**What is Transfer Learning and Finetuning?**

* "Transfer learning is when a model uses what it learned from one task to do a different but similar task."
* "Transfer learning means taking experience from one problem and using it to solve another related problem."
* "It’s reusing what a model already knows to solve a new but similar problem, without starting from scratch."

**Real Life Example:**  
*"It’s like learning to ride a bicycle and then using that skill to learn how to ride a motorcycle faster."*

Computer Scientist took this simple idea and integrated it into deep learning.

Now suppose your specific class is not included among the 1000 classes available in the ImageNet dataset:

* Cat
* Dog
* Vehicle
* Aircraft
* Bird
* Fruit
* Vehicle
* Sports Object
* Musical Instrument
* Human

**Now the ways of doing this Transfer Learning:**

* **Replacing the output layer of pretrain model**
* Or don’t change weights means keep freezing all the weights just retrain the Neural Network Weights
* Or unfreezing some layers of the network and **fine-tuning it** on the new, task-specific dataset

**ILSVRC (ImageNet Large Scale Visual Recognition Challenge) version:**

* Training Images: ~1.2 million
* Validation Images: 50,000
* Test Images: 100,000 (labels hidden, for competition)
* Classes: 1000 (dog, cat, car, airplane, etc.)

**Full ImageNet Dataset (original, maintained by Stanford):**

* 14+ million images
* 21,841 categories (as per WordNet synsets)

**VGG16 MODEL ARCHITECTURE**  
*(Image from screenshot)*

**VGG16 MODEL ARCHITECTURE**  
*(Image from screenshot)*

We train a model on a huge dataset initially, allowing it to **learn primitive features** (like edges, textures, and simple shapes).

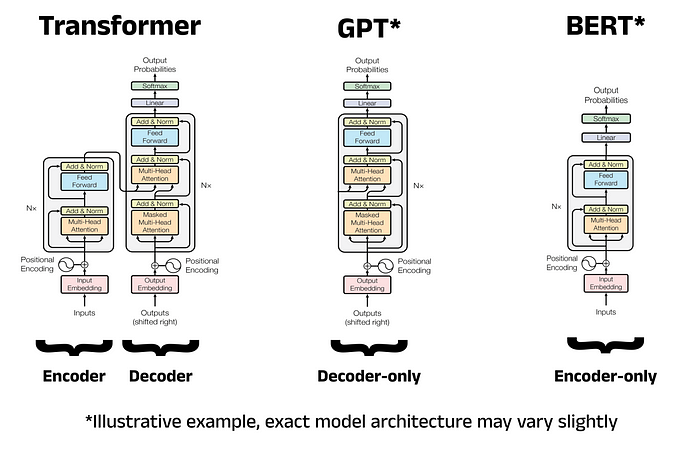
When applying it to a **different dataset**, we don’t retrain the entire model — we only retrain certain layers if the new data is very different.  
Otherwise, we can directly reuse the model’s knowledge.

**Why Do We Use Transfer Learning?**

* Saves training time and resources
* Works well when labeled data is limited
* Gives better performance than training from scratch

**Architectures**

**Transformer vs GPT vs BERT**



**Transformers’ Architecture vs LlaMA Architecture**  
