AI-Based Virtual Gym Assistant for Exercise Guidance and Injury Risk Prevention

BCSE306L – ARTIFICIAL INTELLIGENCE

SLOT – C1 + TC1

Tamilarasi K

REVIEW-2

Santhosh Kumar 21BCE1829

Dennis Asir 21BCE1990

Inban Azhagiri 21BCE5802

Nikhil S. Shinde 21BCE1743

1. Complete Proposed Methodology :

I. Algorithms and Model used:

a. Mediapipe:

MediaPipe is an open-source framework developed by Google Research for building cross-platform applications that process and analyze multimedia data such as video and audio. It provides a comprehensive set of tools and pre-built components for tasks like real-time pose estimation, hand tracking, object detection, and facial recognition. One of its key features is its versatility, enabling developers to easily integrate machine learning models and pipelines into their applications for tasks ranging from simple feature extraction to complex multi-modal analysis. MediaPipe's modular architecture and efficient implementation make it suitable for a wide range of use cases, from mobile and embedded devices to desktop and cloud environments. With its growing popularity and active community support, MediaPipe continues to evolve as a powerful tool for multimedia processing and machine learning applications.

b. LLM(OpenAI):

ChatGPT, a prominent example of Large Language Models (LLMs), epitomizes the transformative potential of these models in natural language understanding and generation. Trained on vast text corpora, ChatGPT harnesses sophisticated machine learning techniques to engage in human-like conversations, comprehend context, and generate coherent responses across a myriad of topics. LLMs, including ChatGPT, leverage their extensive knowledge and linguistic prowess to excel in a wide array of tasks, from text generation to sentiment analysis, revolutionizing human-computer interaction and paving the way for innovative applications across industries. However, their deployment also raises ethical considerations regarding biases and societal impact, underscoring the importance of responsible AI development and deployment. Nonetheless, the remarkable capabilities of ChatGPT and LLMs herald a new era in natural language processing, promising profound advancements in communication, knowledge dissemination, and user interaction.

II. Architecture

III. Equations

Calculating Angles:

The function calculateAngle(a, b, c) calculates the angle formed by the vectors ba and bc where a, b, and c are points in a 2D plane. This calculation is based on the mathematical representation of the angle between two vectors using the arctangent function and trigonometric properties.

Let a = (xa, ya), b = (xb, yb), and c = (xc, yc) be the coordinates of points a, b, and c respectively.

The vectors ba and bc are represented as follows:
ba = (xa - xb, ya - yb)
bc = (xc - xb, yc - yb)

Then, the angle between ba and bc can be calculated using the arctangent function as follows:
radians = atan2((yc - yb), (xc - xb)) - atan2((ya - yb), (xa - xb))

Finally, the angle in degrees is obtained by converting radians to degrees:
angle = abs(radians × 180.0 / π)

If the angle is greater than 180 degrees, it is adjusted to ensure it lies within the range of 0 to
180 degrees:
if angle > 180, then angle = 360 - angle

IV. Pseudocode

Initialize VideoCapture from camera 0
Initialize Pose object from mediapipe with min_detection_confidence and min_tracking_confidence

Set flag bd to False
Initialize start_time with current time

Loop until video capture is open:
    Read a frame from the video capture
    Convert frame to RGB color space
    Process the frame using the Pose object

```
    Convert the frame back to BGR color space

    Try:
        Extract landmarks from the pose results
        Calculate angles for various body parts using calculateAngle function
        Display angles on the frame
        Check if time since start_time is greater than 2 seconds
        If right elbow angle is greater than 110 and bd is False:
            Display "Bad posture in Right ARM" message
            Start a new thread to speak "Bad Posture Detected"
            Set bd to True
            Update start_time with current time
        If right elbow angle is less than 110 and bd is True:
            Set bd to False
        If left elbow angle is greater than 110:
            Display "Bad posture in Left ARM" message
    End Try

    Draw pose landmarks and connections on the frame
    Display the frame

    If 'q' key is pressed:
        Break out of the loop

Release the video capture and close all OpenCV windows
```

2. Results and Implementation :

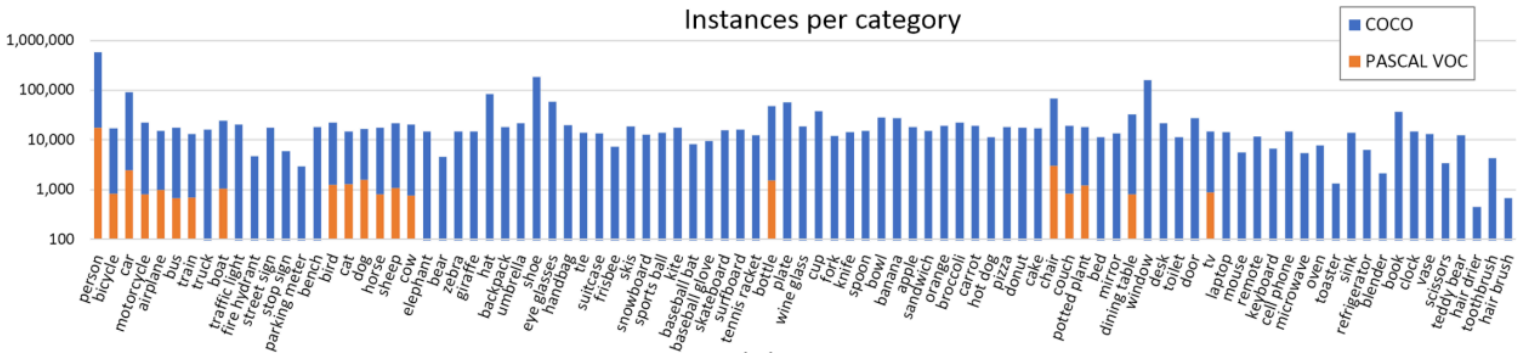
I. Dataset Used:

a. Mediapipe Dataset:

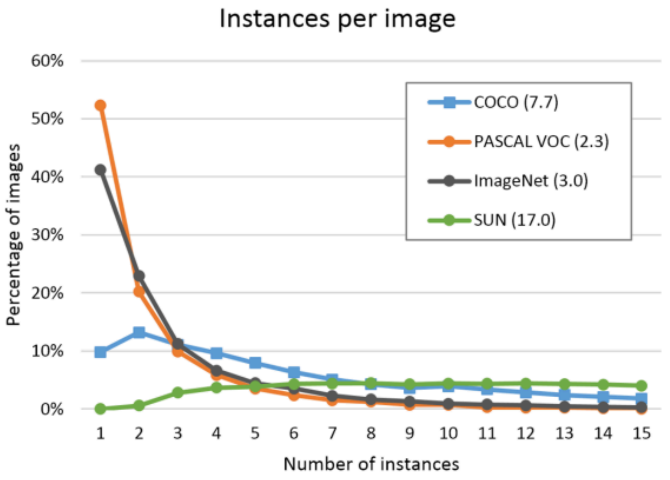
Coco:

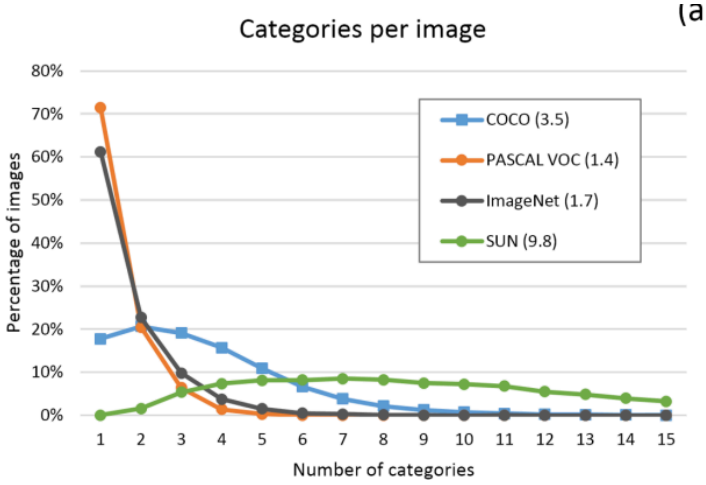
the COCO dataset, a groundbreaking resource designed to propel object recognition and scene understanding to new heights. By situating object recognition within the broader context of scene comprehension, COCO offers a collection of images capturing intricate everyday scenes, each brimming with common objects set in their natural environments. Through meticulous per-instance segmentations, precise object localization is facilitated. With an extensive catalog boasting 91 object types recognizable by a 4-year-old, COCO features a staggering 2.5 million labeled instances across 328,000 images. The dataset's creation involved innovative crowd worker engagement via intuitive interfaces for category detection, instance spotting, and instance segmentation. A comprehensive statistical analysis comparing COCO to benchmarks like PASCAL, ImageNet, and SUN underscores its significance. Additionally, baseline performance evaluations utilizing a Deformable Parts Model illuminate the dataset's potential for bounding box and segmentation detection advancements.

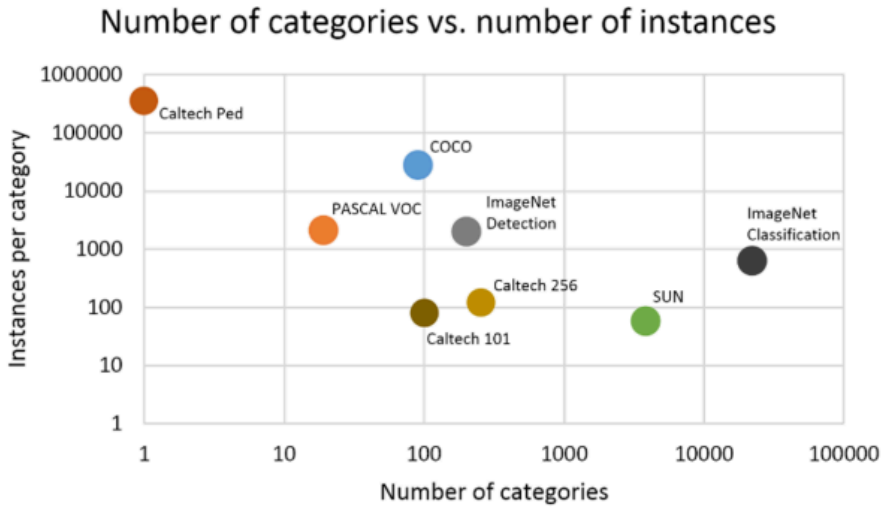
Graphs:

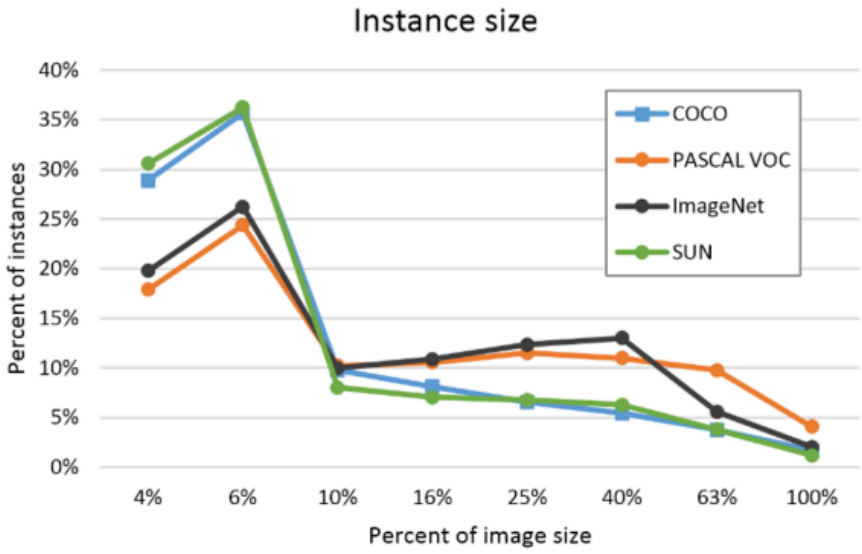


(a)


Instances per image
Percentage of images
Number of instances
COCO (7.7)
PASCAL VOC (2.3)
ImageNet (3.0)
SUN (17.0)


Categories per image
Percentage of images
Number of categories
COCO (3.5)
PASCAL VOC (1.4)
ImageNet (1.7)
SUN (9.8)


Number of categories vs. number of instances
Instances per category
Number of categories
Caltech Ped
COCO
PASCAL VOC
ImageNet Detection
ImageNet Classification
Caltech 256
SUN
Caltech 101

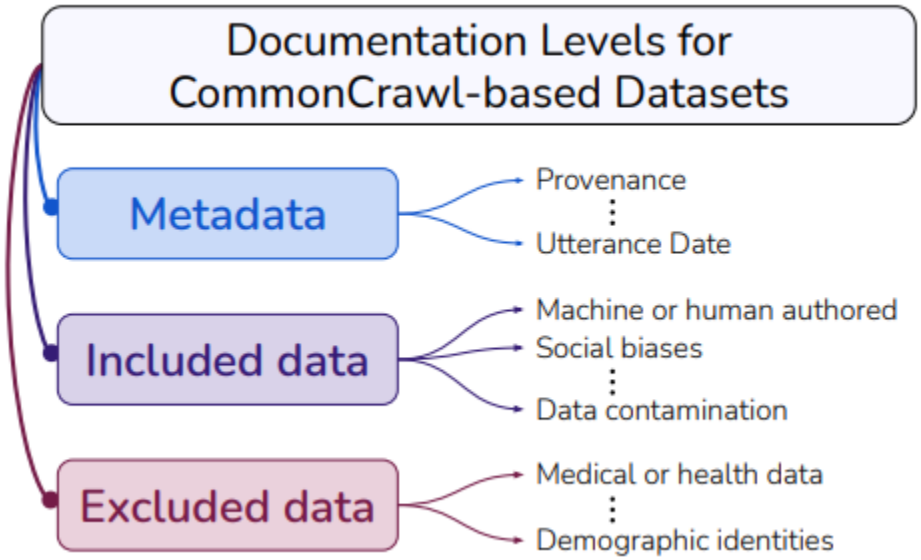

Instance size
Percent of instances
Percent of image size
COCO
PASCAL VOC
ImageNet
SUN

b. ChatGpt Dataset:

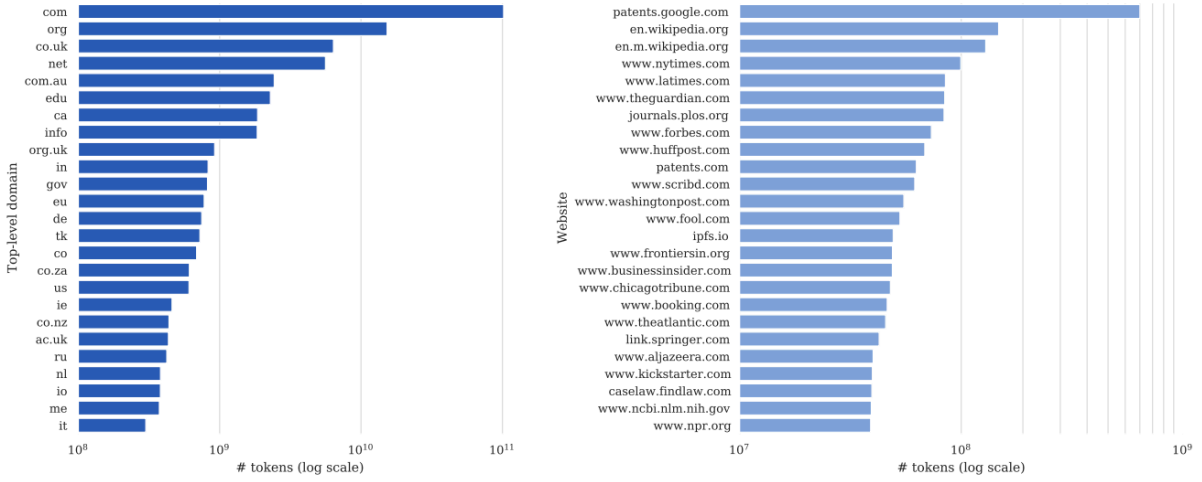
The Common Crawl dataset consists of web pages crawled from the internet. It includes a wide variety of content, such as text, images, and metadata, from different domains and languages. Common Crawl provides this dataset to the public for various purposes, including research, analysis, and machine learning

.

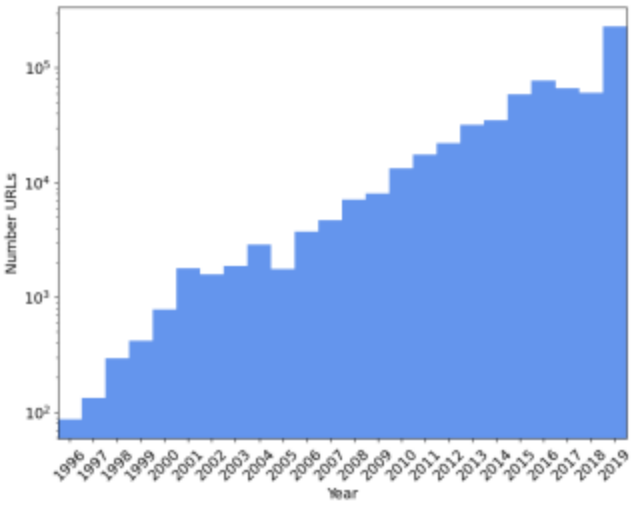
Graphs: :



(Left Figure)We advocate for three levels of documentation when creating web-crawled corpora. On the right, we include some example of types of documentation that we provide for the C4.EN dataset.



(Above Figure)Number of tokens from the 25 most represented top-level domains (left) and websites (right) in C4.EN



(Left Figure)The date URLs were first indexed by the Internet Archive13 before the Common Crawl snapshot was collected

	Dataset	% Matching
Label	LAMA T-REx	4.6
	LAMA Google-RE	5.7
	XSum	15.49
	TIFU-short	24.88
	TIFU-long	1.87
	WikiBio	3.72
	AMR-to-text	10.43
Input	BoolQ	2.4
	CoLA	14.4
	MNLI (*hypothesis*)	14.2
	MNLI (*premise*)	15.2
	MRPC (*sentence 1*)	2.7
	MRPC (*sentence 2*)	2.7
	QNLI (*sentence*)	53.6
	QNLI (*question*)	1.8
	RTE (*sentence 1*)	6.0
	RTE (*sentence 2*)	10.8
	SST-2	11.0
	STS-B (*sentence 1*)	18.3
	STS-B (*sentence 2*)	18.6
	WNLI (*sentence 1*)	4.8
	WNLI (*sentence 2*)	2.1

(Left Table)
The number of exact matches from test sets of various benchmarks in C4.EN. For datasets where the input has multiple components (e.g. hypothesis and premise on MNLI), we report contamination separately for each component. Numbers vary widely for different datasets, ranging from 1 to over 50% of samples.

c. Books used for training LLM regarding Fitness:

1) THE COMPLETE FITNESS
https://www.clemson.edu/business/departments/army-rotc/documents/fitness-handbook.pdf

2) HEALTH AND FITNESS
https://www.uakron.edu/armyrotc/MS1/15.pdf

II. Introduction

MediaPipe is a versatile open-source framework for building multimodal applied machine learning pipelines. In this script, we utilize MediaPipe's Pose Estimation model to detect human poses in real-time, providing invaluable insights into posture dynamics. With its ability to identify 33 landmarks on the human body, this model furnishes a detailed understanding of a person's posture in each frame of the video, enabling real-time analysis and feedback.

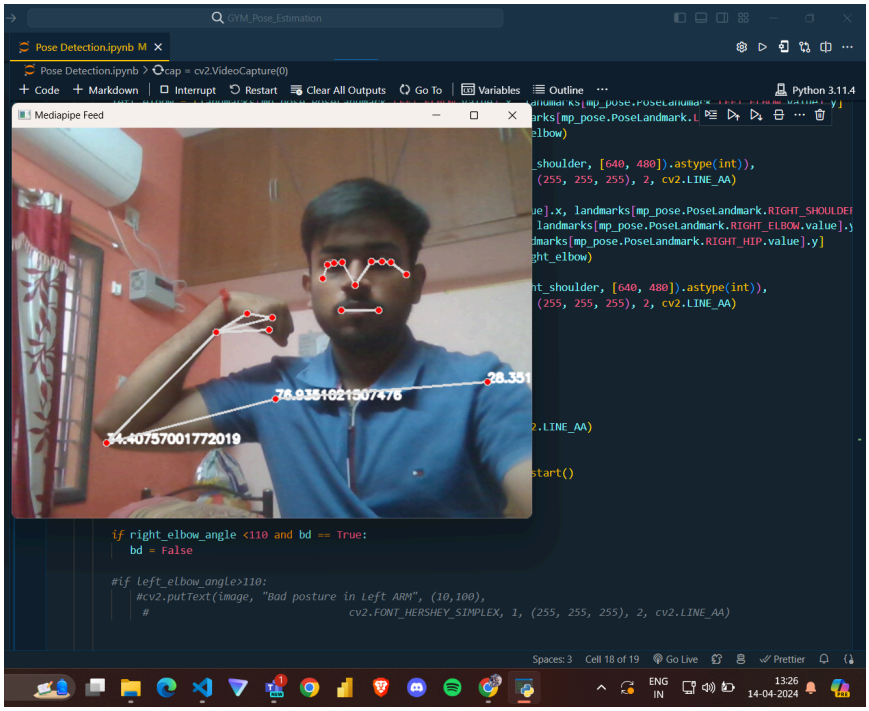
The extracted landmarks serve as inputs to calculate angles at critical joints like elbows and shoulders. If these angles surpass predetermined thresholds, it signals potential posture issues. To promptly notify users about their posture, we employ the pyttsx3 library, which converts text to speech, ensuring immediate feedback and intervention.

To further elevate the efficacy of this system, integrating the ChatGPT API presents a promising avenue. Developed by OpenAI, ChatGPT is renowned for its capability to generate human-like text responses based on given inputs. By feeding the posture data into ChatGPT, we can derive personalized suggestions tailored to the user's specific needs, enhancing their posture correction experience.

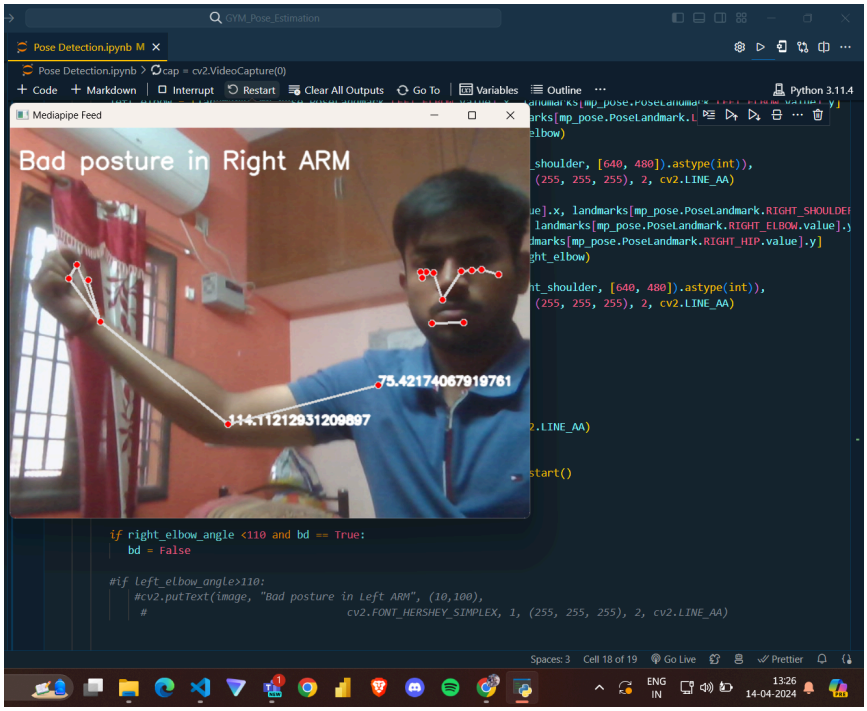
For instance, if the user consistently exhibits a high elbow angle, ChatGPT could provide targeted exercises to improve flexibility or remind the user to take regular breaks from their desk. This transformative integration elevates the system beyond mere posture detection, evolving it into a comprehensive tool for fostering better ergonomics and overall health in both professional and home environments.

III. Implementation:

Good Posture for Bicep Curl:



Bad Posture for Bicep Curl:



IV. Cite some papers

1) Lugaresi, Camillo & Tang, Jiuqiang & Nash, Hadon & McClanahan, Chris & Uboweja, Esha & Hays, Michael & Zhang, Fan & Chang, Chuo-Ling & Yong, Ming & Lee, Juhyun & Chang, Wan-Teh & Hua, Wei & Georg, Manfred & Grundmann, Matthias. (2019). MediaPipe: A Framework for Building Perception Pipelines.
2) Lin, TY. *et al.* (2014). Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48
3) Dodge, Jesse et al. "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus." *Conference on Empirical Methods in Natural Language Processing* (2021).