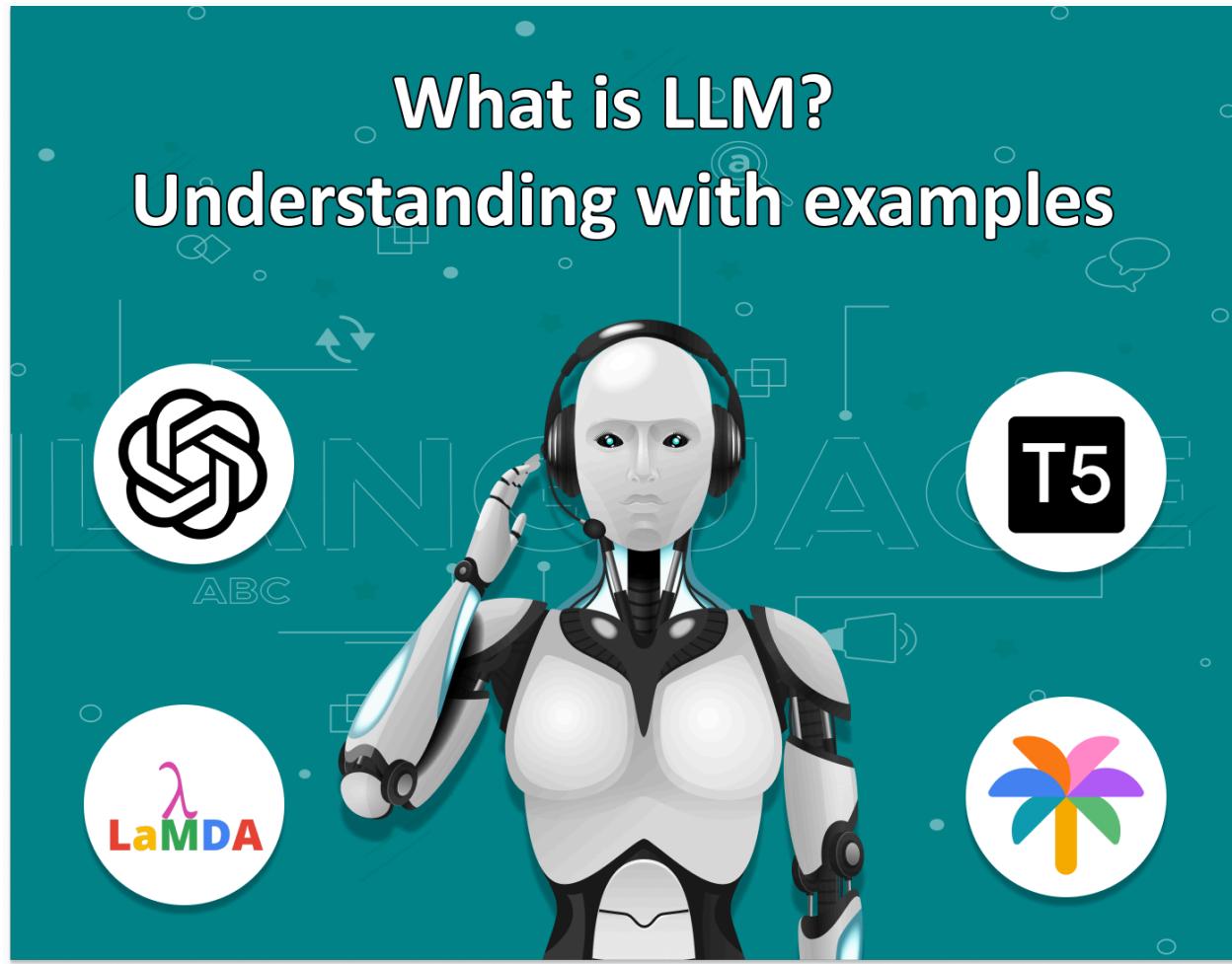
AI assistants: chatbots that answer customer queries, perform backend tasks and provide detailed information in natural language as a part of an integrated, self-serve customer care solution.

Code generation: assists developers in building applications, finding errors in code and uncovering security issues in multiple programming languages, even “translating” between them.

Sentiment analysis: analyze text to determine the customer’s tone in order understand customer feedback at scale and aid in brand reputation management.

Language translation: provides wider coverage to organizations across languages and geographies with fluent translations and multilingual capabilities.

LLMs stand to impact every industry, from finance to insurance, human resources to healthcare and beyond, by automating customer self-service, accelerating response times on an increasing number of tasks as well as providing greater accuracy, enhanced routing and intelligent context gathering.



### CLAUDE :

Artificial intelligence has been revolutionized by the advent of Large Language Models (LLMs), and one of the most notable entrants in this field is CLAUDE. Developed by Anthropic, a leading AI research company, CLAUDE sets new industry benchmarks across various cognitive tasks

#### Capabilities of CLAUDE

CLAUDE is not just a single model but a family of models designed for specific tasks. The family includes three state-of-the-art models: Claude 3 Haiku, Claude 3 Sonnet, and Claude 3 Opus[2].

**Each successive model offers increasingly powerful performance, allowing users to select the optimal balance of intelligence, speed, and cost for their specific application[2].**

**Claude 3 Haiku is the fastest and most cost-effective model for its intelligence category[2]. It excels at executing lightweight actions with industry-leading speed[1].**

**Claude 3 Sonnet is the best combination of performance and speed for efficient, high-throughput tasks[1]. It excels at tasks demanding rapid responses, like knowledge retrieval or sales automation[2].**

**Claude 3 Opus is the most intelligent model, capable of handling complex analysis, longer tasks with multiple steps, and higher-order math and coding tasks[1]. It outperforms its peers on most of the standard evaluation benchmarks for AI systems[2].**

## **Vision Capabilities**

**One of the standout features of the CLAUDE3 models is their sophisticated vision capabilities[2]. They can process various visual formats, including photos, charts, graphs, and technical diagrams[2]. This makes them particularly useful for enterprise customers, who often have up to 50% of their knowledge bases encoded in PDFs, flowcharts, or presentation slides[2].**

## **Security and Accessibility**

**In addition to its impressive capabilities, CLAUDE3 boasts enterprise-grade security and data handling for API[1]. It is SOC II Type 2 certified, with HIPAA compliance options for API[1]. It is accessible through AWS (GA) and GCP (in private preview)[1].**

### **Real-world applications of CLAUDE3**

**CLAUDE3, a Large Language Model (LLM) developed by Anthropic, has many real-world applications. Here are some of them:**

- 1. Content Generation:** CLAUDE3 is excellent at generating content[4]. It can automatically create texts for various purposes, including articles, blog posts, marketing copy, video scripts, and social media updates[4]. Businesses and content creators use these models to streamline content production, saving time and effort in writing [4].
- 2. Natural Language Processing (NLP) Applications:** CLAUDE3 can be leveraged for various NLP tasks, including language translation, text summarization, sentiment analysis, and conversational AI[5].
- 3. Code Generation:** With its powerful AI system and larger context window, CLAUDE3 offers advanced capabilities for code generation[6].
- 4. Image Analysis:** CLAUDE3 can also be used for image analysis[6], making it particularly useful for enterprise customers who often have up to 50% of their knowledge bases encoded in various formats, such as PDFs, flowcharts, or presentation slides[7].

**5. Writing Assistance:** CLAUDE3's language modelling capabilities allow it to comprehend complex sentences, passages, and documents to determine meaning, context and needed actions or answers[8]. It can assist with various everyday tasks, including writing assistance[8].

These applications showcase the versatility and potential of CLAUDE3 in various fields, from content creation to data analysis. As the technology continues to evolve, we can expect to see even more sophisticated and diverse applications of CLAUDE3 in the future.

### **Case study**

I developed notebooks [9,10] and thoroughly tested them in Google Colab to demonstrate CLAUDE3[1,2,3] capabilities.

### **Conclusion**

In conclusion, CLAUDE3 represents a significant advancement in the field of LLMs. With its unique combination of speed, performance, and cost-effectiveness, it offers a compelling choice for enterprises looking to leverage the power of AI. Its sophisticated vision capabilities and robust security measures further enhance its appeal. As we progress, it will be exciting to see how CLAUDE3 continues to shape the landscape of artificial intelligence.



## GPT :

The Generative Pre-trained Transformer (GPT) is a model, developed by Open AI to understand and generate human-like text. GPT has revolutionized how machines interact with human language, enabling more intuitive and meaningful communication between humans and computers.

In this article, we are going to explore more about Generative Pre-trained Transformer.

### Table of Content

- [What is a Generative Pre-trained Transformer?](#)
- [Background and Development of GPT](#)
- [Architecture of Generative Pre-trained Transformer](#)
- [Training Process of Generative Pre-trained Transformer](#)
- [Applications of Generative Pre-trained Transformer](#)
- [Advantages of GPT](#)
- [Ethical Considerations](#)

- **Conclusion**

### **What is a Generative Pre-trained Transformer?**

GPT is based on the transformer architecture, which was introduced in the paper “Attention is All You Need” by Vaswani et al. in 2017. The core idea behind the transformer is the use of self-attention mechanisms that process words in relation to all other words in a sentence, contrary to traditional methods that process words in sequential order. This allows the model to weigh the importance of each word no matter its position in the sentence, leading to a more nuanced understanding of language.

As a generative model, GPT can produce new content. When provided with a prompt or a part of a sentence, GPT can generate coherent and contextually relevant continuations. This makes it extremely useful for applications like creating written content, generating creative writing, or even simulating dialogue.

### **Background and Development of GPT**

The progress of GPT (Generative Pre-trained Transformer) models by OpenAI has been marked by significant advancements in natural language processing. Here’s a chronological overview:

1. **GPT (June 2018):** The original GPT model was introduced by OpenAI as a pre-trained transformer model that achieved state-of-the-art results on a variety of natural language processing tasks. It featured 12 layers, 768 hidden units, and 12 attention heads, totaling 117 million parameters. This model was pre-trained on a diverse dataset using unsupervised learning and fine-tuned for specific tasks.
2. **GPT-2 (February 2019):** An upgrade from its predecessor, GPT-2 featured 48 transformer blocks, 1,600 hidden units, and 25 million parameters in its smallest version, up to 1.5 billion parameters in its largest. OpenAI initially delayed the release of the most powerful versions due to concerns about potential misuse. GPT-2

- demonstrated an impressive ability to generate coherent and contextually relevant text over extended passages.
3. **GPT-3 (June 2020):** GPT-3 marked a massive leap in the scale and capability of language models with 175 billion parameters. It improved upon GPT-2 in almost all aspects of performance and demonstrated abilities across a broader array of tasks without task-specific tuning. GPT-3's performance showcased the potential for models to exhibit behaviors resembling understanding and reasoning, igniting widespread discussion about the implications of powerful AI models.
  4. **GPT-4 (March 2023):** GPT-4 expanded further on the capabilities of its predecessors, boasting more nuanced and accurate responses, and improved performance in creative and technical domains. While the exact parameter count has not been officially disclosed, it is understood to be significantly larger than GPT-3 and features architectural improvements that enhance reasoning and contextual understanding.

### Architecture of Generative Pre-trained Transformer

The transformer architecture, which is the foundation of GPT models, is made up of feedforward neural networks and layers of self-attention processes.

Important elements of this architecture consist of:

1. **Self-Attention System:** This enables the model to evaluate each word's significance within the context of the complete input sequence. It makes it possible for the model to comprehend word linkages and dependencies, which is essential for producing content that is logical and suitable for its context.
2. **Layer normalization and residual connections:** By reducing problems such as disappearing and exploding gradients, these characteristics aid in training stabilization and enhance network convergence.
3. **Feedforward Neural Networks:** These networks process the output of the attention mechanism and add another layer of abstraction and learning capability. They are positioned between self-attention layers.

### Training Process of Generative Pre-trained Transformer

**Large-scale text data corpora are used for unsupervised learning to train GPT algorithms. There are two primary stages to the training:**

1. **Pre-training:** Known as language modeling, this stage teaches the model to anticipate the word that will come next in a sentence. In order to make that the model can produce writing that is human-like in a variety of settings and domains, this phase makes use of a wide variety of internet material.
2. **Fine-tuning:** While GPT models perform well in zero-shot and few-shot learning, fine-tuning is occasionally necessary for particular applications. In order to improve the model's performance, this entails training it on data specific to a given domain or task.

### **Applications of Generative Pre-trained Transformer**

**The versatility of GPT models allows for a wide range of applications, including but not limited to:**

1. **Content Creation:** GPT can generate articles, stories, and poetry, assisting writers with creative tasks.
2. **Customer Support:** Automated chatbots and virtual assistants powered by GPT provide efficient and human-like customer service interactions.
3. **Education:** GPT models can create personalized tutoring systems, generate educational content, and assist with language learning.
4. **Programming:** GPT-3's ability to generate code from natural language descriptions aids developers in software development and debugging.
5. **Healthcare:** Applications include generating medical reports, assisting in research by summarizing scientific literature, and providing conversational agents for patient support.

### **Advantages of GPT**

1. **Flexibility:** GPT's architecture allows it to perform a wide range of language-based tasks.
2. **Scalability:** As more data is fed into the model, its ability to understand and generate language improves.

3. **Contextual Understanding:** Its deep learning capabilities allow it to understand and generate text with a high degree of relevance and contextuality.



## GEMINI :

**Gemini** is Google's latest and most capable AI model family, designed to be a powerful assistant for a wide range of tasks. Here are the key points about Gemini:

Gemini can understand, explain and generate high-quality code in over 20 programming languages, including Python, Java, C++, and Go. It is a multimodal model, meaning it can work with and generate not just text, but also audio, images, and videos.

**Gemini comes in three versions — Gemini Ultra (the flagship model), Gemini Pro (a “lite” version), and Gemini Nano (a smaller model for mobile devices).**

**[Access Google Gemini Now!](#)**

All Gemini models were trained on a large and diverse dataset to give them broad capabilities. Gemini can be used in various ways through Google's products and services

**Gemini Code Assist provides real-time code recommendations, identifies vulnerabilities, and suggests fixes in popular IDEs and developer platforms.**

- **Gemini Cloud Assist helps cloud teams design, deploy, manage, and optimize cloud applications.**
- **Gemini in Security enhances threat detection, investigation, and response for cybersecurity teams.**
- **Gemini in Databases simplifies database management tasks.**

Google has also launched a new Gemini mobile app that allows users to access Gemini's capabilities on the go, such as generating content, answering questions, and controlling smart home devices.

# Gemini AI

- **Unlocking Gemini's Power**
- **AI Redefined: Gemini**
- **Google's AI Revolution**
- **Gemini: The Future AI**

 SoluLab



## Llama :

Meta's Llama 3, the next iteration of the open-access Llama family, is now released and available at Hugging Face. It's great to see Meta continuing its commitment to open AI, and we're excited to fully support the launch with comprehensive integration in the Hugging Face ecosystem.

Llama 3 comes in two sizes: 8B for efficient deployment and development on consumer-size GPU, and 70B for large-scale AI native applications. Both come in base and instruction-tuned variants. In addition to the 4 models, a new version of Llama Guard was fine-tuned on Llama 3 8B and is released as Llama Guard 2 (safety fine-tune).

We've collaborated with Meta to ensure the best integration into the Hugging Face ecosystem. You can find all 5 open-access models (2 base models, 2 fine-tuned & Llama Guard) on the Hub. Among the features and integrations being released, we have:

- [Models on the Hub](#), with their model cards and licenses
- Transformers integration
- [Hugging Chat integration for Meta Llama 3 70b](#)
- Inference Integration into Inference Endpoints, Google Cloud & Amazon SageMaker
- An example of fine-tuning Llama 3 8B on a single GPU with TRL

## Table of contents

- [What's new with Llama 3?](#)
- [Llama 3 evaluation](#)
- [How to prompt Llama 3](#)
- [Demo](#)
- [Using Transformers](#)
- [Inference Integrations](#)
- [Fine-tuning with TRL](#)
- [Additional Resources](#)
- [Acknowledgments](#)

## What's new with Llama 3?

The Llama 3 release introduces 4 new open LLM models by Meta based on the Llama 2 architecture.

They come in two sizes: 8B and 70B parameters, each with base (pre-trained) and instruct-tuned versions.

All the variants can be run on various types of consumer hardware and have a context length of 8K tokens.

- [Meta-Llama-3-8b](#): Base 8B model
- [Meta-Llama-3-8b-instruct](#): Instruct fine-tuned version of the base 8b model
- [Meta-Llama-3-70b](#): Base 70B model
- [Meta-Llama-3-70b-instruct](#): Instruct fine-tuned version of the base 70b model

In addition to these 4 base models, Llama Guard 2 was also released. Fine-tuned on Llama 3 8B, it's the latest iteration in the Llama Guard family. Llama Guard 2, built for production use cases, is designed to classify LLM inputs (prompts) as well as LLM responses in order to detect content that would be considered unsafe in a risk taxonomy.

A big change in Llama 3 compared to Llama 2 is the use of a new tokenizer that expands the vocabulary size to 128,256 (from 32K tokens in the previous version). This larger vocabulary can encode text more efficiently (both for input and output) and potentially yield stronger multilingualism. This comes at a cost, though: the embedding input and output matrices are larger, which accounts for a good portion of the parameter count increase of the small model: it goes from 7B in Llama 2 to 8B in Llama 3. In addition, the 8B version of the model now uses Grouped-Query Attention (GQA), which is an efficient representation that should help with longer contexts.

The Llama 3 models were trained ~8x more data on over 15 trillion tokens on a new mix of publicly available online data on two clusters with 24,000 GPUs. We don't know the exact details of the training mix, and we can only guess that bigger and more careful data curation was a big factor in the improved performance. Llama 3 Instruct has been optimized for dialogue applications and was trained on over 10 Million human-annotated data samples with combination of supervised fine-tuning (SFT), rejection sampling, proximal policy optimization (PPO), and direct policy optimization (DPO).

Regarding the licensing terms, Llama 3 comes with a permissive license that allows redistribution, fine-tuning, and derivative works. The requirement for explicit attribution is new in the Llama 3 license and was not present in Llama 2. Derived models, for instance, need to include "Llama 3" at the beginning of their name, and you also need to mention "Built with Meta Llama 3" in derivative works or services. For full details, please make sure to read the [official license](#).

### Llama 3 evaluation

*Note: We are currently evaluating Meta Llama 3 individually and will update this section as soon as we get the results.*

### Integration with Google Cloud

You can deploy Llama 3 on Google Cloud through Vertex AI or Google Kubernetes Engine (GKE), using [Text Generation Inference](#).

To deploy the Llama 3 model from Hugging Face, go to the [model page](#) and click on [Deploy -> Google Cloud](#). This will bring you to the Google Cloud Console, where you can 1-click deploy Llama 3 on Vertex AI or GKE.



## OPEN SOURCE LLMs :

Large language models, or LLMs, are essential to the present revolution in generative AI. Language models and interpreters are artificial intelligence (AI) systems that are based on transformers, a potent neural architecture. They are referred to as “large” because they contain hundreds of millions, if not billions, of pre-trained parameters derived from a vast corpus of text data.

In this article, we'll look at the *Top 10 open-source LLMs* that will be available in 2024. Even though ChatGPT and (proprietary) LLMs have only been around for a year, the open-source community

has made significant progress, and there are now numerous open-source LLMs available for various applications. Read on to discover the most popular!



*LLM Models open-source*

### Top 10 Open-Source LLM Models

- [1. LLaMA 2](#)
- [2. BLOOM](#)
- [3. BERT \(Bidirectional Encoder Representations from Transformers\)](#)
- [4. Falcon 180B](#)
- [5. OPT-175B](#)
- [6. XGen-7B](#)
- [7. GPT-NeoX and GPT-NeoX](#)
- [8. Vicuna 13-B](#)
- [9. YI 34B](#)
- [10. Mixtral 8x7B](#)

### Top Open-Source Large Language Models For 2024

The basic models of widely used and well-known chatbots, such as [Google Bard](#) and [ChatGPT](#), are LLM. In particular, Google Bard is built on Google's PaLM 2 model, whereas ChatGPT is driven by GPT-4, an LLM created and owned by OpenAI. The proprietary underlying LLM of ChatGPT, Bard, and numerous other well-known chatbots are shared by them. This indicates that they belong to a business and that clients can only use them with a license that they have purchased. Along with rights, that license may also impose limitations on how the [LLM](#) is used and provide access to certain technical details.

However, open-source LLMs are a parallel trend in the Large Language Model that is quickly gaining traction. Open-source LLMs promise to improve accessibility, transparency, and innovation in the rapidly expanding field of generative AI and LMMs in response to growing concerns about the opaque nature and restricted availability of proprietary LLMs, which Big Tech companies like Microsoft, Google, and Meta primarily control.

## 1. LLaMA 2

Most top LLM firms developed their programs discreetly. Meta stands out. Meta provided crucial information about LLaMA 2 and its powerful, open-source alternative. [LLaMA 2](#), a 7–70 billion-parameter generative text model, finished in July 2023. This model is for business and study. The RLHF improved it. Construct and train this text generation model to teach the chatbot natural language. Meta provides open, customizable LLaMA 2, Chat, and Code Llama.

### Features:

- **EleutherAI Development:** LLaMA 2 is a product of the EleutherAI community, known for its commitment to open-source AI research.

- **Improved Performance:** Building upon its predecessor, LLaMA, this model incorporates advanced techniques to enhance language understanding and generation capabilities.
- **Versatility:** LLaMA 2 demonstrates proficiency across various NLP tasks, making it a versatile choice for researchers and developers alike.

## 2. BLOOM

In 2022, Flourish developed BLOOM, an autoregressive Large Language Model (LLM) that generates text by extending a prompt using large amounts of textual data. Over 70 countries' experts and volunteers developed the project in one year. The open-source LLM BLOOM model includes 176 billion parameters. It writes fluently and cohesively in 46 languages and 13 programming languages. BLOOM execution, evaluation, and improvement with training data and source code are public. Hugging Face users use [\*\*BLOOM\*\*](#) free.

### Features:

- **Efficiency:** Developed by Google Research, BLOOM emphasizes efficiency and scalability, making it suitable for large-scale NLP applications.
- **Advanced Algorithms:** BLOOM leverages cutting-edge algorithms to achieve remarkable performance in tasks such as text summarization and language translation.
- **Community Support:** With support from the open-source community, BLOOM continues to evolve, driving innovation in the field of NLP.

## 3. BERT (Bidirectional Encoder Representations from Transformers)

LLM technology relies on [\*\*BERT\*\*](#) (Bidirectional Encoder Representations from Transformers) neural architecture. Google researchers released “Attention is All You Need.” in 2017. BERT was an early transformer test. The 2018 Google Language Model BERT is available as open-source software. It swiftly mastered natural language processing tasks.

Bert's advanced early LLM development capabilities and open-source nature make it a popular Language Model (LLM). With Bert in 2020, Google Search is available in over 70 languages. Many pre-trained [Bert models](#) are open-source. These models help detect harmful comments, clinical notes, and sentiments.

Features:

- **Contextual Understanding:** BERT excels in understanding context and semantics within textual data, owing to its bidirectional architecture.
- **Pretrained Models:** Google's BERT comes with pretrained models for various languages and domains, facilitating transfer learning and adaptation to specific tasks.
- **Wide Applicability:** BERT finds applications in tasks such as sentiment analysis, question answering, and named entity recognition, showcasing its broad utility.

#### 4. Falcon 180B

The new Falcon 180B indicates that the difference between proprietary and open-source large language models is fast narrowing if the Falcon 40B, which ranked #1 on Hugging Face's scoreboard for big language models, wasn't already impressive to the open-source LLM community. Falcon 180B, which was made available by the Technology Innovation Institute of the United Arab Emirates in September 2023, is being trained using 3.5 trillion tokens and 180 billion parameters. Hugging Face indicates that Falcon 180B can compete with Google's PaLM 2, the LLM that runs Google Bard, given its amazing processing capacity. Falcon 180B has already surpassed LLaMA 2 and [GPT-3.5](#) in some NLP tasks.

It's crucial to remember that Falcon 180B needs significant processing power to operate while being free for usage in both commercial and research settings.

Features:

- **Massive Parameter Size:** Falcon 180B boasts a large parameter size, enabling it to capture intricate linguistic patterns and nuances.
- **Enhanced Learning:** With extensive training data and robust architecture, Falcon 180B demonstrates superior learning capabilities, particularly in language-related tasks.
- **Scalability:** Despite its size, Falcon 180B maintains efficiency and scalability, making it suitable for deployment in diverse NLP applications.

## 5. OPT-175B

In 2022, Meta achieved a significant milestone with the publication of the Open Pre-trained Transformers Language Models (TLM), which was part of their aim to use open source to free the LLM race. OPT consists of a set of pre-trained transformers, decoder-only, with parameters ranging from 125M to 175B. The most potent brother is OPT-175B, an open-source LLM that is among the most sophisticated on the market and performs similarly to GPT-3. The public can access both the source code and the pre-trained models. But, you'd best think of another option if you're planning to build an AI-driven business with LLMs, as OPT-175B is only available under a non-commercial license that permits the model's use for research use cases.

**Features:**

- **Precision and Efficiency:** OPT-175B is designed to handle complex language processing tasks with precision and efficiency, catering to the demands of modern NLP applications.
- **Fluency and Coherence:** This model showcases remarkable fluency and coherence in generating text across various domains, reflecting its advanced training techniques.
- **State-of-the-Art Performance:** OPT-175B achieves state-of-the-art performance in language-related benchmarks, solidifying its position as a leading open-source LLM.

## 6. XGen-7B

Businesses are entering the LLM race at an increasing rate. Salesforce was among the latest to enter the market, with the release of its XGen-7B LLM in July 2023. The authors claim that the majority of open-source LLMs concentrate on offering lengthy responses with scant details (i.e., brief prompts with little context). [XGen-7B](#) is an attempt to create a tool that can handle larger context windows. Specifically, the most sophisticated variation of XGen (XGen-7B-8K-base) supports an 8K context window—that is, the whole amount of text in both the input and output.

While XGen only utilizes 7B parameters for training—much fewer than most powerful open-source LLMs like LLaMA 2 or Falcon—efficiency is another top objective. Even though XGen is small in size, it can nonetheless produce excellent results. With the exception of the XGen-7 B-{4K,8K}-inst version, which was trained using instructional data and RLHF and is made available under a noncommercial license, the model is available for both commercial and research use.

#### Features:

- **Versatility:** XGen-7B demonstrates versatility and adaptability, leveraging advanced transformer architectures to excel in diverse NLP tasks.
- **Text Classification:** This model is proficient in tasks such as text classification, sentiment analysis, and document summarization, showcasing its broad applicability.
- **Ease of Use:** With user-friendly interfaces and extensive documentation, XGen-7B is accessible to both novice and experienced practitioners in the field of NLP.

## 7. GPT-NeoX and GPT-NeoX

Generated by scientists at the nonprofit AI research center EleutherAI, GPT-NeoX and GPT-J are two excellent open-source substitutes for GPT. There are 20 billion parameters in GPT-NeoX and 6 billion in GPT-J. These two LLMs are able to produce findings with a high degree of accuracy, even though the majority of advanced LLMs can be trained using more than 100 billion parameters. They can be used in many different domains and application situations because they were trained

**on 22 high-quality datasets from a variety of sources. GPT-NeoX and GPT-J, in contrast to GPT-3, have not been trained using RLHF.**

**GPT-NeoX and GPT-J can be used for any natural language processing activity, including research, marketing campaign planning, sentiment analysis, and text generation. With the NLP Cloud API, you can get both LLMs for free.**

**Features:**

- **Community-Driven Development:** GPT-NeoX/J is a community-driven effort to develop an open-source LLM capable of rivaling proprietary models in performance and scalability.
- **Innovative Architecture:** With its innovative architecture and extensive training data, GPT-NeoX/J delivers impressive results across diverse NLP tasks, including text generation and dialogue systems.
- **Continuous Improvement:** Supported by a vibrant community of researchers and developers, GPT-NeoX/J undergoes continuous improvement, incorporating the latest advancements in NLP research.

## **8. Vicuna 13-B**

**Using user-shared conversations collected from ShareGPT, the LLaMa 13B model was refined to create the open-source conversational model Vicuna-13B. Vicuna-13B is an intelligent chatbot with a plethora of uses; a few are shown below in various industries, including customer service, healthcare, education, finance, and travel/hospitality. According to an initial assessment using GPT-4 as a judge, [Vicuna-13B](#) surpassed other models such as LLaMa and Alpaca in more than 90% of cases, attaining over 90% quality of [ChatGPT](#) and Google Bard.**

**Features:**

- **Efficiency and Accuracy:** Vicuna 13-B prioritizes efficiency and accuracy, making it well-suited for applications requiring nuanced understanding of textual data.
- **Focused Development:** Developed with a focus on specific NLP tasks, Vicuna 13-B delivers robust performance in areas such as language modeling and text completion.
- **Customization Options:** With customizable parameters and fine-tuning capabilities, Vicuna 13-B offers flexibility to adapt to diverse use cases and domains.

## 9. YI 34B

YI 34B China's 01 AI developed a new language model called [YI 34B](#). Right now, this model holds the top spot on the Hugging Face Open LLM leaderboard. The company's goal is to develop bilingual models that are capable of speaking Chinese and English. The model may now be trained on up to 32K tokens, compared to its original 4K token context window.

It's impressive that the company recently released a 200,000 token version of the 34B model. These models can be licensed for commercial usage and are available for research purposes. With 3 trillion tokens under its belt, the 34B model excels in arithmetic and coding. Benchmarks for both the supervised fine-tuned conversation models and the base models have been made available by the company. There are multiple 4-bit and 8-bit versions available for the model.

### Features:

- **Massive Parameter Size:** YI 34B boasts a massive parameter size, enabling it to capture intricate linguistic nuances and generate contextually relevant text.
- **Robust Architecture:** With its robust architecture and extensive fine-tuning, YI 34B demonstrates superior performance in language-related tasks, including text generation and sentiment analysis.
- **Scalability:** Despite its size, YI 34B maintains scalability and efficiency, making it suitable for deployment in resource-constrained environments.

## 10. Mixtral 8x7B

**Mixtral 8x7B**, unveiled by Mistral AI in December 2023, is a decoder-only sparse mixture-of-experts network licensed under Apache 2.0. It outperforms LLaMA 2 and GPT 3.5 on various benchmarks despite having a smaller parameter size. With only 12.9 billion parameters per token out of a total of 46.7 billion, Mixtral achieves comparable processing rates to a 12.9B model.

It's ranked among the top 10 LLMs by the Hugging Face Open LLM Leaderboard, excelling in benchmarks like ARC, HellaSwag, MMLU, and TruthfulQA. Mixtral offers 6 times faster inference than LLaMA 2 70B and outperforms GPT 3.5 in most areas except for the Mt Bench score. It exhibits less bias on the BBQ benchmark and boasts multilingual capabilities in English, French, Italian, German, and Spanish. Mistral AI continually enhances Mixtral's linguistic capabilities to cater to a diverse range of applications and users.

#### Features:

- **Blend of Performance and Efficiency:** Mixtral 8x7B offers a compelling blend of performance and efficiency, making it an attractive choice for various NLP applications.
- **Innovative Training Strategies:** Leveraging innovative training strategies and diverse datasets, Mixtral 8x7B achieves impressive capabilities in language understanding and generation.
- **Accessibility:** With accessible documentation and pretrained models, Mixtral 8x7B is accessible to a wide range of users, facilitating experimentation and research in the field of NLP.

## Cyber Security Basics

Fundamental principles and practices for protecting computer systems and networks from cyber threat.

## **Devices:**

- Use strong passwords for all your devices atleast 12 characters
- Set your software's to update automatically this includes apps, web browsers, operating systems.
- Backup your important files offline like cloud or external hard drive. If your devices contain personal information make sure they are **encrypted** and require **Multi-factor Authentication** to access areas of your network with sensitive information. This requires additional steps beyond logging in with a password- Like a temporary code on a smart phone or a key that's inserted into a computer
- Secure your router by changing its default name & password, turn off remote management, Login out as the administrator. Make sure that your router is using WPA2(or)WPA3 encryption which protects your information sent over your network so it can't be read by outsiders...

## **Types of Cyber Crimes**

**Phishing:** Phishing is a type of social engineering attack that targets the user and tricks

them by sending fake messages and emails to get sensitive information about the user or

trying to download malicious software and exploit it on the target system.

**Ransomware:** [Ransomware attacks](#) are a very common type of cybercrime. It is a type of malware that has the capability to prevent users from accessing all of their personal data on the system by encrypting them and then asking for a ransom in order to give access to the encrypted data.

**Internet Fraud:** Internet fraud is a type of cybercrimes that makes use of the internet and it can be considered a general term that groups all of the crimes that happen over the internet like spam, banking frauds, theft of service, etc.

**CIA Triad :**

**Definition for CIA Triad:** A model designed to guide policies for information security within an organization.

**Confidentiality:** Ensuring that information is accessible only to those authorized to have access.

**Integrity:** Maintaining the accuracy and completeness of data.

**Availability:** Ensuring that authorized users have access to information and resources when needed.

**Importance of CIA Triad:** Confidentiality protects sensitive information from unauthorized access. Integrity ensures data is trustworthy and accurate. Availability guarantees that information and resources are accessible when needed by authorized users.

**AAA Framework :** The AAA (Authentication, Authorization, Accounting) framework is a fundamental security architecture widely used in network and computer security to manage and enforce access control and auditing. Here's an overview of each component of the AAA framework:

### **Authentication:**

**- Definition:** Authentication is the process of verifying the identity of a user, device, or entity. It ensures that the entity requesting access is indeed who or what it claims to be.

**- Methods:**

- Password-based: Users provide a username and password.
- Multi-Factor Authentication (MFA): Combines two or more independent credentials (e.g., password, biometric, smart card).
- Biometric: Uses unique biological characteristics (e.g., fingerprints, retina scans).
- Certificate-based: Utilizes digital certificates issued by a trusted Certificate Authority (CA).

## Authorization:

- **Definition:** Authorization determines what an authenticated user or entity is allowed to do. It involves granting or denying permissions based on predefined policies.

### - Methods:

- **Role-Based Access Control (RBAC):** Permissions are assigned based on user roles within an organization.

- **Attribute-Based Access Control (ABAC):** Permissions are granted based on attributes (e.g., user characteristics, resource types, environmental conditions).

- **Access Control Lists (ACLs):** Lists that specify which users or system processes are granted access to objects and what operations are allowed.

## Accounting:

- **Definition:** Accounting, also known as auditing, involves tracking and recording user activities and resource usage. This helps in monitoring, analysis, and reporting.

### - Methods:

- **Logging:** Recording user actions and access events in logs for future reference.

- Usage Monitoring: Keeping track of resource consumption and usage patterns.
- Auditing: Regularly reviewing logs and usage reports to ensure compliance and detect anomalies.

### Application of the AAA Framework:

#### 1. Network Security:

- Remote Access: AAA is crucial for secure remote access through VPNs and remote desktop services.
- Wireless Networks: Used in securing Wi-Fi networks (e.g., WPA2-Enterprise) by verifying users and devices before allowing network access.

#### 2. Cloud Services:

- Identity and Access Management (IAM): Cloud providers use AAA principles to manage user identities, permissions, and track activity.
- Multi-tenant Environments: Ensures that users can only access their own data and applications.

#### 3. Application Security:

- Web Applications: AAA is implemented to control user access to various parts of an application based on user roles and permissions.
- APIs: Ensures that only authenticated and authorized applications can interact with APIs.

#### 4. Compliance and Auditing:

- Regulatory Compliance: Organizations use accounting mechanisms to comply with regulations like GDPR, HIPAA, and PCI-DSS, which require detailed logs and access records.
- Security Audits: Regular audits based on accounting data help in identifying security issues and ensuring compliance with security policies.

#### AAA Protocols and Technologies:

- RADIUS (Remote Authentication Dial-In User Service): A networking protocol providing centralized Authentication, Authorization, and Accounting management for users who connect and use a network service.
- TACACS+ (Terminal Access Controller Access-Control System Plus): Provides detailed accounting and auditing, separating all three components of AAA, often used in enterprise networks.
- LDAP (Lightweight Directory Access Protocol): Used for accessing and maintaining distributed directory information services, often used in conjunction with AAA for

authentication purposes.

- Kerberos: A network authentication protocol designed to provide strong authentication for client/server applications by using secret-key cryptography. 7980 The AAA framework is essential for maintaining robust security in various systems, ensuring that only authorized users gain access, their activities are appropriately controlled, and all actions are logged for accountability and auditing purposes.

**5. OWASP(Open Web Application Security Project)** OWASP, the Open Web Application Security Project, is a non-profit organization dedicated to improving the security of software. OWASP provides a wealth of free and open resources related to application security, including tools, documentation, and forums for discussing security issues. Here are some key aspects and resources provided by OWASP:

### Key Projects and Resources

#### 1. OWASP Top Ten:

- **Definition:** The OWASP Top Ten is a regularly updated list of the top ten most critical web application security risks.

- **Purpose:** Raises awareness and provides guidance on how to mitigate these risks.

- Current Top Ten Risks (as of the latest update):

1. Broken Access Control

2. Cryptographic Failures

3. Injection

4. Insecure Design

5. Security Misconfiguration

6. Vulnerable and Outdated Components

7. Identification and Authentication Failures

8. Software and Data Integrity Failures

9. Security Logging and Monitoring Failures 8182

10. Server-Side Request Forgery (SSRF)

1. **Broken Access Control:** A weakness that allows an attacker to gain access to user

accounts. The attacker in this context can function as a user or as an administrator in the system.

2. **Cryptographic Failures:** Cryptographic failures refer to vulnerabilities and security risks

arising from the incorrect implementation, configuration, or use of cryptographic systems. These failures can compromise the confidentiality, integrity, and authenticity of data. They are among the most critical security risks.

**3.Injection:** Injection is a class of vulnerabilities where an attacker can send untrusted input to an application, which is then processed in an unsafe manner, leading to unintended behavior. This can include accessing or modifying data, executing arbitrary commands, or otherwise compromising the application's security. Injection flaws are among the most critical security risks.

**4.Software and Data Integrity Failures:** Software and data integrity failures occur when software or data is compromised due to improper implementation, configuration, or lack of validation and verification mechanisms. These failures can lead to unauthorized code execution, data corruption, or system compromise.

**Conclusion:**

OWASP plays a crucial role in the field of web application security by providing essential resources, fostering community collaboration, and promoting best practices to mitigate security risks

## **.6.SQL Injection**

## What is SQL injection (SQLi):

SQL injection (SQLi) is a web security vulnerability that allows an attacker to interfere with the queries that an application makes to its database. This can allow an attacker to view data that they are not normally able to retrieve. This might include data that belongs to other users, or any other data that the application can access. In many cases, an attacker can modify or delete this data, causing persistent changes to the application's content or behavior. In some situations, an attacker can escalate a SQL injection attack to compromise the underlying server or other back-end infrastructure. It can also enable them to perform denial-of-service attacks.

## SQL Injection (SQLi):

- **Description:** Occurs when an attacker is able to manipulate SQL queries by injecting malicious input into them.

## Some of the common types of SQL Injection:

### 1. Classic SQL Injection:

- If a login form directly includes user input as a SQL query:

```
SELECT * FROM users WHERE username = 'admin' AND password = 'password';
```

### 2. Blind SQL injection:

When an application doesn't return error messages but behavior changes can be inferred. An attacker might use conditional statements to extract data.

```
SELECT * FROM users WHERE id = 1 AND (SELECT SUBSTRING(password, 1, 1)
FROM users WHERE username='admin') = 'a';
```

### 3.Error-Based SQL Injection:

Uses database error messages to gather information about the database structure.

```
SELECT * FROM users WHERE id = 1' AND 1=CONVERT(int, (SELECT
@@version)); --'
```

## 7.Cross Site Scripting(XSS)

Cross-Site Scripting (XSS) attacks are a type of injection, in which malicious scripts are injected into otherwise benign and trusted websites. XSS attacks occur when an attacker uses a web application to send malicious code, generally in the form of a browser side script, to a different end user. Flaws that allow these attacks to succeed are quite widespread and occur anywhere a web application uses input from a user within the output it generates without validating or encoding it. An attacker can use XSS to send a malicious script to an unsuspecting user. The end user's browser has no way to know that the script should not be trusted, and will execute the script. Because it thinks the script came from a trusted source, the malicious script can access

any cookies, session tokens, or other sensitive information retained by the browser and used with that site. These scripts can even rewrite the content of the HTML page.

Types of Cross site-scripting:

**1. Stored XSS (Persistent XSS):**

o **Description:** Malicious script is permanently stored on the target server,

such as in a database, a message forum, or a comment field.

o **Example:** An attacker posts a malicious script in a forum comment, which is then viewed by any user accessing that comment.

**2. Reflected XSS (Non-Persistent XSS):**

o **Description:** Malicious script is reflected off a web server, such as in an error message, search result, or any other response that includes input sent to the server.

o **Example:** An attacker sends a victim a link with a malicious script embedded in the URL. When the victim clicks the link, the script is executed in their browser.

**3. DOM-Based XSS:**

- o **Description:** The vulnerability exists in client-side scripts that modify the DOM (Document Object Model) based on user input.

- o **Example:** An attacker manipulates a client-side script to execute arbitrary code by altering the DOM environment in the victim's browser.

## How XSS Attacks Work

### **1. Injection:**

- o The attacker injects malicious code into a web application. This can be through input fields, URL parameters, or other means of sending data to the application.

### **2. Execution:**

- o The injected code is executed in the context of the user's browser when they view the affected page.

### **3. Impact:**

- o The malicious script can perform any actions that the user can perform, access any data that the user can access, and manipulate the content displayed to the user.

## Consequences of XSS Attacks

- **Session Hijacking:** Stealing session cookies to impersonate the user.
- **Credential Theft:** Stealing user credentials or other sensitive information.
- **Defacement:** Altering the content of the website.
- **Phishing:** Redirecting users to malicious sites that mimic the appearance of legitimate ones.
- **Malware Distribution:** Delivering malicious payloads to users' browsers.

## 8. Firewall

A firewall is a network security device or software that monitors and controls incoming and outgoing network traffic based on predetermined security rules. Its primary purpose is to establish a barrier between a trusted internal network and untrusted external networks, such as the internet, to block malicious traffic and prevent unauthorized access. Here's a detailed look at firewalls, including their types, functions, and configurations.

### Types of Firewalls

#### 1. Packet-Filtering Firewalls:

- o **Description:** These firewalls operate at the network layer and make decisions based on the source and destination IP addresses, ports, and protocols.

- o **Advantages:** Fast and efficient.
- o **Disadvantages:** Limited to basic filtering; cannot inspect the payload of packets.

## 2. Stateful Inspection Firewalls:

- o **Description:** These firewalls track the state of active connections and make decisions based on the context of the traffic.
- o **Advantages:** More secure than packet-filtering firewalls as they understand the state of network connections.
- o **Disadvantages:** More resource-intensive.

## 3. Proxy Firewalls (Application-Level Gateways):

- o **Description:** These firewalls act as an intermediary between end users and the internet, making requests on behalf of users.
- o **Advantages:** Can inspect the entire application layer, providing a high level of security.
- o **Disadvantages:** Can be slower and more complex to configure.

## 4. Next-Generation Firewalls (NGFW):

- o **Description:** These firewalls combine traditional firewall capabilities with advanced features like deep packet inspection, intrusion prevention systems (IPS), and application awareness.
- o **Advantages:** Comprehensive security, capable of detecting and blocking complex attacks.
- o **Disadvantages:** Can be expensive and require significant resources.

## Functions of Firewalls

- **Traffic Filtering:** Blocking unauthorized access while permitting authorized communications.
- **Monitoring and Logging:** Keeping records of network activity for security analysis and compliance.
- **Intrusion Detection and Prevention:** Identifying and preventing potential threats in real-time.
- **VPN Support:** Securing remote access via Virtual Private Networks.
- **Content Filtering:** Controlling access to websites and content based on policies. *Firewall allowing Good Traffic 8990 Firewall Blocking bad traffic*

## Advantages of using Firewalls:

• Firewalls play an important role in the companies for security management. Below are some of the important advantages of using firewalls.

• It provides enhanced security and privacy from vulnerable services. It prevents unauthorized users from accessing a private network that is connected to the internet.

• Firewalls provide faster response time and can handle more traffic loads.

• A firewall allows you to easily handle and update the security protocols from a single authorized device.

• It safeguards your network from phishing attacks

## 9. Vulnerability Scanner

A vulnerability scanner is a tool designed to identify and assess vulnerabilities in computer systems, networks, and applications. These scanners perform automated scans to detect security weaknesses that could be exploited by attackers, such as outdated software, misconfigurations, and missing patches.

### Acunetix:

Acunetix is a comprehensive web vulnerability scanner designed to identify security

weaknesses in web applications and services. It automates the process of detecting and reporting on a wide range of vulnerabilities, including SQL injection, cross-site scripting (XSS), and other security issues specific to web applications. Acunetix is widely used by security professionals to ensure web applications are secure and comply with industry standards.

## Benefits of Using Acunetix

- **Comprehensive Coverage:** Acunetix provides extensive vulnerability detection for a wide range of web application vulnerabilities.
- **Ease of Use:** User-friendly interface and automation capabilities make it accessible for both security professionals and developers.
- **Accurate Results:** Advanced scanning technology reduces false positives and ensures accurate vulnerability detection.
- **Integration:** Seamlessly integrates with development and DevOps tools, enhancing workflow and collaboration.
- **Compliance:** Helps organizations maintain compliance with various security standards and regulations.

## **Chapter 3 Conclusion**

**1. Computer Vision:**

- Techniques and applications for enabling machines to interpret and process visual information.
- Understanding of image processing techniques.
- Development and implementation of vision-based solutions.

**2. Convolutional Neural Networks (CNN):**

- Proficiency in building and training CNN models.
- Knowledge of CNN architecture and applications in image recognition and classification tasks.

**3. Image Classification:**

- Experience using Google Teachable Machine for image classification.
- Understanding the workflow from image collection to model training and evaluation.
- Skills in categorizing and labeling images based on specific rules.

**4. Image Object Detection:**

- Ability to develop object detection models.
- Knowledge of algorithms such as YOLO, SSD, and Faster R-CNN.
- Practical applications of object detection in various domains.

**5. YOLO (You Only Look Once) - Object Detection:**

- Proficiency in using YOLO for real-time object detection.
- Experience with domain-specific datasets in medical, agriculture, drones, and traffic.
- Integration of YOLO models in real-world applications.

**6. Medical Image Analysis and Labelling:**

- Skills in using Roboflow for image labeling.
- Understanding the importance of accurate labeling in medical image analysis.
- Proficiency in developing AI models for medical applications.

**7. Human Pose Estimation:**

- Experience using Google Teachable Machine for human pose estimation.
- Understanding techniques for detecting and tracking human figures and their poses in images or videos.

**8. Mediapipe Studio:**

- Knowledge of building multimodal applied machine learning pipelines.
- Experience using Mediapipe Studio for hand gesture recognition and other applications.

**9. OpenCV Basics:**

- Understanding fundamental concepts and functionalities of OpenCV.
- Practical skills in using OpenCV for various computer vision tasks.

**11. Google Dialogflow:**

- Proficiency in using Google Dialogflow for natural language understanding.
- Skills in developing and deploying conversational agents.

**12. Generative AI:**

- Techniques for generating new content such as music, text, and images.
- Experience with models for music generation, text generation, and image generation.

**13. AI Models:**

- Knowledge of various AI models used for different applications.
- Skills in summarization, fill-mask models, and transformers.

**14. Visual Question & Answering:**

- Development of models that answer questions about images.
- Integration of visual and textual data for question answering.

**17. Large Language Models (LLMs):**

- Knowledge of advanced language models like Claude, GPT, Gemini, LLaMA3, and Open LLMs.
- Experience in text generation and language understanding.

**18. Other Topics:**

- Implementation of Google's Vision API for image analysis.
- Understanding and using small language models (SLMs) like BERT and GPT.
- Skills in deploying and managing AI models using Ultralytics Hub.
- Development of lightweight models for mobile and embedded devices using TensorFlow Lite.
- Proficiency in sentiment analysis and creating deepfakes.

**Cyber Security Skills Acquired**

**1. Cyber Security Basics:**

- Fundamental principles and practices for protecting computer systems and networks from cyber threats.

**2. Types of Cyber Crimes:**

- Understanding various forms of illegal activities conducted via the internet.

**3. CIA Triad:**

- Core principles of cybersecurity—Confidentiality, Integrity, and Availability.

**4. AAA Framework:**

- Knowledge of Authentication, Authorization, and Accounting framework for managing and securing identities and their access.

**5. OWASP:**

- Familiarity with the Open Web Application Security Project and its focus on improving software security.

**6. SQL Injection:**

- Understanding of SQL injection techniques and prevention methods.

**8. Firewall:**

- Knowledge of network security systems that monitor and control incoming and outgoing network traffic based on predetermined security rules.

**9. Vulnerability Scanner:**

- Proficiency in using tools like Acunetix for identifying and addressing vulnerabilities in systems and application