# General AI/ML

Unit 2: Optimization,
Fine-tuning, Transfer Learning

# 2.2.2

# Pre-trained Models

Introduction to common pre-trained models

TIL-AI
TODAY I LEARNED-AI

# Pre-Trained Models for NLP

# BERT (Bidirectional Encoder Representations from Transformers)

- Developed by Google and trained on a large dataset of unannotated text

- A transformer-based model, that uses self-attention mechanisms to process input text.

- Processes the textual sequence bidirectionally, considering both left and right context simultaneously for each word. This bidirectional context is used understand the meaning of each word in its surrounding context.

TIL-AI
TODAY I LEARNED-AI

# GPT-4 (Generative Pre-trained Transformer 4)

- Developed by OpenAI and trained on trained on ~13T tokens, including both text-based and code-based data
- Exceptional at understanding and generating human-like text based on the context it's provided.
- Powers advanced conversational AI, creative content generation, language translation, and more, demonstrating unparalleled versatility in text generation and understanding.
- **Vision Multi-Modal:** GPT-4 includes a vision encoder for autonomous agents to read web pages and transcribe images and videos.

TIL-AI
TODAY I LEARNED-AI

# ELMO (Embeddings from Language Models)

- A deep contextualized word representation model developed by researchers at the Allen Institute for Artificial Intelligence
- Trained on a large dataset of unannotated text and can be fine-tuned for a wide range of natural language processing (NLP) tasks
- ELMo word vectors are generated through a two-layer bidirectional language model, featuring both forward and backward passes in each layer
- Has the capability to produce distinct embeddings for the same word deployed in diverse contexts across different sentences

TIL-AI
TODAY I LEARNED-AI

# Transformer-XL

- Language representation model developed by researchers at Carnegie Mellon University and Google Brain
- Expansion of the original Transformer model to better handle long-term dependencies
- Uses a segment-level recurrence mechanism and a novel positional encoding scheme

# RoBERTa

- A variant of BERT developed by Facebook AI
- Iterates on BERT's approach with optimizations improvements that have led it to outperform BERT on a wide range of benchmarks
- Also trained on a larger dataset and fine-tuned on a variety of natural language processing (NLP) tasks, making it a more powerful language representation model than BERT

# Pre-Trained Models for Computer Vision

TIL-AI
TODAY I LEARNED AI

# MobileNetV2

- Computer vision model open-sourced by Google
- Uses depthwise convolutions to significantly reduce the number of parameters compared to other networks, resulting in a lightweight deep neural network.
- Powers real-time object detection and classification on mobile devices, enabling advanced vision capabilities in resource-constrained environments.

TIL-AI
TODAY I LEARNED AI

# Vision Transformer

- Applies the principles of the Transformer model to the domain of computer vision
- Treats image patches as sequences, similar to words in a sentence, allowing it to learn contextual relationships between different parts of an image
- Used in image segmentation, object detection, and areas requiring detailed image analysis.

TIL-AI
TODAY I LEARNED AI

# YOLO (You Only Look Once)

- A real-time object detection system
- Processes images in a single evaluation, making it extremely fast while maintaining high accuracy
- Utilizes a single neural network to predict multiple bounding boxes and class probabilities for those boxes
- Widely used in surveillance systems, autonomous vehicles, and any application requiring real-time detection and classification of objects.

# Pre-Trained Models for Automatic Speech Recognition

# Whisper

- A robust, multilingual speech recognition system developed by OpenAI.
- Trained on a diverse dataset of spoken language from the internet, enabling it to recognize speech accurately across different contexts.
- Exceptional at understanding speech in various languages and accents, robust to background noise

TIL-AI
TODAY I LEARNED-AI

# Wave2Vec2.0

- A self-supervised learning framework for speech recognition developed by Facebook AI.
- Utilizes raw audio waves to learn speech representations without needing labeled data.
- Employs a contrastive task that predicts the current audio frame from past context, significantly reducing the reliance on annotated resources.

TIL-AI
TODAY I LEARNED-AI

# DeepSpeech

- Open-source speech recognition engine developed by Mozilla trained by machine learning techniques based on Baidu's Deep Speech research paper
- Leverages a sophisticated deep learning architecture, incorporating recurrent neural networks (RNNs) with connectionist temporal classification (CTC) for efficient decoding of audio streams into text
- Processes audio in real-time with excellent accuracy and is designed for high-performance speech-to-text conversion

TIL-AI
TODAY I LEARNED AI