

Course 1: Generalities about AI



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

What is AI?

- **Intelligence:** ability to **extract knowledge** from observations
- This knowledge is used to **solve tasks in different contexts and environments** (automation)

Old way: Memorize

- Human experts code the machines
- Goods: we know what we are doing.
- Bads: requires **explicit** solutions (not available for some problems).

Modern way: Learning

- Let machines teach themselves how to solve a problem (**implicit**).
- Goods: universally applicable
- Bads: lack of understandability/robustness.
- Requires **training**.

Memorizing (explicit) vs Learning (implicit)

Global overview...

What is AI?

- **Intelligence:** ability to **extract knowledge** from observations
- This knowledge is used to **solve tasks in different contexts and environments** (automation)

Old way: Memorize

- Human experts code the machines
- Goods: we know what we are doing.
- Bads: requires **explicit** solutions (not available for some problems).

Modern way: Learning

- Let machines teach themselves how to solve a problem (**implicit**).
- Goods: universally applicable
- Bads: lack of understandability/robustness.
- Requires **training**.

Memorizing (explicit) vs Learning (implicit)

Global overview...

What is AI?

- **Intelligence:** ability to **extract knowledge** from observations
- This knowledge is used to **solve tasks in different contexts and environments** (automation)

Old way: Memorize

- Human experts code the machines
- Goods: we know what we are doing.
- Bads: requires **explicit** solutions (not available for some problems).

Modern way: Learning

- Let machines teach themselves how to solve a problem (**implicit**).
- Goods: universally applicable
- Bads: lack of understandability/robustness.
- Requires **training**.

Memorizing (explicit) vs Learning (implicit)



What is AI?

- **Intelligence:** ability to **extract knowledge** from observations
- This knowledge is used to **solve tasks in different contexts and environments** (automation)

Old way: Memorize

- Human experts code the machines
- Goods: we know what we are doing.
- Bads: requires **explicit** solutions (not available for some problems).

Modern way: Learning

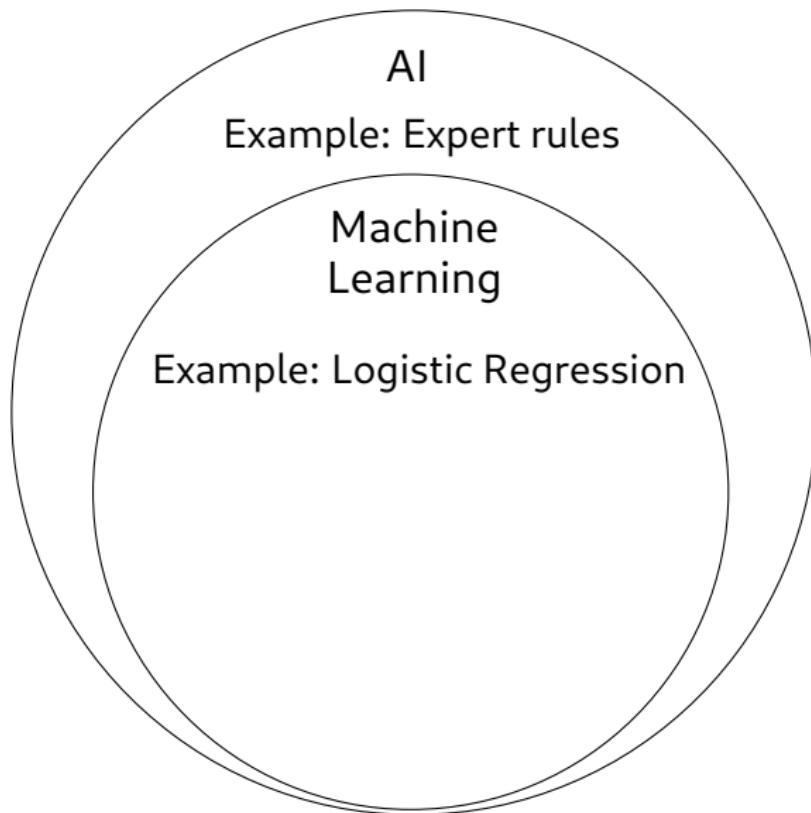
- Let machines teach themselves how to solve a problem (**implicit**).
- Goods: universally applicable
- Bads: lack of understandability/robustness.
- Requires **training**.

Memorizing (explicit) **vs** Learning (implicit)

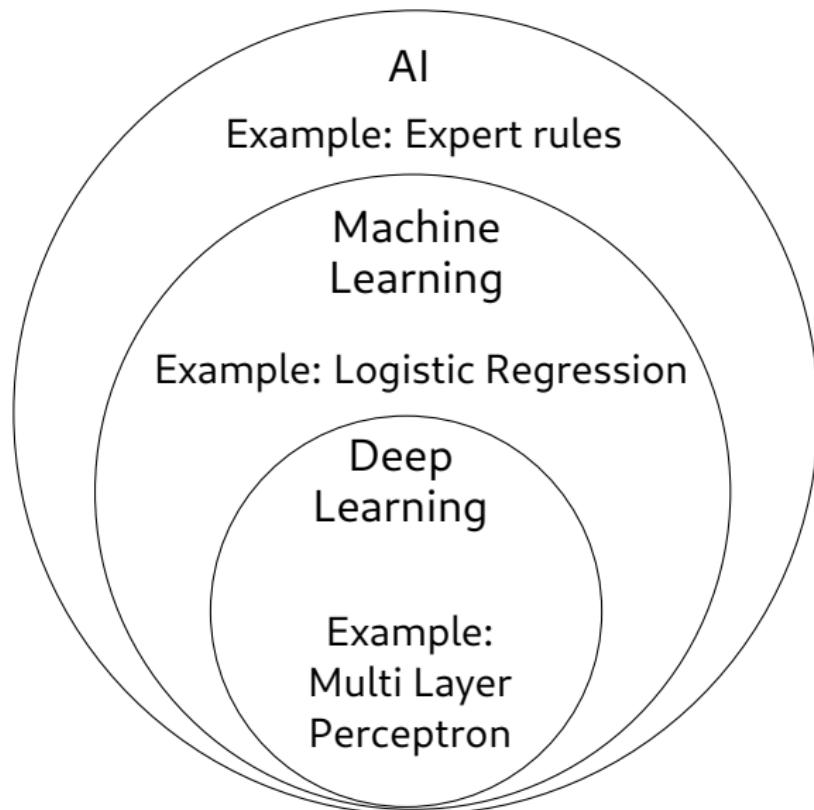
AI

Example: Expert rules

AI, Machine Learning & Deep Learning



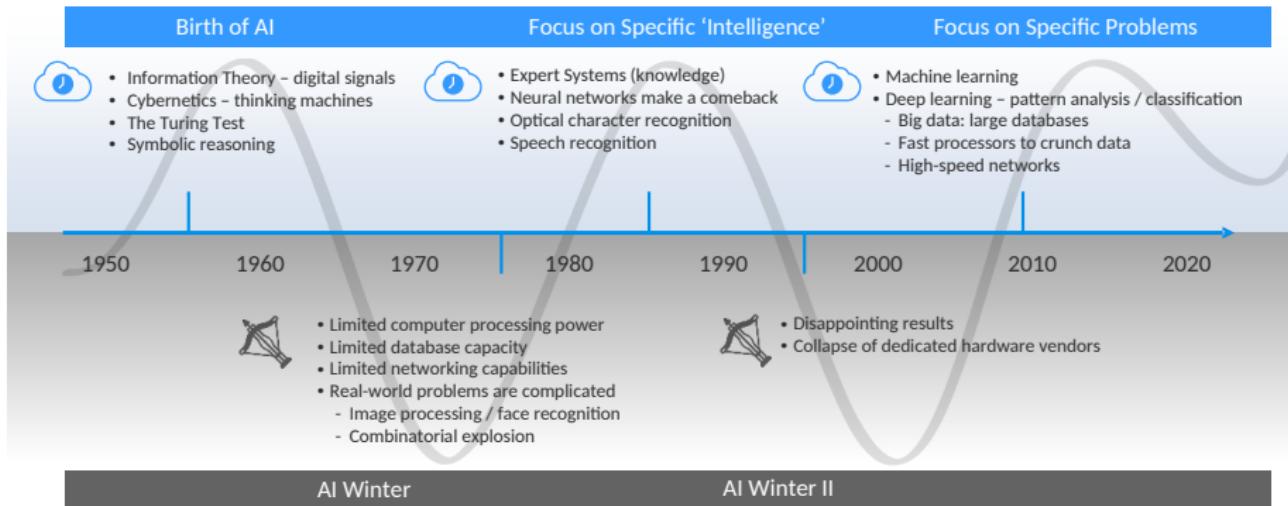
AI, Machine Learning & Deep Learning



Adapted from *Deep Learning*, 2016 by A. Courville, I. Goodfellow and Y. Bengio

An AI Timeline

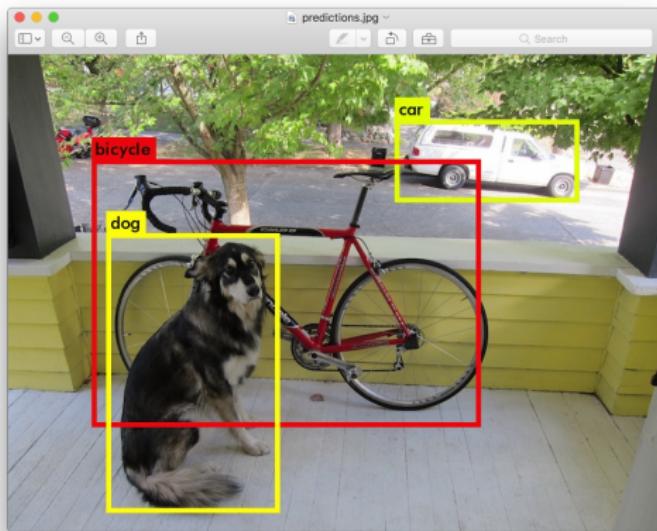
Source:  harmon.ie



Traditional application domains of AI

Vision

- Object/face recognition,
- Detection,
- Autonomous vehicles,
- Automatic diagnostic,
- Defects identification,
- Video applications...



Traditional application domains of AI

Natural Language Processing (NLP)

- Automatic assistant,
- Voice-to-text,
- Automatic translation,
- Automatic summarizing,
- Sentiment analysis,
- Text indexing...

Speak now

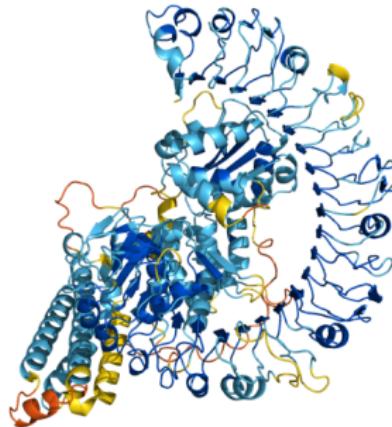
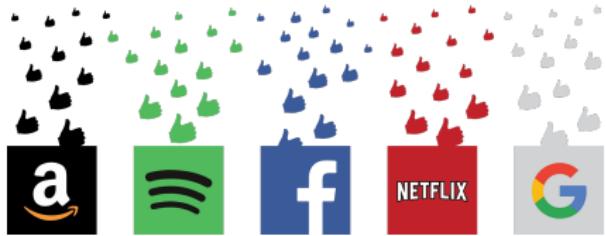


Cancel

Traditional application domains of AI

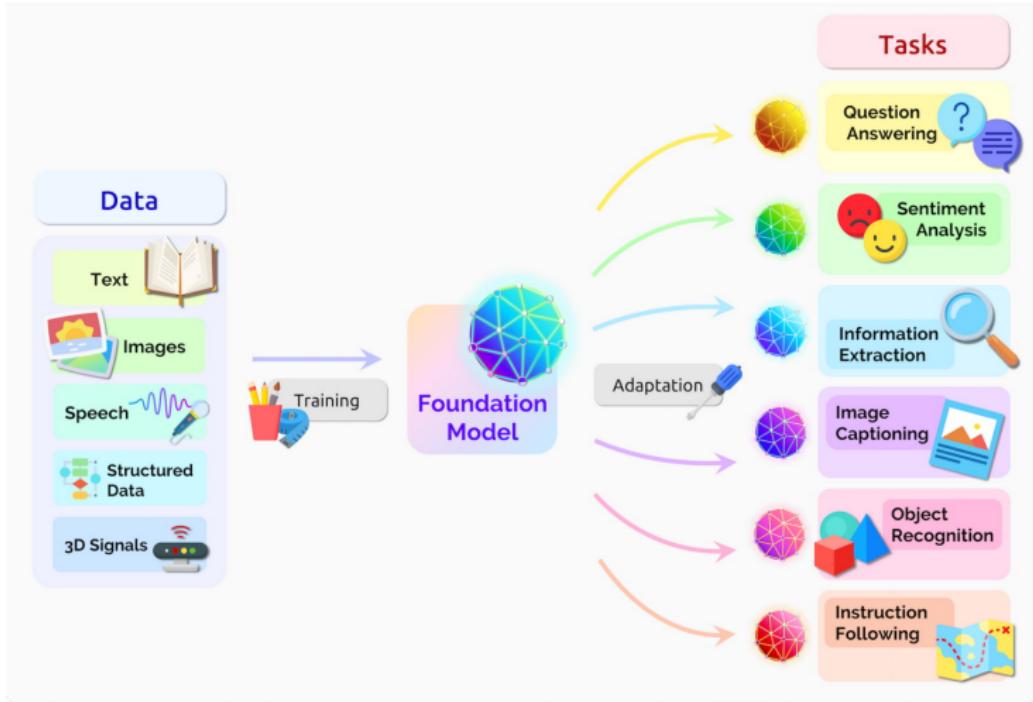
Tons of other domains...

- Medical imaging,
- Decision aid,
- Data mining,
- Visualization,
- Recommender systems,
- Market analysis...



<https://alphafold.ebi.ac.uk/>

Foundation models: a game changer



Source: <https://blogs.nvidia.com/blog/what-are-foundation-models/>

Generative AI: a recent breakthrough

Text Synthesis with Large Language Models

- GPT-3, 175 billion parameters trained without explicit supervision on over 600GB of text <https://arxiv.org/abs/2005.14165>
⇒ Model behind ChatGPT <https://openai.com/blog/chatgpt/>
- LLaMA (Meta), Gemini (Google), Mistral 8x7B (MistralAI), ... And many others!

 Write an introduction to a master course on artificial intelligence for an engineering school

 Welcome to the Master's course on Artificial Intelligence (AI) at our engineering school. In this program, you will gain a comprehensive understanding of the latest advances in the field of AI and machine learning. You will learn to design and implement intelligent systems using a variety of techniques and algorithms, and you will learn to critically evaluate the performance and limitations of these systems. The course will cover a wide range of topics including supervised and unsupervised learning, deep learning, natural language processing, computer vision, and robotics. You will also have the opportunity to work on real-world projects and to collaborate with other students and faculty members on cutting-edge research. By the end of the program, you will have the skills and knowledge needed to excel in a career in AI or to pursue further research in the field.

Generative AI: a recent breakthrough

Image Synthesis

- Stable Diffusion Models (<https://arxiv.org/pdf/2403.03206>)
- DALL-E (<https://labs.openai.com/>)
- Midjourney (<https://docs.midjourney.com/>)

Possible thanks to **multimodality**: encoding both text and image (ex: CLIP (<https://arxiv.org/abs/2103.00020>))

The screenshot shows the DALL-E web interface. At the top, there are navigation links for 'History' and 'Collections'. Below that is a text input field with placeholder text 'Edit the detailed description'. Underneath is another text input field containing the prompt: 'a watercolor style painting of a teacher in the classroom with students'. To the right of this prompt are three buttons: 'Surprise me', 'Upload', and 'Generate'. Below the input fields, four generated image thumbnails are displayed, each showing a different watercolor-style painting of a teacher interacting with students in a classroom setting. At the bottom of the interface, there is a set of standard web navigation icons.

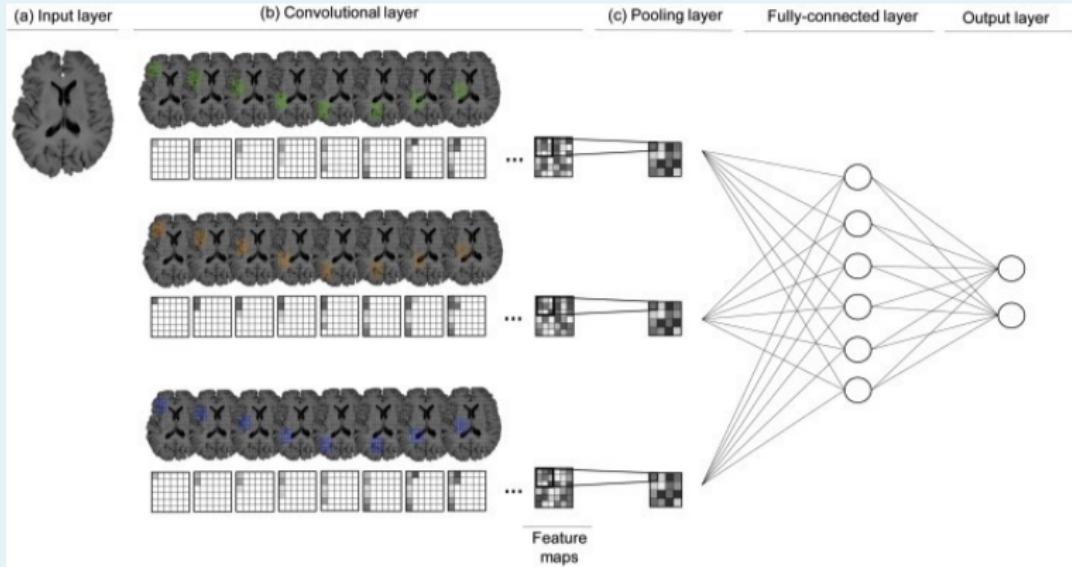
Generative AI: a recent breakthrough

Many other modalities...

- Speech (Many actors, ex: Kyutai)
- Music (StableAudio, Udio, Suno, ...)
- Video (StableVideo, Sora (OpenAI), ...)
- ...
- **Multimodal** models are arising

Some key open challenges (core AI research)

Interpretability

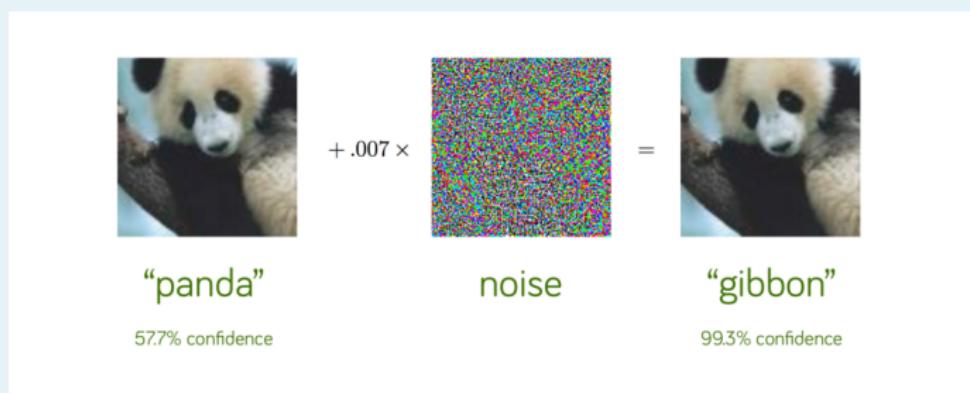


A trained model might be very accurate, but how does it take its decision ?

"Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications", Vieira et al. 2017.

Some key open challenges (core AI research)

Learning what should be learned (robustness / adversarial attacks)



Random noise added to input images can dramatically change the result.

"Intriguing properties of neural networks", Arxiv research report, 2013.

Some key open challenges (core AI research)

Computational and memory footprints

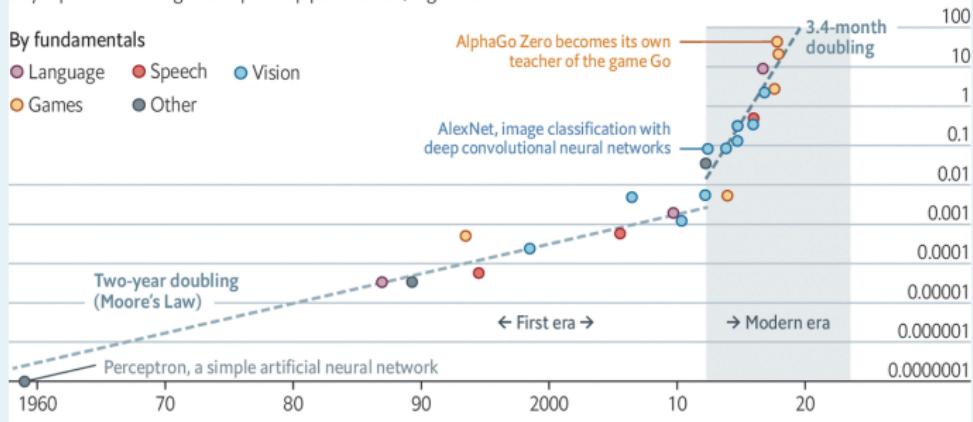
Deep and steep

Computing power used in training AI systems

Days spent calculating at one petaflop per second*, log scale

By fundamentals

- Language
- Speech
- Vision
- Games
- Other



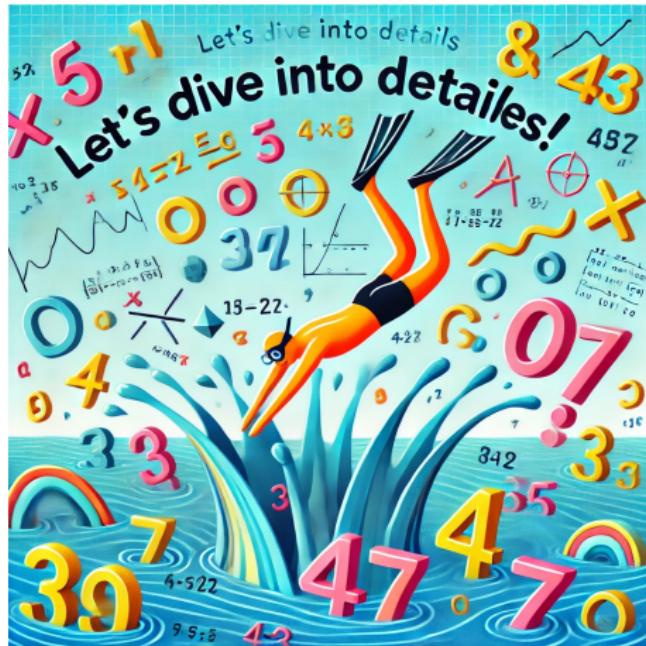
Source: OpenAI

The Economist

*1 petaflop = 10^{15} calculations

Training a large algorithm: thousands to millions of parameters using Gigabytes of data.

Let's dive into details!



Machine Learning

Examples

- Learning to play chess through playing games,
- Learning to recognize dogs and cats in images from annotated examples ...

Machine learning

- Supervised: Learning from **inputs** and **human annotations**
- Unsupervised: Learning from **inputs** only (patterns)
- Self-supervised: Learning by reconstructing the **inputs** from distorted versions

Generalization

- Generalization refers to the ability to infer **good decisions or representations from examples.**
- Goal: transfer knowledge from a dataset to another

Machine Learning

Examples

- Learning to play chess through playing games,
- Learning to recognize dogs and cats in images from annotated examples ...

Machine learning

- **Supervised:** Learning from **inputs** and **human annotations**
- **Unsupervised:** Learning from **inputs** only (patterns)
- **Self-supervised:** Learning by reconstructing the **inputs** from distorted versions

Generalization

- Generalization refers to the ability to infer **good decisions or representations from examples.**
- Goal: transfer knowledge from a dataset to another

Machine Learning

Examples

- Learning to play chess through playing games,
- Learning to recognize dogs and cats in images from annotated examples ...

Machine learning

- **Supervised:** Learning from **inputs** and **human annotations**
- **Unsupervised:** Learning from **inputs** only (patterns)
- **Self-supervised:** Learning by reconstructing the **inputs** from distorted versions

Generalization

- Generalization refers to the ability to infer **good decisions or representations** from **examples**.
- Goal: transfer knowledge from a dataset to another

Input/output

- **Goal:** infer a function of parameters \mathbf{W} from an input (often tensor) space to an output (often tensor) space, $\mathbf{y} = f(\mathbf{x}, \mathbf{W})$.
- **Example:** input can be an image, output a vector where the largest value indicate the category the image belongs to.

Error/Loss

- **Loss \mathcal{L} :** nonnegative measure of the discrepancy between expected output $\hat{\mathbf{y}}$ and obtained output \mathbf{y} .
- **Example:** output should be $[0, 1]$ but is $[0.2, 0.8]$.

Parameters

- $f(\cdot, \mathbf{W})$ contains **parameters \mathbf{W}** to be trained,
- In most cases, an ideal $f(\cdot, \mathbf{W})$ exists but is **hard to find in practice**,
- Learning is a **regression ill-posed** problem.

Input/output

- **Goal:** infer a function of parameters \mathbf{W} from an input (often tensor) space to an output (often tensor) space, $\mathbf{y} = f(\mathbf{x}, \mathbf{W})$.
- **Example:** input can be an image, output a vector where the largest value indicate the category the image belongs to.

Error/Loss

- **Loss \mathcal{L} :** nonnegative measure of the discrepancy between expected output $\hat{\mathbf{y}}$ and obtained output \mathbf{y} .
- **Example:** output should be $[0, 1]$ but is $[0.2, 0.8]$.

Parameters

- $f(\cdot, \mathbf{W})$ contains **parameters \mathbf{W}** to be trained,
- In most cases, an ideal $f(\cdot, \mathbf{W})$ exists but is **hard to find in practice**,
- Learning is a **regression ill-posed** problem.

Input/output

- **Goal:** infer a function of parameters \mathbf{W} from an input (often tensor) space to an output (often tensor) space, $\mathbf{y} = f(\mathbf{x}, \mathbf{W})$.
- **Example:** input can be an image, output a vector where the largest value indicate the category the image belongs to.

Error/Loss

- **Loss \mathcal{L} :** nonnegative measure of the discrepancy between expected output $\hat{\mathbf{y}}$ and obtained output \mathbf{y} .
- **Example:** output should be $[0, 1]$ but is $[0.2, 0.8]$.

Parameters

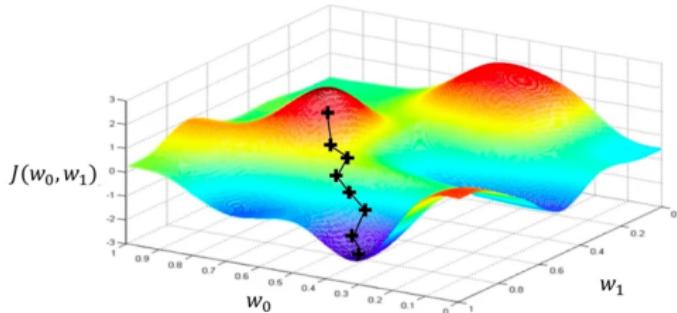
- $f(\cdot, \mathbf{W})$ contains **parameters \mathbf{W}** to be trained,
- In most cases, an ideal $f(\cdot, \mathbf{W})$ exists but is **hard to find in practice**,
- Learning is a **regression ill-posed** problem.

- Loss: $J(\mathbf{W}) = \sum_i \mathcal{L}(f(\mathbf{x}^{(i)}, \mathbf{W}), \mathbf{y}^{(i)})$, $i = \text{examples}$
- Model parameters: $\mathbf{W}^* = \text{argmin}(J(\mathbf{W}))$

Training Algorithm

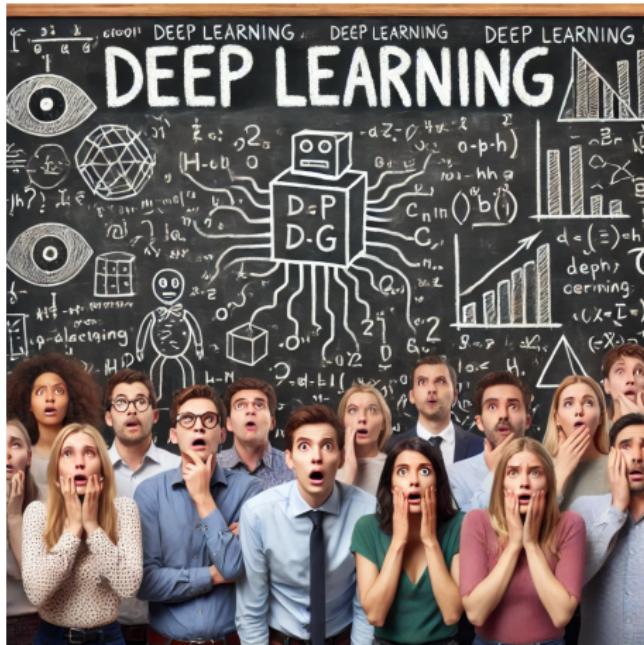
Gradient Descent:

- Randomly Initialize model weights
- Compute Gradient of the Loss $\frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$
- Update weights $\mathbf{W} \leftarrow \mathbf{W} - \eta \frac{\partial J(\mathbf{W})}{\partial \mathbf{W}}$
- Repeat until convergence



from MIT course introtodeeplearning.com

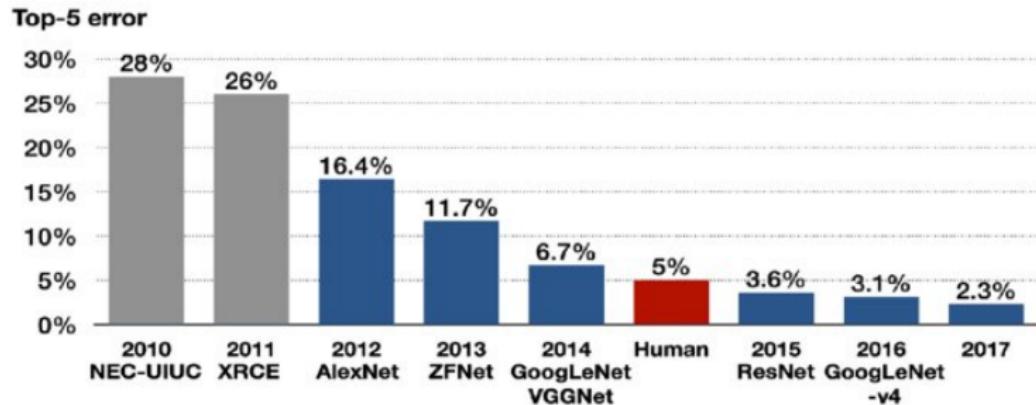
Deep Learning



A turning point: AlexNet (2012)

Started the reign of Deep Learning

- Using deep neural networks
- A major breakthrough in image classification:



Source: Kang, D. Y., Duong, H. P., & Park, J. C. (2020). Application of deep learning in dentistry and implantology. Journal of implantology and applied sciences, 24(3), 148-181.

Details for the human evaluation: Russakovsky, Dieg et al.. ImageNet Large Scale Visual Recognition Challenge,
<https://arxiv.org/pdf/1409.0575.pdf>

The great elders of Deep Learning (Turing Prize 2018)

Geoffrey Hinton



- Cognitive psychologist and computer scientist,
- Prof. at University of Toronto and works for Google,
- Known for back-propagation and Boltzmann machines.

Yoshua Bengio



- Computer scientist,
- Prof. at Université de Montréal and head of MILA,
- Known for his work on deep learning.

Yann le Cun



- Computer scientist,
- Prof. at New York University then he joins FAIR,
- Known for his work on back-propagation and CNNs.

Where did the Deep Learning revolution came from?

- The use of GPUs for computation.
- The share of huge datasets on Internet.
- HuggingFace / Github / Arxiv: easy and open-source share of research.
- The rise of representation learning.



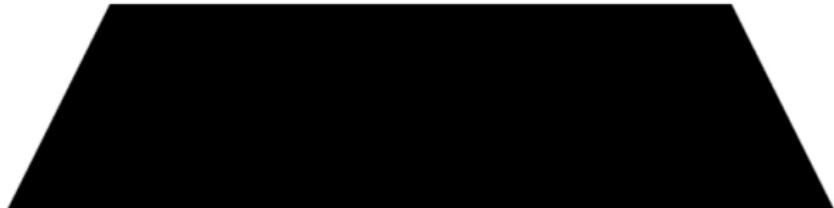
Where did the Deep Learning revolution came from?

- The use of GPUs for computation.
- The share of huge datasets on Internet.
- HuggingFace / Github / Arxiv: easy and open-source share of research.
- The rise of representation learning.



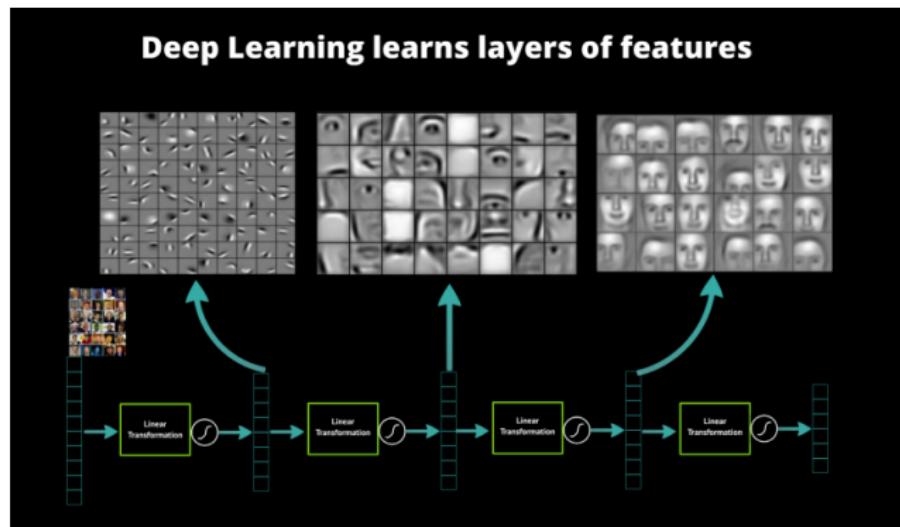
Where did the Deep Learning revolution came from?

- The use of GPUs for computation.
- The share of huge datasets on Internet.
- HuggingFace / Github / Arxiv: easy and open-source share of research.
- The rise of representation learning.

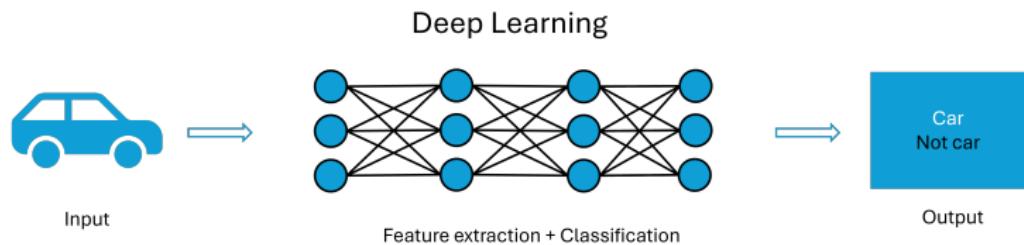
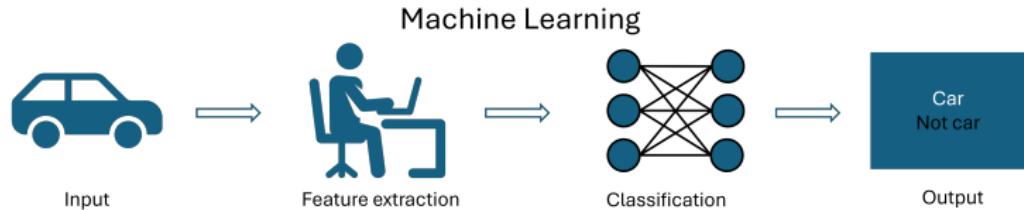


Where did the Deep Learning revolution came from?

- The use of GPUs for computation.
- The share of huge datasets on Internet.
- HuggingFace / Github / Arxiv: easy and open-source share of research.
- The rise of representation learning.



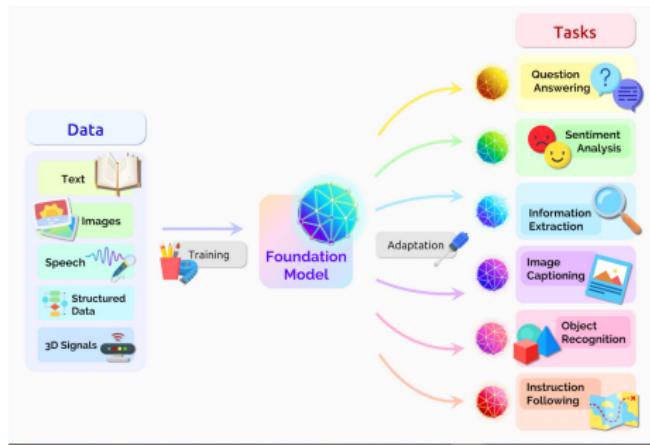
Traditional Deep Learning (before 2020)



Inspired from: <https://www.softwaretestinghelp.com/data-mining-vs-machine-learning-vs-ai/>

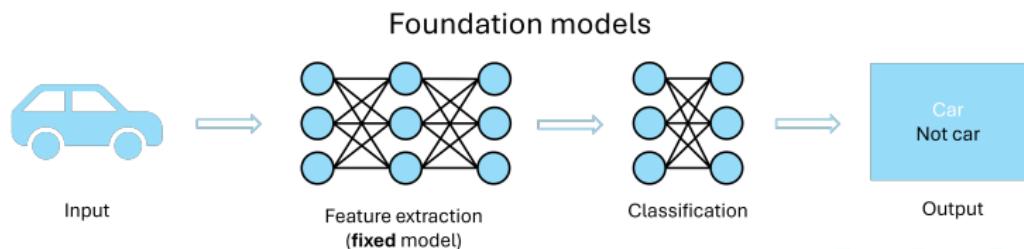
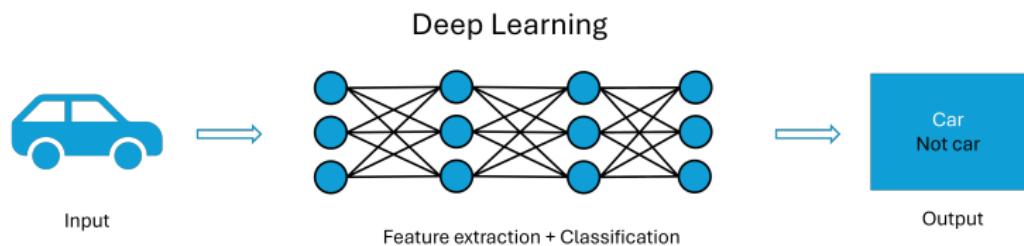
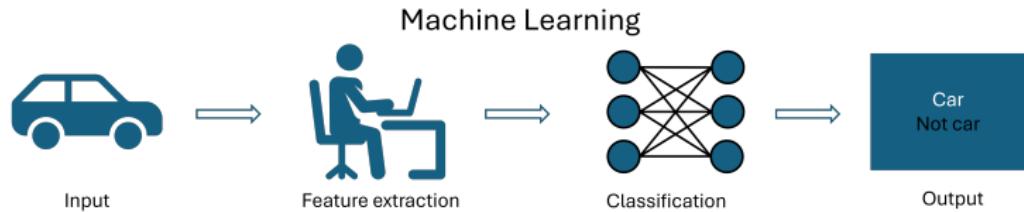
Deep Learning today - Foundation models (1/3)

- Model trained on an Internet scale dataset
- Self-Supervised training: pretext tasks
- Generalization is not a problem anymore! All is about **particularization**



Source: <https://blogs.nvidia.com/blog/what-are-foundation-models/>

Deep Learning today - Foundation models (2/3)

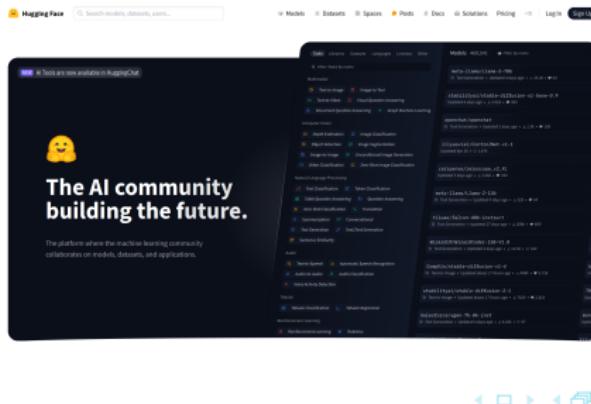


Deep Learning today - Foundation models (3/3)

- Foundation models are really hard to train...
 - Requires a lot (a lot!) of data
 - Requires a lot of computing power
 - Requires a lot of time
- ... But they can be used as powerful feature extractor!
 - Some open-source Foundation models exist, and are available on HuggingFace <https://huggingface.co/>
 - We will use such algorithms during our labs sessions.

Deep Learning today - Foundation models (3/3)

- Foundation models are really hard to train...
 - Requires a lot (a lot!) of data
 - Requires a lot of computing power
 - Requires a lot of time
 - ... But they can be used as powerful feature extractor!
 - Some open-source Foundation models exist, and are available on HuggingFace <https://huggingface.co/>
 - We will use such algorithms during our labs sessions.



Lab Session 1

Take-home messages (for now):

- Foundation models allow to meaningful and rich information in data ;
- They can be used to represent a set of data, and extract features.

Lab assignments:

- Introduction to Python, Environments, Numpy, etc,
- Introduction to data visualization for Machine Learning,
- Choose one modality (Text, Image, Audio)
- Tests on some examples of each modality (embeddings are pre-computed for you).

Link to the lab: <https://mee-labs.gitlab-pages.imt-atlantique.fr/intro2ai/lab1/>

Presentation - session 2

By groups of **four students**:

A task on your modality

You will have to prepare a 15 minutes presentation (for session 2) on a chosen modality (Text, Image or Audio). Choose **two** tasks:

■ Text:

- 1 Language Understanding with LLM (Large Language Models)
- 2 Using LLM to generate images, video, music, etc..

■ Image:

- 1 Semantic Segmentation
- 2 3D generation from 2D images

■ Audio:

- 1 Source Separation
- 2 Speech generation

Present the tasks, principles, and tools used to solve it.