Hewlett Packard
Enterprise

HPE AI : Strategy and Portfolio

HPC&AI Center of Excellence (CoE)

March, 2023

Christophe Menichetti
AI solutions Architect
Christophe.menichetti@hpe.com

WORKSHOP : AGENDA

AI BASICS:
WHAT IS AI ? BASICS AND CONCEPTS

Part **1**

AI CLIENT AND MARKET:
CLIENT PROFILING AND HPE STRATEGY

Part **2**

HPE PORTFOLIO:
HARDWARE AND SOFTWARE HPE SOLUTIONS FOR AI

Part **3**

AI BASICS

INTRO: VOCABULARY AND HISTORY

ARTIFICIAL INTELLIGENCE :

ANY IT SOLUTIONS THAT ENABLES MACHINES TO IMITATE HUMANS BEHAVIOUR

MACHINE LEARNING:

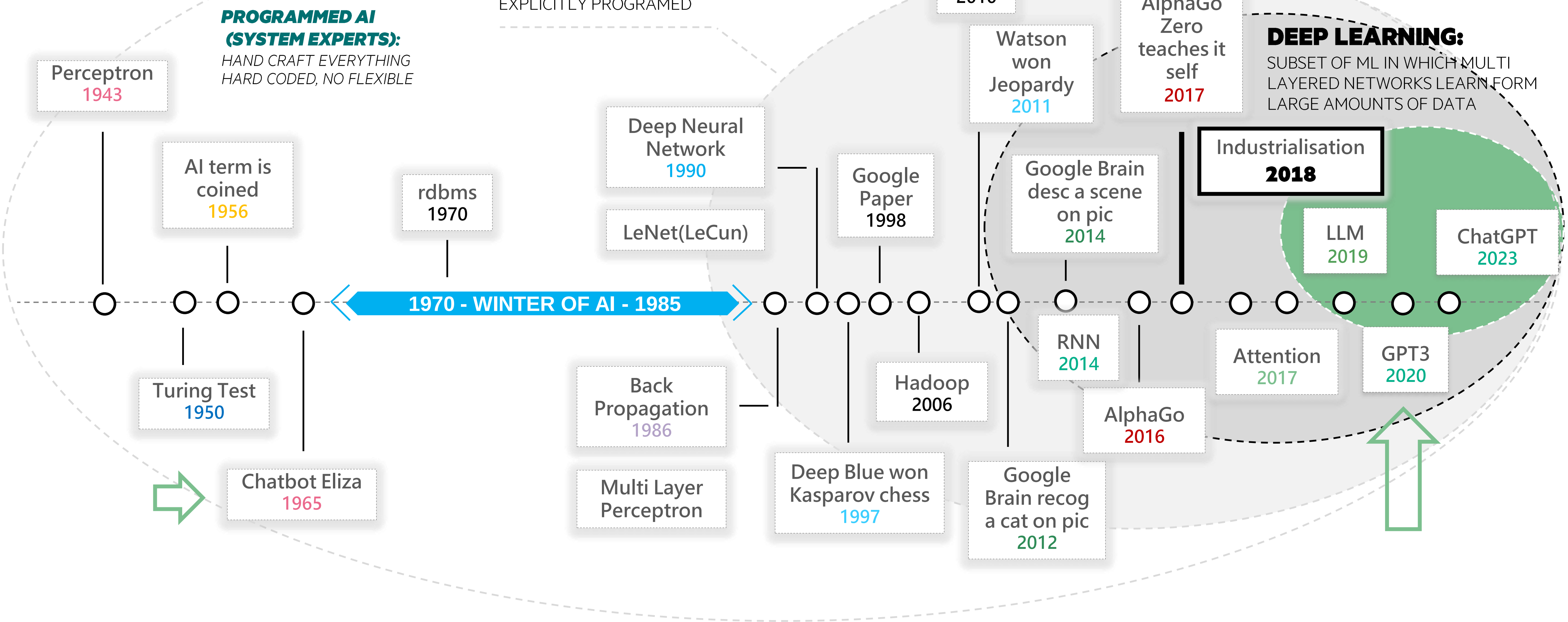
SUBSET OF AI WHICH USES STATISTICAL METHODS TO LEARN WITHOUT BEING EXPLICITLY PROGRAMED

FOUNDATION MODELS:

FOUNDATION MODELS ARE LARGE AI MODELS TRAINED ON ENORMOUS QUANTITIES OF UNLABELED DATA. THIS PROCESS RESULTS IN GENERALIZED MODELS CAPABLE OF A WIDE VARIETY OF TASKS

DEEP LEARNING:

SUBSET OF ML IN WHICH MULTI LAYERED NETWORKS LEARN FORM LARGE AMOUNTS OF DATA



ML/DL: INTRODUCTION TO THE BASICS

Real Estate agency willing to automate (software) house's price estimation

The business purpose is simple : estimate the right (most profitable) price

But the business rules / logic is complex, depending on many parameters

Name	Surface	Rooms	Pool	Price
House1	2,000	4	0	270,000
House5	3,500	6	1	510,000
House12	1,500	4	0	240,000

```
Sales_price = 0;
if (city == Montpellier)
  if (block == "Odyseum") price_sqft = 160;
  else (if (block == "Comedie") price_sqft = 120;
          else(if …)
                else (if)
        )
….
```

ML/DL BEHIND THE COVER : ALL ABOUT MATHS

Machine Learning =
1) Data (and Data Analysis)
2) Mathematics
3) Prediction

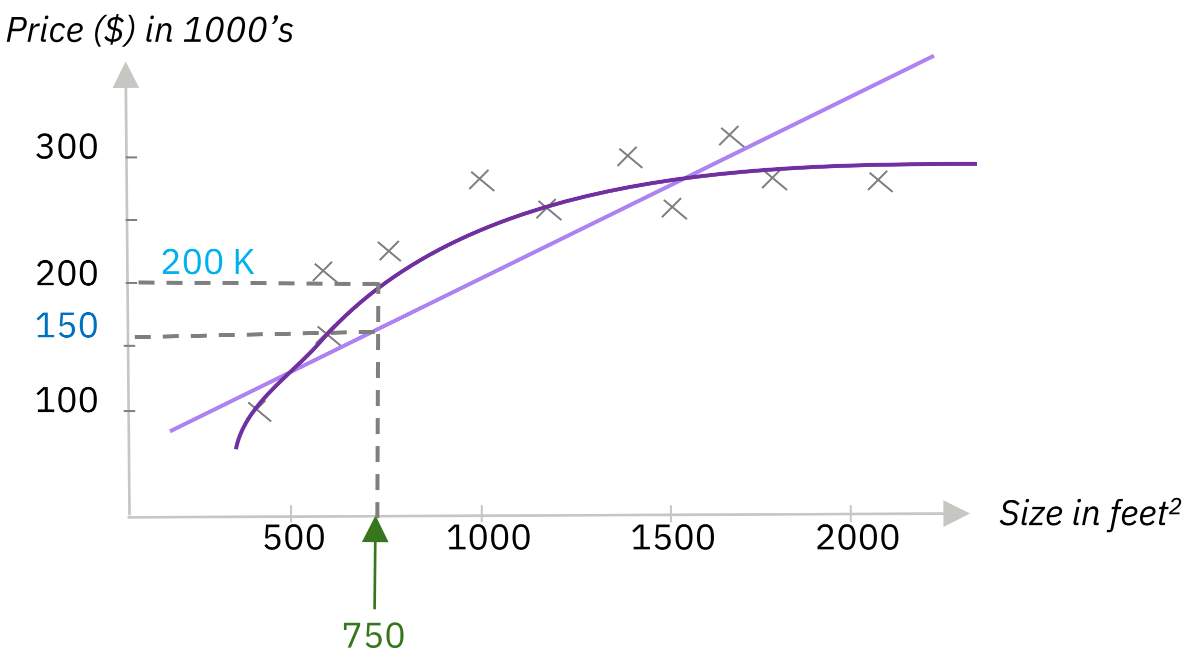
F(X) = Y

Name	Surface	Rooms	Pool	Price
House1	2,000	4	0	270,000
House5	3,500	6	1	510,000
House12	1,500	4	0	240,000

Features (attributes) x1,x2,x3

Label (target) = Y

*Many functions (**predictive models**) could apply on those data to make a prediction (more or less accurate) – without hard coding the business rule (**without explicitly programmed**)*



Linear Model ax+ b
Parameters are called W (weights) and B (biases)

polynomial model: ax^2 + bx + c

ML/DL BEHIND THE COVER : ALL ABOUT MATHS

Name	Surface	Rooms	Pool	Price
House1	2,000	4	0	270,000
House5	3,500	6	1	510,000
House12	1,500	4	0	240,000

Algorithm
Mathematical function

Algorithm
Result = weight1*param1 + weight2*param2 + weight0

Model
Algorithm with weighted parameters

Model
Price in $US = 100*(square_feet) + 10000*(bedrooms) + 100,000

Consider an example first pass. Record 1 is a house of 1500 square feet with three bedrooms. The training process submits that data to the untrained model, which might be:

Price in $US = 100*(area in square feet) + 10,000*(bedrooms) + 100,000

The output is $280,000.

Because you are using supervised training, the record included a label: $314,000. The training process assesses the difference between the output label and the original label. It then adjusts the model weights in an attempt to do better. It might decide to weigh area and bedrooms more heavily and to increase the bias—as just one example:

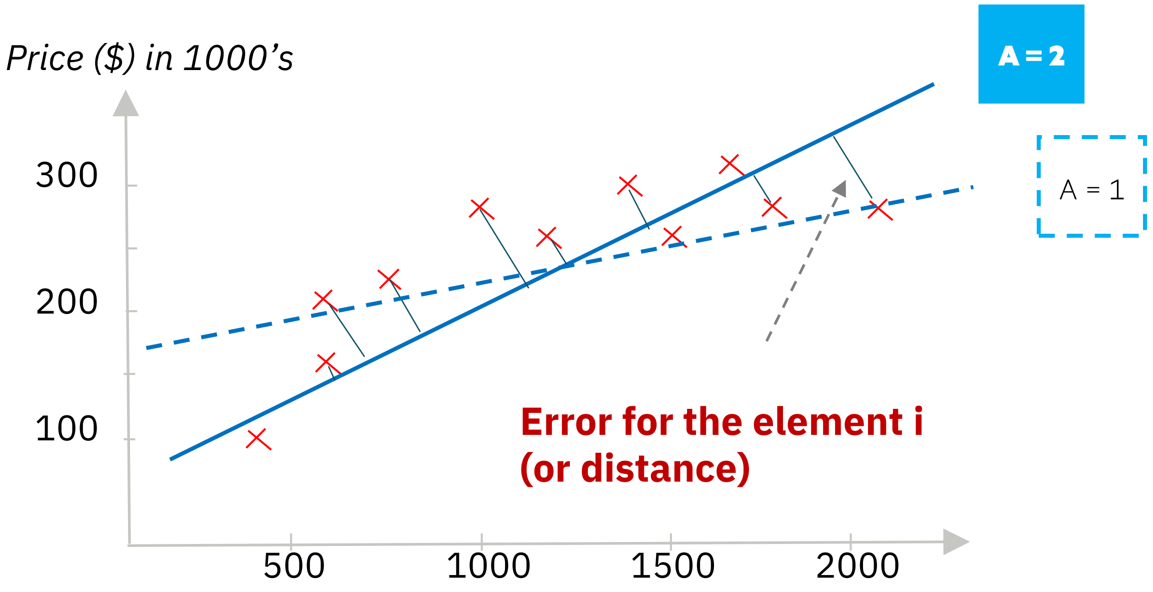
Price in $US = 110*(area in square feet) + 11,000*(bedrooms) + 110,000

After adjusting the model, the training process conducts a second pass on a *new* record. Again, it assesses the result and its difference from the accurate label. It makes a further adjustment to the model weights.

The process continues for pass after pass until the model has been trained on the desired amount of data. The model is now considered "trained," and it should be performing very well.

*The goal of machine learning is to define and minimize the error (**the cost function)** – so in other word, maximize the accuracy of the predictive model*

We have to define the error function and how to minise the error (algo = optimiser)

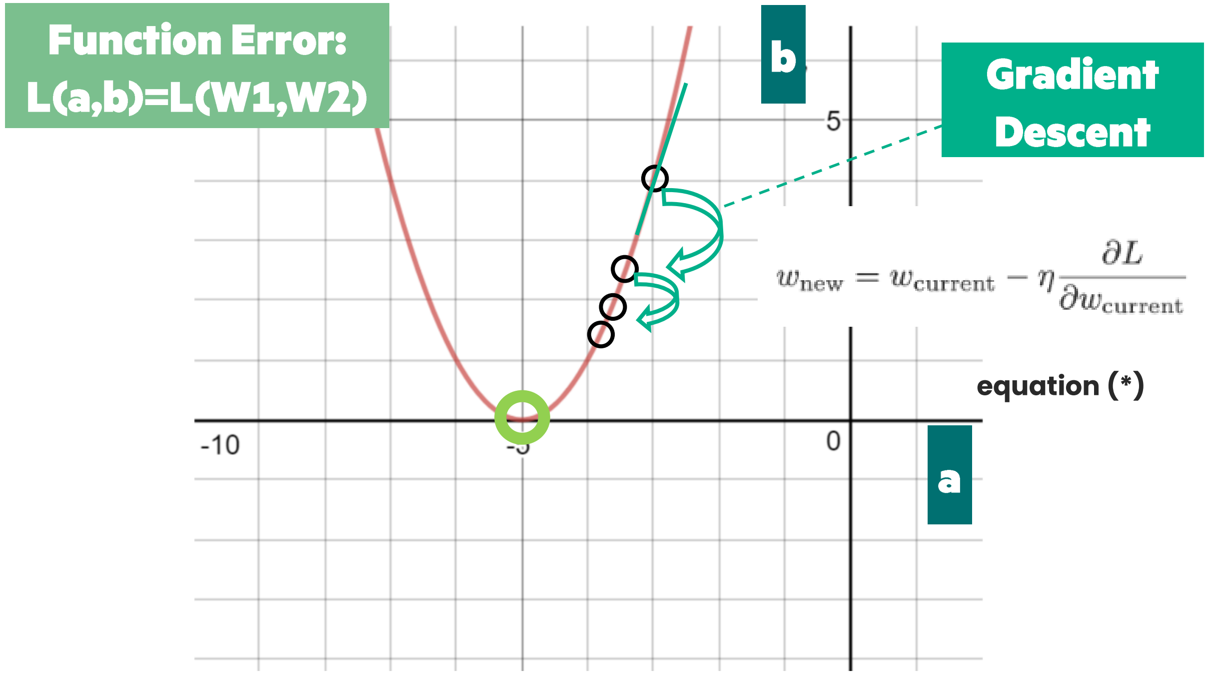


ERROR (COST) = $\min \ 1/n \sum_{i=1}^{n} (y_i - f(x_i))^2$

ML/DL SECRET SAUCE : TRAINING - OPTIMISER AND LEARNING RATE

Example by hand :

Question : Find the local minima of the function $y=(x+5)^2$ starting from the point $x=3$



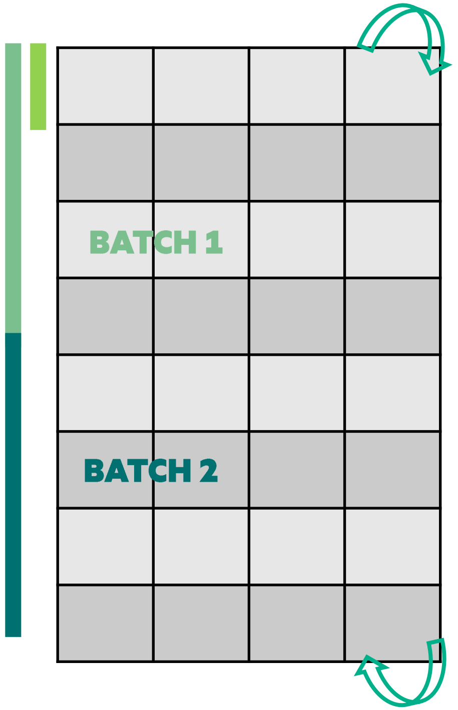
$$w_{\text{new}} = w_{\text{current}} - \eta \frac{\partial L}{\partial w_{\text{current}}}$$

equation (*)

You can analytically solve the solution of finding the minimum of Error function but it could be time consuming.

The idea is to use a fastest way (iterative algorithm) to solve it : **gradient descent** (derivative with respect to parameters and bias)

It is basically iteratively updating the values of ***a*** and ***b*** using the value of gradient, as in this **equation (*)**



EPOCH (iteration over training dataset)

10 000 **RECORDS**

Calculate the error

Assess loss

Correcting weights (parameter) to decrease error using optimiser (gradient descent for instance)

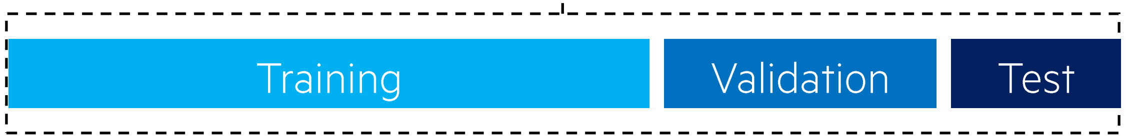
Correcting after training over the full data set (epoch) or after piece of data (batches)

Alpha = learning rate

A(n) = A(n=1) = alpha x d(Error)/dA

DIFFERENT DATASET FROM HISTORICAL DATABASE

dataset that is independent of the training to test generalisation



*algorithm learns relationships between the **features** and the **target** variable.*

*dataset of examples used to **tune the hyperparameters** of a model (dev set)*

ML/DL DOUBLE SAUCE : HYPERPARAMETERS

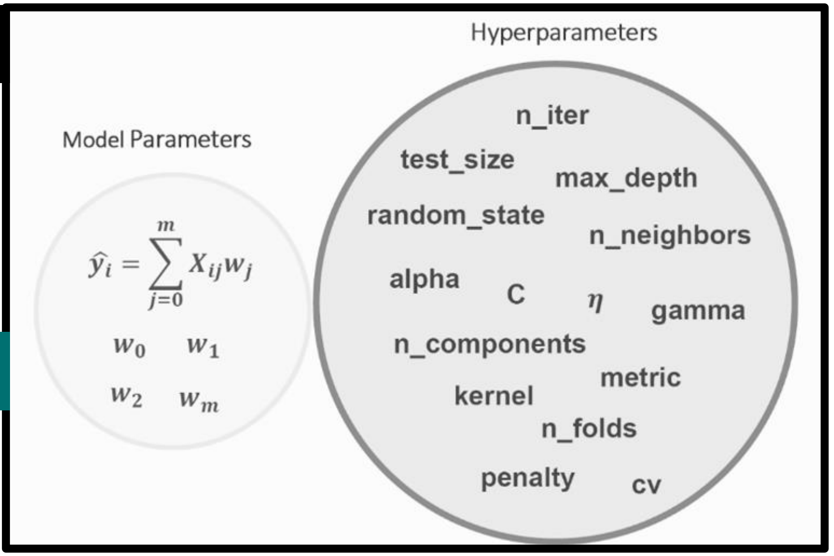
WHAT IF ? What if **we change manually the size** of training dataset and **test data set** ? **Which impact** on my model parameter precision ?

		training	SCORE
Training	Test size = 40%	Epochs = **100**	**60%**
Training	Test size = 20%	Epochs = **100**	**90%**
Training	10%	Epochs = **100**	**70%**

HP are defined as the parameters that are explicitly defined by the user to control the learning process

Hyperparameters are **parameters whose values control the learning process** and determine the values of model parameters

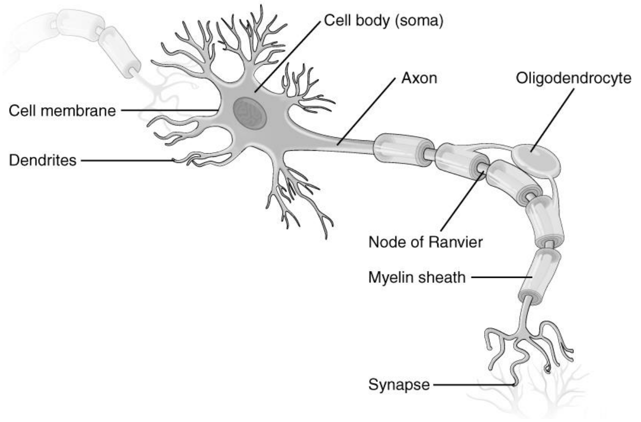
The prefix 'hyper_' suggests that they are **'top-level' parameters** that control the learning process and the model parameters that result from it.



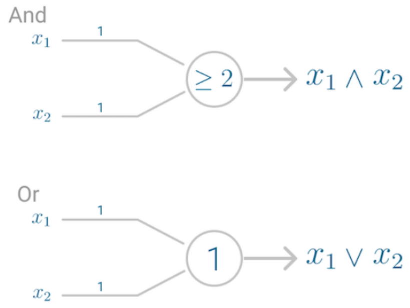
(MODEL) PARAMETER ?	**HYPERPARAMETER ?**
Model parameters are configuration variables that are internal to the model, and a model learns them on its own. These are usually not set manually	*Hyperparameters are those parameters that are explicitly (manually or thru HPO) defined by the user to control the learning process*

DEEP LEARNING : ZOOM

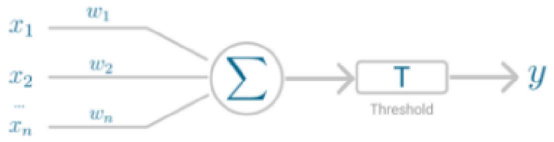
How to Learn ? Imitating Brain ?



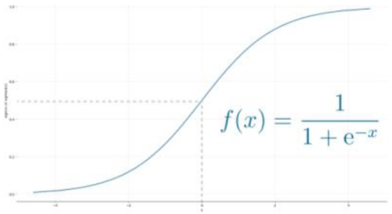
Artificial Neuron = first application = **logic gate**



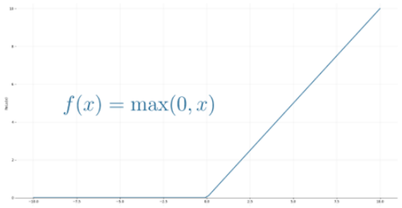
Then, **Perceptron** = algorithm that could learn the weights in order to generate an output.



(For classification for image perception)

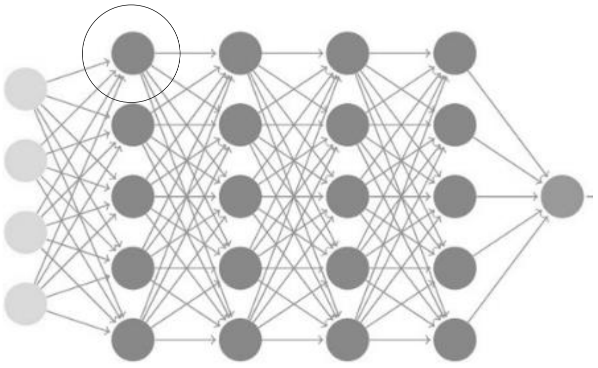


Sigmoid function (Image by author).



ReLU function. (Image by author)

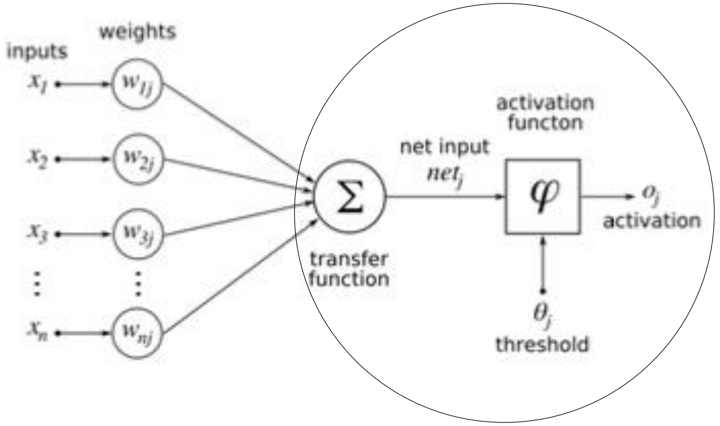
Artificial Neuron Network (ANN)



MLP (**Multi Layer Perceptron**) is one kind of neural networks, where the activation function is sigmoid, and error term is cross-entropy(logistics) error

MLP is one of the several kinds of **Artificial Neuron Networks** (ANN)

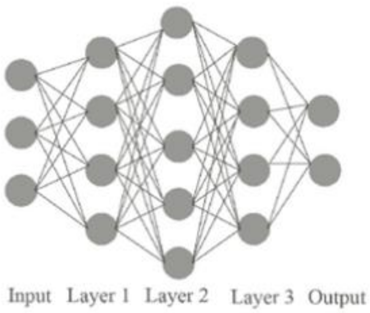
MLP is fully connected Feed Forward (FF) network but you can find CNN, RNN, …



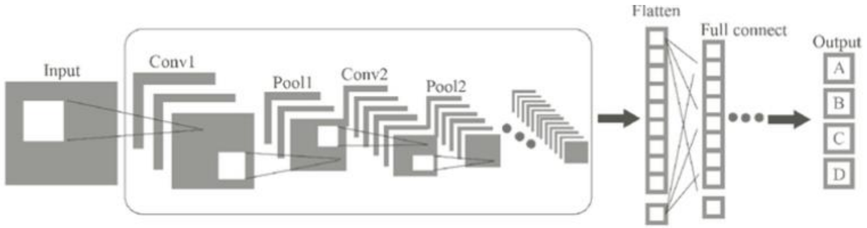
DNN : DIFFERENT ARCHITECTURES FOR DIFFERENT NEEDS

There are many types of AI or deep learning models. For natural language processing (NLP) we will turn to language models : RNN, Encoders, Transformers

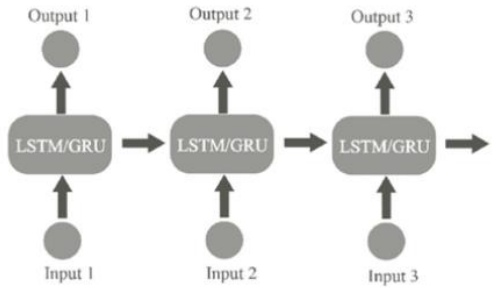
DNN (FFN) : FOR TABULAR DATA

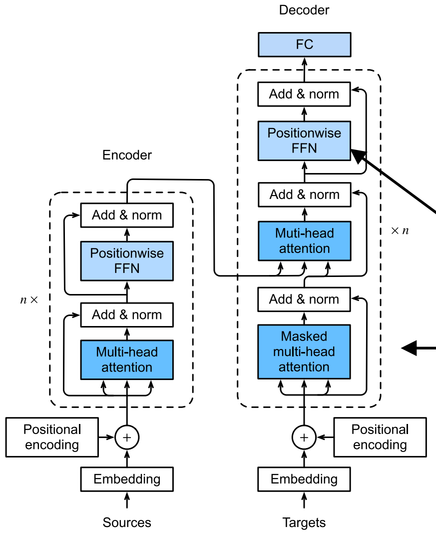


CNN : FOR IMAGES / VIDEOS



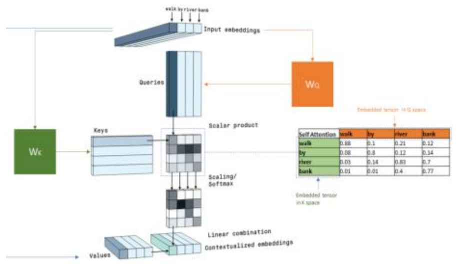
RNN : FOR TIME SERIES AND TEXT





LANGUAGE MODEL / LARGE LANGUAGE MODEL (LLM)

If it's predicting the next word in the sequence, it's called **next-token-prediction**; if it's predicting a missing word in the sequence, it's called **masked language modeling**.



Text Corpus		Task: Predict from past
Nothing is impossible. Even the word impossible says I'm possible	→	Nothing Nothing is Nothing is impossible ...

Randomly masked	A quick [MASK] fox jumps over the [MASK] dog
Predict	A quick brown fox jumps over the lazy dog

NLP: WHAT IS NLP ? BASICS

DEFINITION

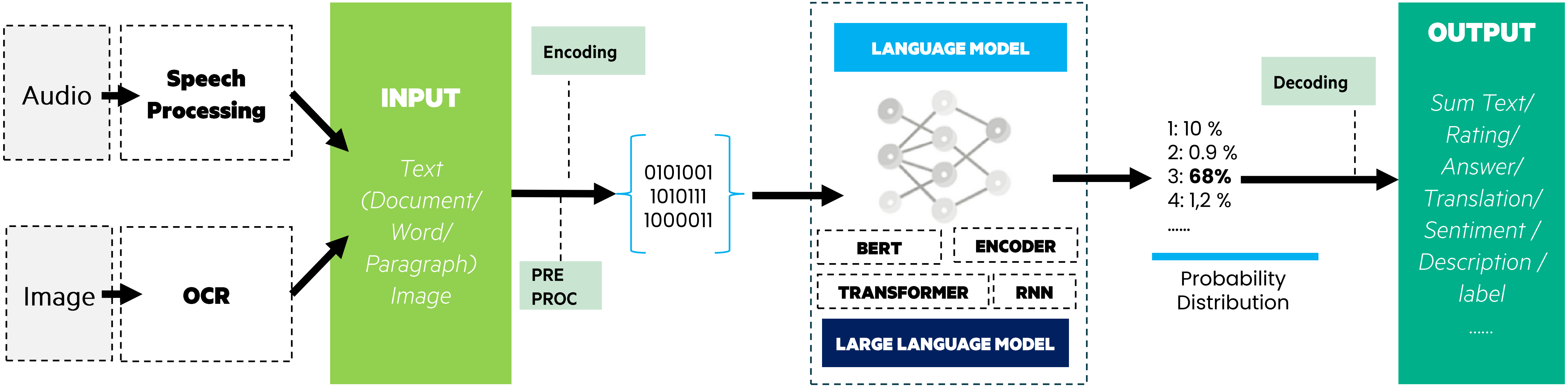
“ NLP strives to build machines that **understand and respond to text or voice data** – and respond with text or speech of their own – **in much time the same way humans do**” (by IBM)

USE CASES

NLG : Natural Language Generation

NLU : Natural Language Understanding

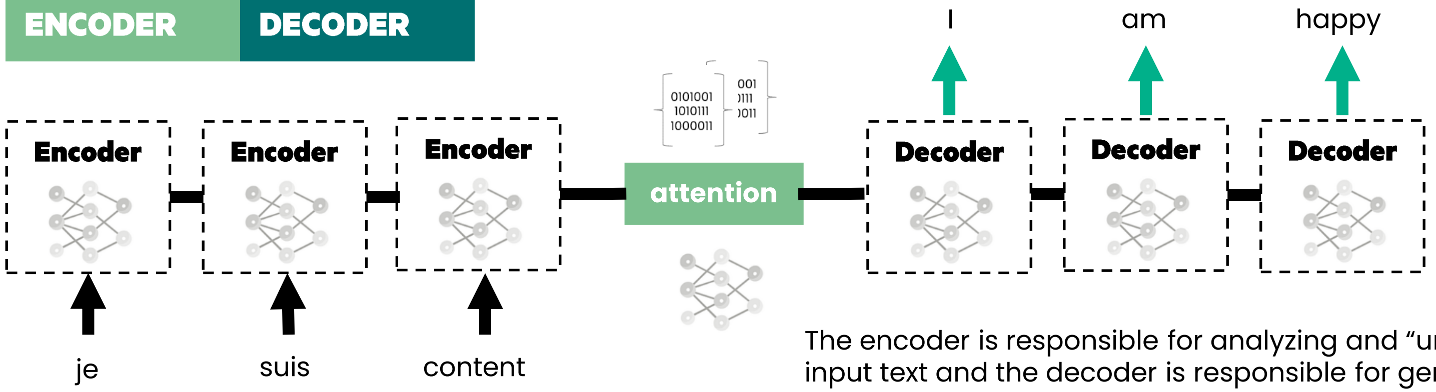
Information retrieval	query (+corpus)	→ document
Information extraction	query (+corpus)	→ fact (tuple)
Machine translation	source text	→ translation
Speech recognition	sounds	→ words
Question answering	question	→ answer
Summarization	text	→ summary
Conversational agents	prompt	→ response (a



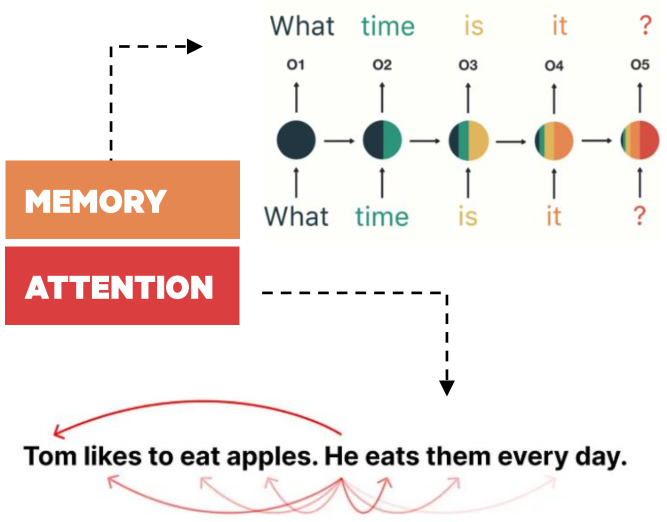
TRANSFORMER: ATTENTION IS ALL YOU NEED

HUMANS : Reading and capabilities to understand concepts and the world around us are all about : MEMORY and ATTENTION

ENCODER DECODER



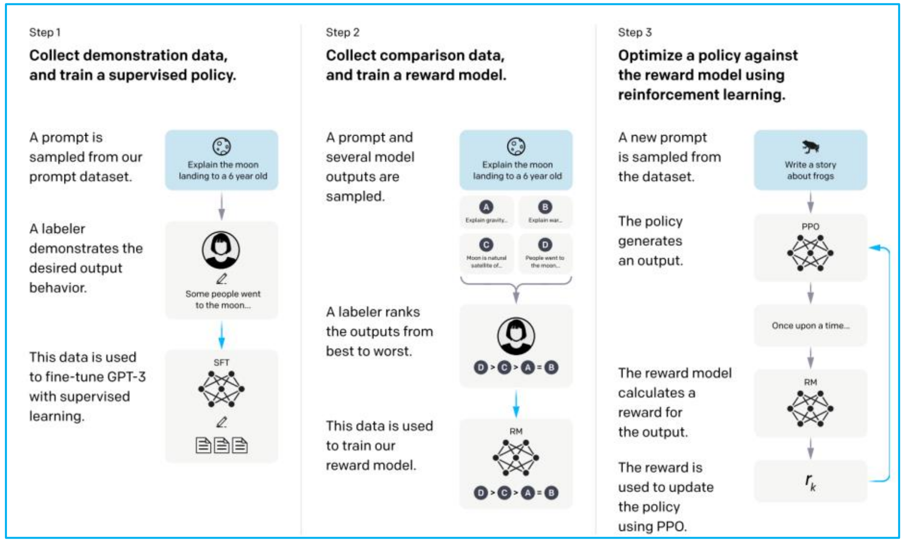
The encoder is responsible for analyzing and "understanding" the input text and the decoder is responsible for generating output

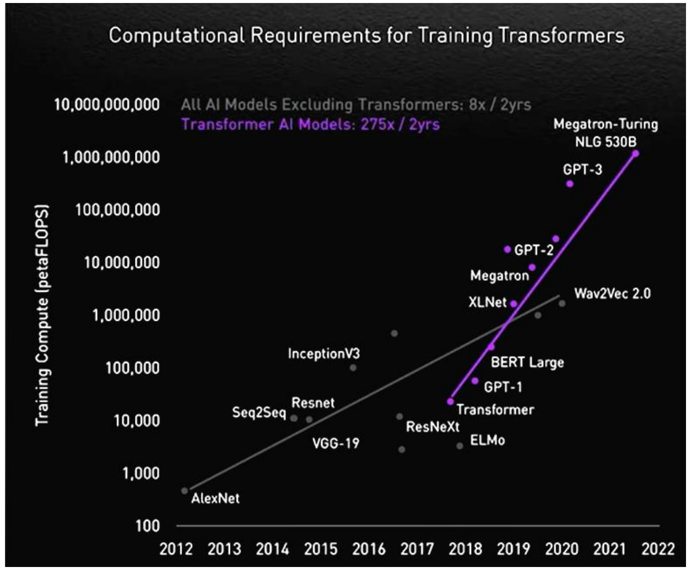


LLM CHATGPT

GPT 3 training on :

- **570 GB** of data from books, wikipedia, research articles, webtexts, websites ~ 300 billion words were fed into the system / 800 GB of memory to train it
- **175B of parameters , 2048 tokens as input**
- 96 decoder layers
- Training hardware: Access to a supercomputer with ~10,000 GPUs and ~285,000 CPU cores.
- **= 5M$ and 2 years training**





LLM : « CHATGPT » COMPETITION

- **OpenAI**
  - ChatGPT — 175B
  - Microsoft — 1B + 10B
- **BAIDU**
  - ERNIE — 260B
  - GA : Q2 — voice
- **Anthropic**
  - CLAUDE — xxxxB
  - GOOGLE — 10%-300m
- **META**
  - LaMDa — 137B
  - xxxx — 1B + 10B
- **GOOGLE**
  - BARD — xxx
  - GA : Q3 — > 2021
- **ALPEH ALPHA**
  - LUMINOUS — 200B
  - EUROPEAN — 5 LANGUAGES
- **Hugging Face**
  - BLOOM — 176B
  - Jean ZAY (FR) — > 59 languages
- **Nvidia**
  - Megatron — 530B
  - Microsoft — > MT NLG

ALPEH ALPHA: LUMINOUS, EUROPEAN CHATGPT

Copywriting

You are in need of some marketing material? Briefly describe your product or service in the text field below. Luminous will then create an advertising slogan from the information provided!

Text:

Olympique de Marseille also known simply as Marseille or by the abbreviation OM is a French professional men's club based in Marseille.Founded in 1899, the club plays in Ligue 1 and have spent most of their history in the top tier of French football. The club has won ten Ligue 1 titles, ten Coupes de France and three Coupes de la Ligue. In 1993, coach Raymond Goethals led the team to become the first and only French club to win the UEFA Champions League, defeating Milan 1–0 in the final, the first under the UEFA Champions League branding of the tournament.

Slogan:

We're not just the best team in France, we're the best team in Europe. #om

View Settings Open in Playground Reset Submit

Text to Table

Structuring information is usually a tedious task. That's why Luminous can do it for you. Just enter any text and we will create a table.

Text:

Marseille's home ground is the 67,394-capacity Stade Vélodrome in the southern part of the city, where they have played since 1937. The club has a large fan-base, having regularly averaged the highest attendance in French football. Marseille's average home gate for the 2018–19 season was 50,361, the highest in Ligue 1. The stadium underwent renovation from 2011 to 2014, increasing its capacity to 67,000 ahead of France's hosting of UEFA Euro 2016. In 2015, the club was ranked 23rd globally in terms of annual revenue, generating €130.5 million.

Table:

Home ground	Capacity	Attendance	Revenue
Stade Vélodrome	67,394	50,361	€130.5 million
Stade Velodrome	67,394	50,361	€130.5 million

View Settings Open in Playground Reset Submit

Tasks	Sample Prompts
Text Generation	Generate a story about a futuristic world.
Question Answering	Who was the first President of the United States?
Summarization	Summarize the plot of the movie "Inception".
Translation	Translate "Bonjour" from French to English.
Conversational Modeling	Can you have a conversation with me about movies?
Text Classification	Is this article about sports or politics?
Text Completion	Complete the sentence: "The sun rises in the east and sets in the..."

Protestors holding up a sign that reads:

Completion

We want one Germany

ALEPH ALPHA

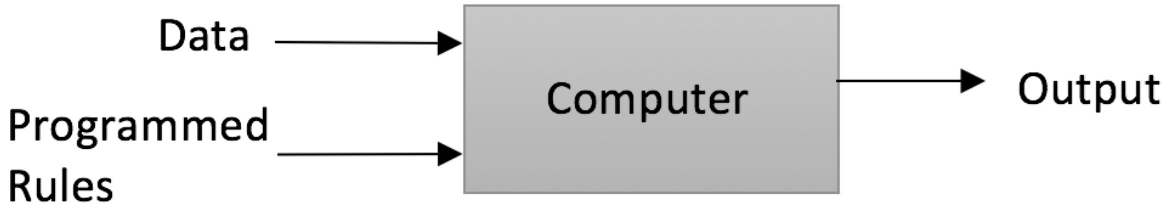
Tasks	Sample Prompts
Sentiment Analysis	What is the sentiment of the following text: "I love this product."
Named Entity Recognition	Who is the author of "To Kill a Mockingbird"?
Text Similarity	Which is more similar: "dog" or "cat"?
Text-to-Speech	Can you convert the following text to speech: "Hello, world."
Speech-to-Text	Can you transcribe the following audio clip: "Hello, world."
Image Captioning	Generate a caption for this image of a sunset over a beach.

Tasks	Sample Prompts
Image Classification	What type of animal is in this picture?
Text-to-Code	Generate code for a program that prints "Hello, world." in Python.
Code Generation	Generate a function in Python that calculates the factorial of a number.
Chatbot	Can you serve as a customer support agent and answer my questions?
Poem Generation	Generate a poem about love.
Song Lyrics Generation	Generate the lyrics for a sad love song.

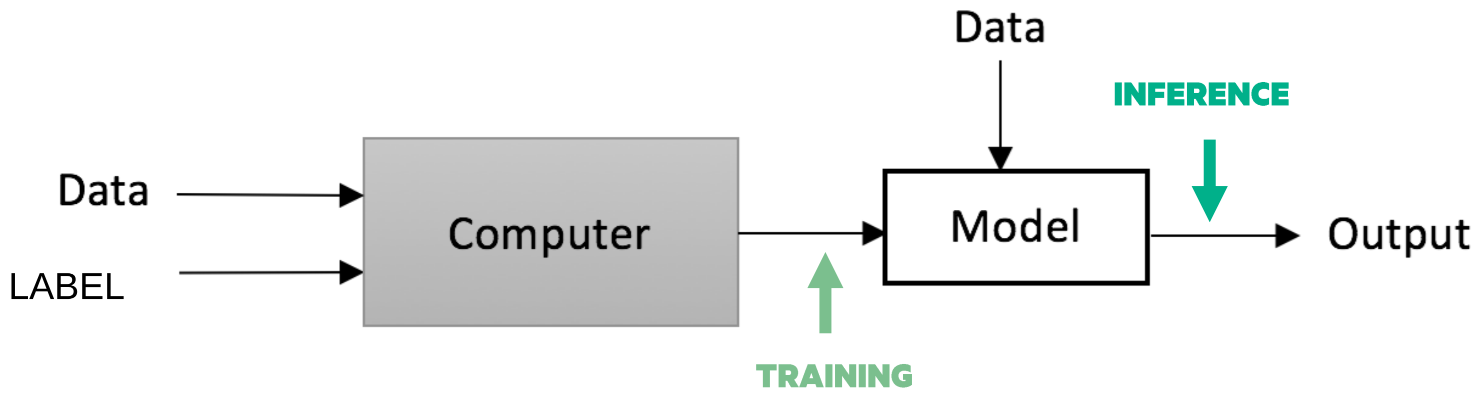
ML/DL POINT : CHANGING PARADIGM

Difference traditional programming vs AI

From deterministic programming to probabilistic programming



Traditional Programming



Machine Learning

ML/DL : WHAT IS DEEP LEARNING ?

Task: predict house prices based on school rating (s), # of bedrooms (be), # of bathrooms (ba), ft² (f)

Rule-based AI

SME defines a set of rules, these rules are explicitly programmed:

```
if (s==9 and be==2
    and ba==2 and f==1000)
then
    price = $1000000;

else if (…) then …
else if (…) then …
```

Machine Learning

Collect a “labeled dataset”: example of houses with prices

```
House 1: s = 9, be = 2,
         ba = 2, f = 1000,
         price = $1000000
House 1: s = 4, be = 2,
         ba = 1, f = 700,
         price = $600000
```

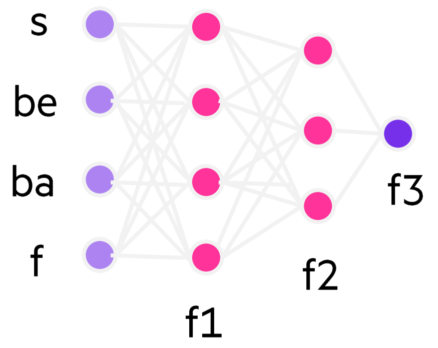
Define a function (model):

```
F: (s,be,ba,f) -> price
price = F(s,be,ba,f) =
w1*s + w2*be + w3*ba + w4*f
```

Train a model: run a program to find the best values of w1, w2, w3, w4

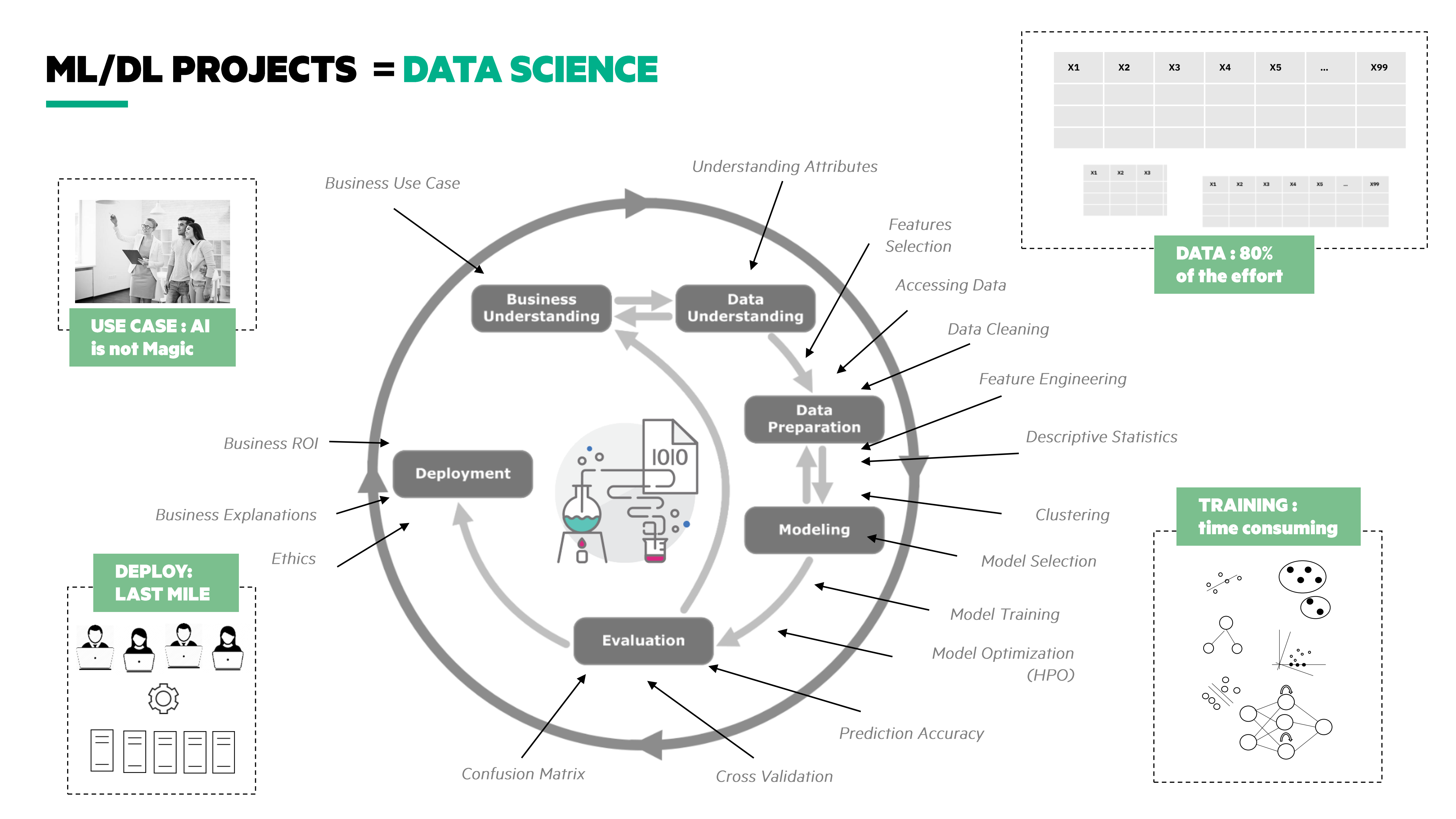
Deep Learning

As traditional ML, but a function is more complex – a function of functions




```
F: (s,be,ba,f) -> price
price = F(s,be,ba,f) =
f3(f2(f1(s,be,ba,f)))
```

AI CLIENTS AND CHALLENGES


ML/DL PROJECTS = DATA SCIENCE
X1 X2 X3 X4 X5 ... X99
DATA : 80% of the effort
USE CASE : AI is not Magic
Business Use Case
Understanding Attributes
Features Selection
Accessing Data
Data Cleaning
Feature Engineering
Descriptive Statistics
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Business ROI
Business Explanations
Ethics
Clustering
Model Selection
Model Training
Model Optimization (HPO)
Prediction Accuracy
Cross Validation
Confusion Matrix
TRAINING : time consuming
DEPLOY: LAST MILE

ML/DL CHALLENGES : DATA SCIENCE IS A TEAM WORK

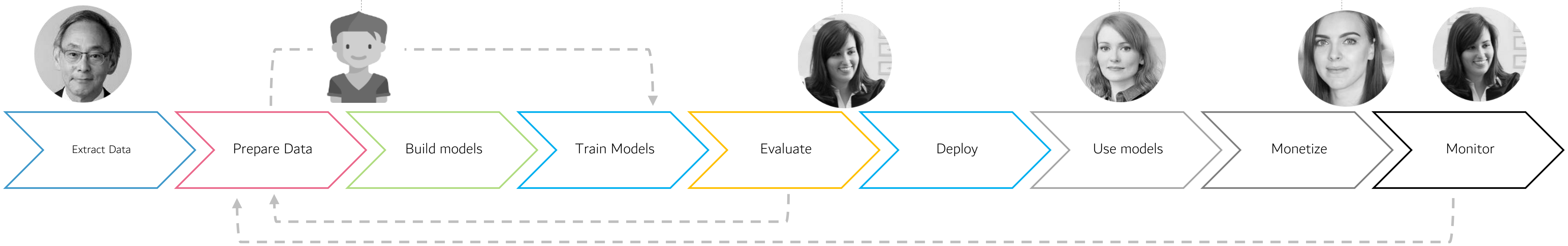
Data Engineer
Design install and maintain data management systems

Data Scientist
Understand business problem
Builds and Trains Models

Business Process Owner
Designs biz processes, leads a team.
Continuous Improvement & Health

App Dev
Builds & supports apps
Fairness & Robustness

Biz Analyst – Ops for AI
Ensures ongoing health and transparency of AI in production



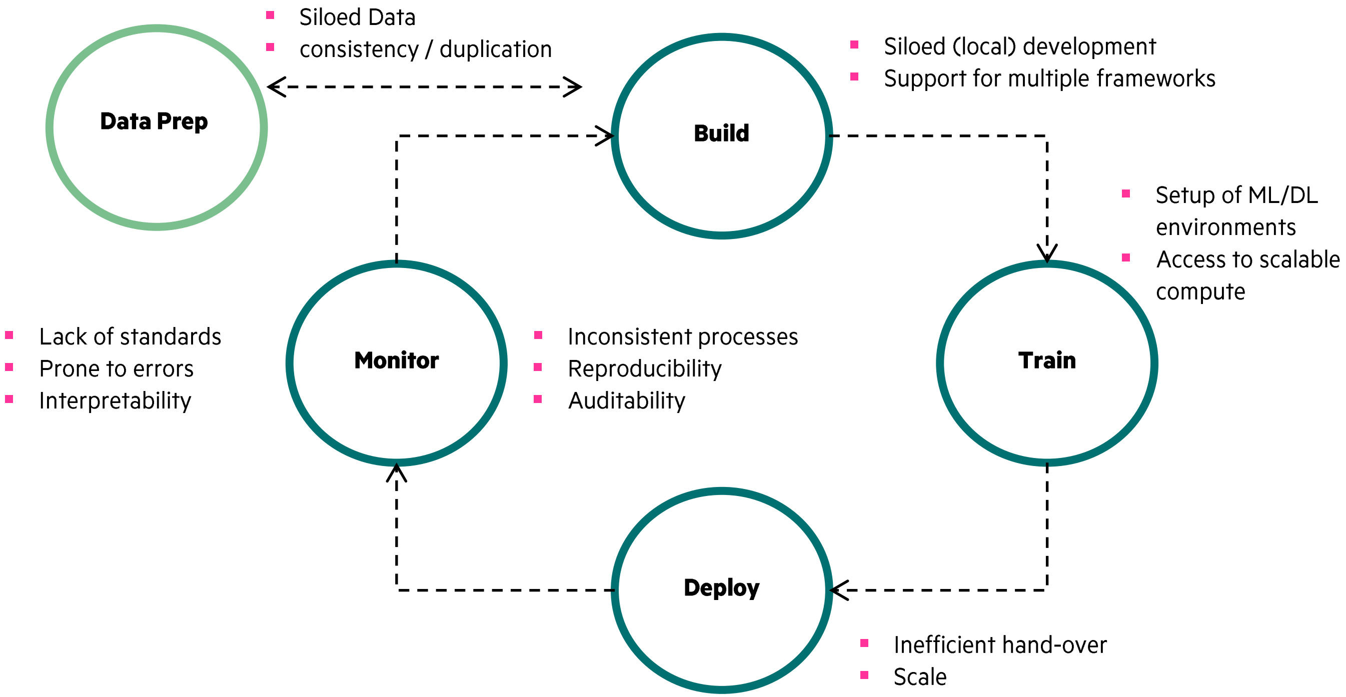
IT TEAM / INFRASTRUCTURE

Data Architect
Create blueprint for Data management Systems
Define, Integrate and maintain data sources/flow

Project Managers
Team Leaders

Executive Managers
Business Sponsors

ML/DL CHALLENGES : MANY TOOLS / PRODUCTION / BUSINESS



Machine learning/stats	Scientific computing	Distributed systems	Analytics	Web
Data scientists/modelers	Research/computational scientists	Data engineers/architects	Data/business analysts	Developers
R, scikit-learn, PyMC, statsmodels, PyBrain, theano	NUMBA, SciPy, xray, NumPy, HDF, Cython, PyTables	Spark, kafka, mrjob, Luigi, STORM, hadoop	PostgreSQL, pandas, SQLAlchemy, xlwings, Excel	Scrapy, bokeh, Flask, django, jupyter, RDFLib

CUSTOMER AI JOURNEY

60%

Early

- How
- Data
- People

30%

Proof of Concept

- Decide
- Infrastructure
- Other Use cases

10%

Production

- Scale
- Ethics
- Strategy

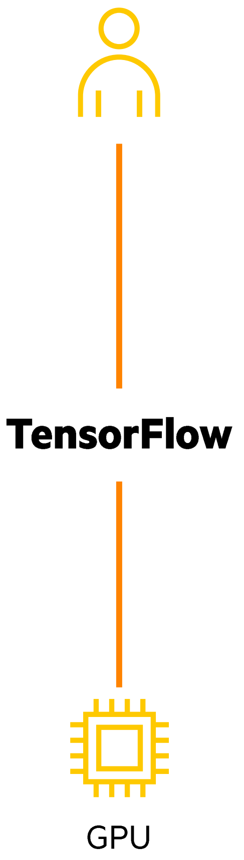
90%

10%

HPE AI NEXT : HPC&AI

AI at scale Challenges

On day one…small and simple



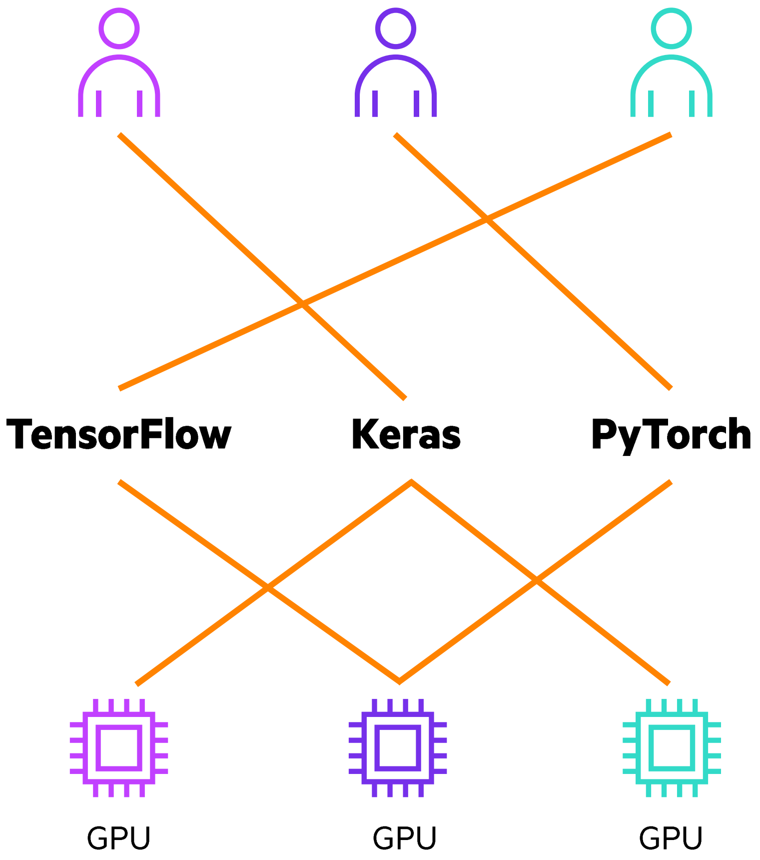
Small experiments on a single machine

Simple infrastructure to manage

No need to share resources or productionize results

Ad hoc experiment tracking is sufficient

On day two….complex and complicated



GPU sharing and infrastructure management

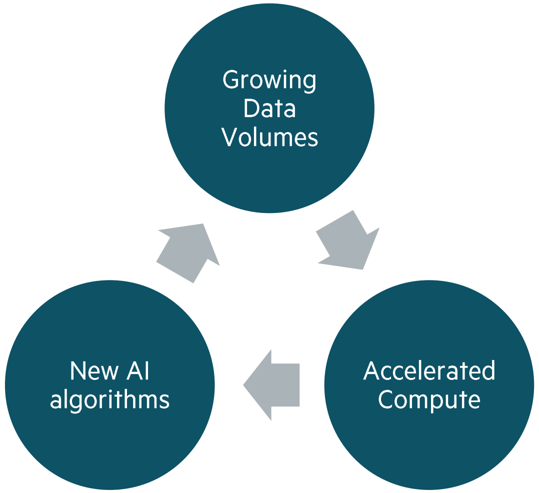
Hyperparameter search ad hoc, time-consuming, often no tool support

Experiment tracking for reproducibility and collaboration

Distributed training; hard to configure, fragile, not multi-user by default

AI at scale market requires new software platform

Emerging AI Mega Trends



Difficult Ecosystem Choices

DIY using hundreds of point solutions – most without commercial support

or

Adopt CSP or accelerator-vendor provided technologies that create lock-in

Few Successful Implementers

Big Tech Companies ✓
(e.g., Alphabet, Meta)

AI Native Companies ✓
(e.g., Open AI, Cruise, Aleph Alpha)

Majority of companies ✗

Market Requirements

Edge-to-cloud AI lifecycle software based on open technologies and built for scale

End-to-end capabilities:
- Data Acquisition & Preparation
- Development & Training
- Deployment & Inference
- Governance & Performance Management

Common user experience with deployments from edge to cloud

Optimized performance across heterogenous compute

AI-at-Scale Platform

Industry-Specific Workload Solutions

Curated solutions, training & inference-related platforms, and reference configurations for key industry workloads

Manufacturing	Financial Services & Insurance	Health Care & Life Sciences	Government

AI Data Management at Scale

Manage and data lineage, features, augmentation and pipelines in a high performance, distributed fashion

AI Development & Training at Scale

Train large-scale machine learning models faster while hiding the complexity of underlying heterogeneous infrastructure

AI Deployment & Inference at Scale

Deploy & manage models and run inference on heterogeneous infrastructure from data center to edge

Optimized Infrastructure

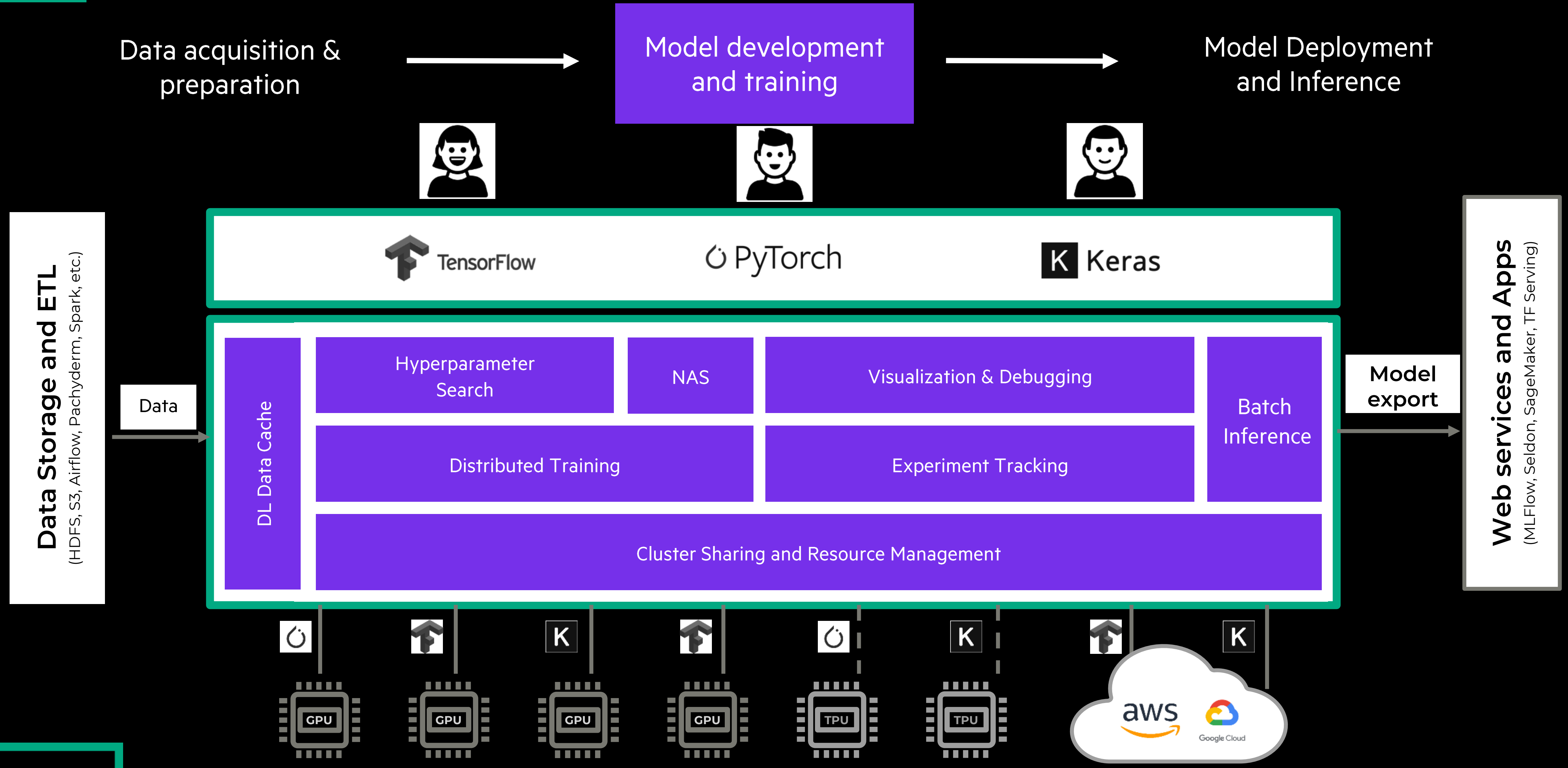
Choice of optimal infrastructure for any at Scale AI workload

AI Compute	AI Storage	AI High-Performance Fabric	AI Accelerators

Across On-Premises, Private Cloud and Public Cloud

HPE ML DEVELOPMENT ENVIRONMENT

DETERMINED AI TRAINING PLATFORM



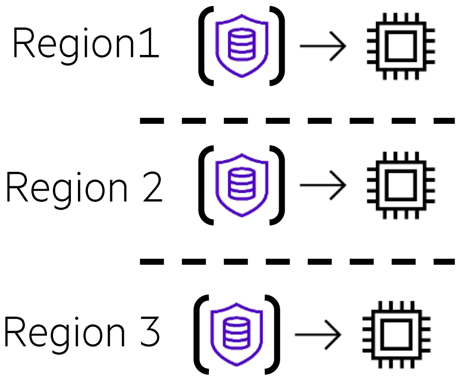
HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT AND DETERMINED OPEN SOURCE—COMPARED

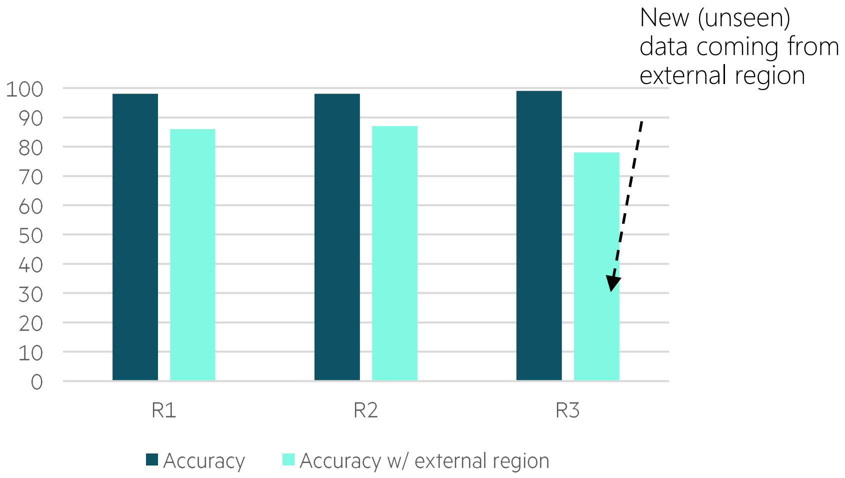
	Open source Determined Software	HPE Machine Learning Development Environment
Distributed training	✓	✓
Model optimization	✓	✓
Metadata tracking	✓	✓
Cluster resource management	✓	✓
GPU cost management	✓	✓
Collaboration and experiment tracking	✓	✓
Security		
• Single sign on (SSO)	X	✓
• Automated user provisioning	X	✓
Premium dedicated support	X	✓

HPE SWARM LEARNING

SWARM LEARNING INTRODUCTION: CONTEXT

Localized and disconnected

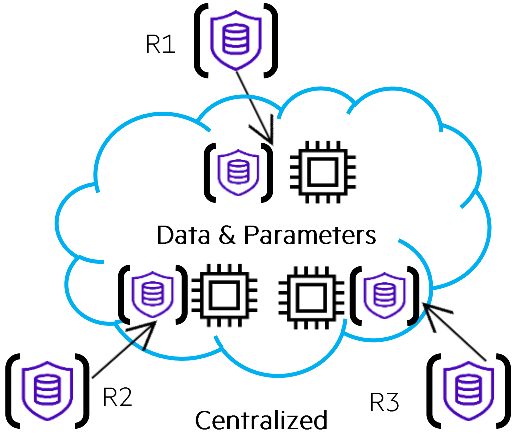


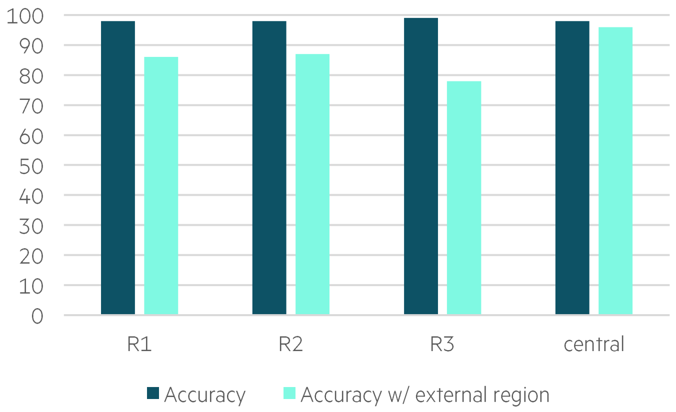


Biased Model due to localized data

A machine learning model is only as good as the (regional) data it is fed

Centralized





Good accuracy .... BUT

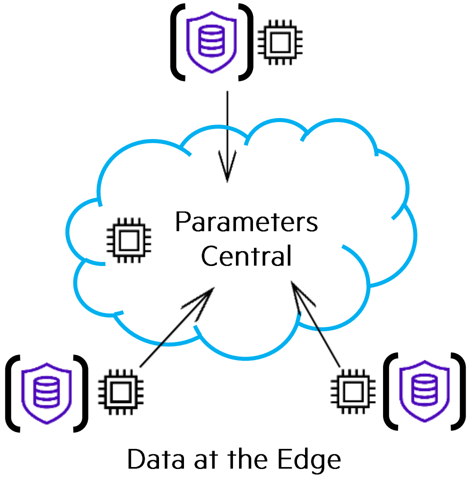
LOW EFFICENCY : Multiple sites send raw data over the network; need high bandwidth

Lack of DATA PRIVACY: Privacy acts like GDPR prevent moving data to a central repository

LACK OF COLLABORATION: Data generated in silos (e.g. data centers, sensors, vehicles)

LACK OF MONETIZATION: Data is new currency – owners look for ways to monetize the data

SWARM LEARNING INTRODUCTION: SOLUTION ?



Federated Learning

- Models are trained at Edge; Only Learning (parameters) are shared
- Data stays at Edge – ensures data privacy
- Parameters (Weights) are merged by Central Coordinator

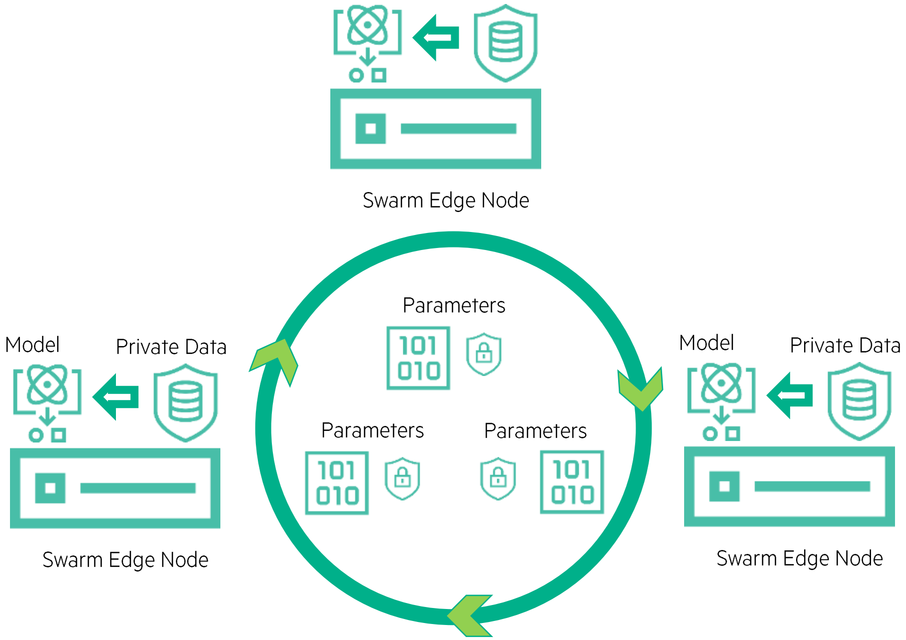
Federated Learning solves privacy issues

But it still need a central custodian for coordination and merging the local learnings

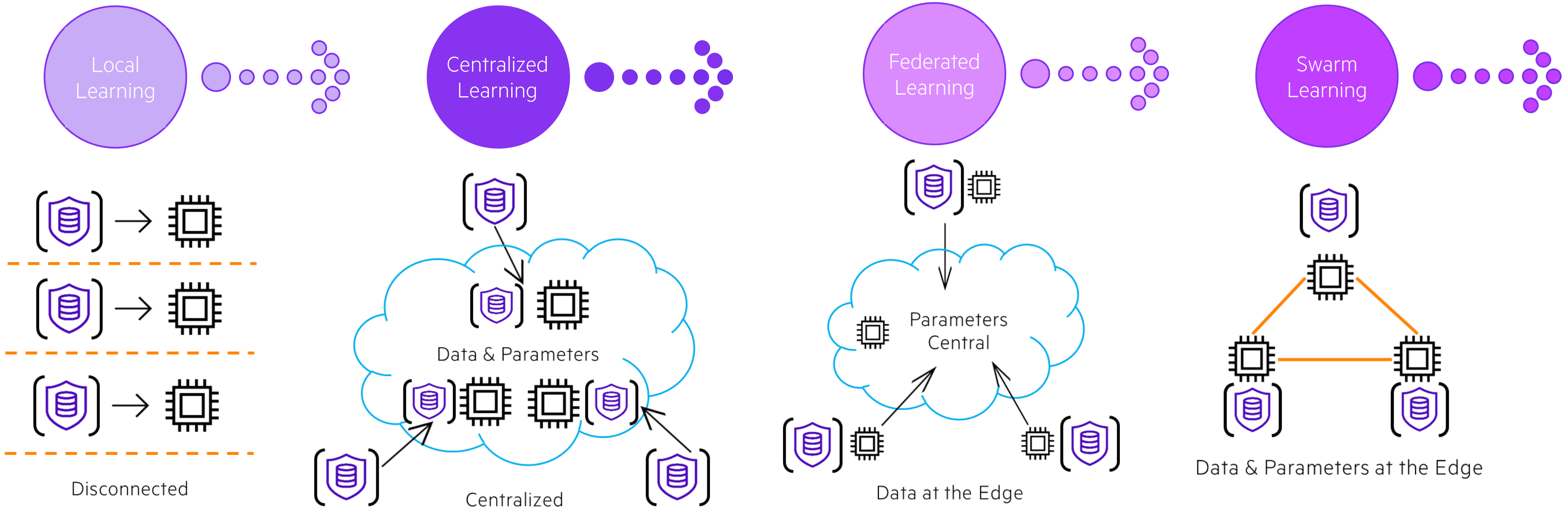
Swarm Learning

Democratic Machine Learning

- Equal and like-minded partners in the network
- Ownership of the data remains local
- Data protection and data security solved locally
- Collaborative learning makes model less susceptible to bias

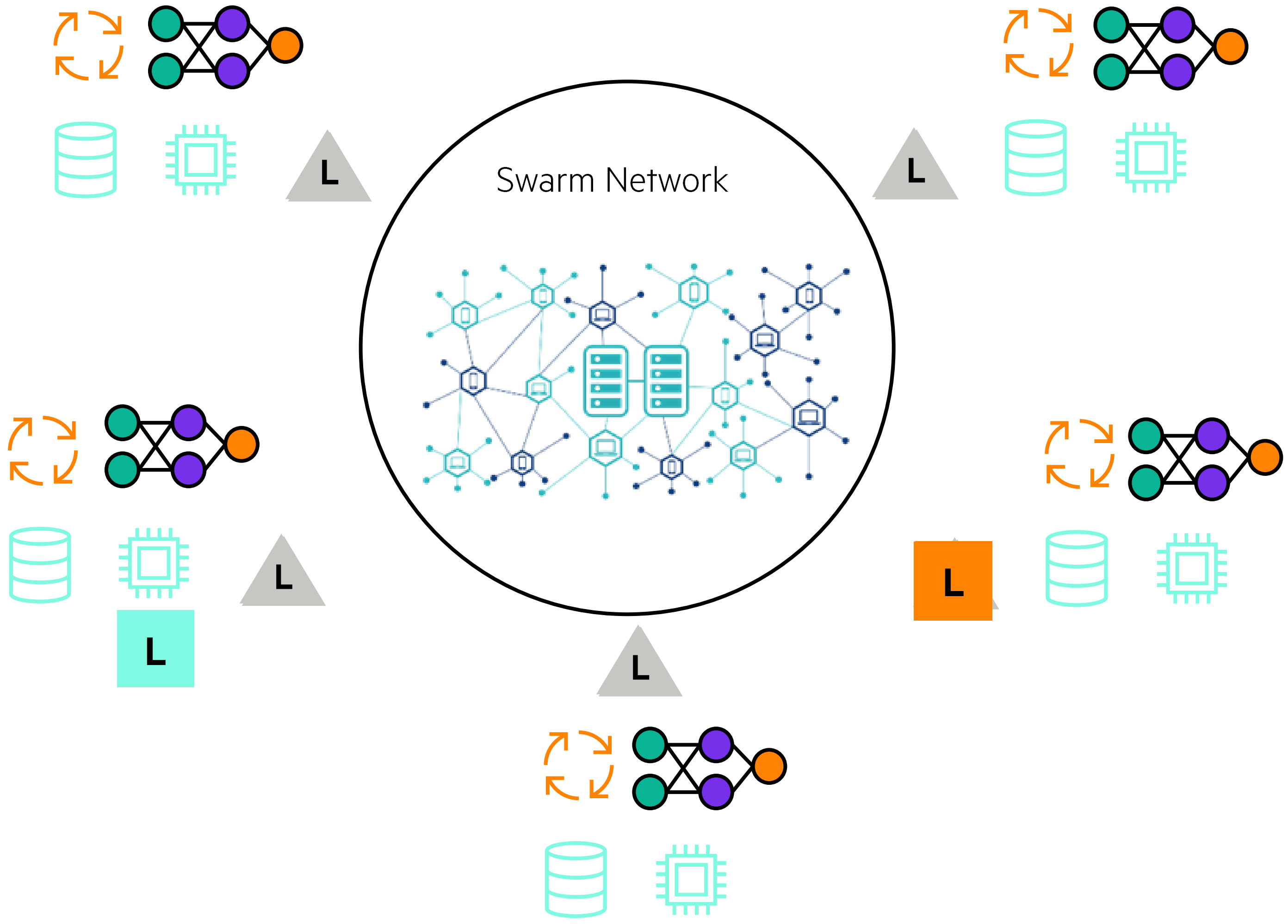


AI ON THE EDGE: MACHINE LEARNING JOURNEY



Swarm Learning enables privacy-preserving, secure and collaborative machine learning by treating all participants equally

SL INTERNALS: HIGH LEVEL PRINCIPLE



0. On boarding (done offiline)

1. **Register**
   Nodes register to Swarm Network and receive ML model

2. **Train**
   Nodes train the model on local data for a time-window (epoch)

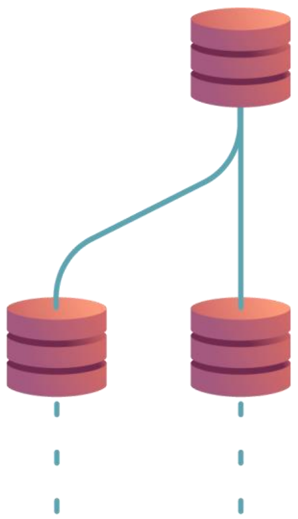
3. **Merge**
   Nodes share and merge the trained models

4. **Repeat**
   Repeat 1 & 2 till desired accuracy is achieved

PACHYDERM

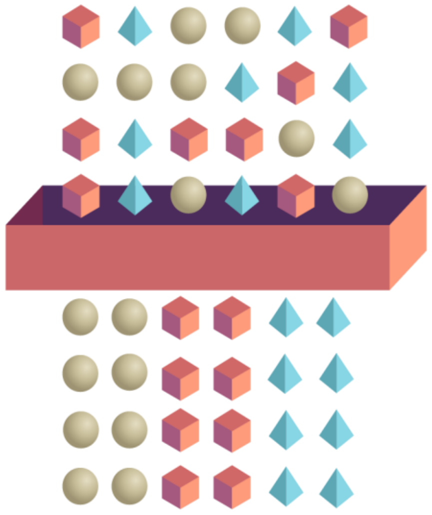
KEY ASPECTS OF MLDM

Introduction to Machine Learning Data Management (aka Pachyderm)



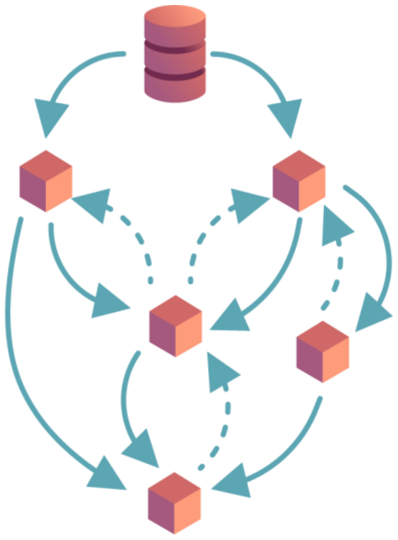
Data Versioning

Manage data with the same production practices as code



Data Pipelines

Developers need to be empowered with choice, not restricted



Data Lineage

Be able to instantly reconstruct any past output/decision

OSS VS. ENTERPRISE

	COMMUNITY EDITION For small teams who prefer to build and support their own software.	Enterprise Edition Organizations that require advanced features and unlimited potential.
Console	✓	✓
Notebook Support	✓	✓
Immutable Data Lineage	✓	✓
Native Data Version Control	✓	✓
Deduplication	✓	✓
Data-Driven Pipelines	16	Unlimited
Parallel Processing (Parallel Workers)	8	Unlimited
Role Based Access Controls (RBAC)	-	✓
Pluggable Auth – Login with your IdP	-	✓
Enterprise Support	-	✓

Hewlett Packard Enterprise

THANK YOU