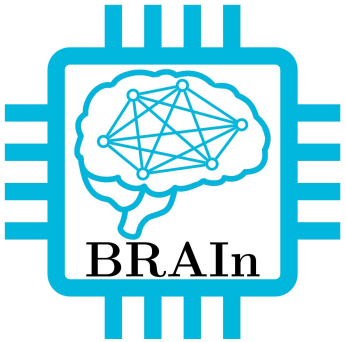
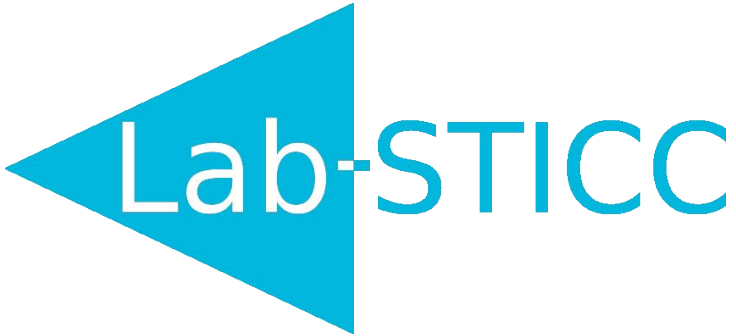
Tutorial day: Foundation Models

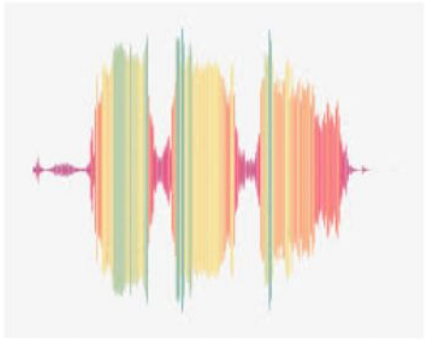
Équipe BRAIn

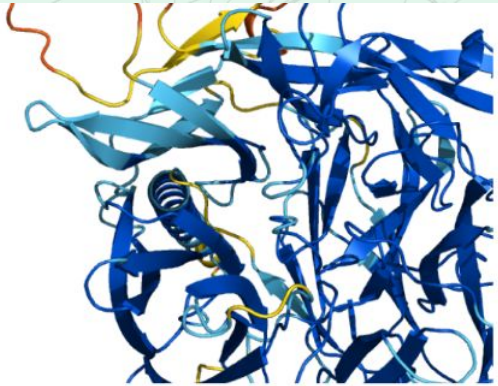


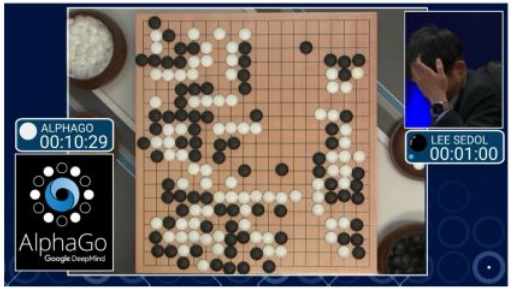
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom



About Deep Learning





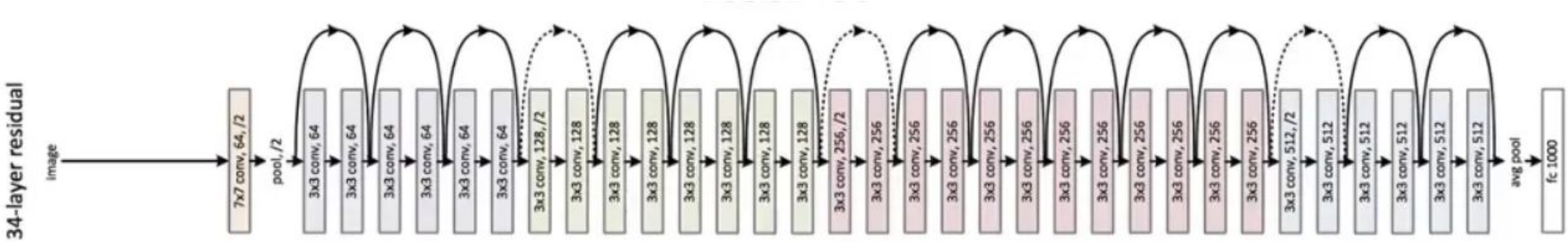


Your AI pair programmer

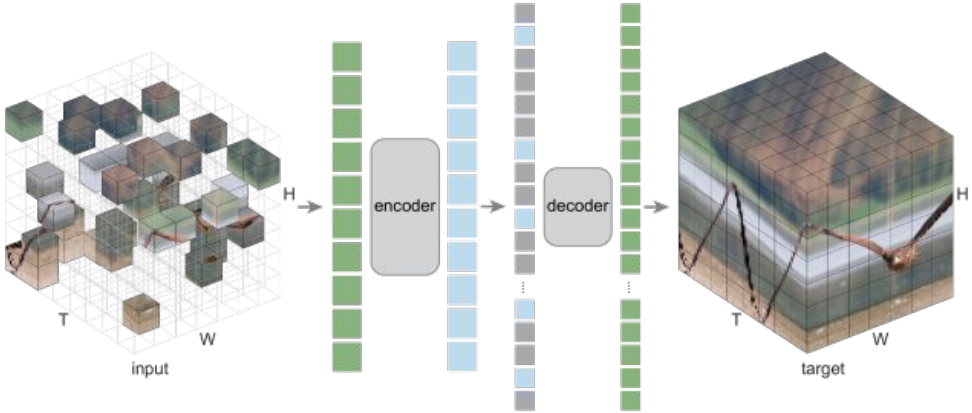
With GitHub Copilot, get suggestions for whole lines or entire functions right inside your editor.

Deep Learning

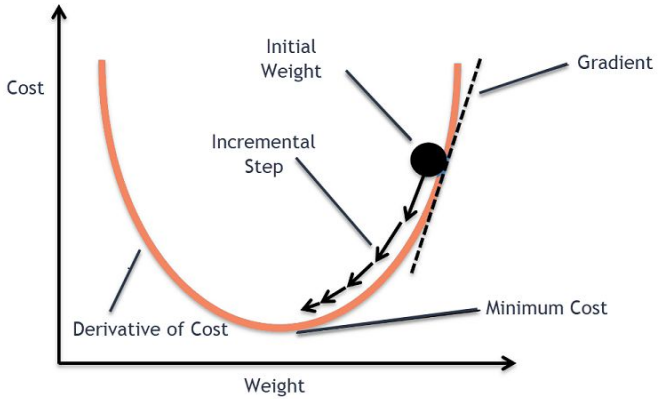
- Compositional approach:



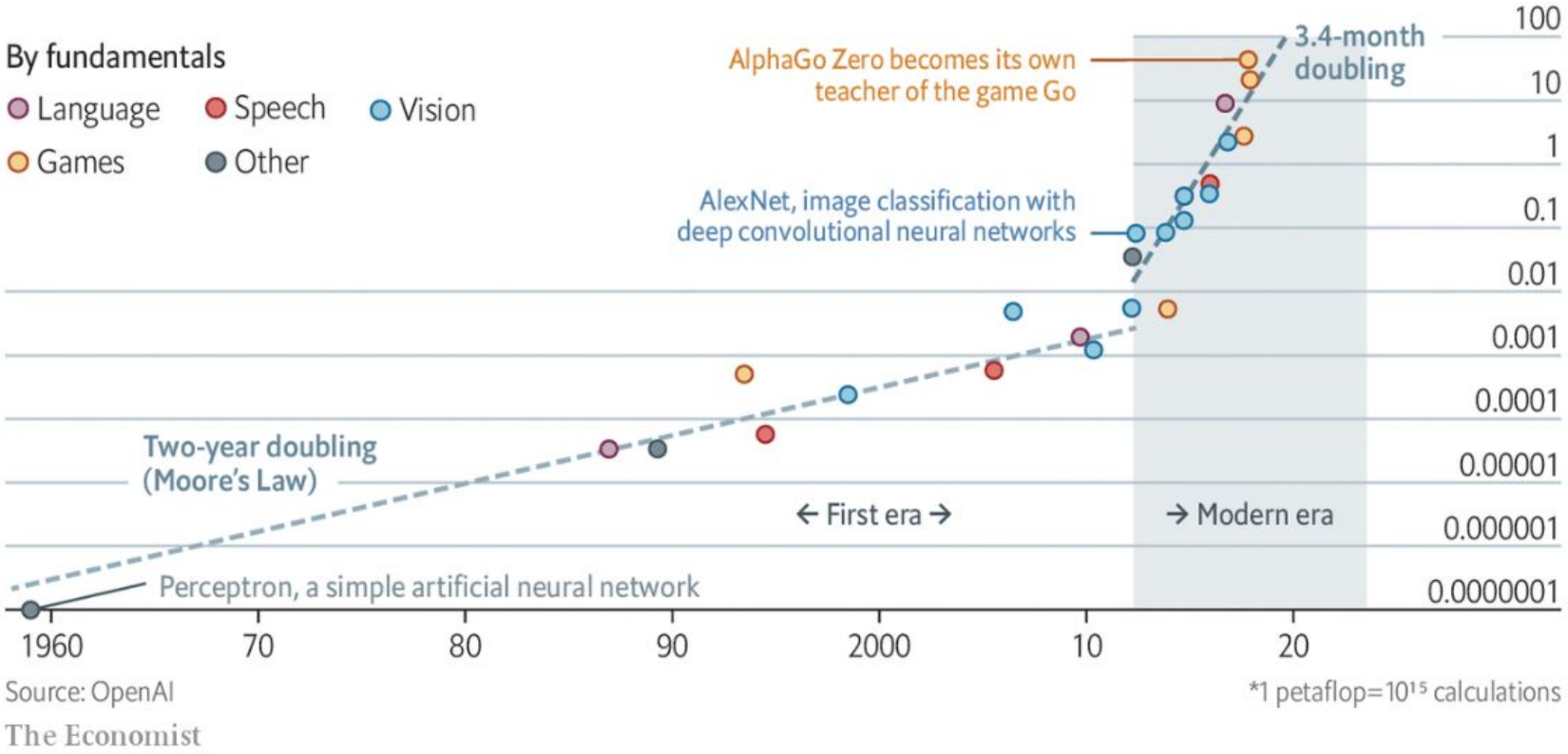
- End to end learning:



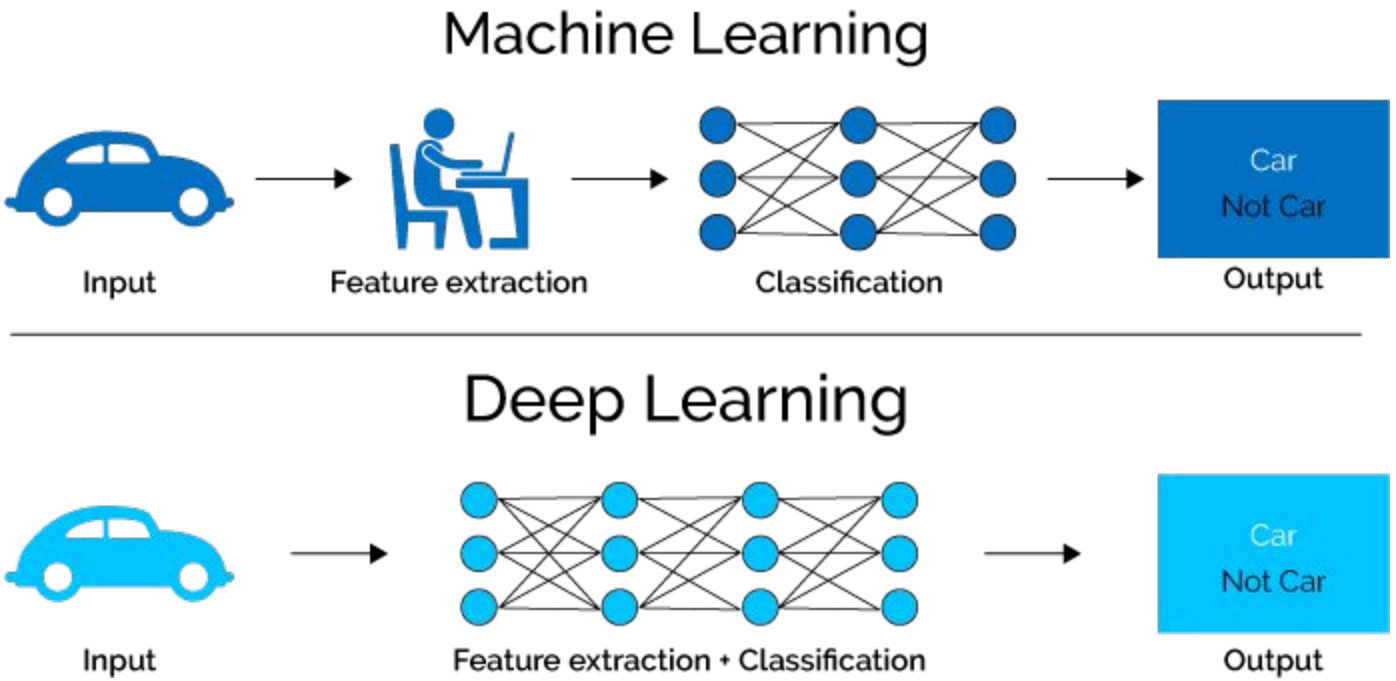
- Differential programming:



Why it works?

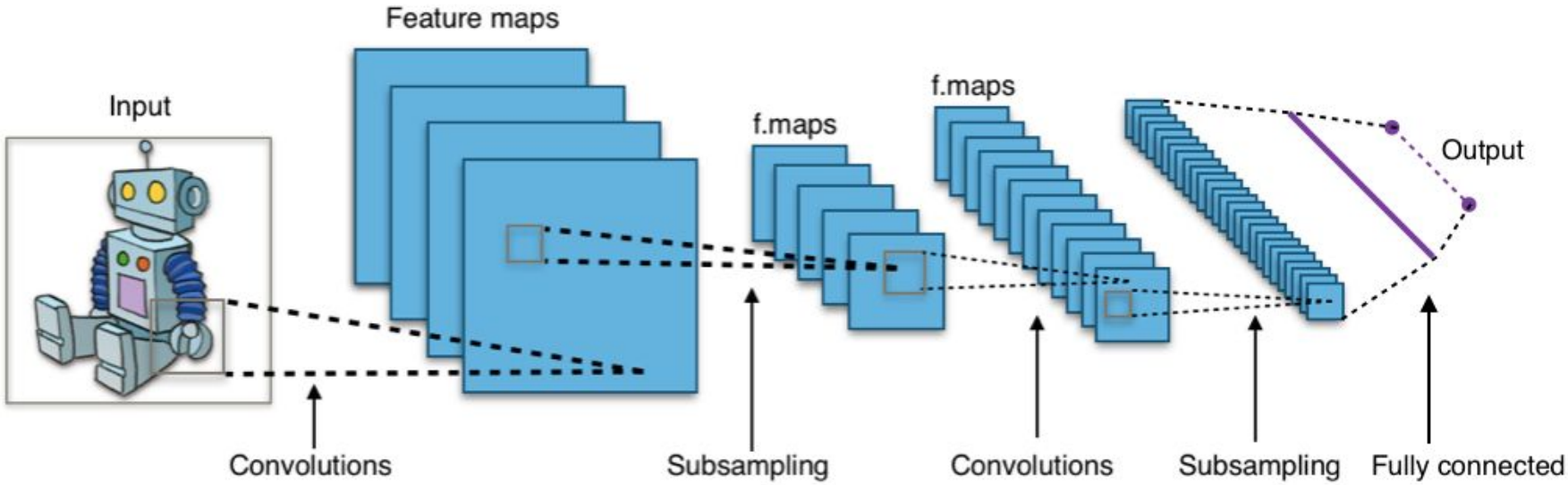

By fundamentals
Language
Speech
Vision
Games
Other
AlphaGo Zero becomes its own teacher of the game Go
3.4-month doubling
AlexNet, image classification with deep convolutional neural networks
Two-year doubling (Moore's Law)
← First era →
→ Modern era
Perceptron, a simple artificial neural network
100
10
1
0.1
0.01
0.001
0.0001
0.00001
0.000001
0.0000001
1960
70
80
90
2000
10
20
Source: OpenAI
*1 petaflop=10¹⁵ calculations
The Economist

From handcrafted features to handcrafted hyperparameters

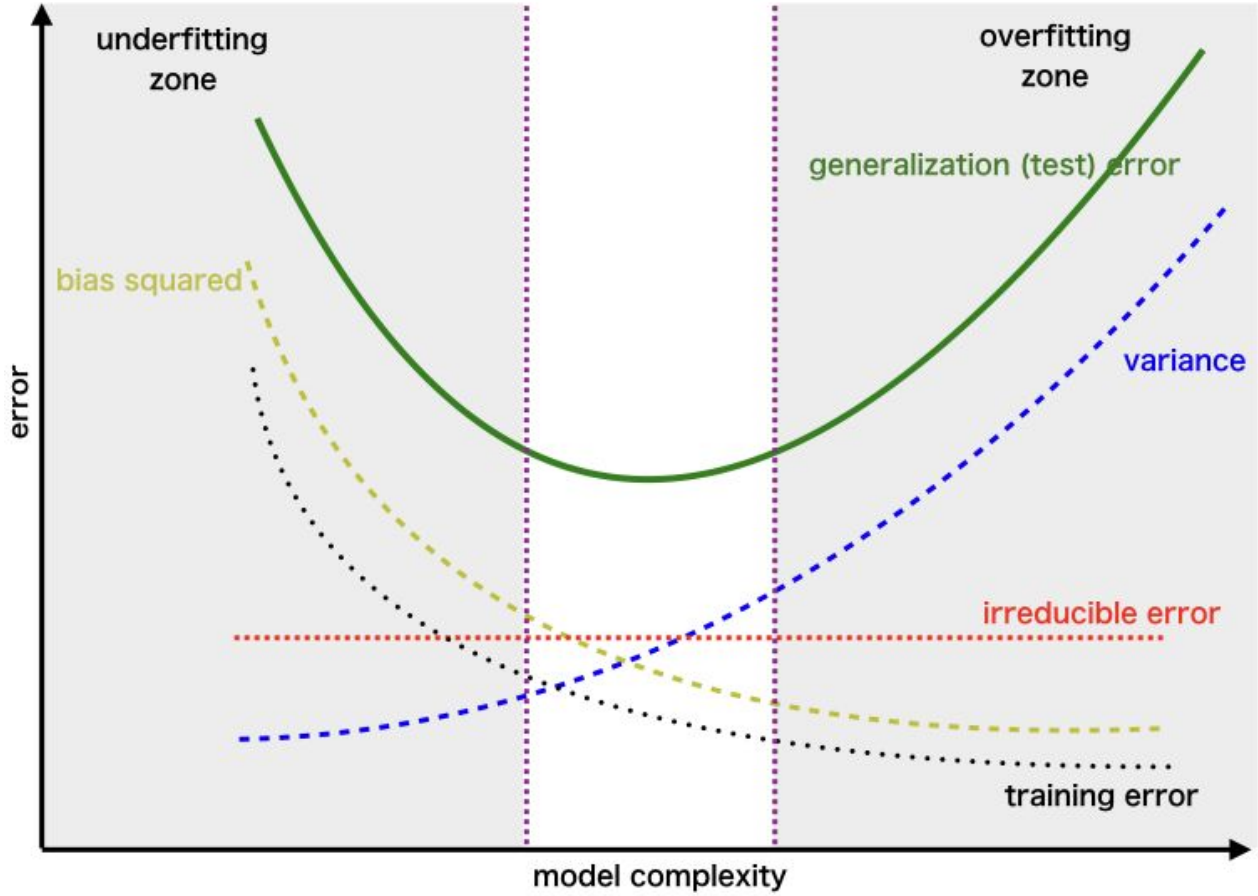


Transformers

About the inductive bias



About the inductive bias



Transformers

- No inductive bias
- Best generalization in many domains:
  - On par with Convolutions for image
  - SOTA for natural language processing
- Few hyperparameters
- Require large datasets to train

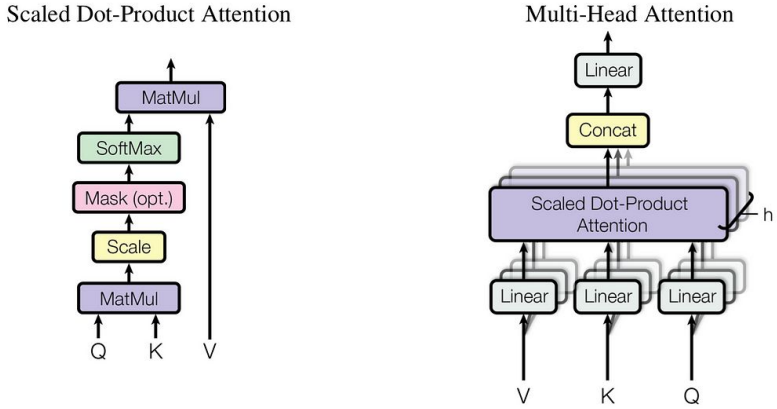
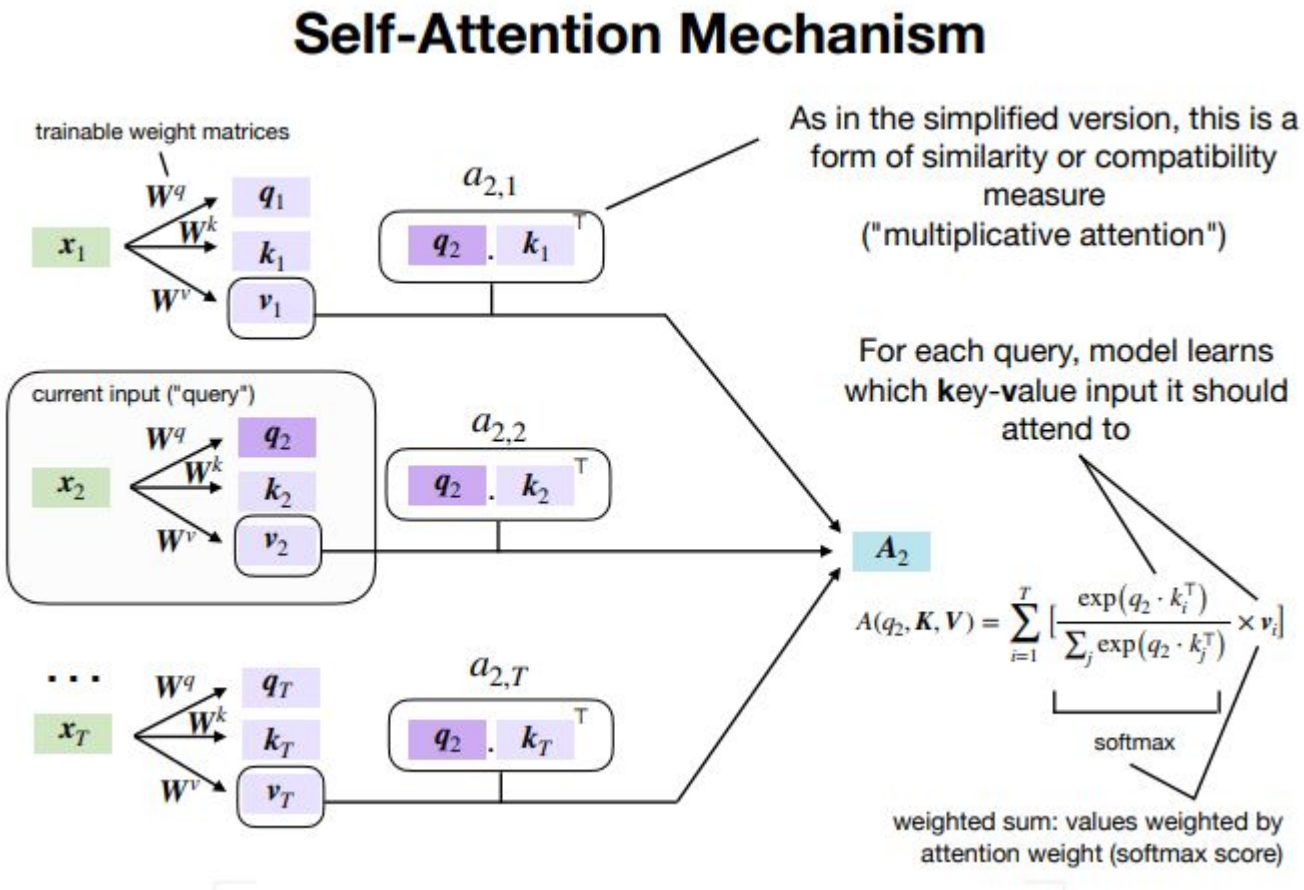


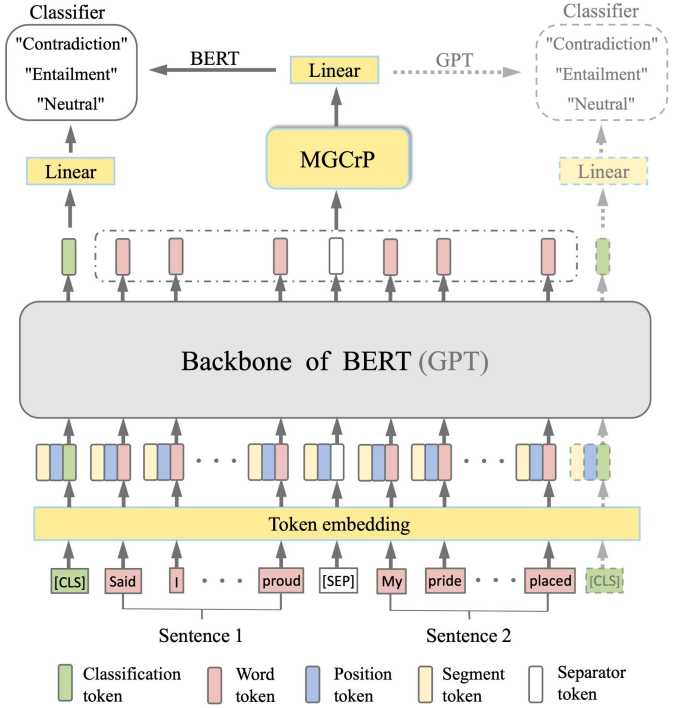
Figure 2: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel.

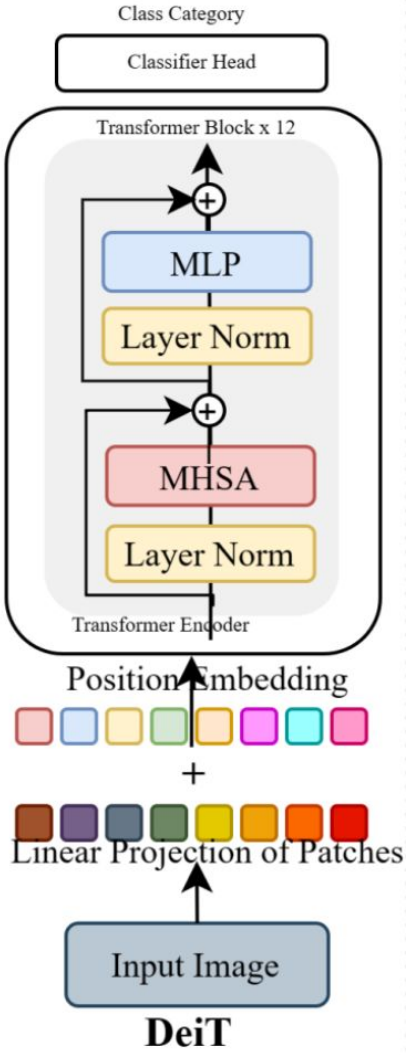
Self-attention



Transformer architectures

- Hyperparameters:
  - Inner dim
  - Number of heads
  - Number of Transformer blocks
- Position embedding helps incorporate a notion of structure
- Inputs are fixed in size

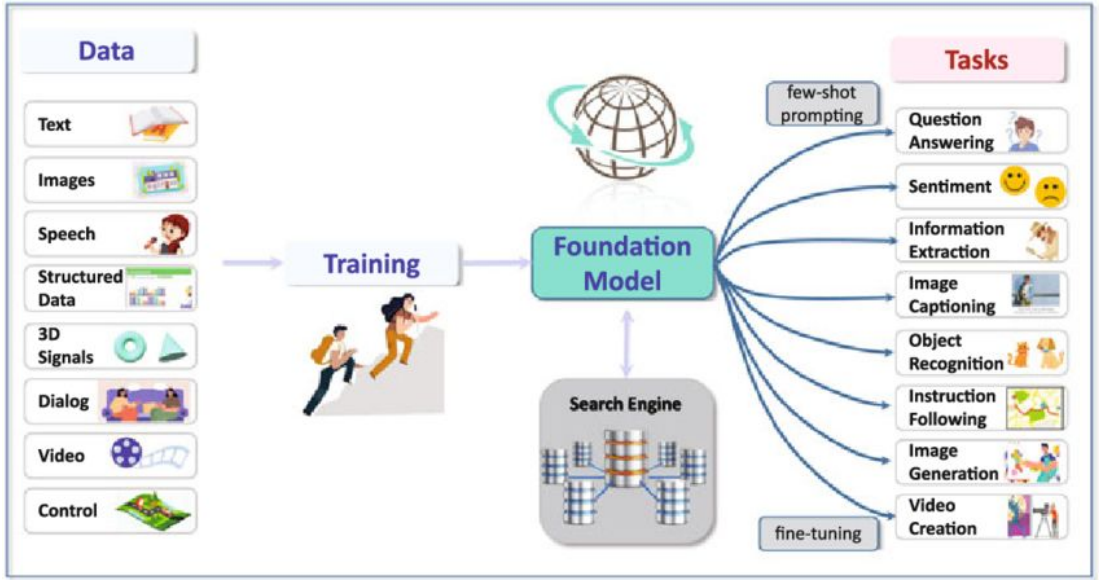




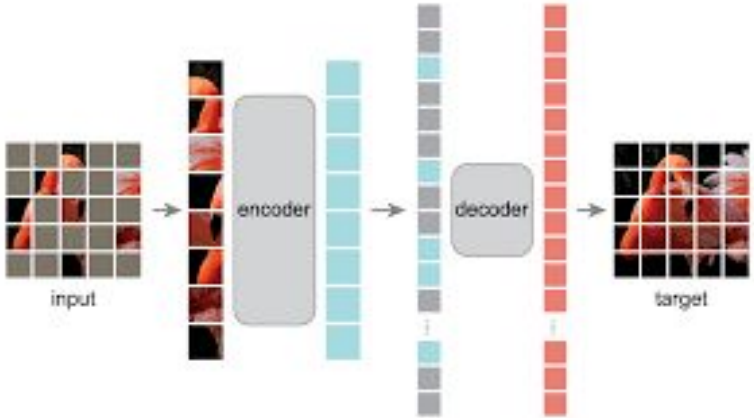
From generalization to particularization: Foundation Models

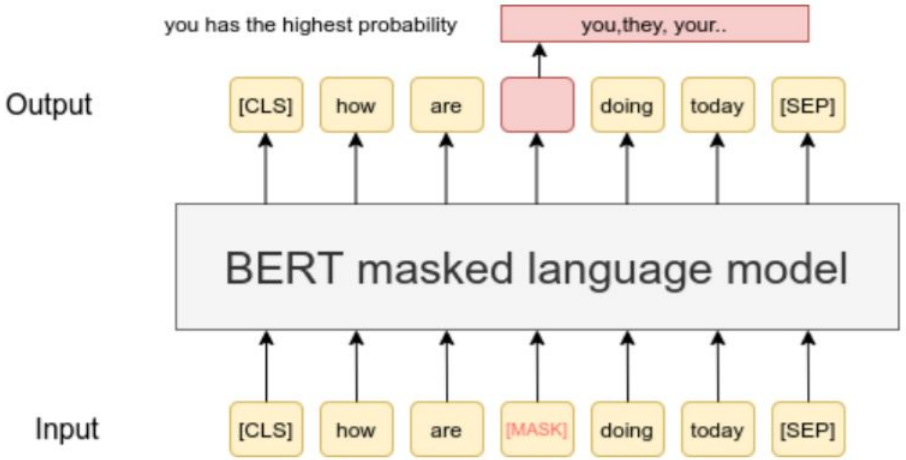
What is a foundation model?

- Model trained on an internet-scale dataset
- Training task is typically not straightforward:
  - Self-supervision (MAE, sequence prediction…)
  - Pretext tasks
- They are multipurpose and generic
- Particularization vs. generalization

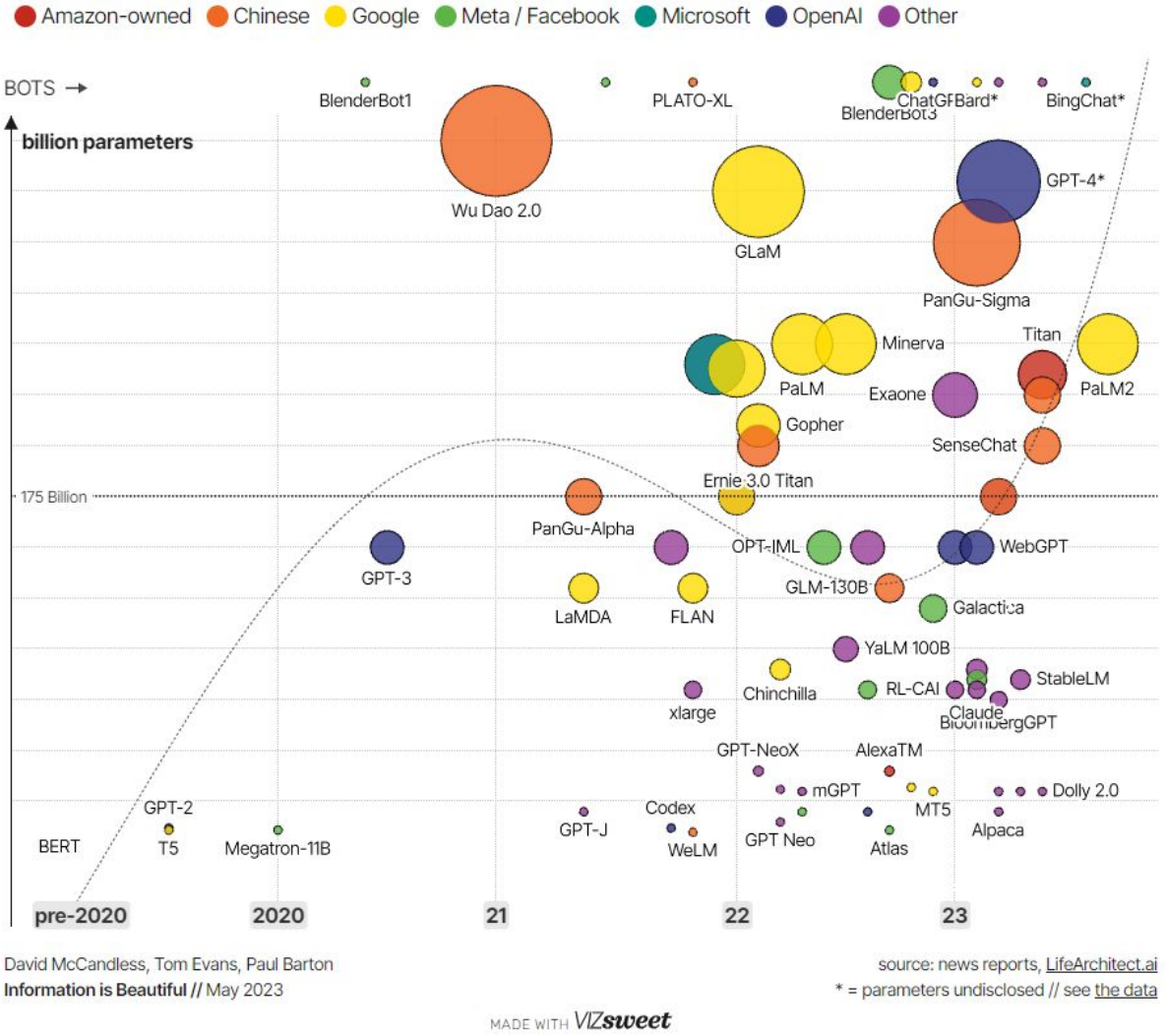


Self supervision to leverage huge datasets





Foundation models are rising (very) fast



Quick notes before diving into the details

101 vocabulary

- ***Hyperparameters:*** values to be chosen before setting up an architecture and training it
- ***Token:*** unit vector to describe data
- ***Freezing weights:*** fixing some parameter values during training, so that they cannot be updated
- ***Fine-tuning:*** training slowly to find a way to adapt without overfitting
- ***Adapter:*** small complement architecture to be trained to adapt a large model with few parameters
- ***Latent space:*** representation of the input data deep into the considered architecture

The price of foundation models

- **Can we train foundation models?**
  - Short answer: no
- **Can we use foundation models?**
  - Most of them on standard desktop machines (with a strong GPU inside)
  - Some of them using only CPU (but slow)
  - Some of them using expensive hardware (e.g. A100)
- **Can we download training datasets?**
  - Often, but not always

Use cases and limitations

- **Use cases:**
  - Process and automatically organize collections of data
  - Use as components in systems
  - Quick prototyping

- **Limitations:**
  - Black box
  - Biases
  - May require heavy computations

Program of the day

Now

10h: Global introduction

11h: Vision with DINOv2 + SAM

12h: Lunch break

14h: Natural Language Processing with Llama2

15h: MultiModal vision/text withCLIP

16h: Coffee break

16h30: MultiModal audio/text with CLAP