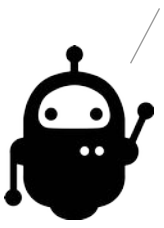
Principes et limites de **l'IA**, pour **soi**, au **travail** et dans la **société**

v1.7

GEPhyX 2025 - Grenoble
Jean-Luc.Parouty@univ-grenoble-alpes.fr

Merci
beaucoup pour
votre invitation !



Formation

Introduction au Deep Learning

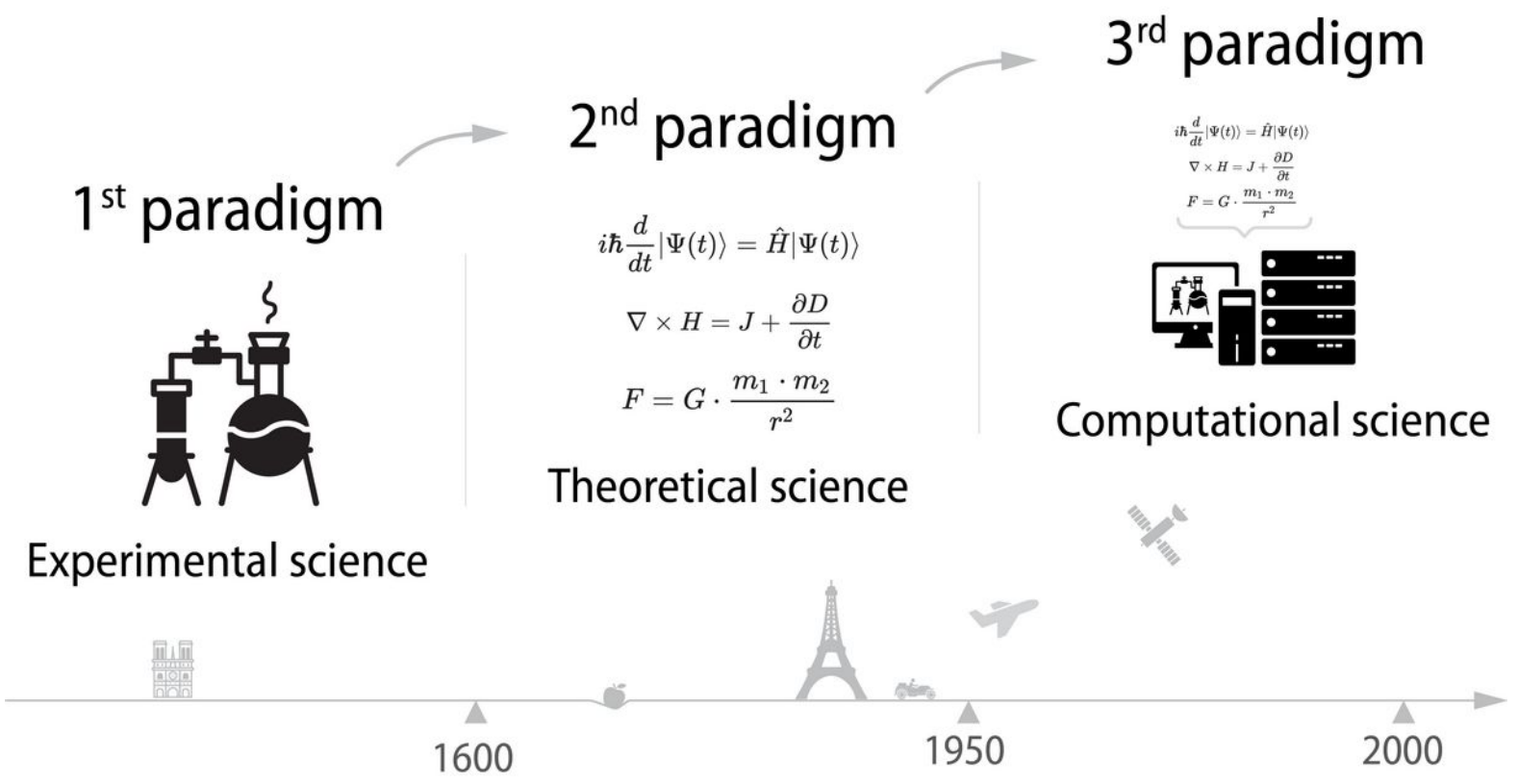
YouTube

→ https://fidle.cnrs.fr

FIDLE est portée par le **CNRS**, l'institut d'Intelligence Artificielle **MIAI Cluster** de Grenoble et l'**Université Grenoble Alpes** (UGA)

[Science]

Scientific paradigms



Scientific paradigms

1st paradigm

Experimental science

2nd paradigm

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$$

$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$

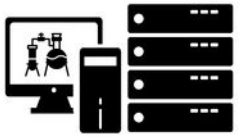
Theoretical science

3rd paradigm

$$i\hbar\frac{d}{dt}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle$$

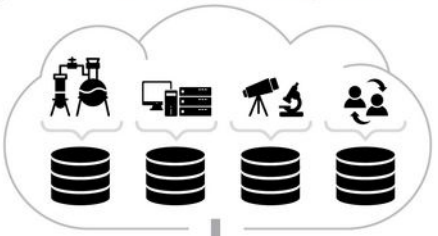
$$\nabla \times H = J + \frac{\partial D}{\partial t}$$

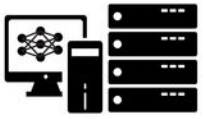
$$F = G \cdot \frac{m_1 \cdot m_2}{r^2}$$



Computational science

4th paradigm[1]



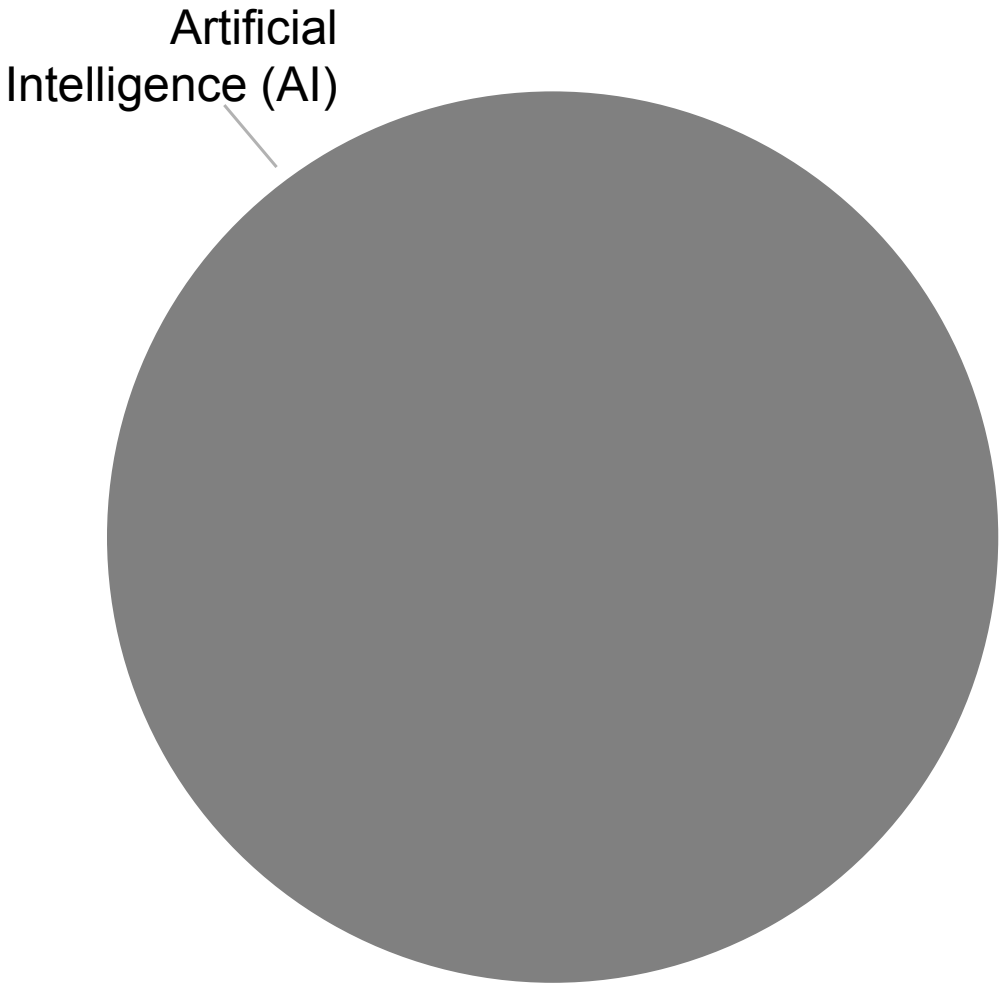


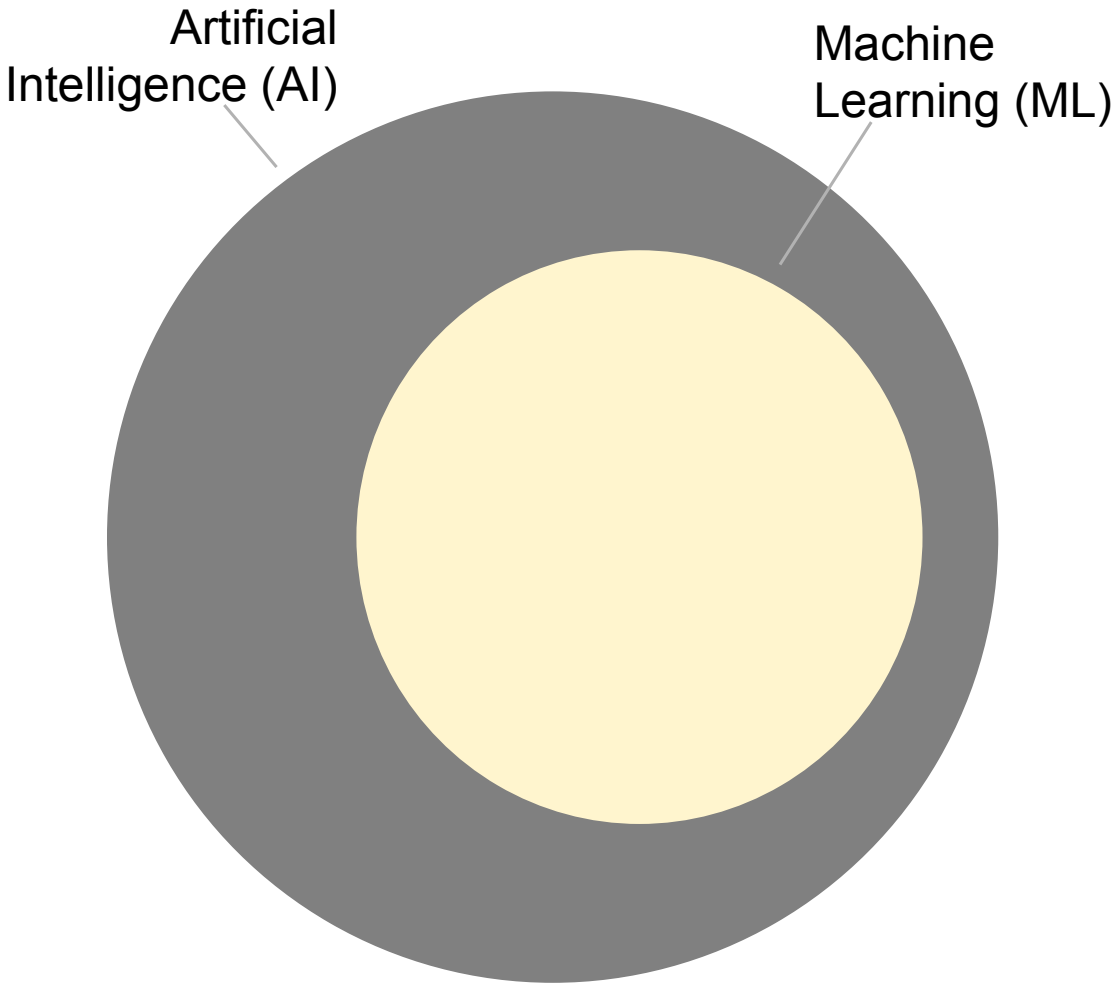
Data-driven science

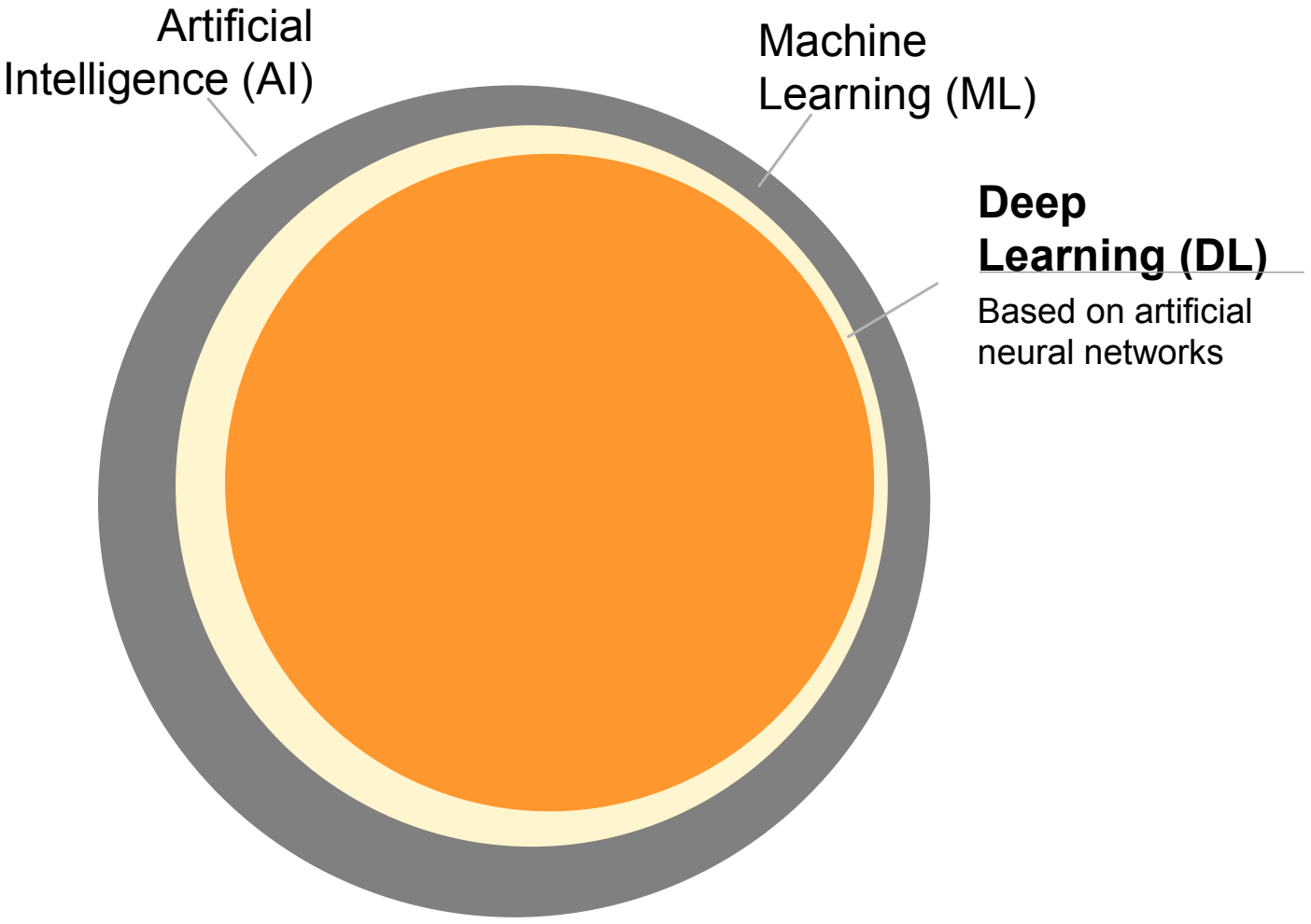
1600 1950 2000

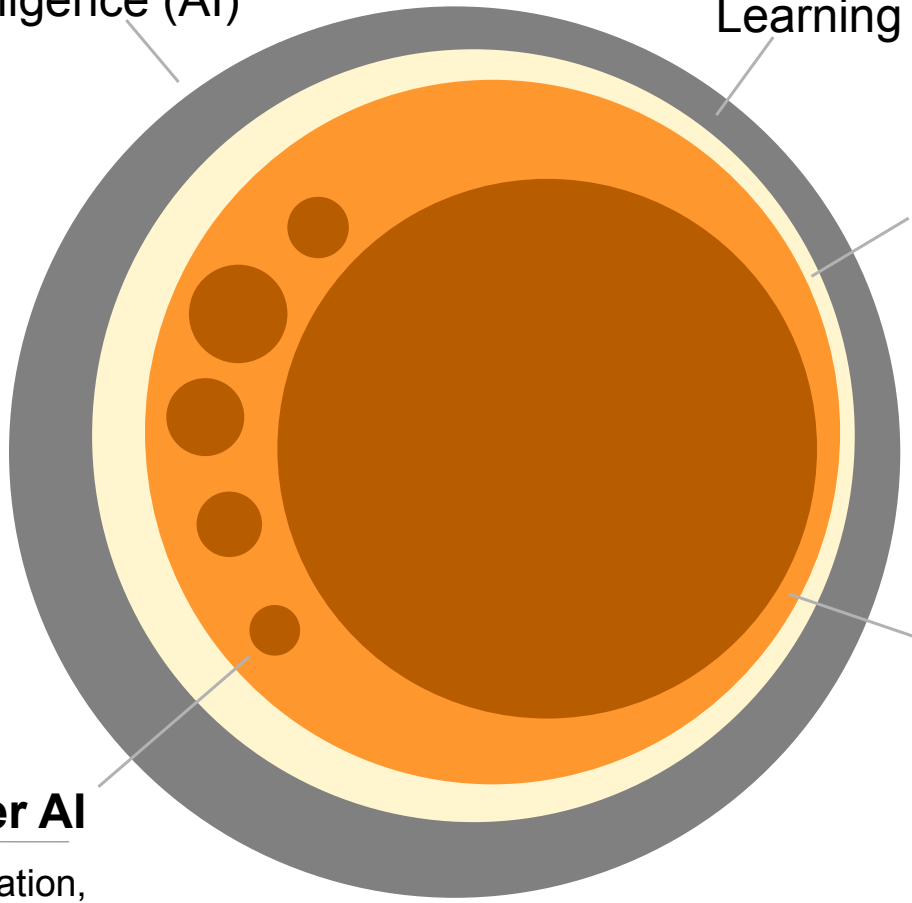
[1] Jim Gray, 2007


[IA]
?


Artificial
Intelligence (AI)


Artificial
Intelligence (AI)
Machine
Learning (ML)


Artificial Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)
Based on artificial neural networks


Artificial Intelligence (AI)
Machine Learning (ML)
Deep Learning (DL)
Based on artificial neural networks
Generative AI
Large Language Model (LLM)
Diffusion Model, etc.
Other AI
Classification, regression, etc.

Science

Social signal learning of the waggle dance in honey bees

Science, 9 mars 2023
https://www.science.org/doi/10.1126/science.ade1702

Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions

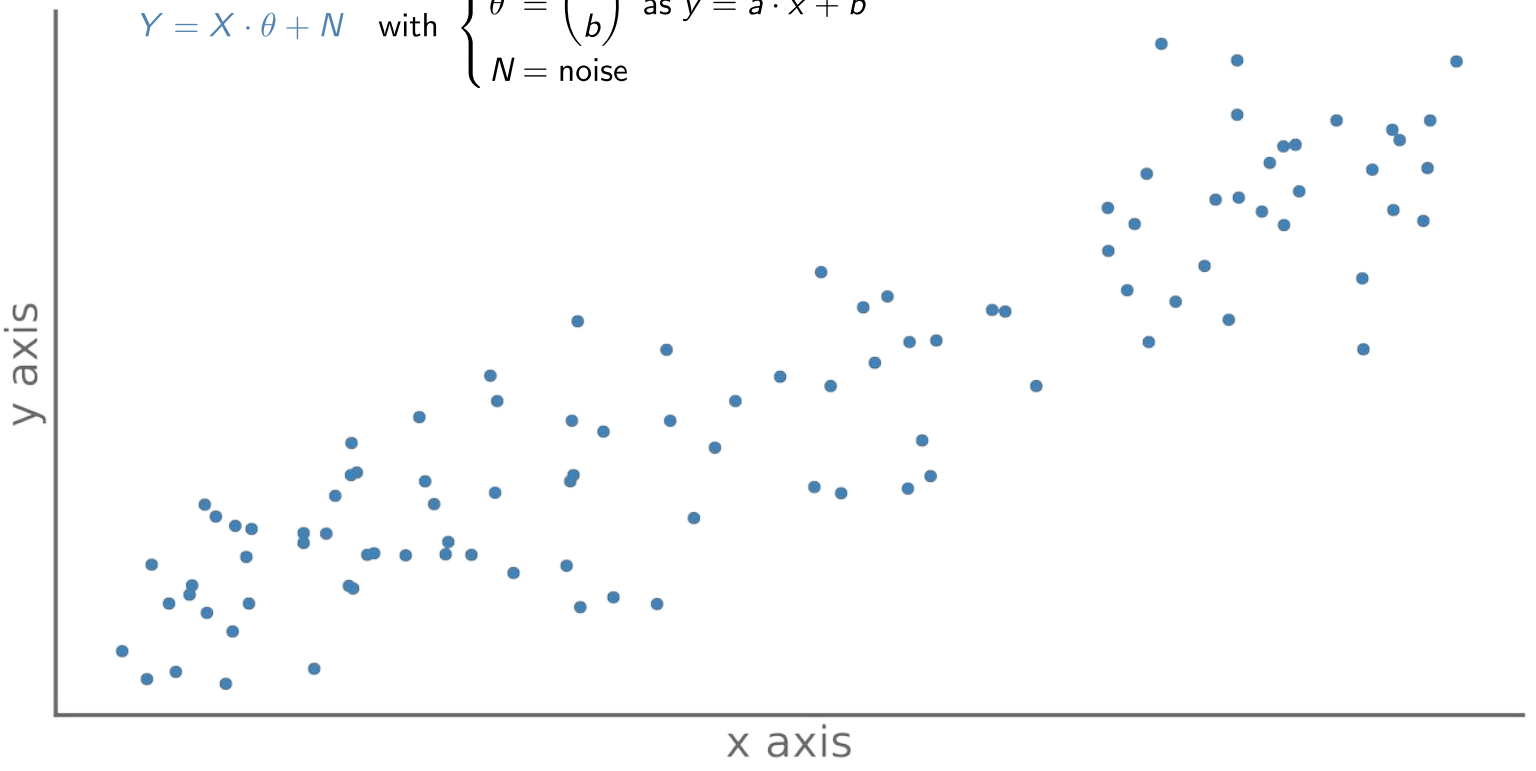
Science, 30 novembre 2018
https://doi.org/10.1126/science.aat1590

[Artificial Neuron]

Linear regression

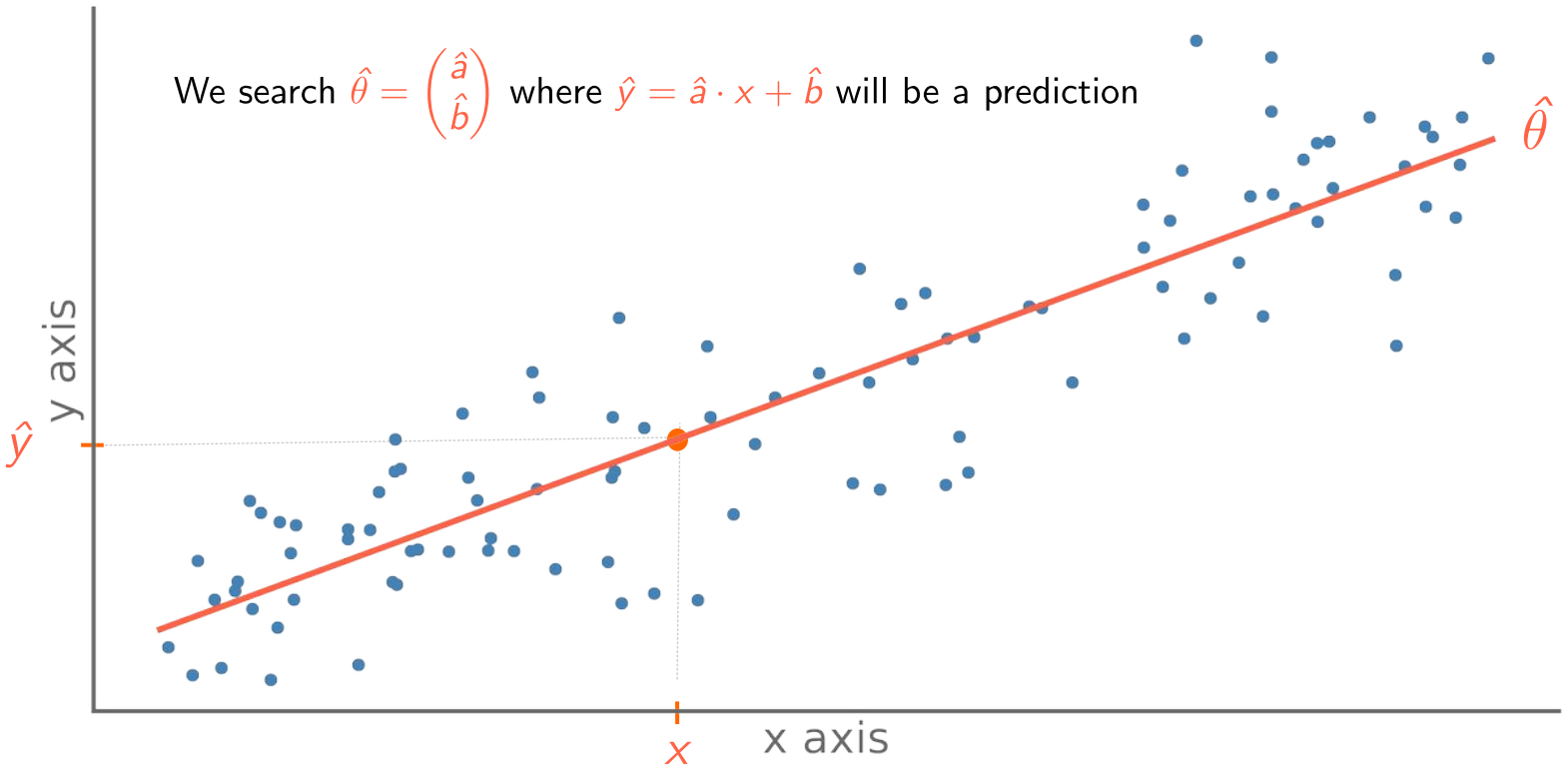
We have a phenomenon, for which we have **observations**

$$Y = X \cdot \theta + N \quad \text{with} \quad \begin{cases} \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ as } y = a \cdot x + b \\ N = \text{noise} \end{cases}$$



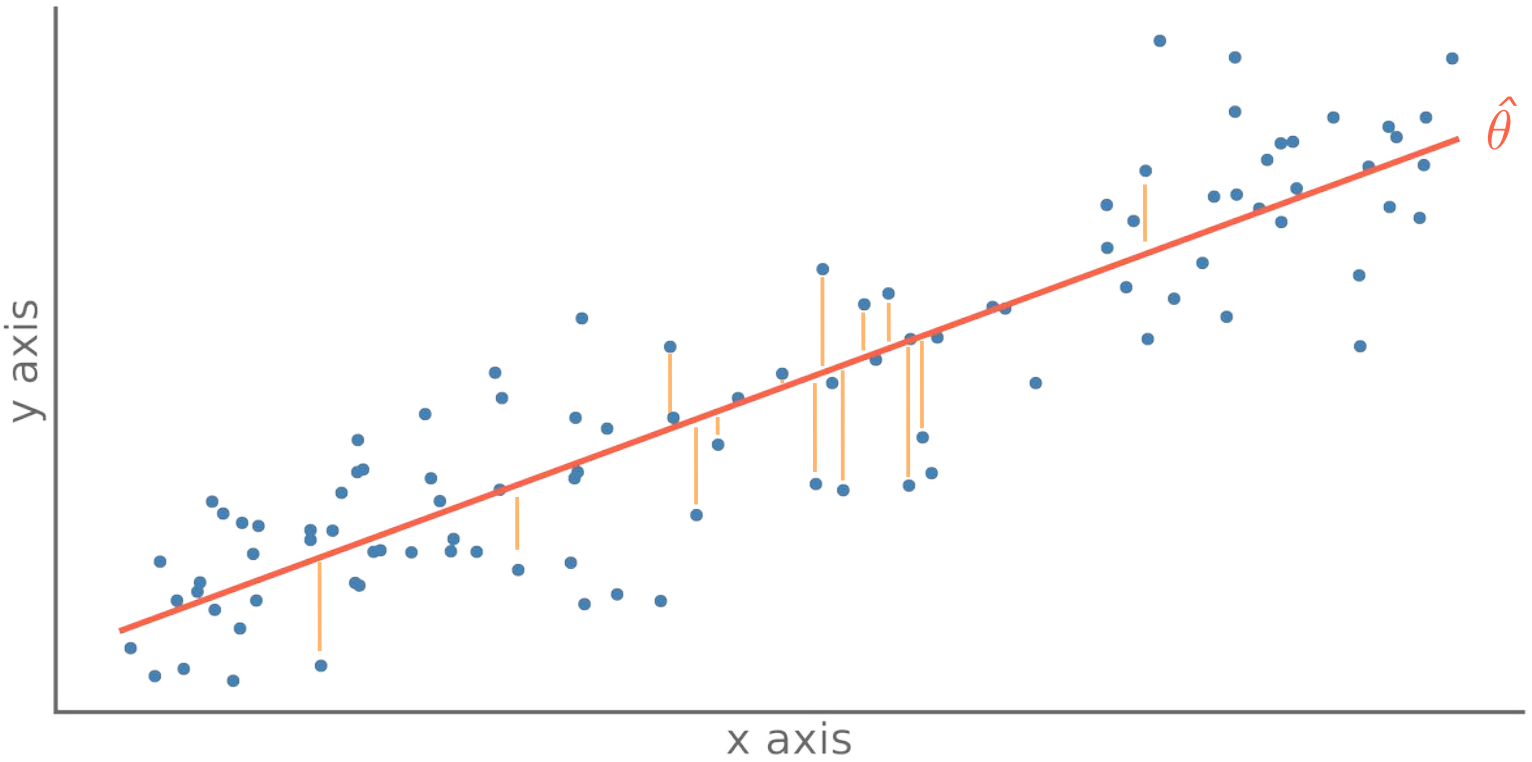
Linear regression

We are looking for a straight line that passes « as close as possible » to our points :



Linear regression

"As close as possible" means "minimize the distance" between the line and our points (observations).

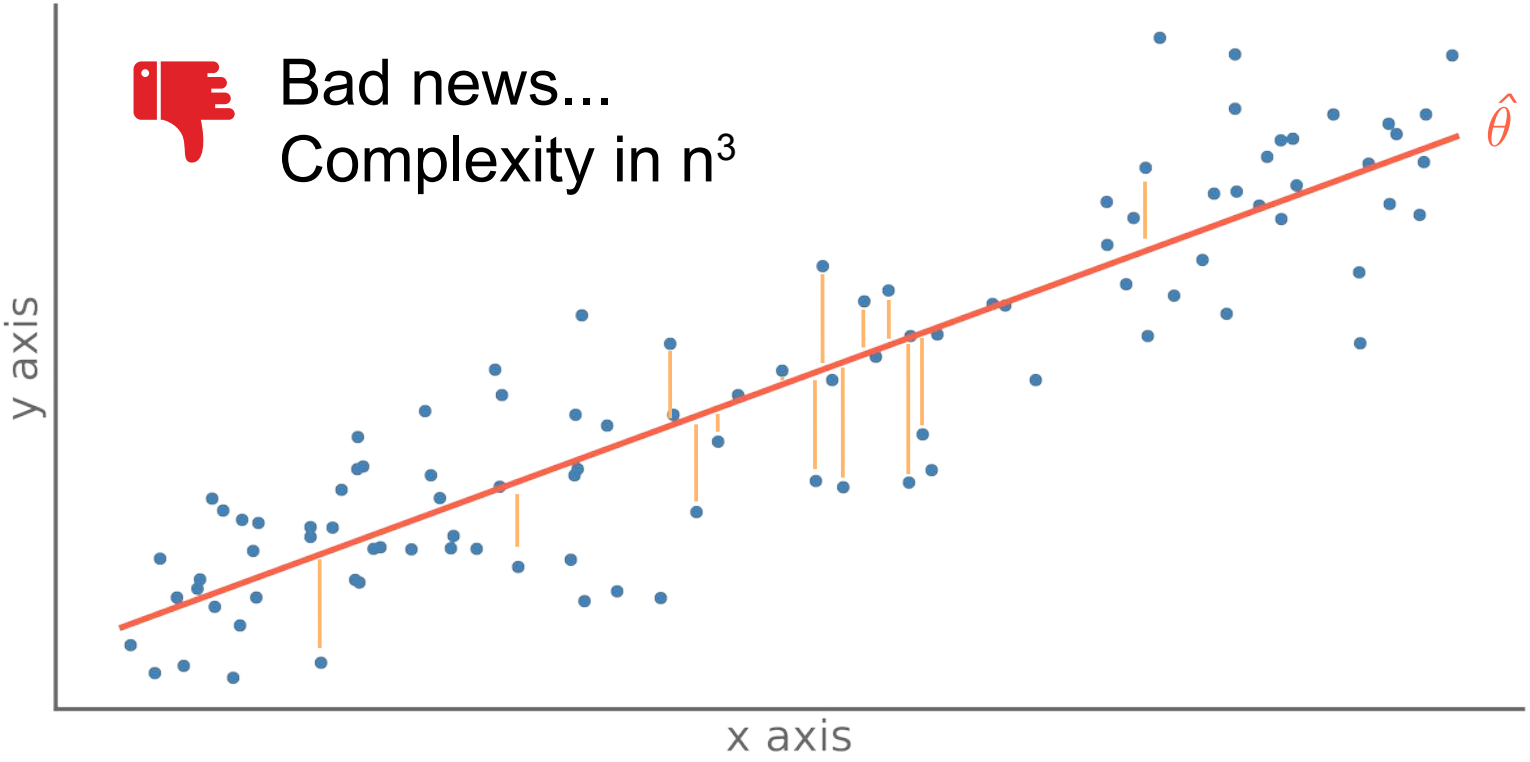


Linear regression

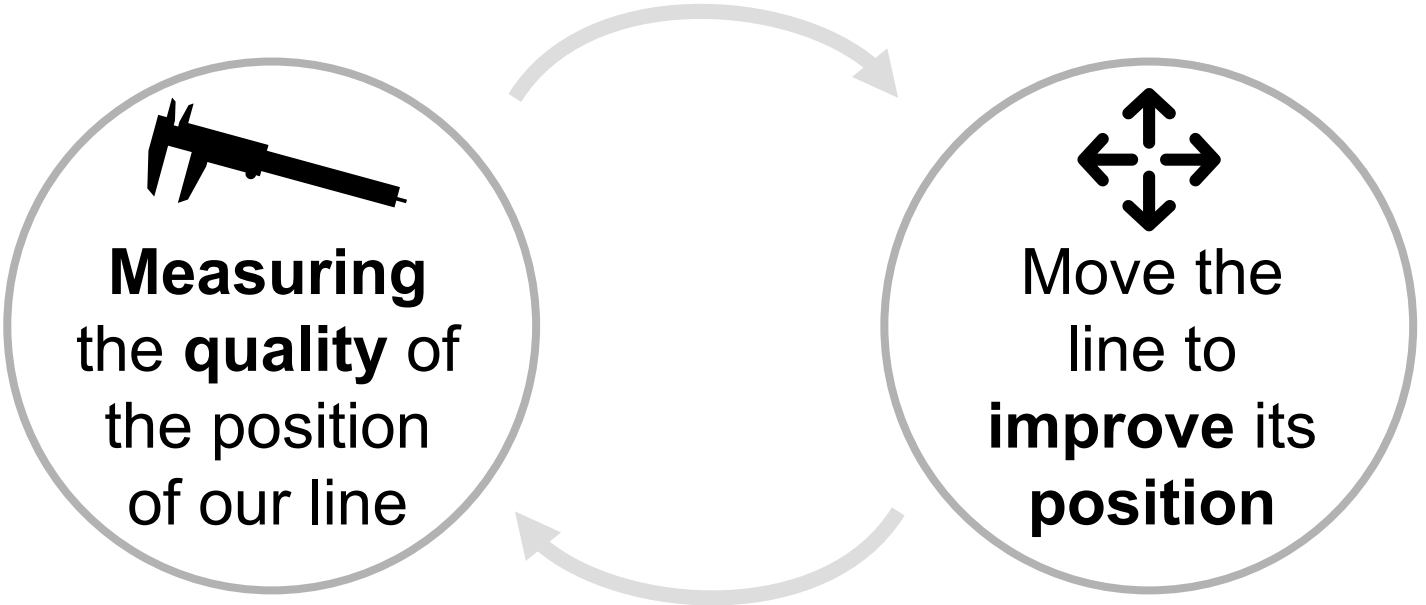
$$\hat{\theta} = (X^{-T}.X)^{-1}.X^{-T}.Y$$

Good news !
We have a direct solution !

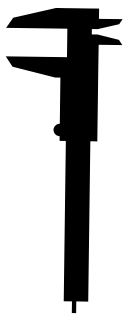
Bad news...
Complexity in n^3



Linear regression



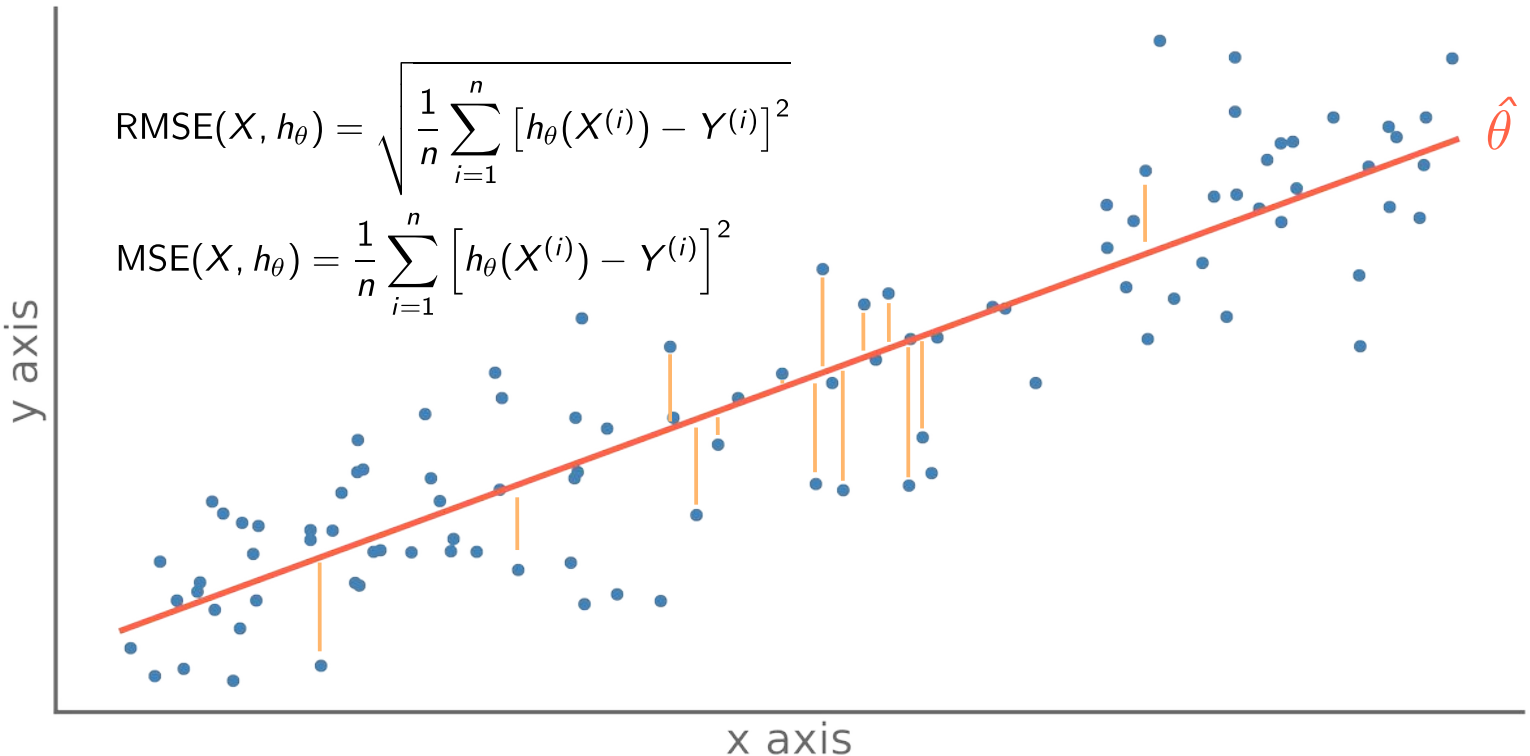
Linear regression



We'll use a **«loss function »**, which we'll try to **minimize**.

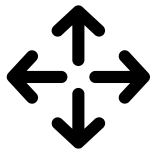
$$\mathrm{RMSE}(X, h_\theta) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2}$$

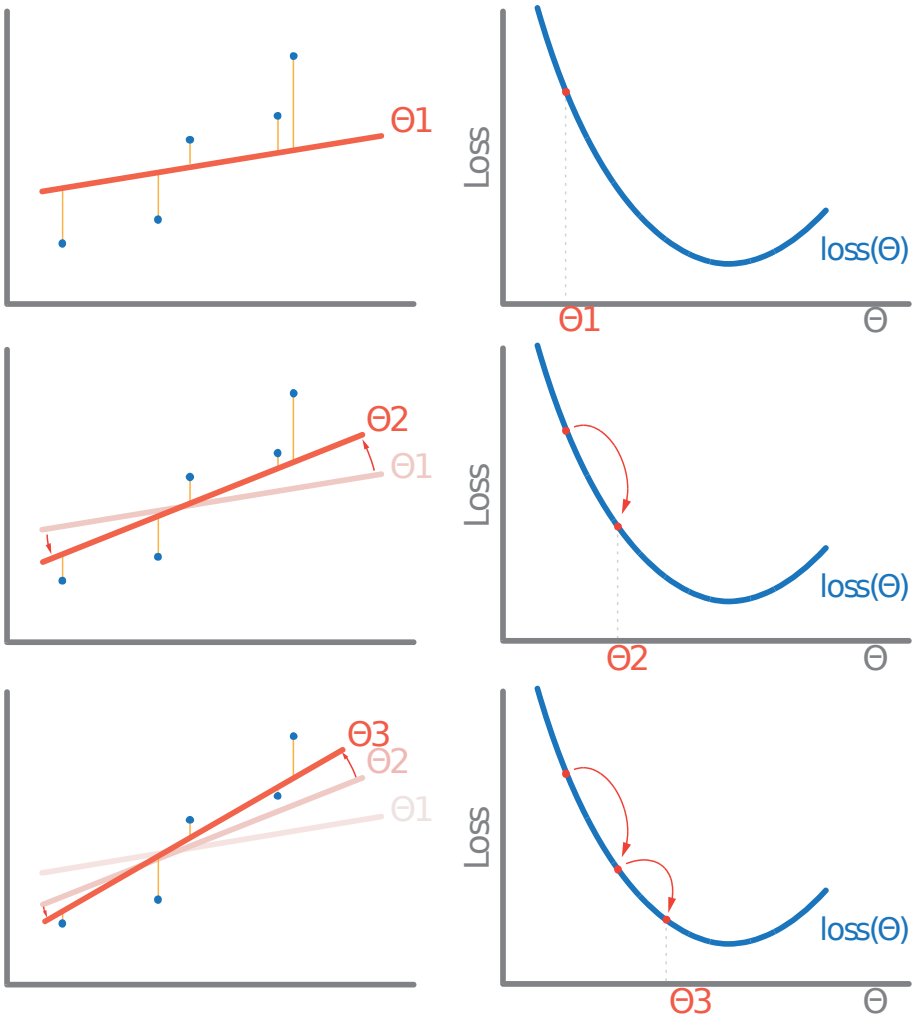
$$\mathrm{MSE}(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$



RMSE : Root Mean Square Error - Erreur quadratique moyenne
MSE : Mean Squared Error - Moyenne du carré des erreurs

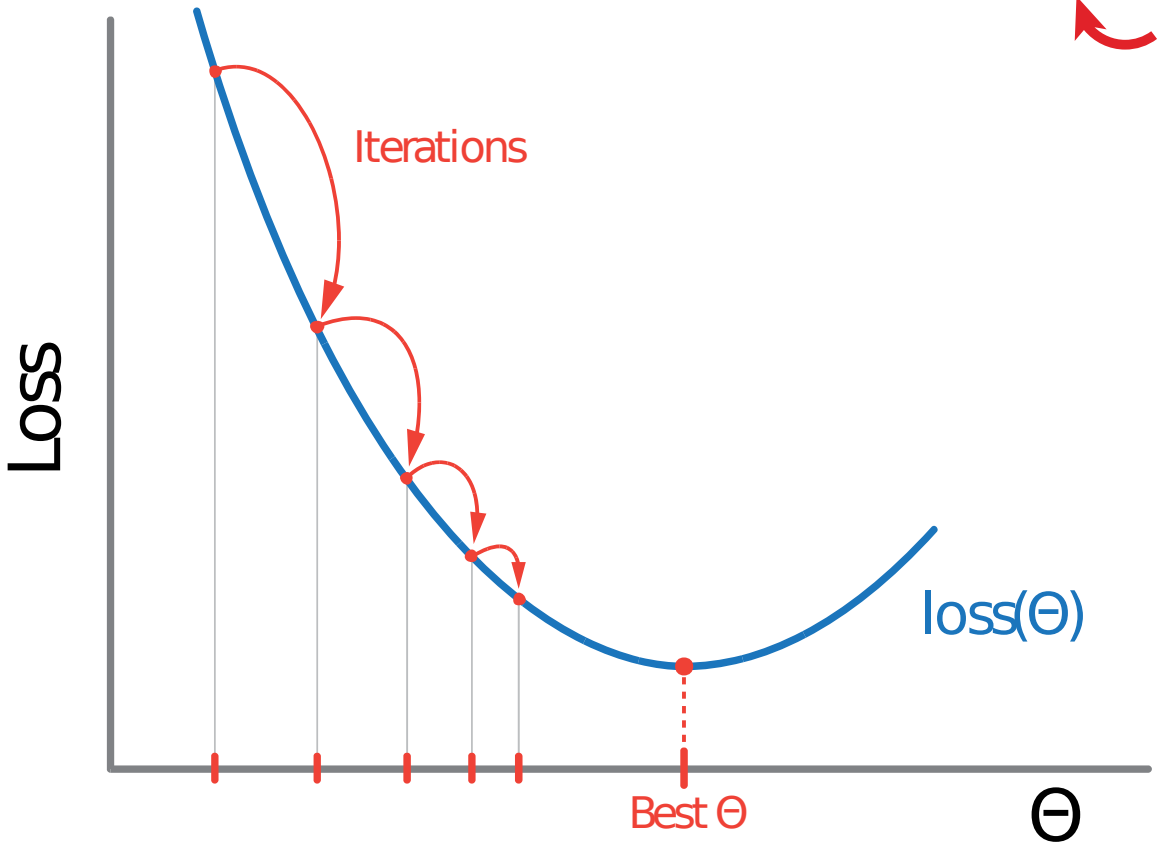
Linear regression





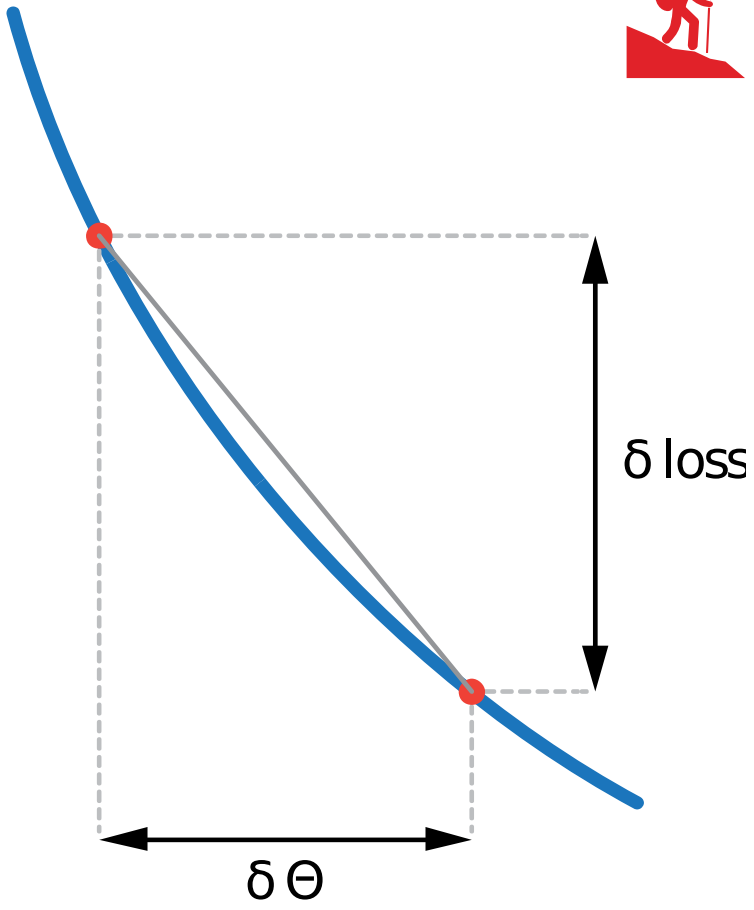
We will iteratively look for the best position of our line, by varying its parameters (Θ). This is called **optimization**

Optimization



We will iteratively look for the best position of our line, by varying its parameters (Θ). This is called **optimization**

Gradient descent



By changing Θ from δΘ
We improve loss(Θ) of δloss
The gradient is the slope we will follow to minimize our loss function.

$$\text{gradient} = \frac{\delta \text{loss}}{\delta \theta}$$

One iterative solution is : $\theta \leftarrow \theta - \eta \cdot \frac{\delta \text{loss}}{\delta \theta}$

where η is the learning rate

This process is called **gradient descent**

Gradient descent

$$MSE(X, h_\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[h_\theta(X^{(i)}) - Y^{(i)}\right]^2$$

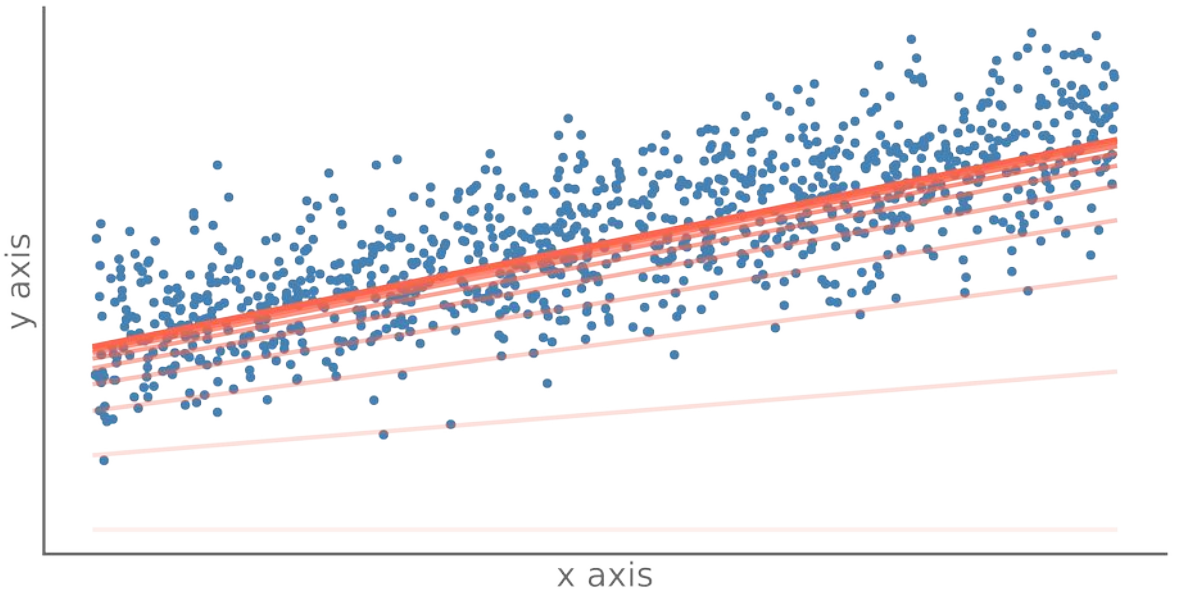
$$\nabla_\theta MSE(\Theta) = \begin{bmatrix} \frac{\partial}{\partial\theta_0} MSE(\Theta) \\ \frac{\partial}{\partial\theta_1} MSE(\Theta) \\ \vdots \\ \frac{\partial}{\partial\theta_m} MSE(\Theta) \end{bmatrix} = \frac{2}{n} X^T \cdot (X \cdot \Theta - Y)$$

Iterative solution is : $\Theta \leftarrow \Theta - \eta \cdot \nabla_\theta MSE(\Theta)$

where η is the learning rate

n : number of observations

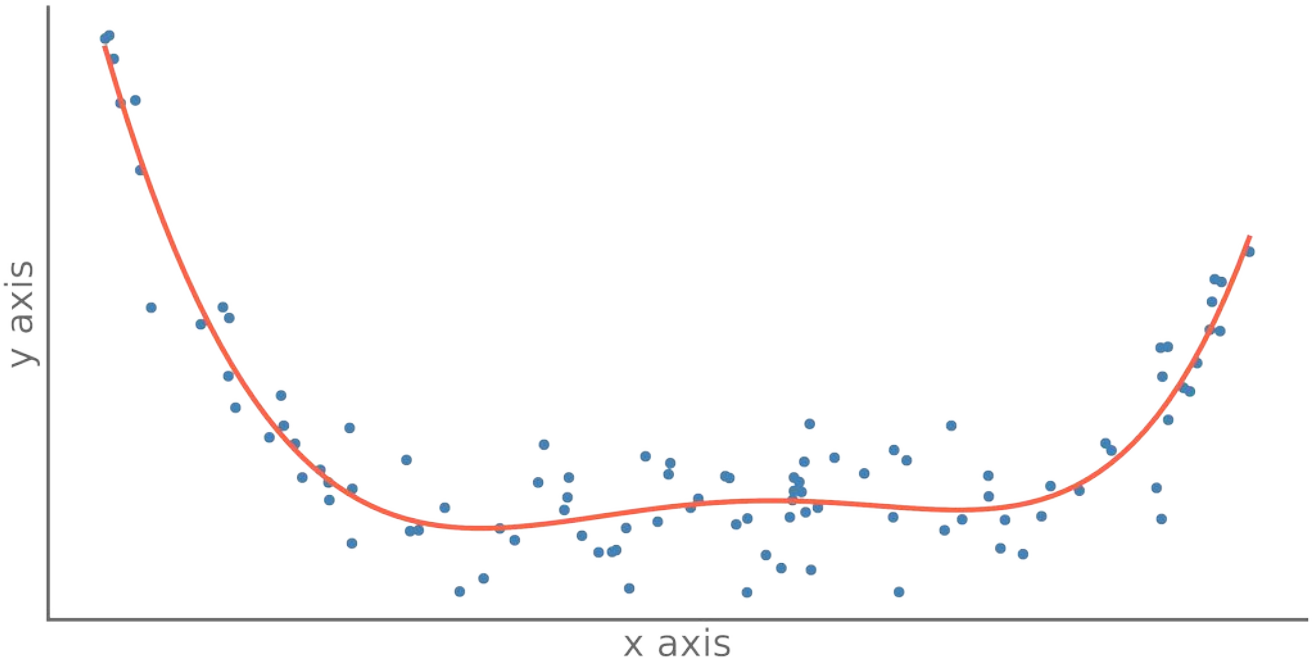
m : number of characteristics



#i	Loss	Gradient		Theta	
0	+12.481	-6.777	-1.732	-3.388	+0.000
20	+4.653	-4.066	-1.039	-2.033	+0.346
40	+1.835	-2.440	-0.624	-1.220	+0.554
60	+0.821	-1.464	-0.374	-0.732	+0.679
80	+0.455	-0.878	-0.224	-0.439	+0.754
100	+0.324	-0.527	-0.135	-0.263	+0.799
120	+0.277	-0.316	-0.081	-0.158	+0.826
140	+0.260	-0.190	-0.048	-0.095	+0.842
160	+0.253	-0.114	-0.029	-0.057	+0.851
180	+0.251	-0.068	-0.017	-0.034	+0.857
200	+0.250	-0.041	-0.010	-0.020	+0.861

Fidle notebook [GRAD1] - https://fidle.cnrs.fr

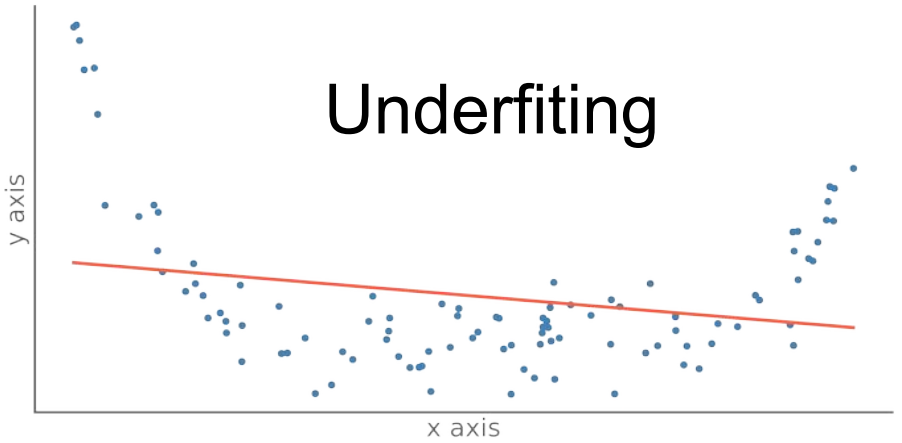
Polynomial regression

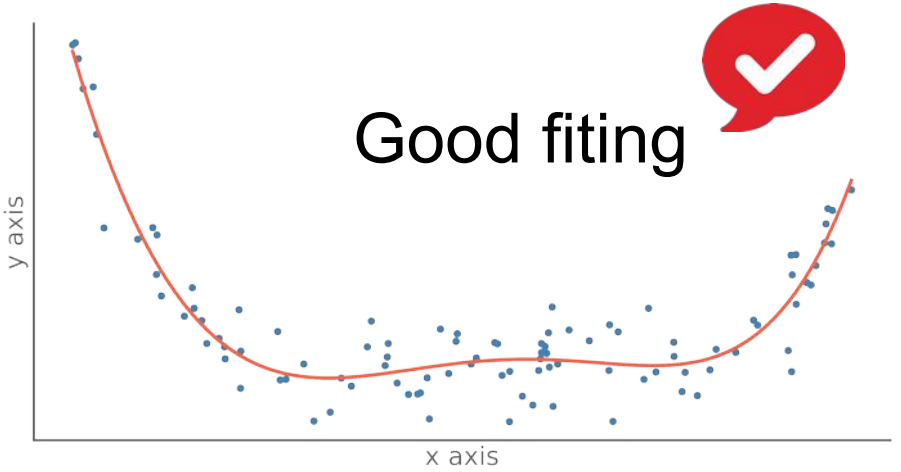


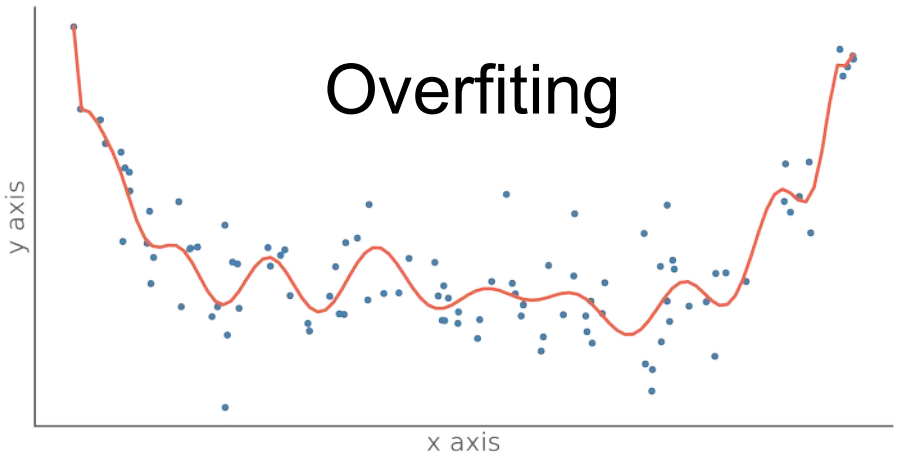
$$P_n(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \cdots + a_n \cdot x^n = \sum_{i=0}^{n} a_i \cdot x^n$$

Fidle notebook [POLR1] - https://fidle.cnrs.fr

Underfiting and overfiting





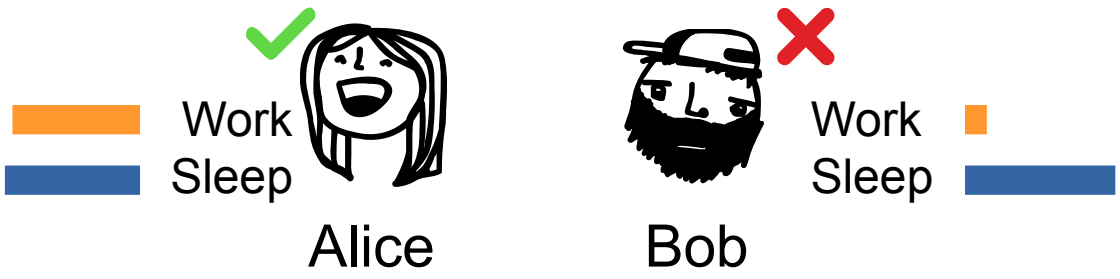


Fidle notebook [POLR1] - https://fidle.cnrs.fr

Great, but where's my first **artificial neuron**?

Okay, let's look at an example !

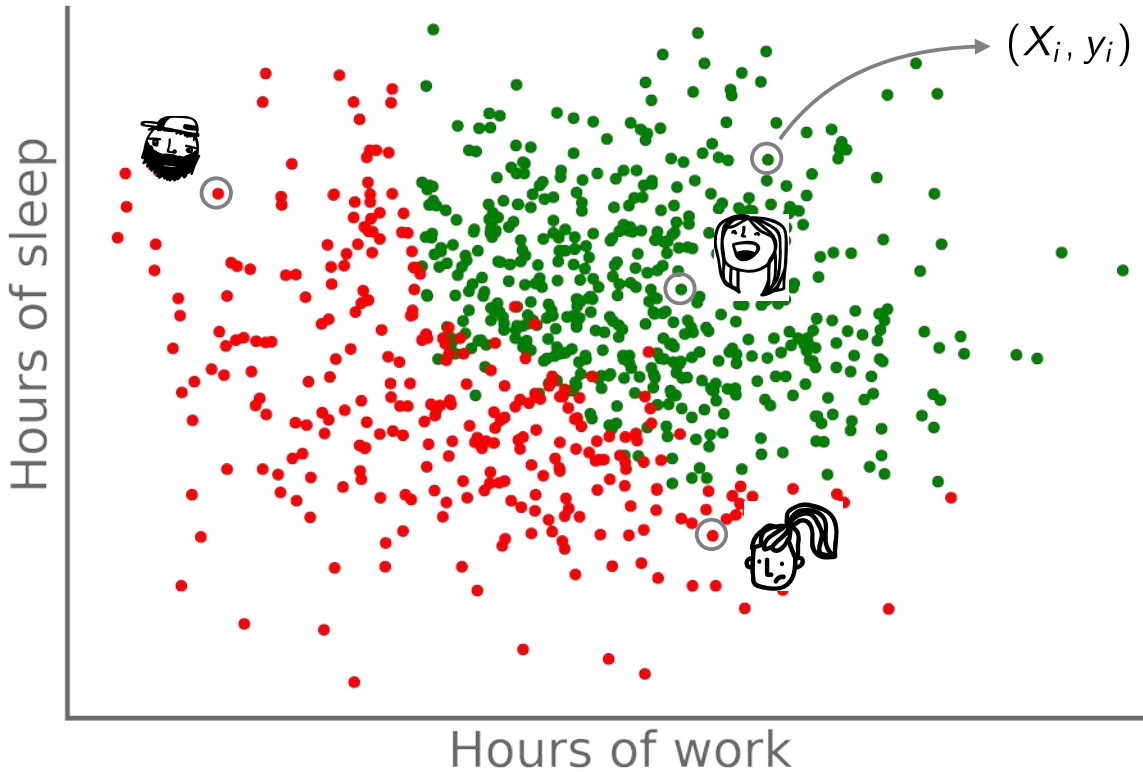
Let's try to predict our students' **success in an exam**, based on their **work time** and **sleep time**:



Logistic regression

Dataset : X Observations

y Classe

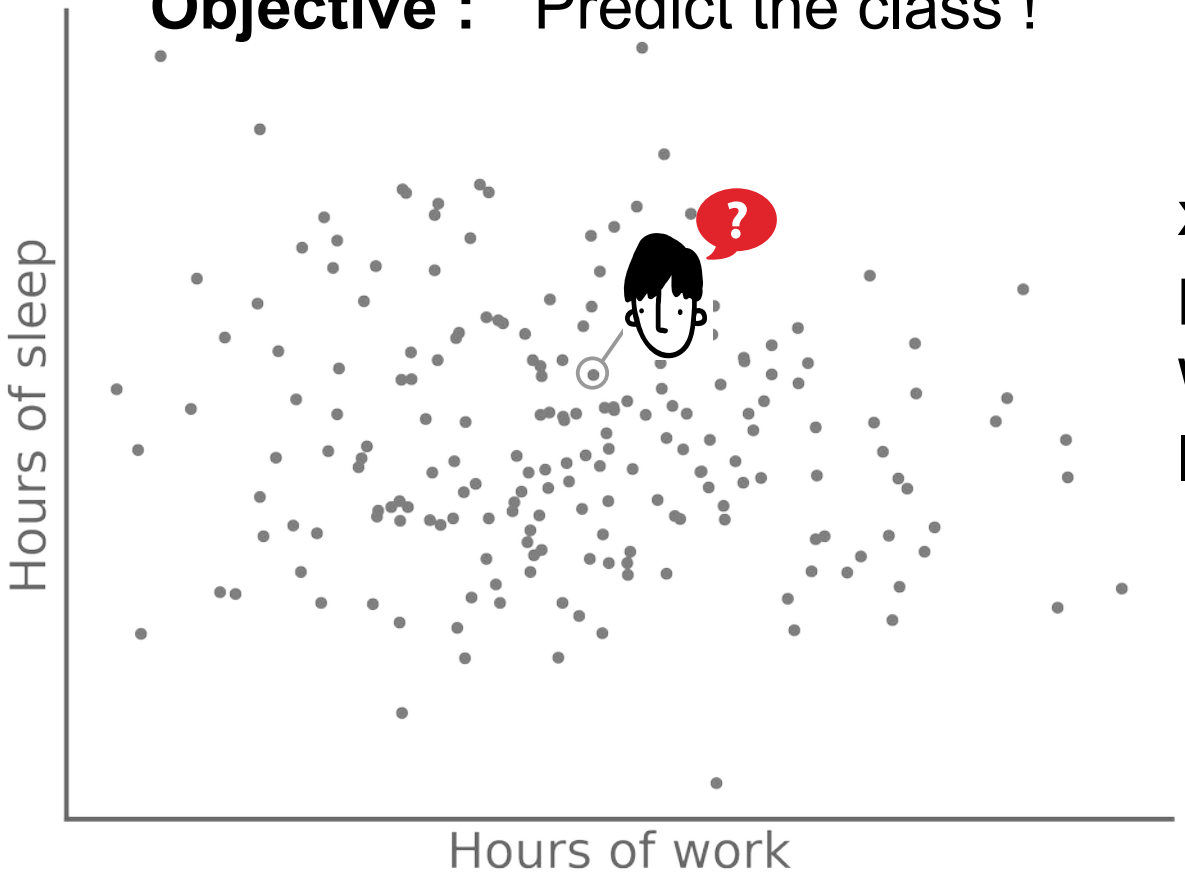


$$\begin{cases} X_i = \begin{pmatrix} x_{i1} = \text{Hours of work} \\ x_{i2} = \text{Hours of sleep} \end{pmatrix} \\ y_i = \begin{cases} 1 & \text{Successful examination} \\ 0 & \text{Failed exam} \end{cases} \end{cases}$$

Fidle notebook [LOGR1] - https://fidle.cnrs.fr

Logistic regression

Objective : Predict the class !

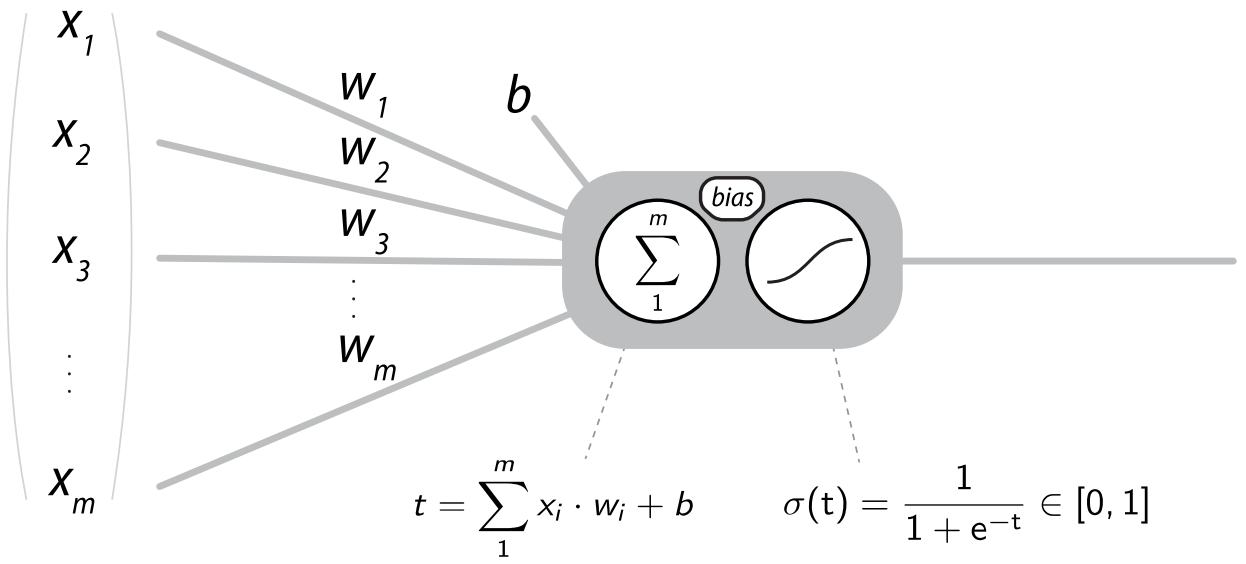


x given, we want to predict : $\mathbf{y_{pred} = f(x)}$

Where f(x) is a linear function

Logistic regression

$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$

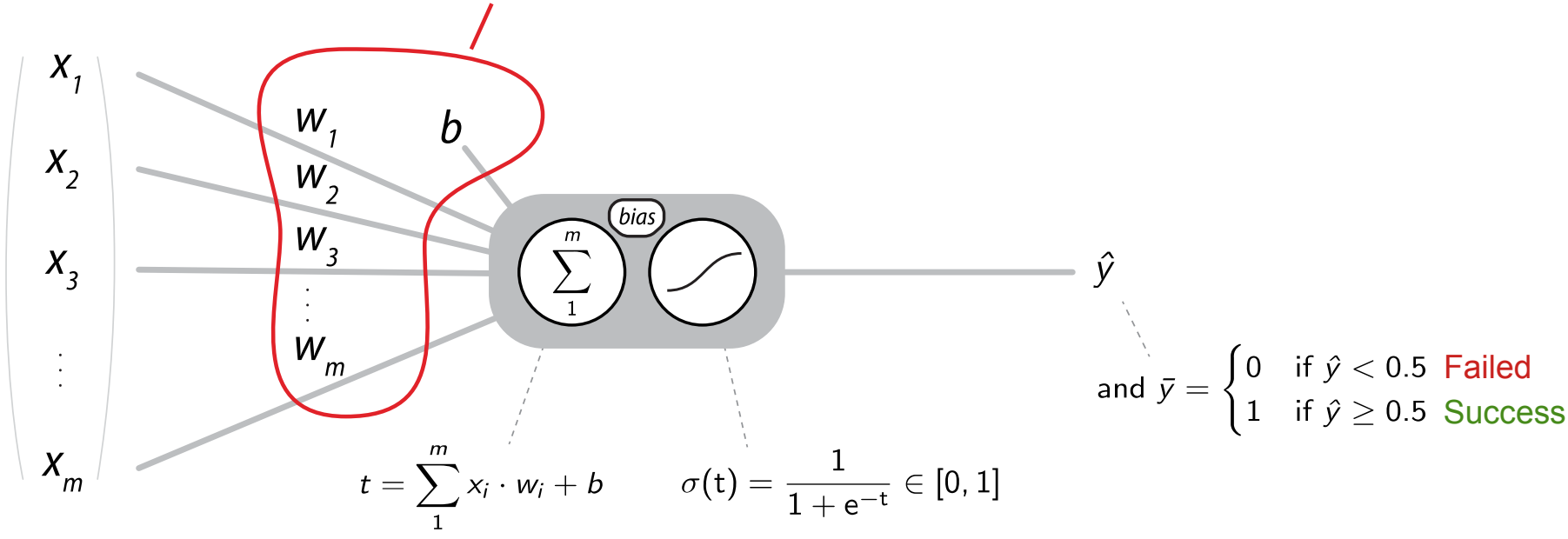


$$\text{and } \bar{y} = \begin{cases} 0 & \text{if } \hat{y} < 0.5 \text{ Failed} \\ 1 & \text{if } \hat{y} \geq 0.5 \text{ Success} \end{cases}$$

Input	Bias / Weight	Activation function	Output
X	Θ	$\sigma(t)$	$\hat{y}$

Logistic regression

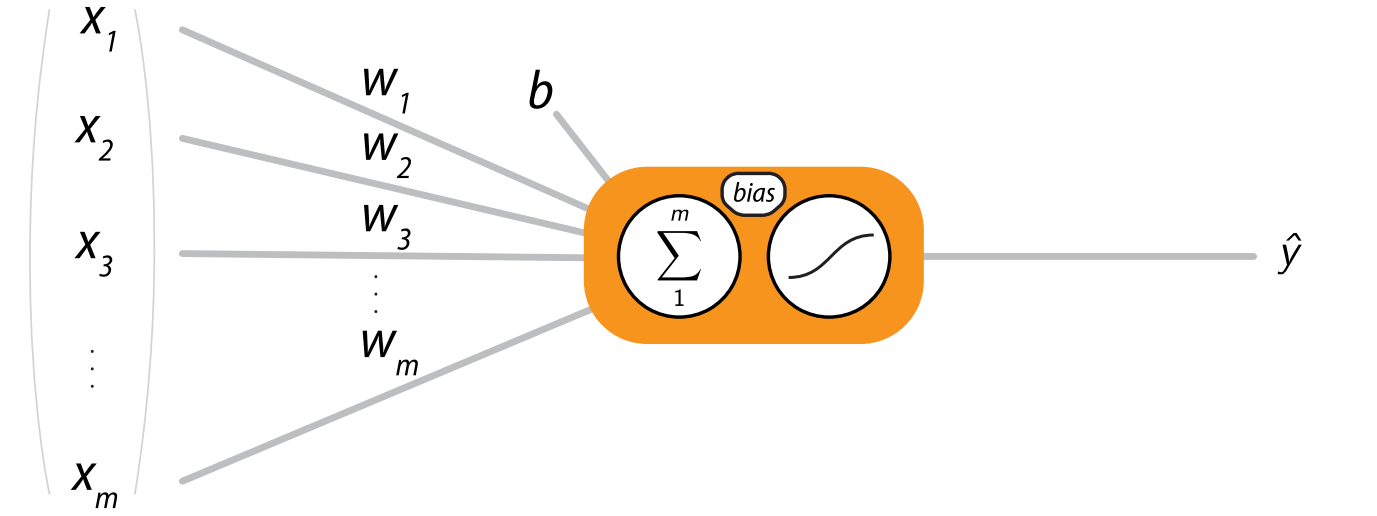
Determined by the minimisation of a cost function J(Θ)



Input	Bias / Weight	Activation function	Output
X	Θ	$\sigma(t)$	$\hat{y}$

Logistic regression

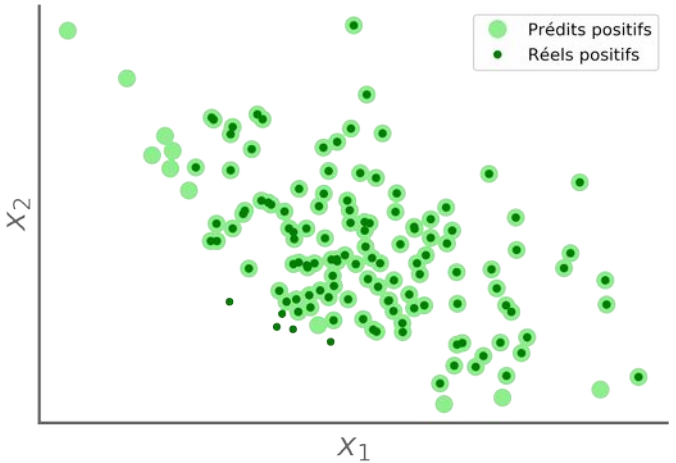
$$\hat{y} = \sigma(\Theta^T \cdot X + b)$$

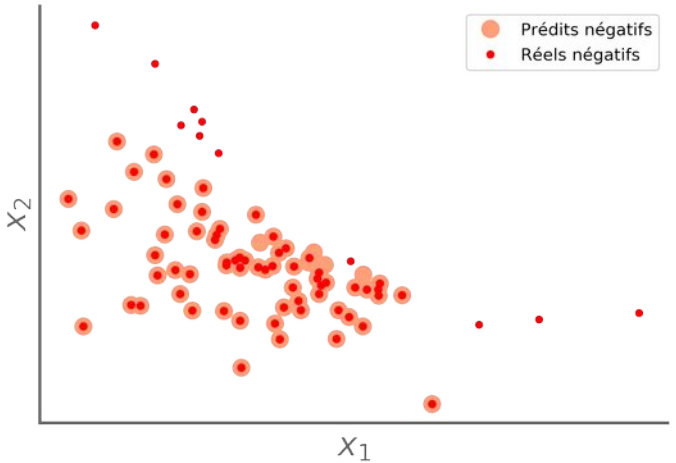


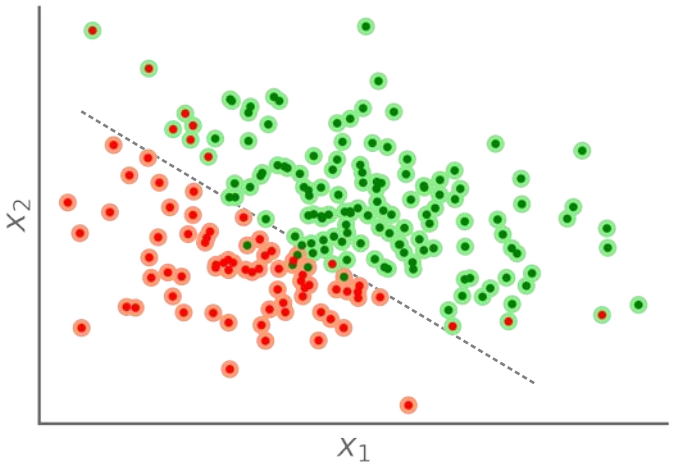
That's an « **artificial neuron** » !

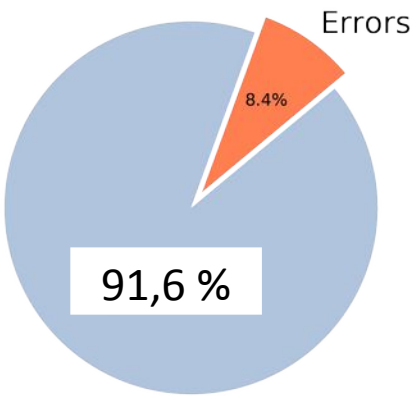
So, we have a neural network of… 1 neuron !

Logistic regression









Fidle notebook [LOGR1] - https://fidle.cnrs.fr

[intelligence]

[intelligence]

« Capacité à percevoir ou à déduire des **informations** et à les conserver en tant que **connaissances** pour les appliquer à des **comportements adaptatifs** dans un environnement ou un contexte.»

*« Ability to perceive or infer **information** and to retain it as **knowledge** to be applied to **adaptive** behaviors within an environment or context. »*[*]

[*] Wikipedia

[intelligence]

« Ensemble des **fonctions mentales** ayant pour objet la connaissance **conceptuelle** et **rationnelle** »*

« Set of mental functions aimed at conceptual and rational knowledge »

* https://www.larousse.fr

The big Controversy

Modéliser le cerveau :

Penser s'apparente à un **traitement du signal** massivement parallèle, via des **fonctions élémentaires**.

Forger une opinion :

« Penser, consiste à itérer sur des **symboles de haut niveau**, porteurs d'une **représentations symbolique**.

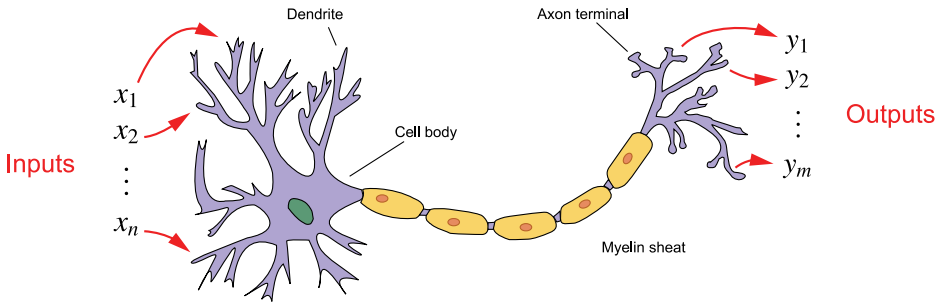
Connectionnism

Modelling the brain

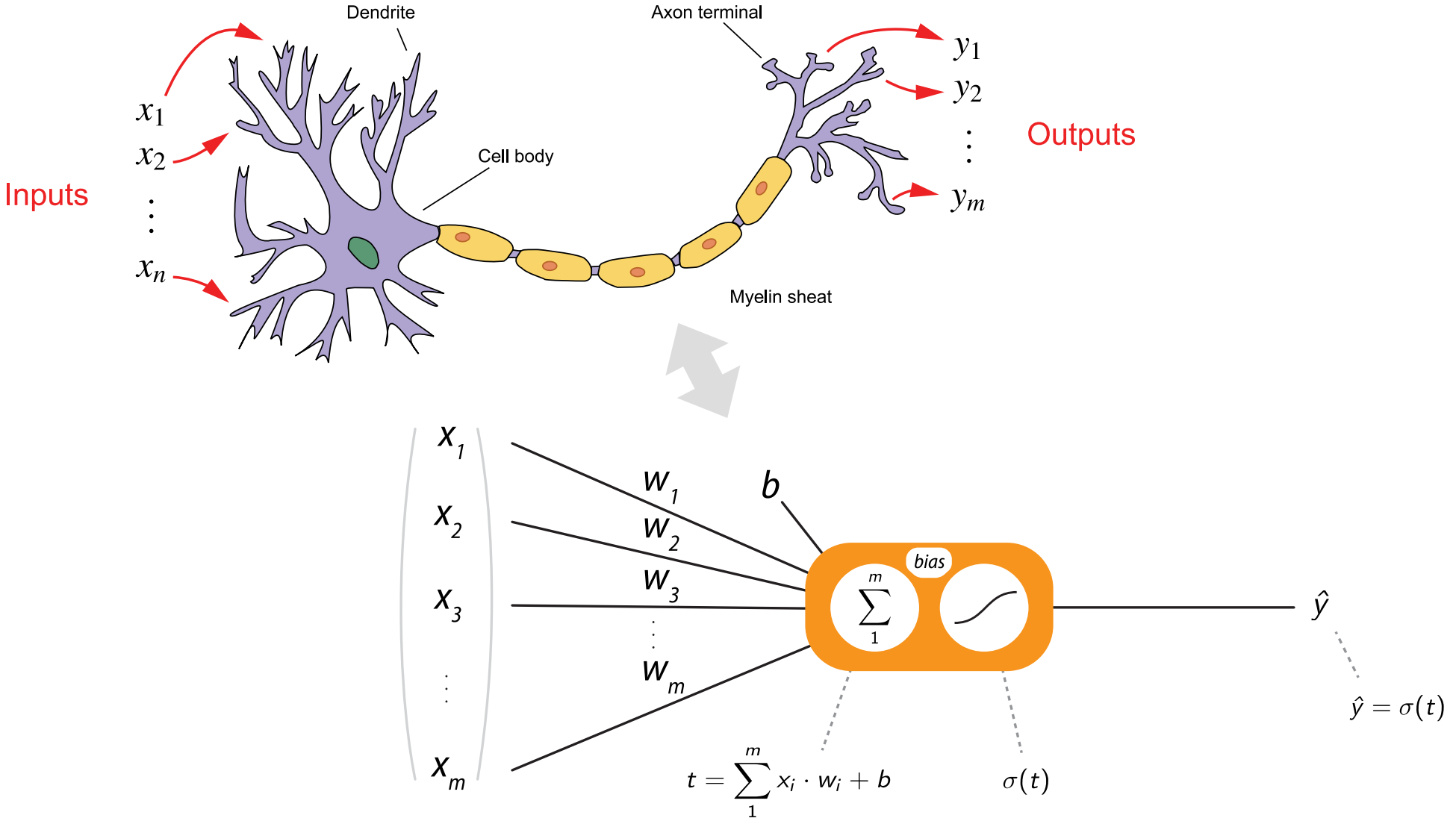
vs

Symbolic

Making a mind



Tout [homme] est [mortel]
[Socrate] est un [homme]
Donc [Socrate] est [mortel]


Inputs
x_1
x_2
x_n
Dendrite
Cell body
Axon terminal
Myelin sheat
y_1
y_2
y_m
Outputs
x_1
x_2
x_3
x_m
w_1
w_2
w_3
w_m
b
bias
$\sum_{1}^{m}$
$\hat{y}$
$\hat{y} = \sigma(t)$
$t = \sum_{1}^{m} x_i \cdot w_i + b$
$\sigma(t)$

The big Controversy

Modéliser le cerveau :

Penser s'apparente à un **traitement du signal** massivement parallèle, via des **fonctions élémentaires**.

Forger une opinion :

« Penser, consiste à itérer sur des **symboles de haut niveau**, porteurs d'une **représentations symbolique**.

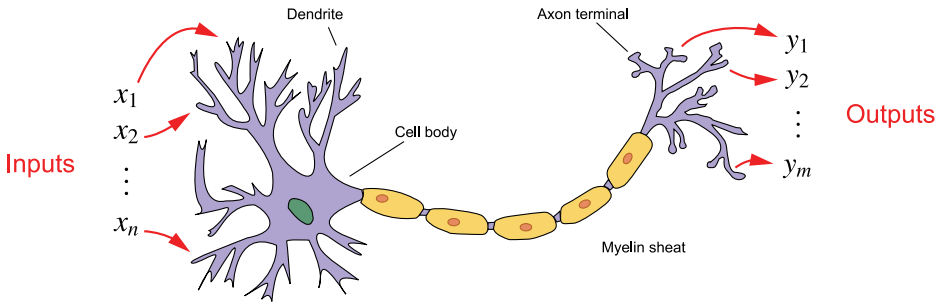
Connectionnism

Modelling the brain

vs

Symbolic

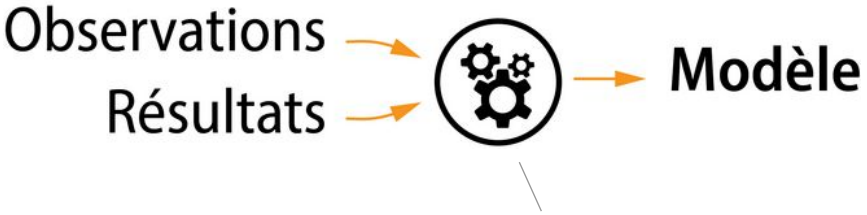
Making a mind



Tout [homme] est [mortel]
[Socrate] est un [homme]
Donc [Socrate] est [mortel]

The big Controversy

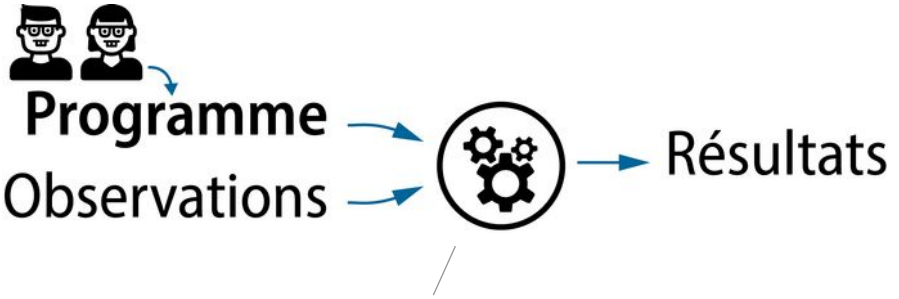
Inductive approach



Connectionnism

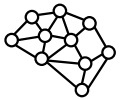
vs

Deductive approach



Symbolic

Facts ▶ Rules and laws



Model

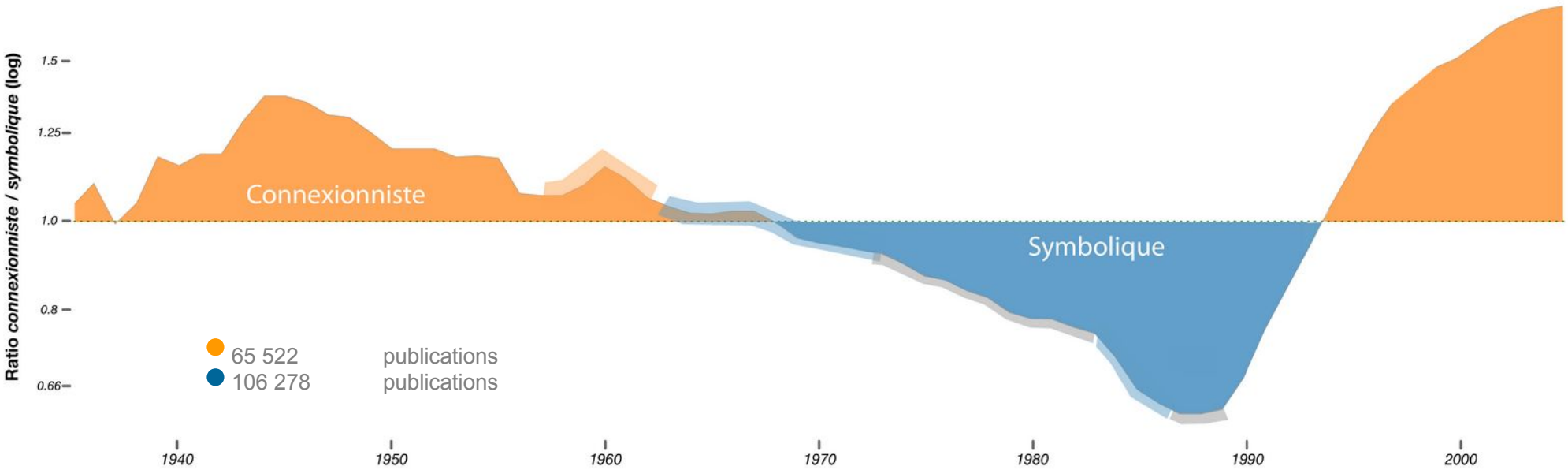
Expert

Rules and laws ▶ Special case

The big Controversy

Evolution of the academic influence of connexionist and symbolic approaches[1]

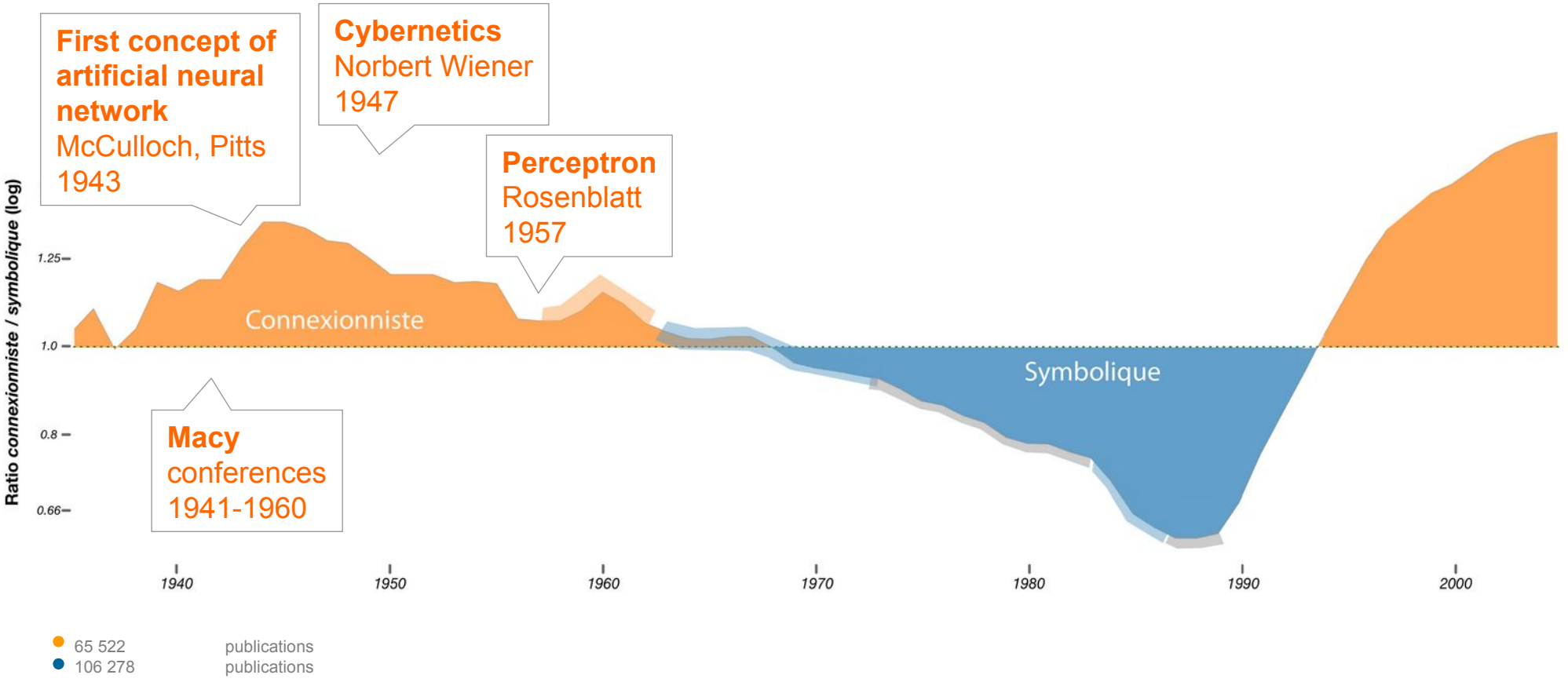
Ratio of publications between connexionists and symbolists



[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

The big Controversy

Evolution of the academic influence of connexionist and symbolic approaches[1]



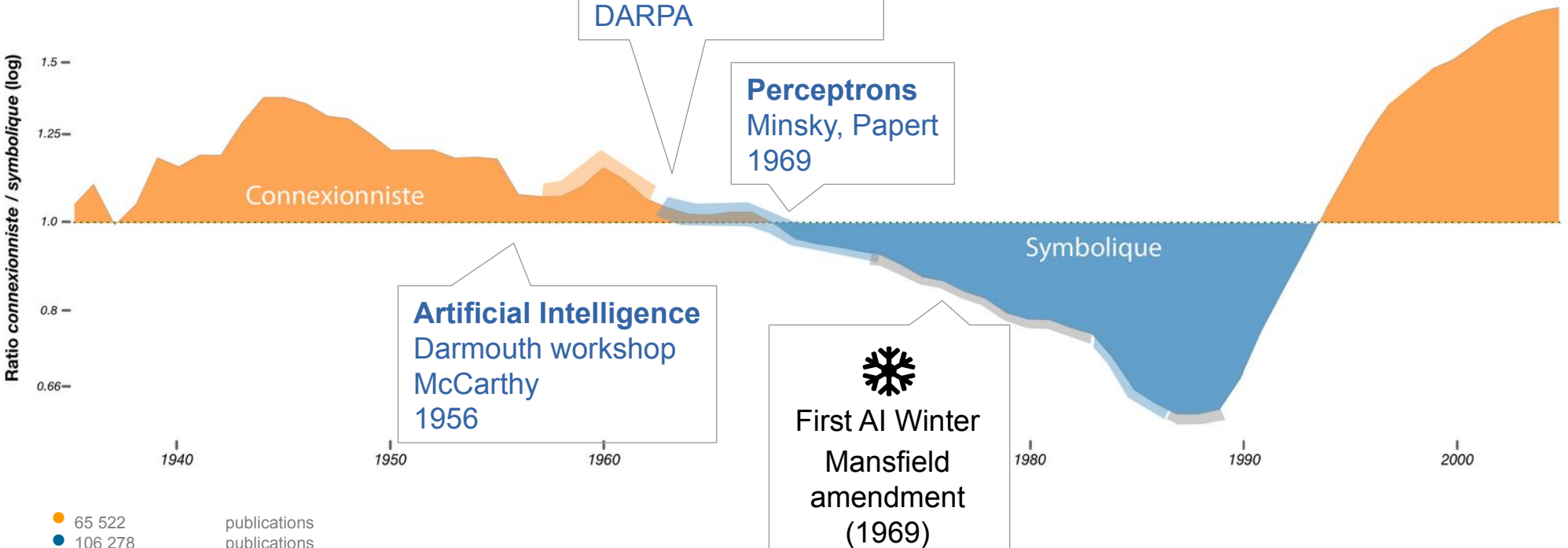
[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

The big Controversy

Evolution of the academic influence of connexionist and symbolic approaches[1]

MIT (Minsky, Papert)
CM (Simon, Newel)
Stanford (McCarthy)
DARPA

Perceptrons
Minsky, Papert
1969



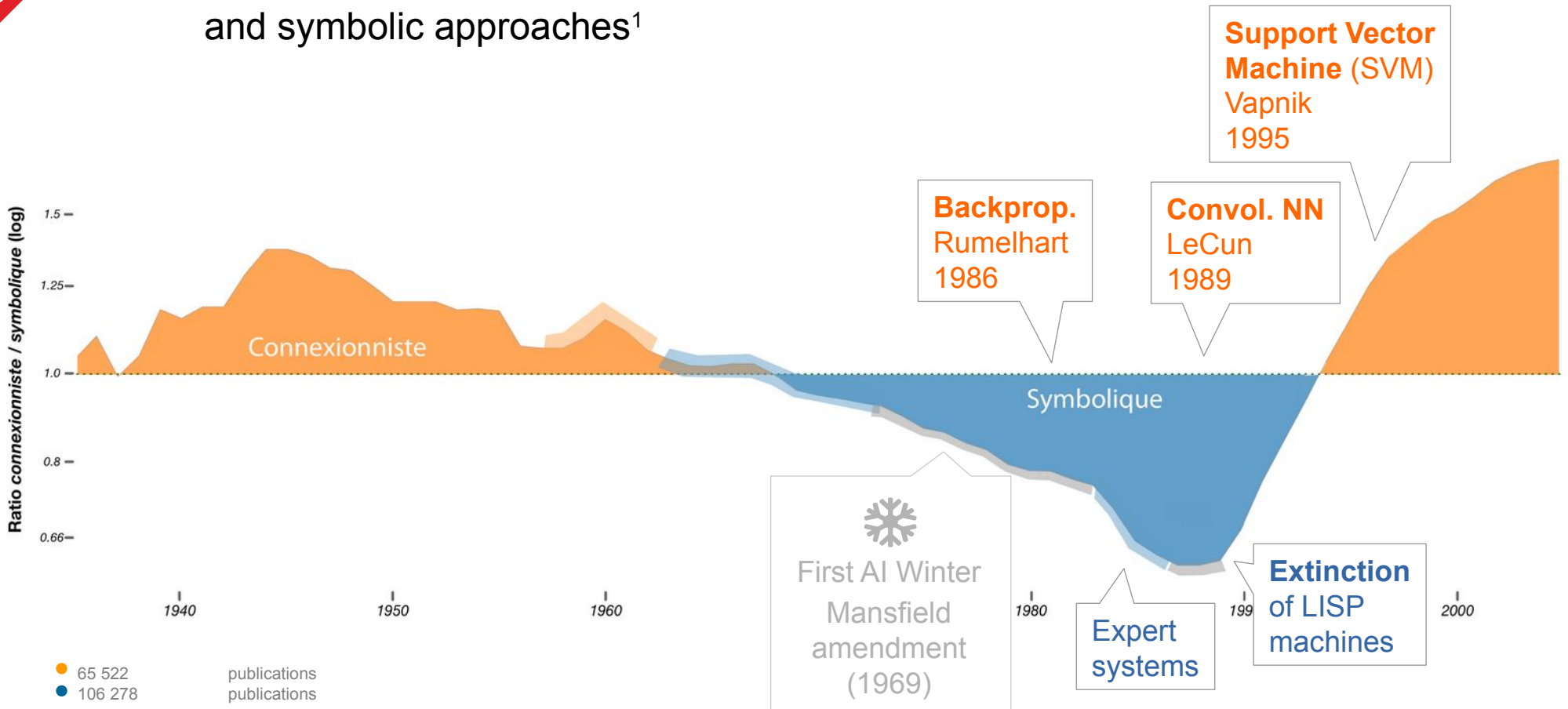
Artificial Intelligence
Darmouth workshop
McCarthy
1956

First AI Winter
Mansfield amendment (1969)
Lighthill report (1973)

[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

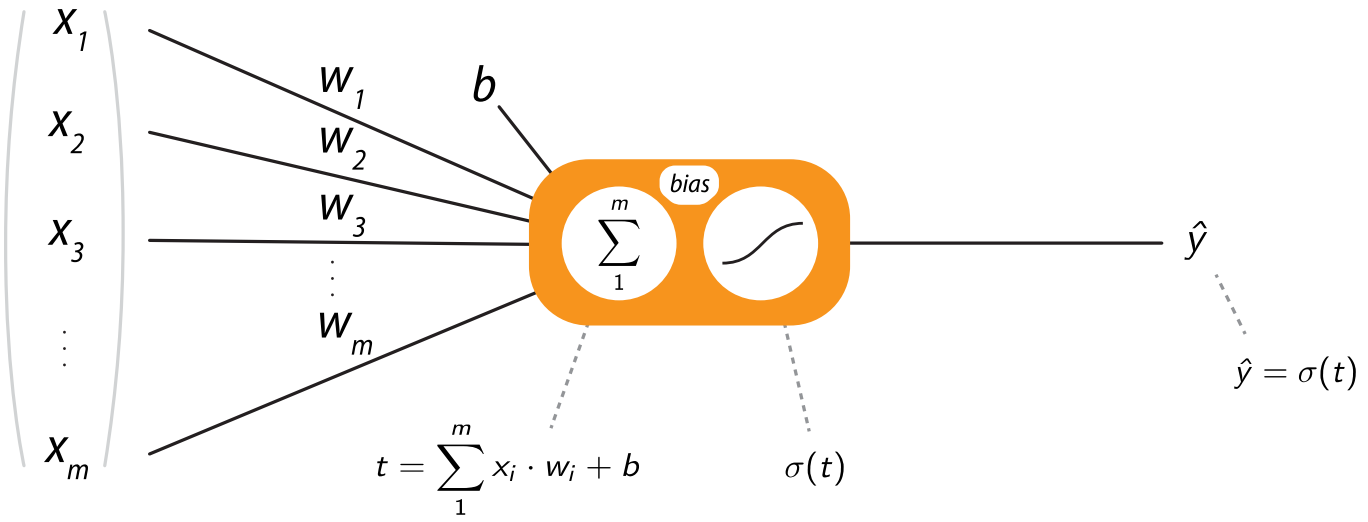
The big Controversy

Evolution of the academic influence of connexionist and symbolic approaches[1]



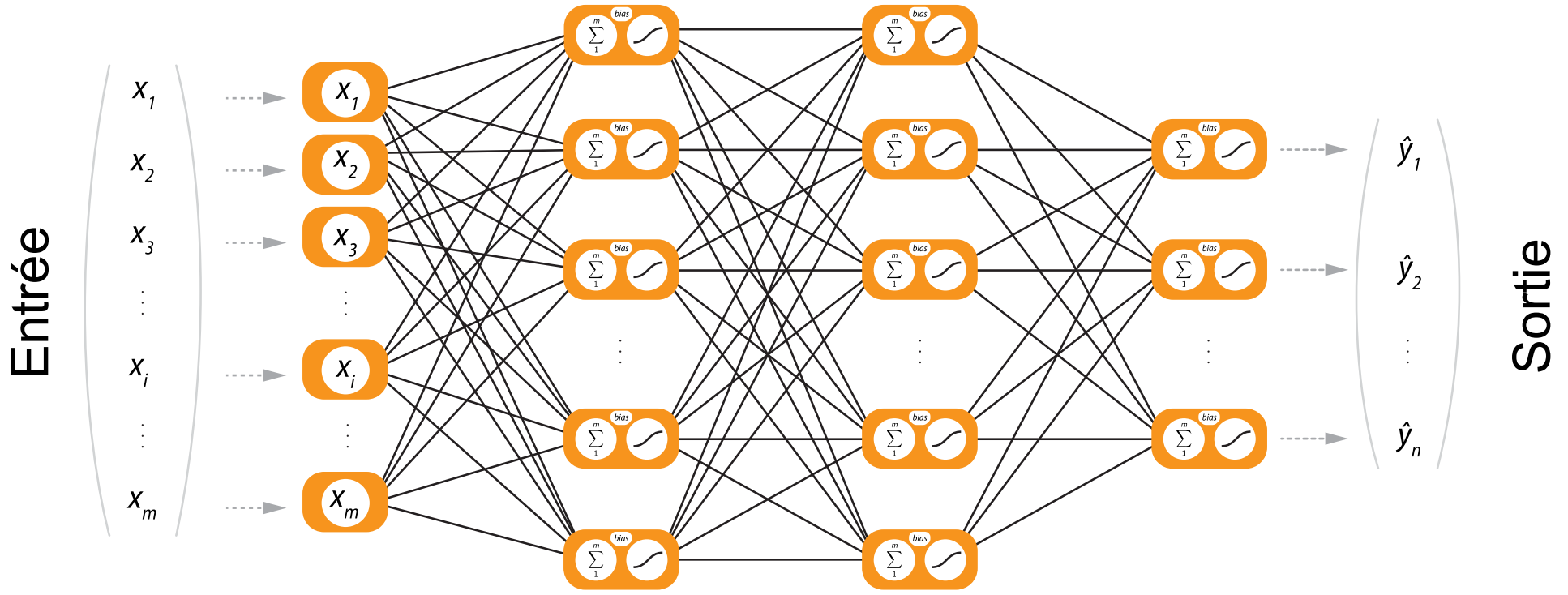
[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

Deep Neural Networks



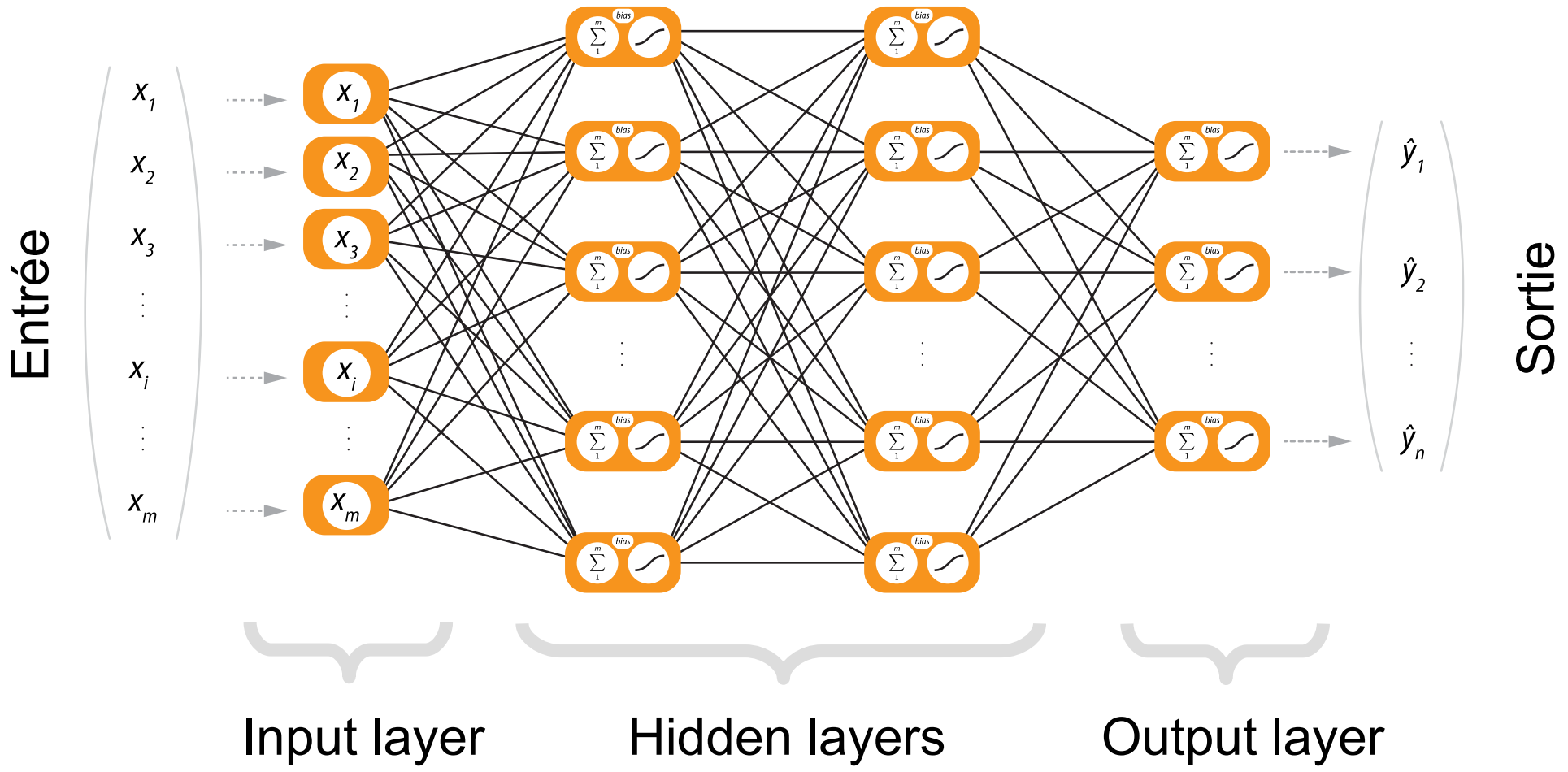
1957 : A single neuron...

Deep Neural Networks



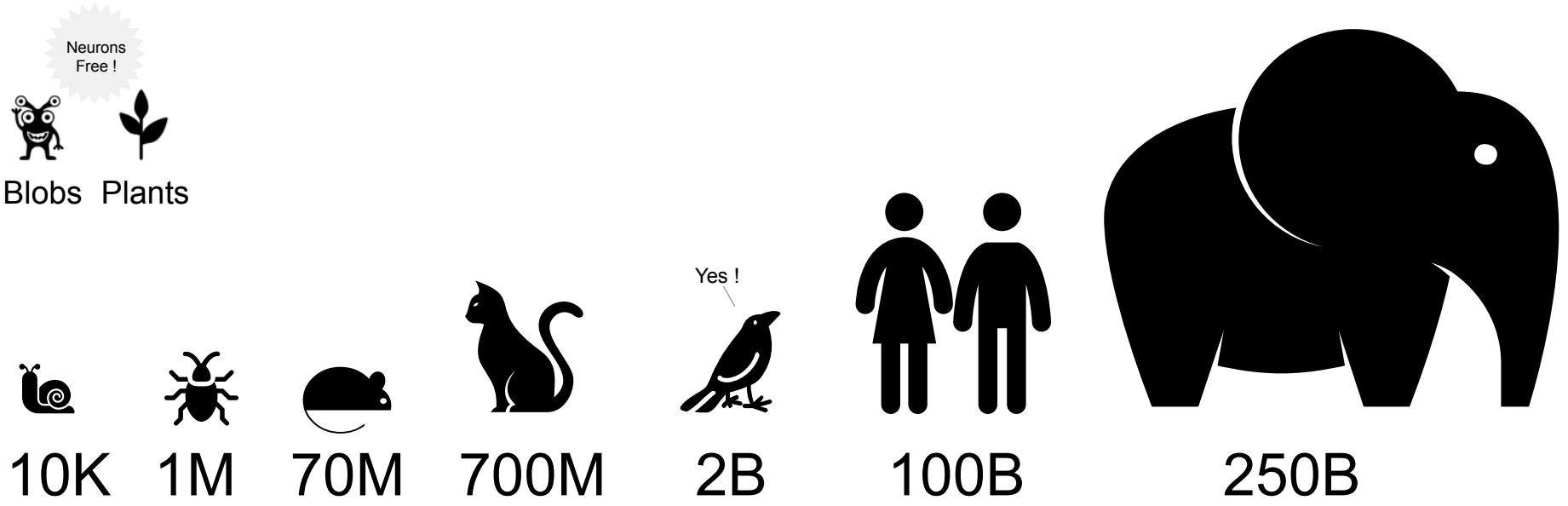
1986 : A neural network !

Deep Neural Networks



Short Disgression !

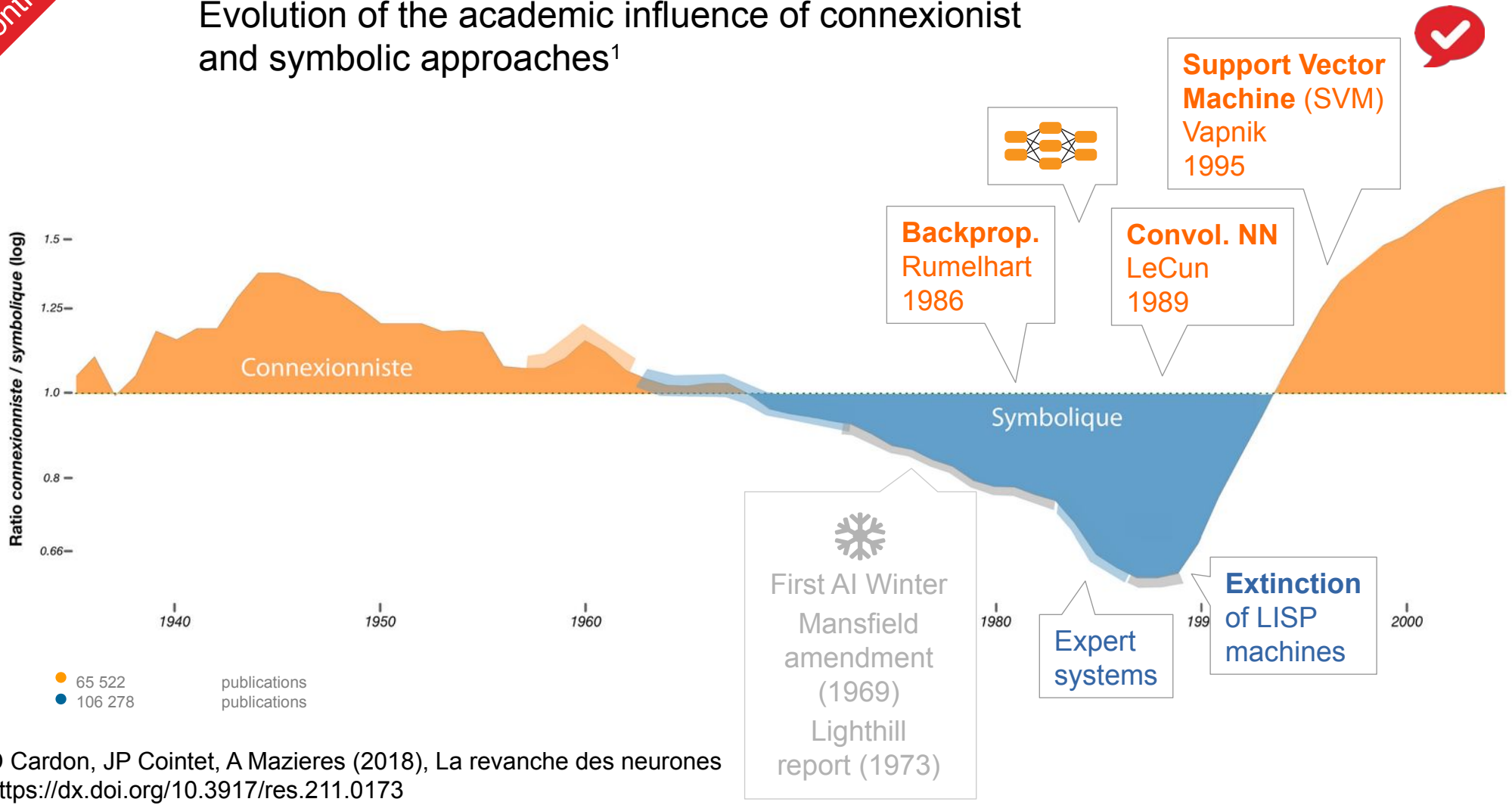
One neuron is **good**... but more than one is **better** !



Short Disgression !

The big Controversy

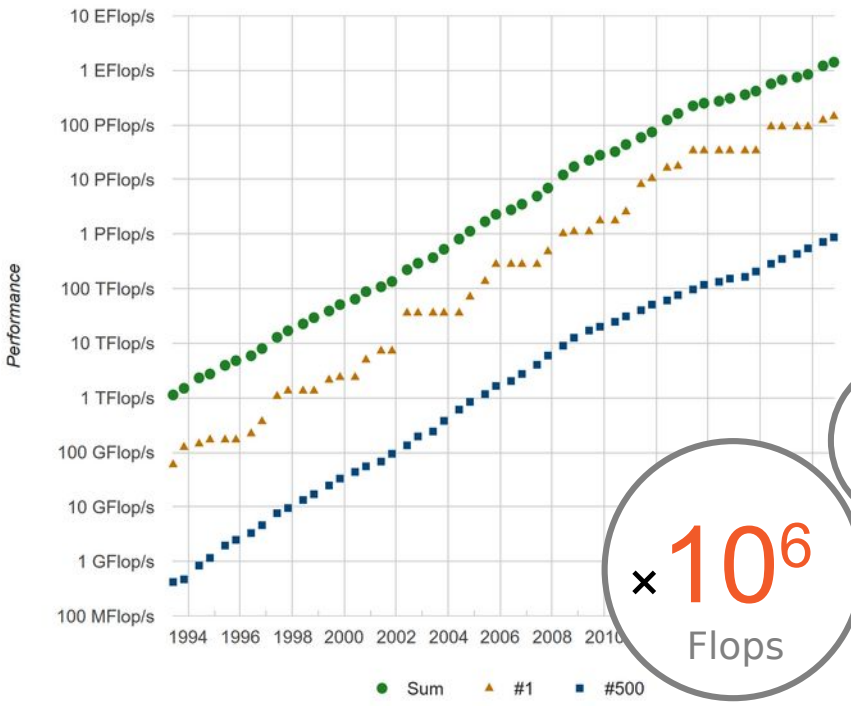
Evolution of the academic influence of connexionist and symbolic approaches[1]



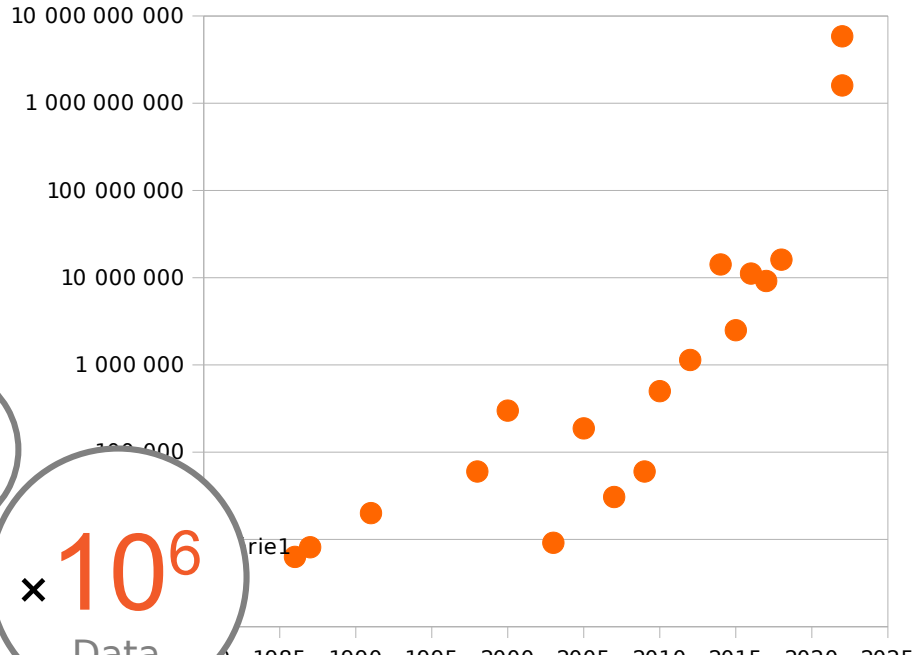
[1] D Cardon, JP Cointet, A Mazieres (2018), La revanche des neurones https://dx.doi.org/10.3917/res.211.0173

The big Controversy

Performance Development[1]



Datasets for machine-learning[2]



25 ans

× 10^6 Flops

× 10^6 Data

Laboratoire
Cas particulier

➡

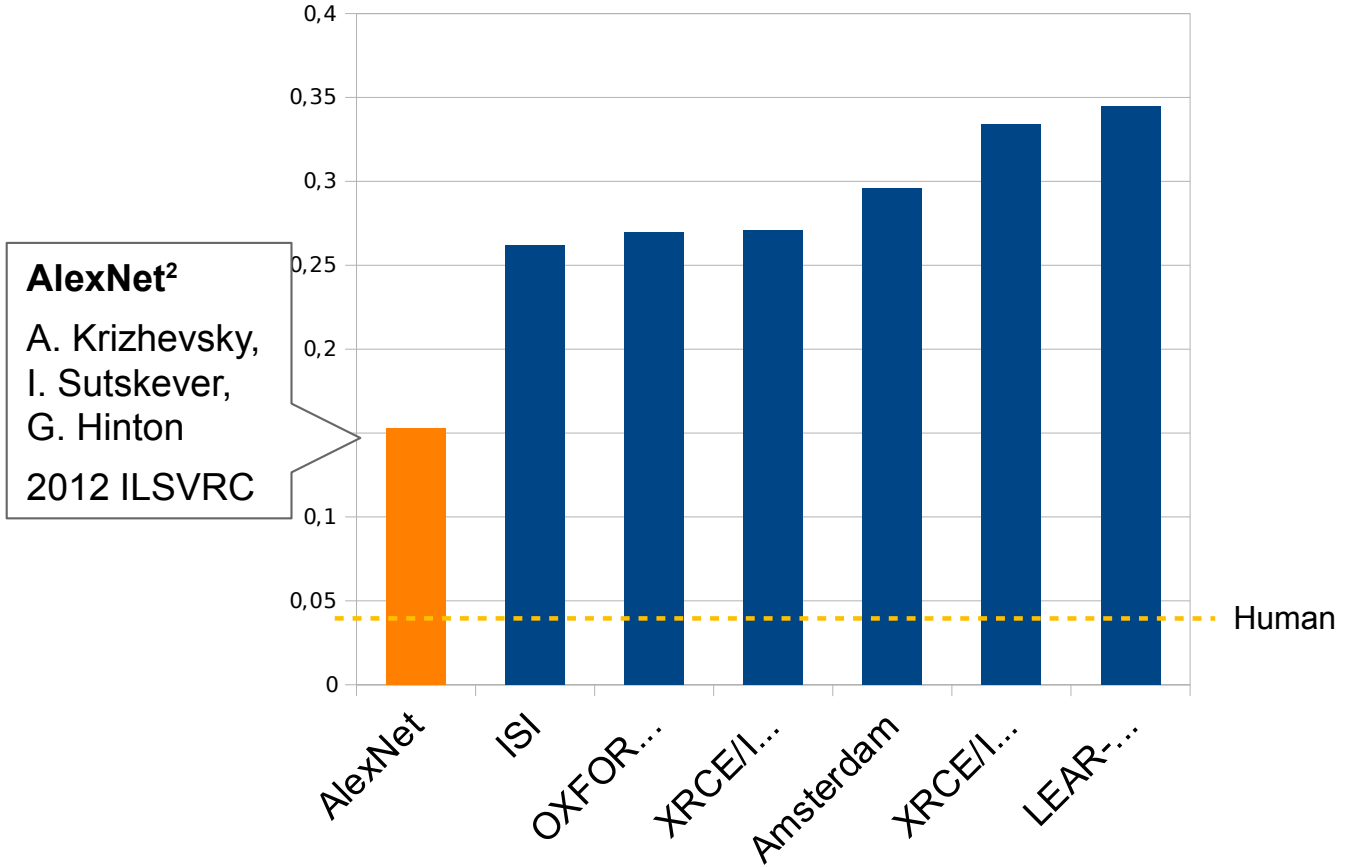
Monde réel

[1] TOP500 List - https://www.top500.org/

[2] https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research

The big Controversy

Images classification - Top 5 error at ILSVRC 2012[3,4]

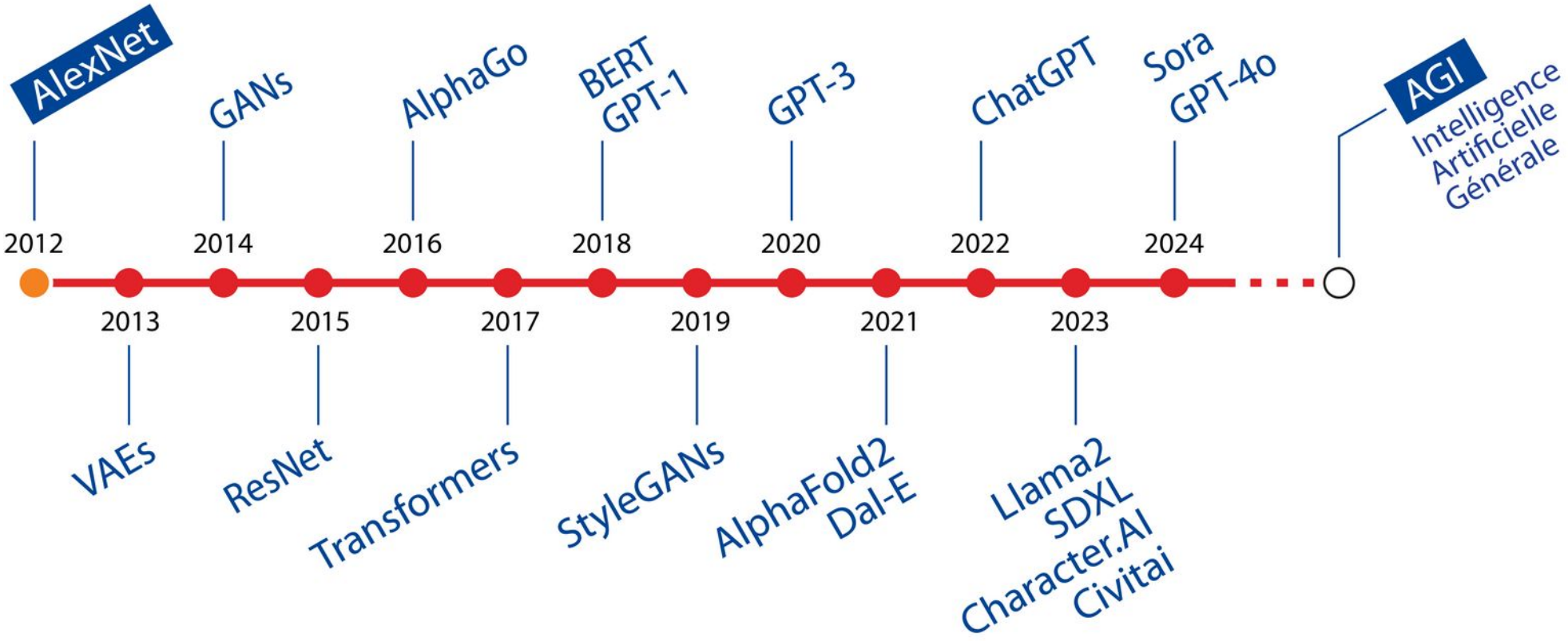


Without mathematical guarantee, DNN have proven to be **more effective** in the face of the **complexity of the real world !**

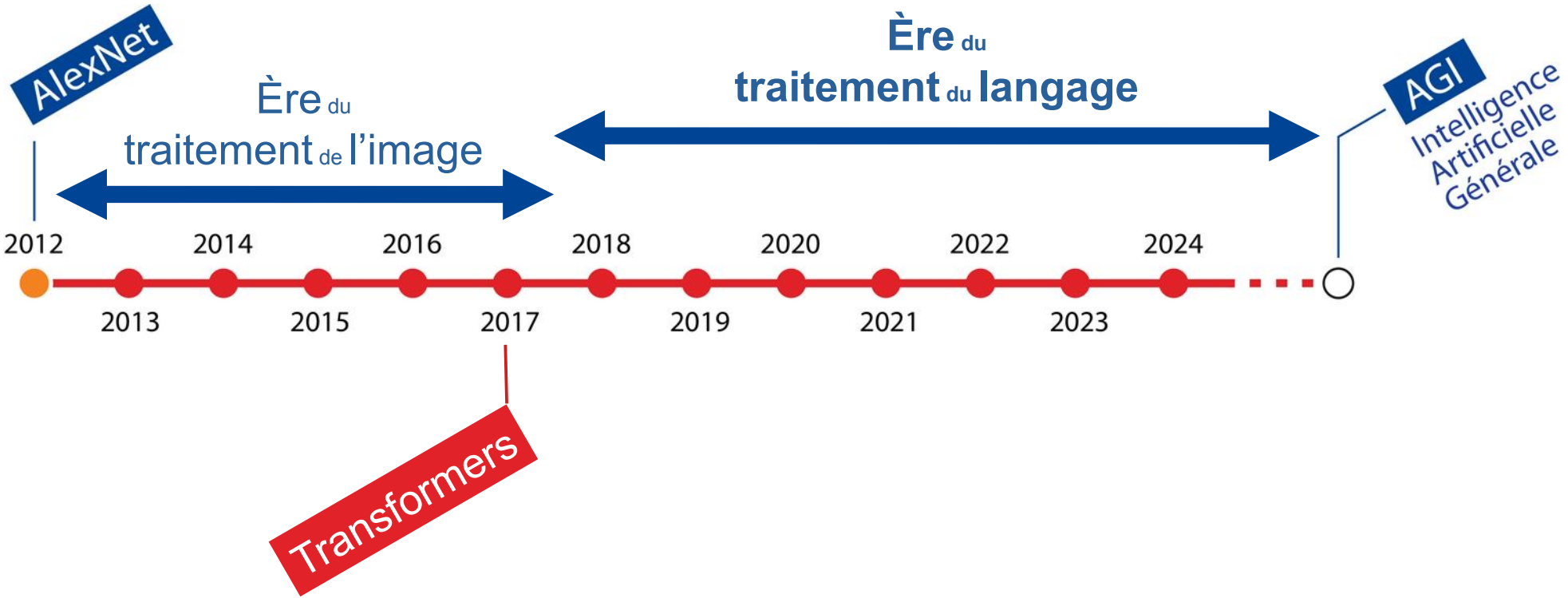
[2] AlexNet at ILSVRC, https://image-net.org/challenges/LSVRC/2012/results.html
[3] ImageNet Large Scale Visual Recognition, https://www.image-net.org/challenges/LSVRC/

[4] Similar evolution in NLP, translation, board games, etc.

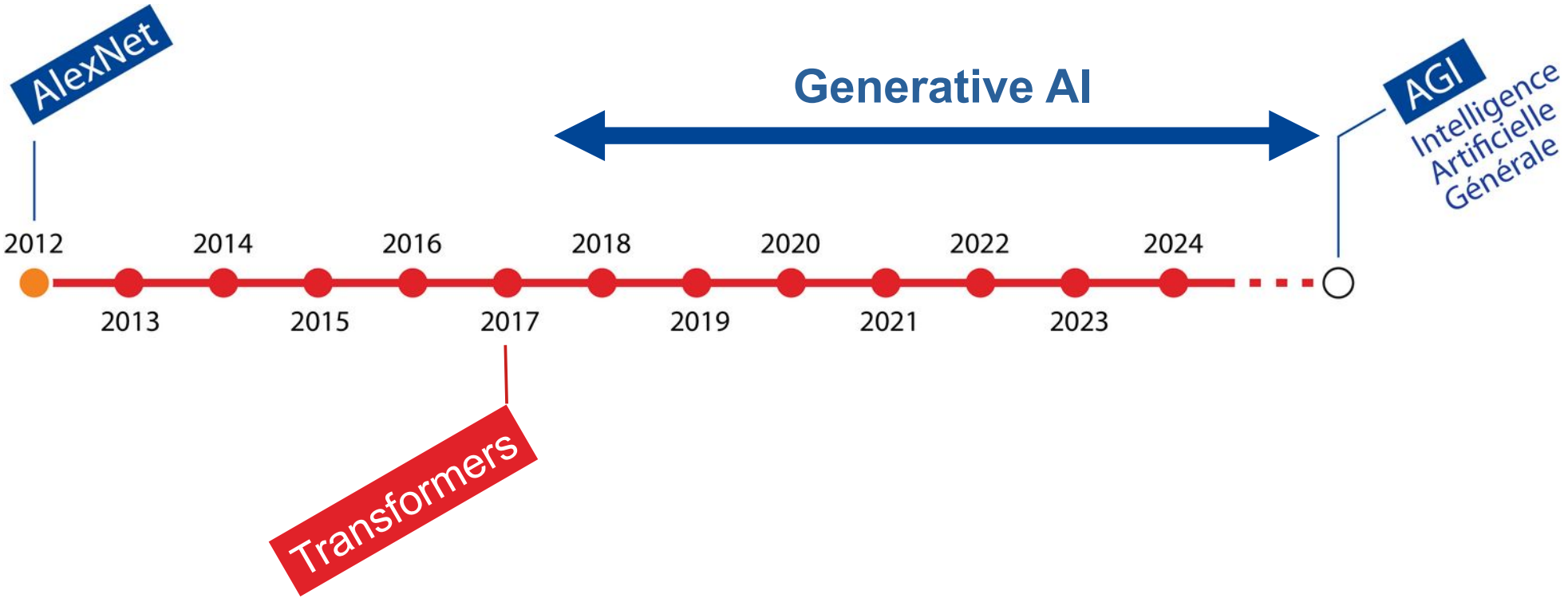
From AlexNet to... ?



From AlexNet to... ?



From AlexNet to... ?



« Small AI »

10 M parameters

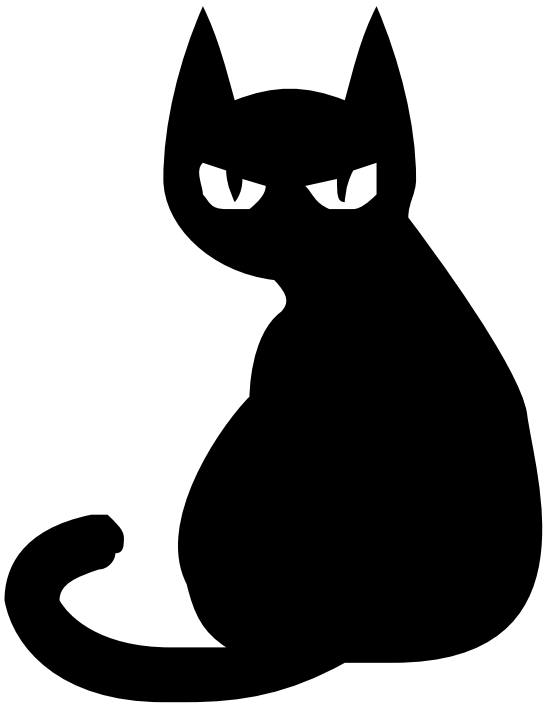
x 100.000

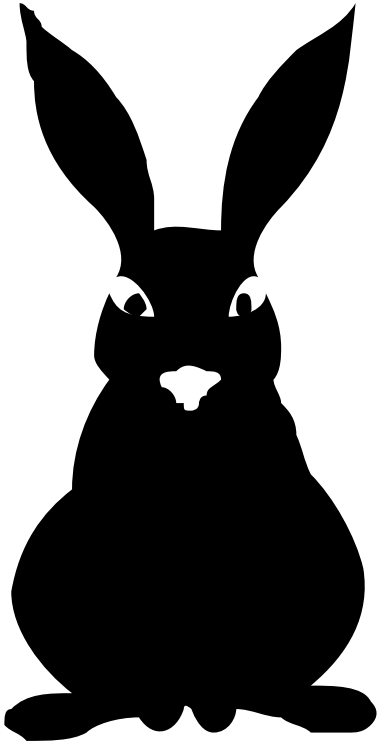
« Large AI »

1000 G parameters

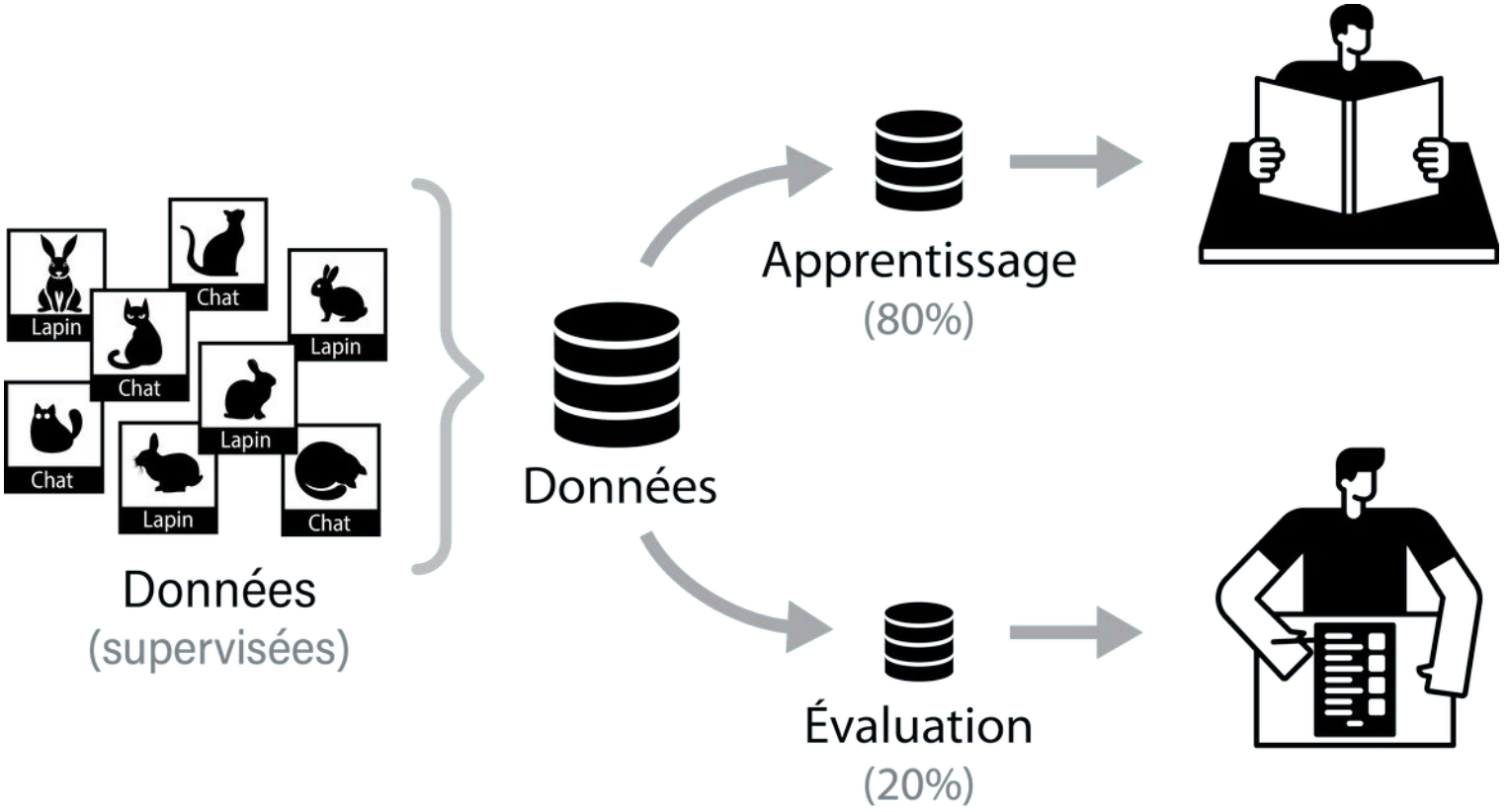
Un exemple de « petite » IA :

Imaginons que l'on veuille **classer** des photos de **chats** et de **lapins** :

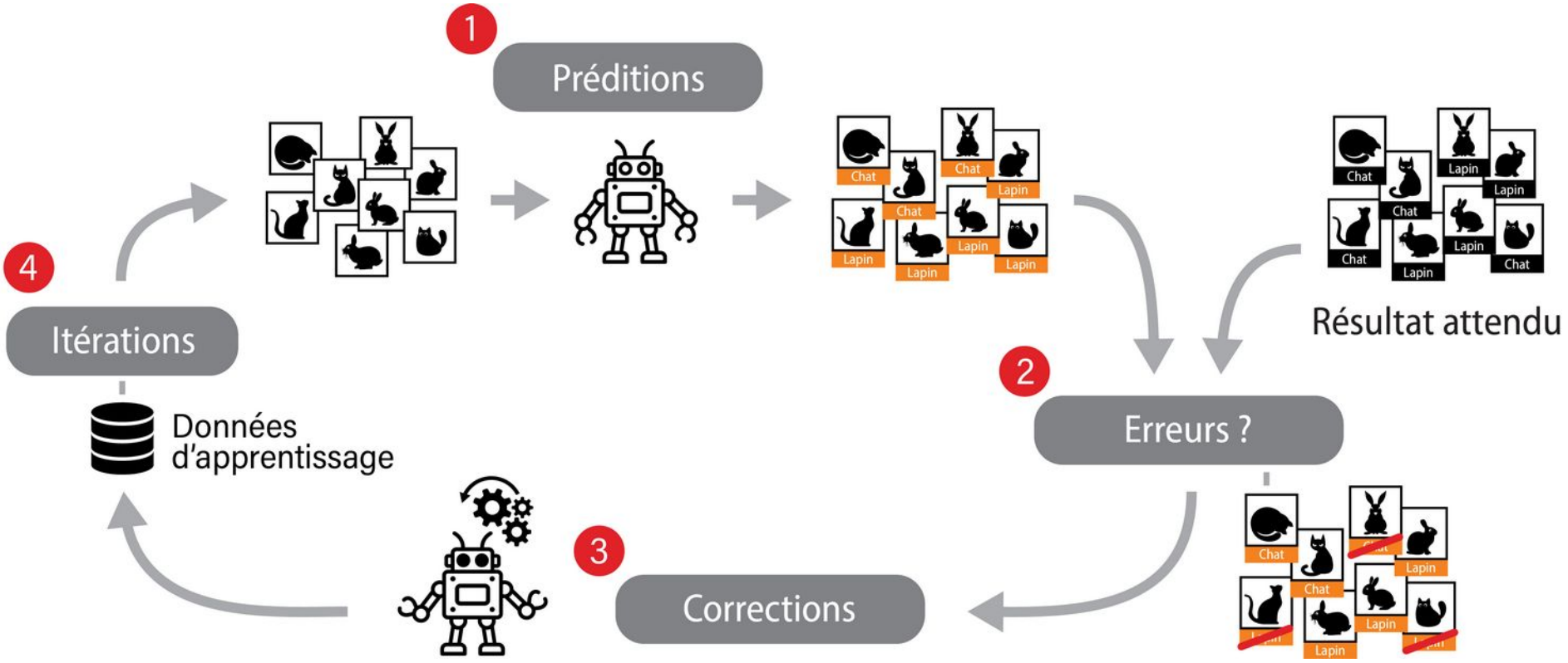




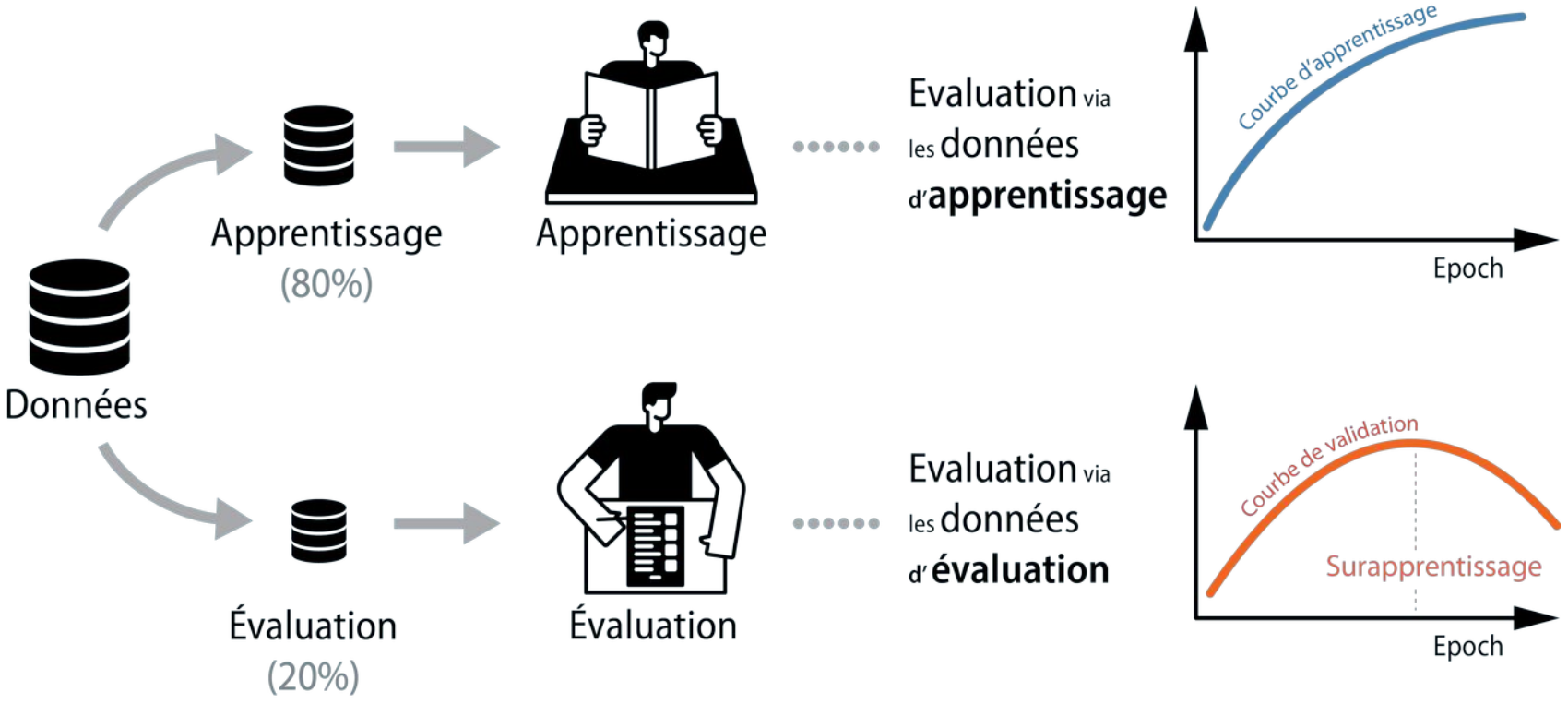
Tout d'abord, il nous faut des **données !**



On va ensuite **entraîner** notre modèle :



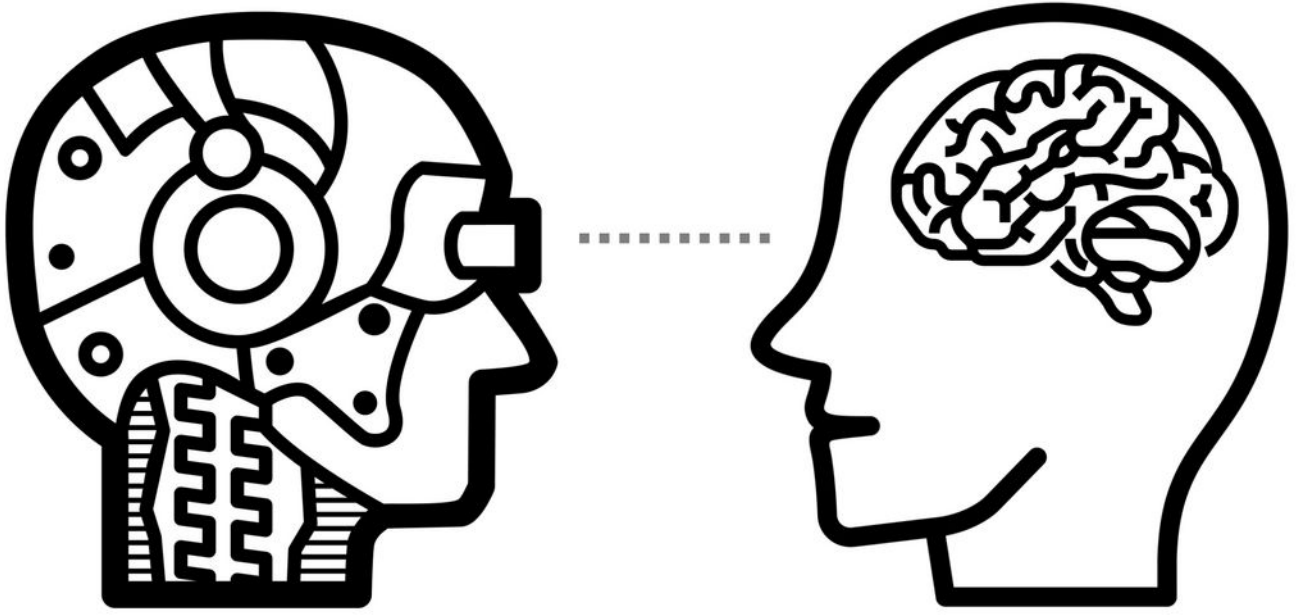
...tout en **suivant** son **apprentissage**



L’éducation d’une « GRANDE » IA

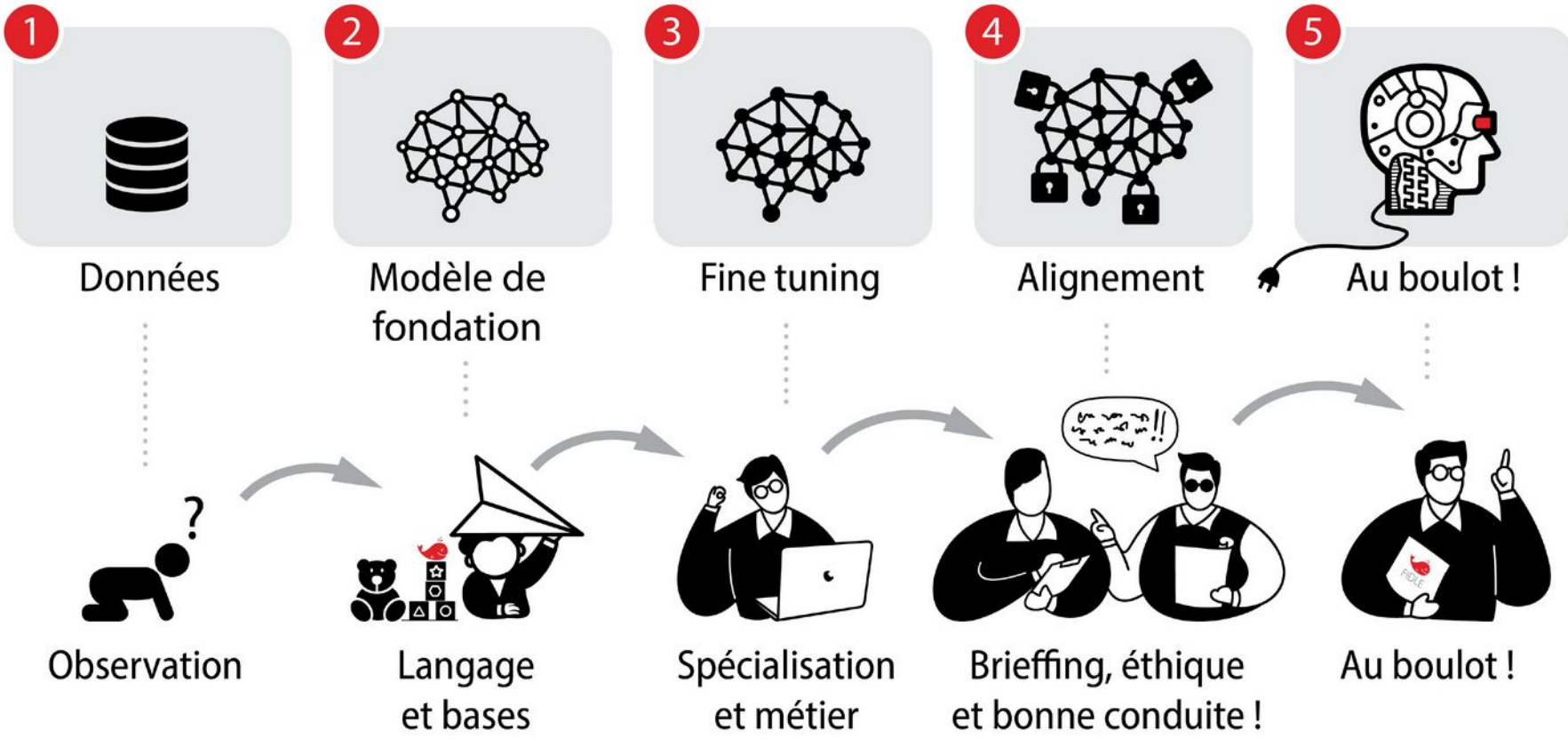
Celles qui sont à la mode !

C’est un peu comme chez nous,les humains !



Pfff...

L'éducation d'une « GRANDE » IA



Micro lexique

Anti-jargon de l'IA

LLM (Large Language Model) :
Modèle de langage ou conversationnel

Diffusion models : Modèles génératifs utilisés pour créer des images, su du son ou des vidéos

Transformer : Architecture de base derrière les LLMs, proposée en 2017.

Prompt : Consigne textuelle (ou orale) donnée à un modèle génératif pour guider la génération.

RAG (Retrieval-Augmented Generation) :
Technique combinant génération (LLM) et recherche de documents (retrieval)

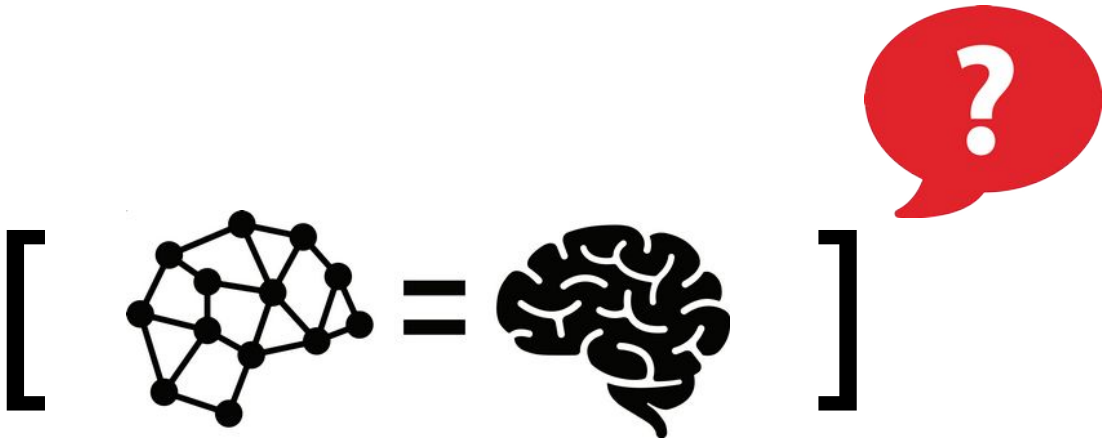
Biais : Préjugés ou discriminations reproduits ou amplifiés par un modèle en raison des données sur lesquelles il a été entraîné.

Temperature : Paramètre qui contrôle la créativité des réponses d'un modèle
(plus élevé = plus créatif)

Context : Quantité maximale de texte (en tokens) que le modèle peut prendre en compte en une fois.

Fine-tuning : entraînement complémentaire d'un modèle préexistant sur des données spécifiques pour l'adapter à une tâche ou un domaine particulier.

Alignement : Processus visant à faire en sorte que les modèles d'IA se comportent conformément aux valeurs humaines, aux intentions de leurs concepteurs et aux attentes des utilisateurs.



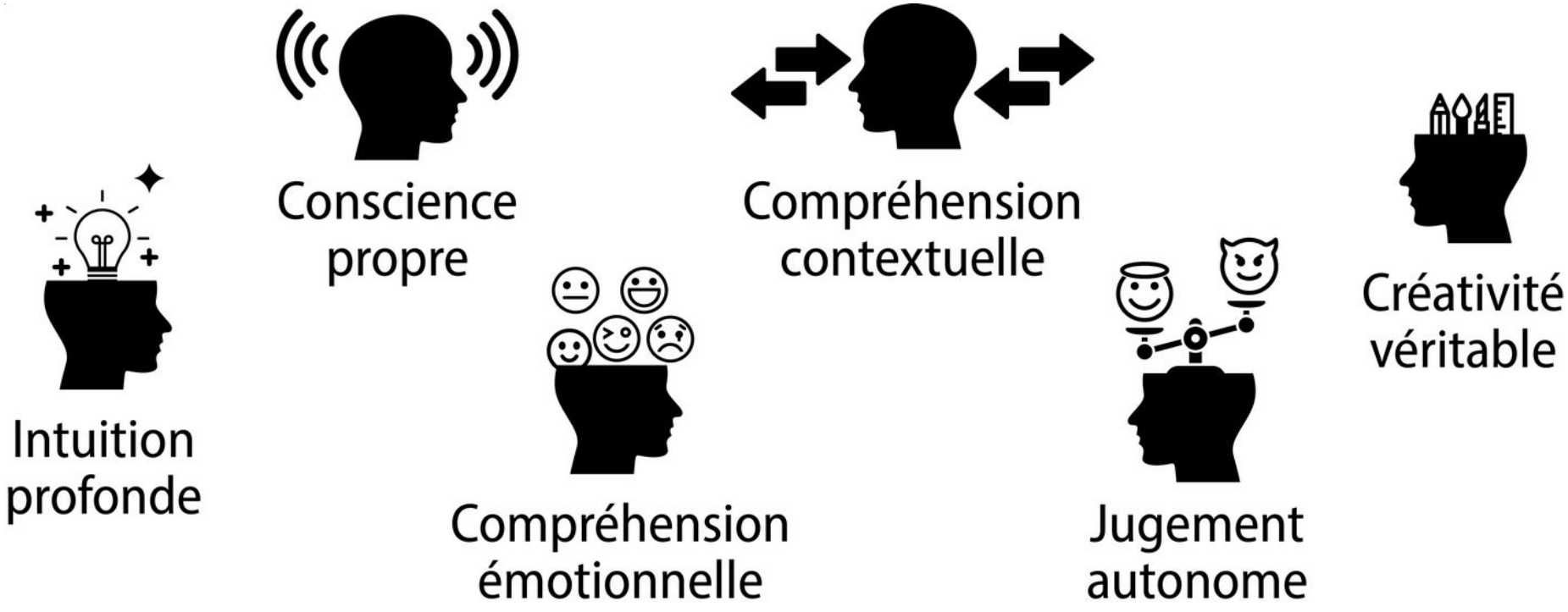
IA vs Humains ?

Une IA est (potentiellement) capable **d'effectuer** la **plupart** de **nos tâches intellectuelles…**

...Même si certaines différences et/ou limites existent encore*.

(*) ...tout au moins, en mai 2025 ;-)

On cite usuellement* :



(*) ...tout au moins, en mai 2025 ;-)

Comment ~~utiliser~~

travailler

~~faire~~ avec

une **IA*** **?**

(*) De type générative / LLM

Comme avec un **collaborateur** interne ou externe… Il est essentiel de savoir :

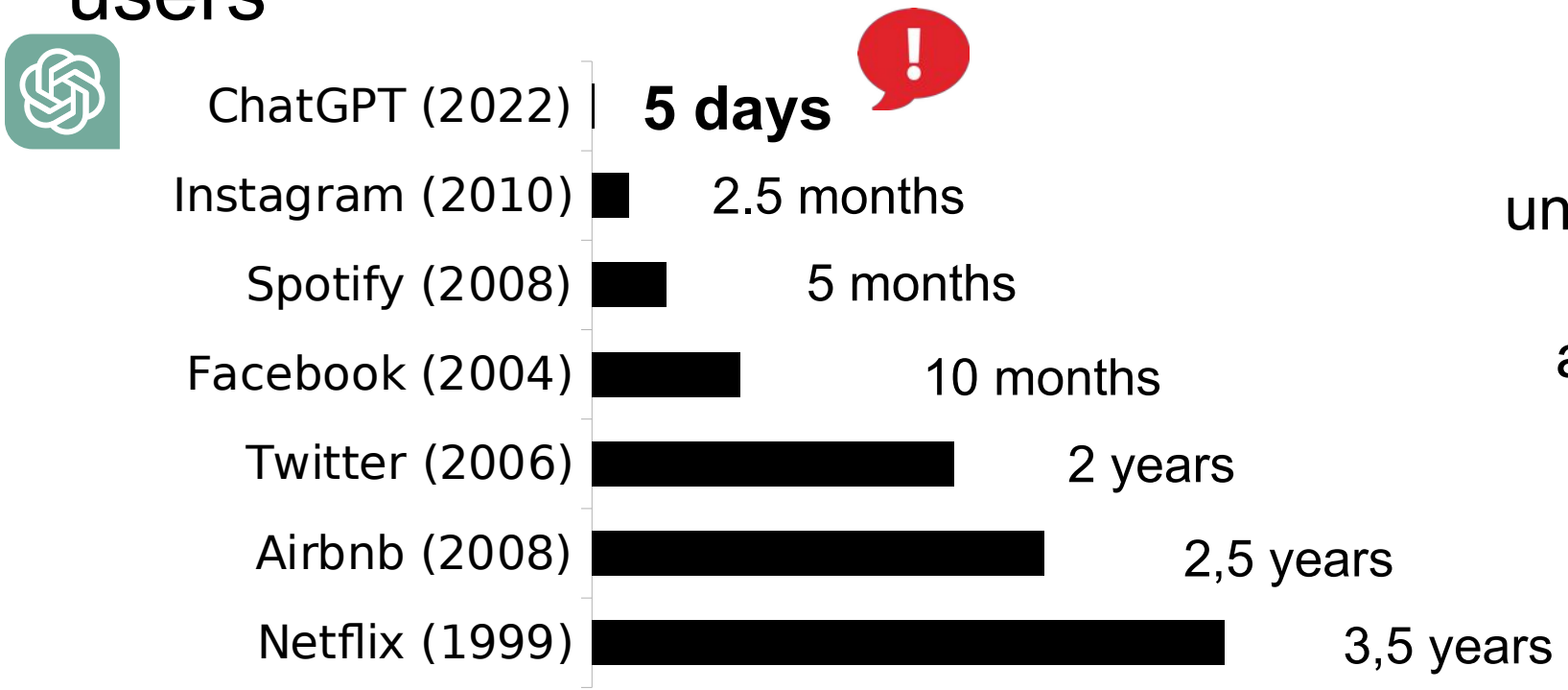
1. Évaluer la **confidentialité** et la **pertinence** de la tâche
2. **Exprimer** précisément ce que l'on **attend** (objectifs, contexte, forme, etc)
3. Évaluer le **résultat** !

[Tout va bien…]

…vraiment ?

Apropriation... ?

Time by service to reach 1 million users

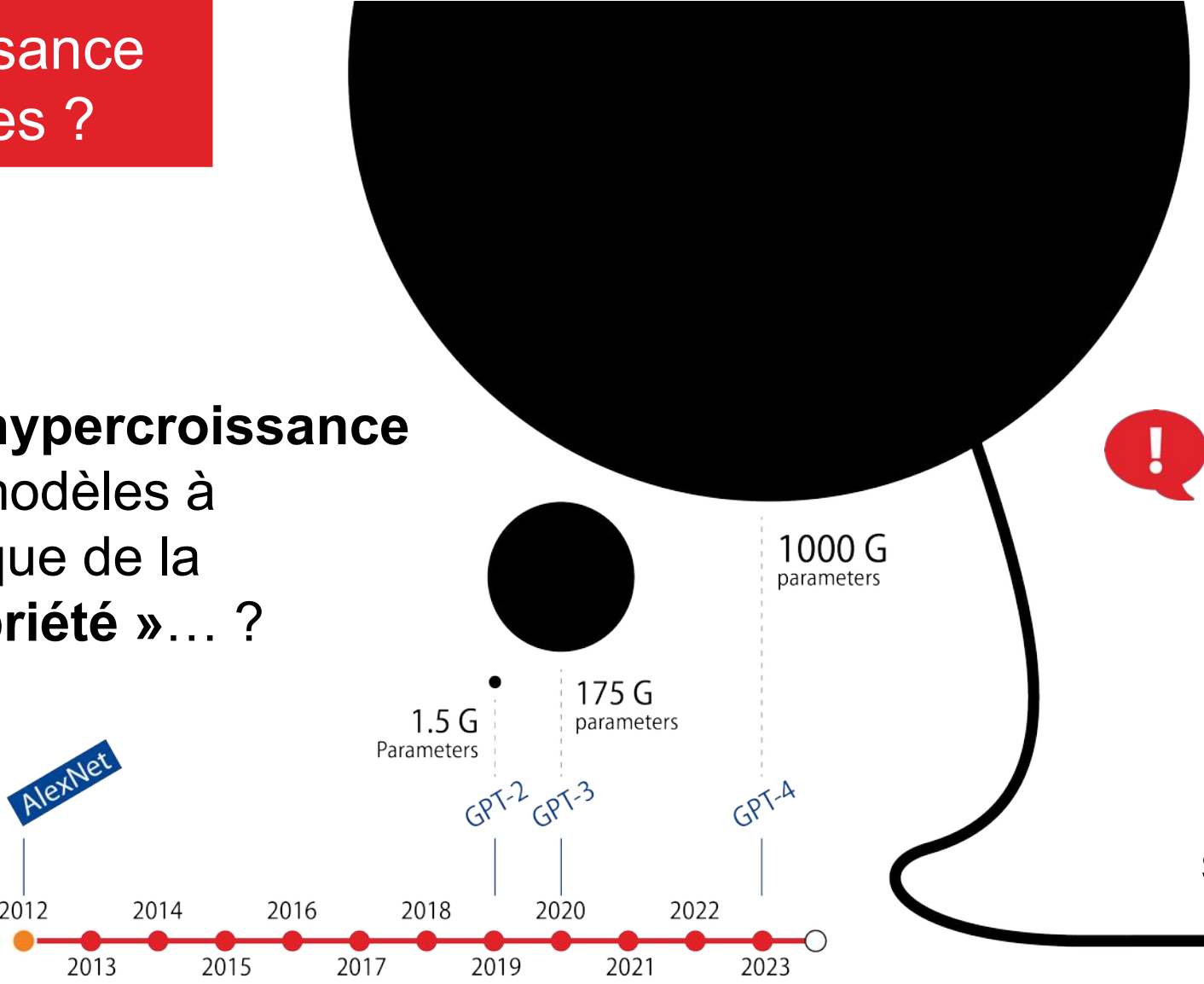


An unprecedented social appropriation

https://www.statista.com/chart/29174/time-to-one-million-users/

Une croissance sans limites ?

Une **hypercroissance** des modèles à l'époque de la **« sobriété »**… ?



Mais ce n'est pas tout...

...et tout n'est pas simple !!

FIDLE PRÉSENTE :

GRAND PROCÈS DE L'IA

L'IA est-elle une amie qui nous veut du bien ?

Grand procès de l'IA

1. Violation de la **vie privée** et du **droit d'auteur**

2. **Biais** et **discrimination**

3. Harcèlement et **Manipulation**

4. Absence de **transparence**

5. **Hallucination** et **mensonges**

6. Perte et dilution de la **responsabilité**

7. **Déshumanisation** des échanges

8. Perte **d'autonomie** humaine

9. Perte de **confiance**

10. Menace pour la **sécurité**

11. Impact **économique**

12. Nouvel ordre **oligarchique**

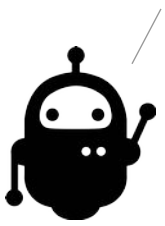
13. Impact **social** et **inégalités** sociales

14. Impact **environnemental**

15. Menace pour nos **démocraties**

16. **Singularité** technologique

Merci beaucoup !



More about :

→ https://fidle.cnrs.fr

FIDLE