Introduction to AI and Open Source AI Definition

by **Giuseppe Aceto**,
Professor of Computer Engineering, University of Napoli Federico II
President of NaLUG - Napoli GNU/Linux Users Group

giuseppe.aceto@unina.it

“AI Beyond the Horizon: Shaping Aerospace's Future”

November 27th, 2024, Napoli

Key principles of Open Source AI

open source initiative®

Open Source AI is built on principles that prioritize accessibility, transparency, and collaboration, which collectively enhance the development and deployment of AI technologies. *The effectiveness of this approach has been proved with Open Source Software.*

Accessibility

Open Source AI promotes accessibility by making **code and resources available** to everyone, allowing developers from various backgrounds to contribute and innovate.

Transparency

Transparency in Open Source AI ensures that **algorithms and data** used in AI models are **open for scrutiny, fostering trust** and enabling users to **understand how decisions are made**.

Community Collaboration

Community collaboration is central to Open Source AI, as it encourages **collective problem-solving and knowledge sharing**, leading to faster advancements and improved solutions.

Defining Freedoms of Open Source AI

An *Open Source AI* is an AI system made available under terms and in a way that grant the **freedoms*** to

- **Use** the system for any purpose and without having to ask for permission.
- **Study** how the system works and inspect its components.
- **Modify** the system for any purpose, including to change its output.
- **Share** the system for others to use with or without modifications, for any purpose.

These freedoms apply both

- to a **fully functional system**
- to **discrete elements** of a system

⇨ A **precondition** to exercising these freedoms is to **have access to the preferred form to make modifications** to the system.

*These freedoms are derived from the Free Software Definition

AI ML Definitions

An **AI system** is a machine-based system that, for explicit or implicit objectives, **infers, from the input it receives, how to generate outputs** such as **predictions, content, recommendations, or decisions**, that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

Recommendation of the Council on Artificial Intelligence OECD/LEGAL/0449, Organization for Economic and Co-operation Development (OECD), 2024

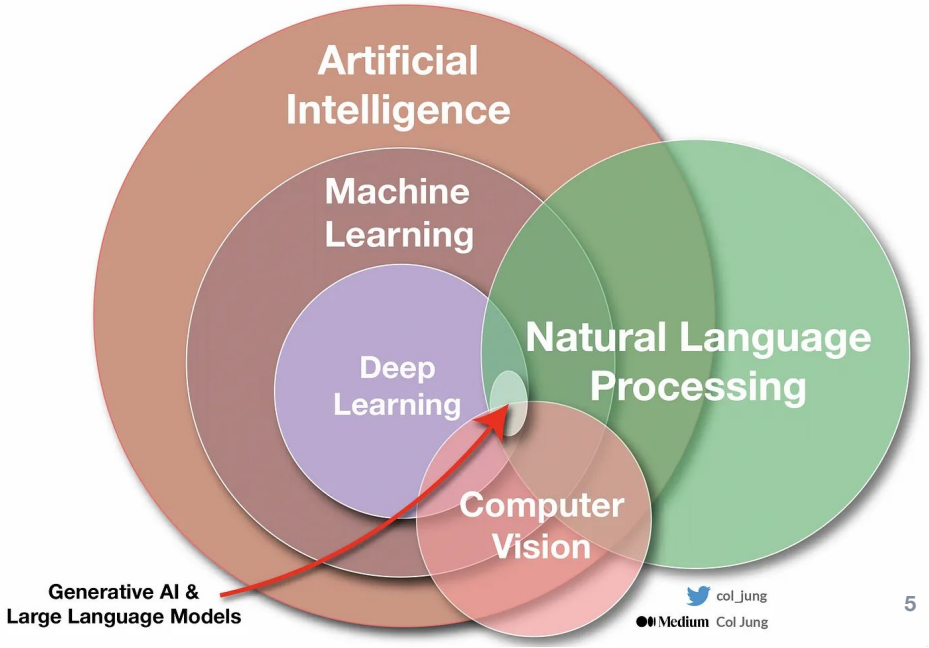
Machine learning (ML) is a set of techniques that allows machines to **improve their performance** and usually **generate models in an automated manner** through **exposure to training data**, which can help identify patterns and regularities rather than through explicit instructions from a human.
The process of **improving a system's performance using machine learning techniques** is known as **"training"**.

Explanatory memorandum on the updated OECD definition of an AI system, OECD Artificial Intelligence Papers, No. 8, OECD Publishing, Paris

Learning strategies

- **Supervised learning** involves training a model on a labeled dataset, which means the output is known and used to guide the learning process.
- **Unsupervised learning** deals with unlabeled data and aims to find hidden patterns or intrinsic structures within the input data.
- **Semi-supervised learning** is a hybrid approach that uses a small amount of labeled data and a large amount of unlabeled data to improve learning accuracy
- **Self-supervised learning** obtains supervisory clues from the data itself, to predict any hidden part (or property) of the input from any unhidden part of the input itself.
- **Reinforcement learning** focuses on training models to make sequences of decisions by rewarding desired behaviors and penalizing unwanted ones.

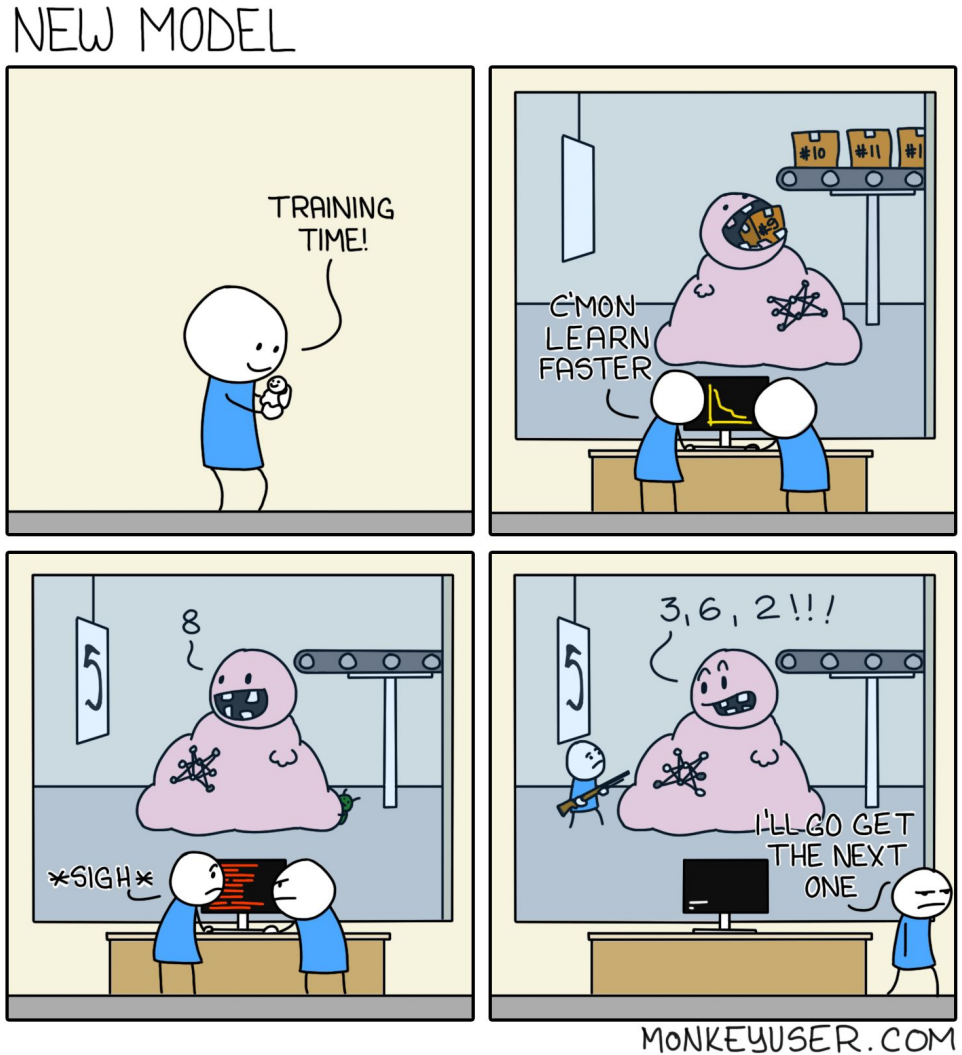
Large Language Models (LLMs) are a category of *foundation models* (trained on immense amounts of data) that "understand" and generate natural language and other types of content.



How to Train a LLM

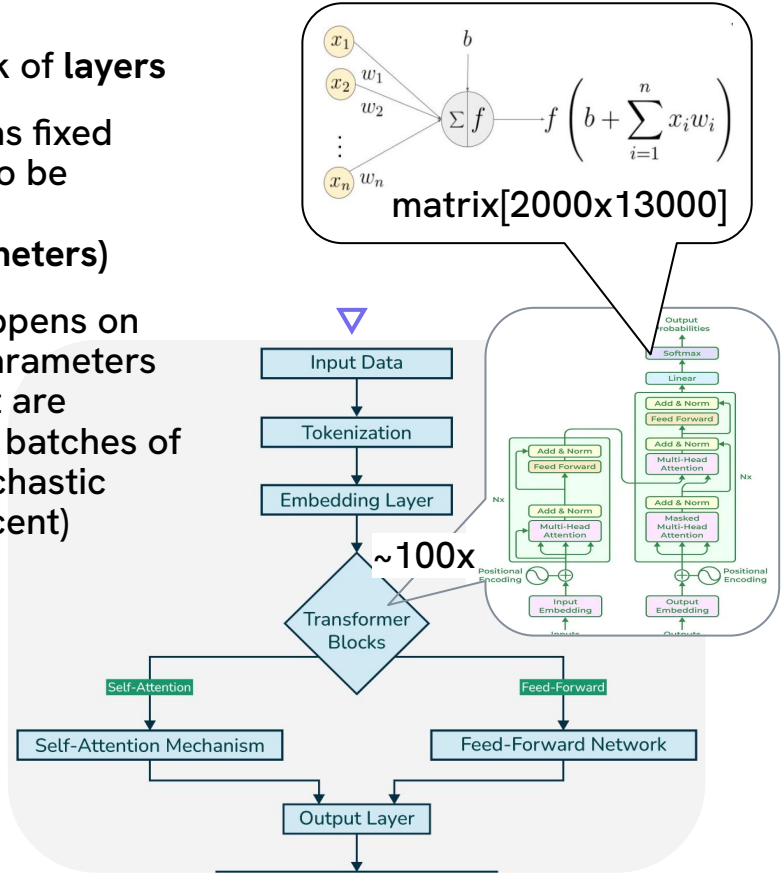
Training Procedure Overview

- **Pre-Training:** self-supervised learning on massive datasets
- **Fine-Tuning:** additional (self-)supervised learning on specialized datasets
  - **Hyper-parameter Tuning** varying the training process parameters (samples, batches, epochs…)
  - **Reinforcement Learning** from **Human Feedback** (**RLHF**) to improve **alignment** of the model output to ethical and legal guidelines



LLM under the hood

- a (deep) stack of **layers**
- **each layer** has fixed parameters to be designed (**model parameters**)
- **"training"** happens on modifiable parameters (**weights**) that are **optimized** on batches of samples (stochastic gradient descent)



Layers of an LLM

- **Tokenization**: text data is cut in sub-words/characters and mapped to numbers
- **Embedding** layer maps tokens into continuous vectors, capturing semantic meanings and relationships between words.
- **Transformer** architecture captures contextual relationships. Turns embeddings into 3 matrices.

  Stacking Transformers learns hierarchical representations, to match complex patterns.
- **Decoding** in the output layer maps back to tokens (extracted from the computed **probability distribution**)

Freedom to modify ML systems

Parameters

The model parameters, such as **weights** or other **configuration settings** (architecture, hyperparameters, optimization checkpoints).

Code

The complete **source code** used to **train** and **run** the system. The Code shall represent the full specification of how the data was processed and filtered, and how the training was done. Code shall be made available under OSI-approved licenses. For example, include code used for

- processing and filtering data
- training (including arguments and settings)
- validation and testing
- supporting libraries (tokenizers. hyperparameters search)...

Freedom to modify ML systems

Data Information

Sufficiently detailed **information about the data** used to train the system so that a skilled person can build a **substantially equivalent system**.

1. the complete description of all data used for training (including unshareable data), in
   a. provenance
   b. scope
   c. characteristics
   d. how was obtained and selected
   e. labeling procedures
   f. data processing and filtering methods
2. a listing of all publicly available training data and where to obtain it;
3. a listing of all training data obtainable from third parties and where to obtain it, including for fee.

Criticism

Training Data is the "source"

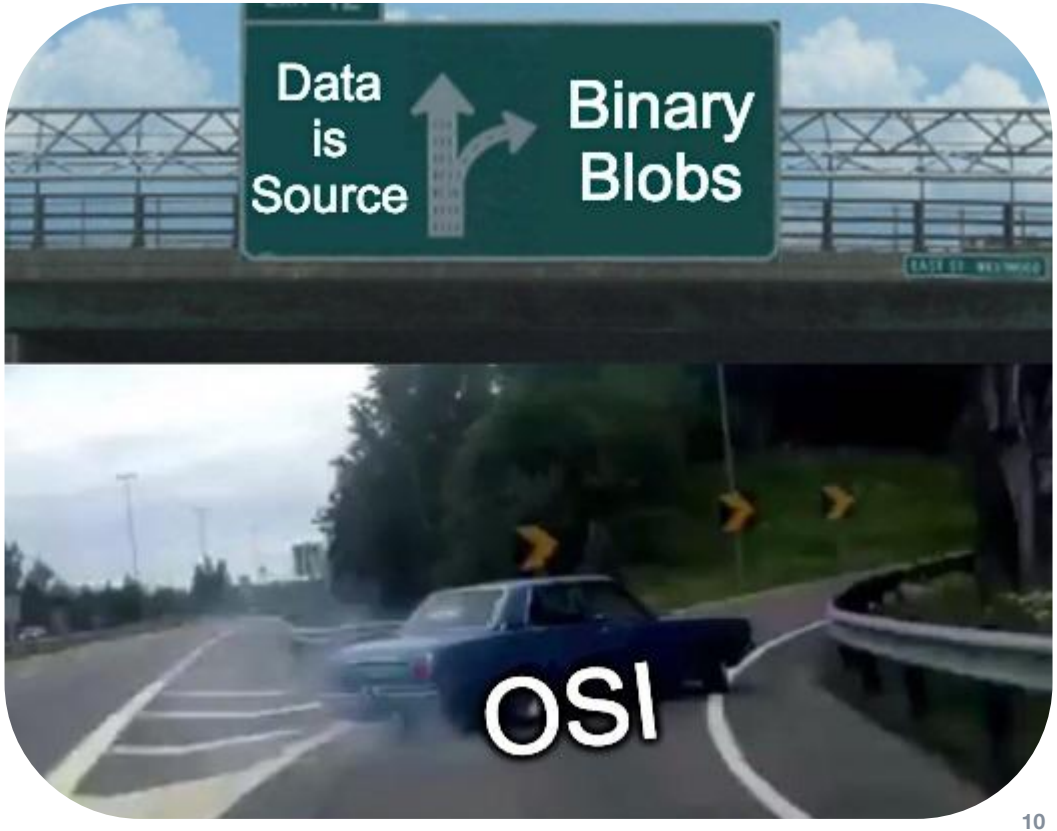
According to many* the **training data is necessary** for the 4 freedoms, so the OSAID is fundamentally incompatible with the OSD.

"Sufficiently detailed **information**" **about the data is not enough.**

OSAID motivation for data exclusion

There are cases in which training data sharing is not legally allowed (or desirable)

- medical data
- privacy
- indigenous knowledge



*Bruce Perens (DFSG & OSD author and OSI founder), Bruce Schneier (fellow and lecturer at Harvard's Kennedy School, a board member of EFF), Debian developers (proposing a General Resolution opposing OSAID), etc..

https://samjohnston.org/2024/10/22/debian-general-resolution-gr-drafted-opposing-osis-open-source-ai-definition-osaid

Hardware

Training costs

Besides **data collection and curation**, just the computing requires **specialized hardware** (GPUs, TPUs) and **distributed computing** (clusters, datacenters, with high bandwidth connections, cooling, high energy costs).

Project of META new AI datacenter in Temple, Texas (datacenterdynamics.com)

Year	Model Name	Model Creators/Contributors	Training Cost (USD) Inflation-adjusted
2017	Transformer	Google	$930
2018	BERT-Large	Google	$3,288
2019	RoBERTa Large	Meta	$160,018
2020	GPT-3 175B (davinci)	OpenAI	$4,324,883
2021	Megatron-Turing NLG 530B	Microsoft/NVIDIA	$6,405,653
2022	LaMDA	Google	$1,319,586
2022	PaLM (540B)	Google	$12,389,056
2023	GPT-4	OpenAI	$78,352,034
2023	Llama 2 70B	Meta	$3,931,897
2023	Gemini Ultra	Google	$191,400,000

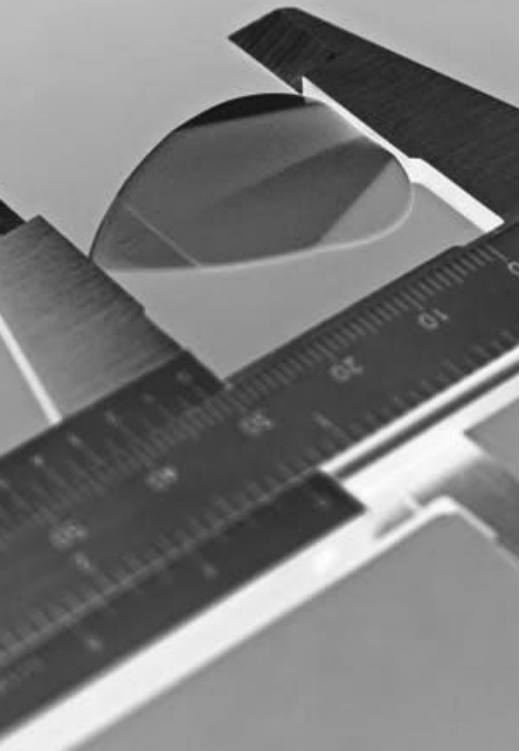
Data by visualcapitalist.com, based on cloud compute rental prices

References and sources

The Open Source AI Definition - 1.0
https://opensource.org/ai/open-source-ai-definition
Debian General Resolution draft opposing OSAID
https://samjohnston.org/2024/10/22/debian-general-resolution-gr-drafted-opposing-osis-open-source-ai-definition-osaid/
Schneier on Security
https://www.schneier.com/blog/archives/2024/11/ai-industry-is-trying-to-subvert-the-definition-of-open-source-ai.html
Data Center infos b7 datacenterdynamics
https://www.datacenterdynamics.com/en/analysis/how-meta-redesigned-its-data-centers-for-the-ai-era/
Training costs for LLM
https://www.visualcapitalist.com/training-costs-of-ai-models-over-time/

Images:
https://media.geeksforgeeks.org/wp-content/uploads/20240826180729/Exploring-the-Technical-Architecture-Behind-Modern-Language-Models.png
https://media.geeksforgeeks.org/wp-content/uploads/20240607170651/Transformer-python.webp
Decorations:
https://www.pexels.com

Which AI systems comply with the OSAID 1.0?



Passing Validation

- **Pythia** by Eleuther AI
- **OLMo** from AI2
- **Amber** and **CrystalCoder** by LLM360
- **T5** from Google

Incompatible legal agreements or lack required components

- **LLAMA2** by Meta
- **Grok** from X/Twitter
- **Phi-2** by Microsoft
- **Mixtral** by Mistral

With Adjustments to license/legal terms

- **BLOOM** by BigScience
- **Starcoder2** from BigCode
- **Falcon** by TII needs

NAPOLI GNU/LINUX USERS GROUP

Associazione di Promozione Sociale

- non a scopo di lucro
- attiva a Napoli e dintorni da fine '90
- favorisce la diffusione della **filosofia GNU** (www.gnu.org)
- www.nalug.tech