

Machine Learning

Concepts

Martin Vielvoye 2021

Pourquoi en parle-t-on autant ?

Une vue d'ensemble en 2021

La loi de Moore

La loi de Moore :

"Le nombre de transistors par puce électronique double tous les deux ans."

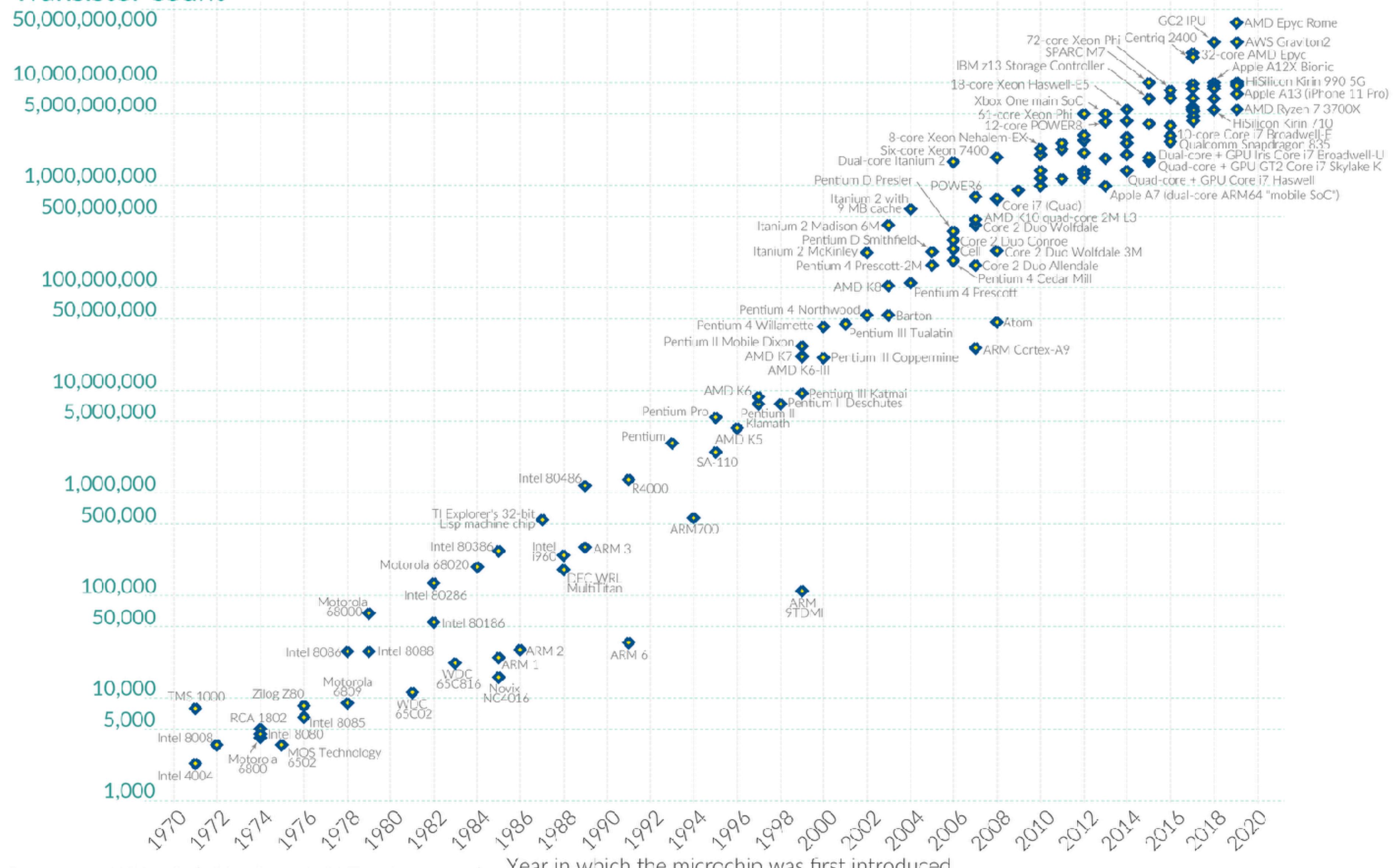
- Des processeurs de plus en plus **puissants/rapides** (*FLOPS*).
- La mémoire est toujours plus **grande**.
- Les ordinateurs seront toujours un peu **moins chers** (relativement à leurs puissance).

Note : La consommation électrique par opération ne réduit pas, par conséquent la consommation totale de chaque machine est de plus en plus grande.

Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Transistor count



Internet et le big data

Global Data

4.4 MM d'utilisateurs à travers le monde.

44 MM de GB par jour, prédis à **463 MM** Gb par jour en 2025.

Youtube

82 années de vidéos uploadées tous les jours et **28 500 années** de vidéos visionnées par jour.

Facebook

Plus de **5 MM** de likes sur Facebook tous les jours.

IoT

De plus en plus d'appareils connectés à Internet.

(Smartphone, montre intelligente, prise connectée, etc)

Donnée et predictions de 2017. Le tout boosté par le confinement mondial en 2020.

<https://www.slideshare.net/Micro-Focus/growth-of-internet-data-2017>

Qu'est-ce-qu'une IA performante ?

- Des utilisations **pratiques** des algorithmes.
- Une **performance** qui s'approche ou qui dépasse des humains.
- Des applications **utiles** aux entreprises.
- Une **utilisation simplifiée** pour des utilisateurs plus natifs avec les nouvelles technologies.
- Des perspectives d'utilisations très ouvertes aux potentiels **innovateurs**.

Supervisé vs. Non Supervisé

Let's Classify!

Est-ce-que la photo représente un chien ?

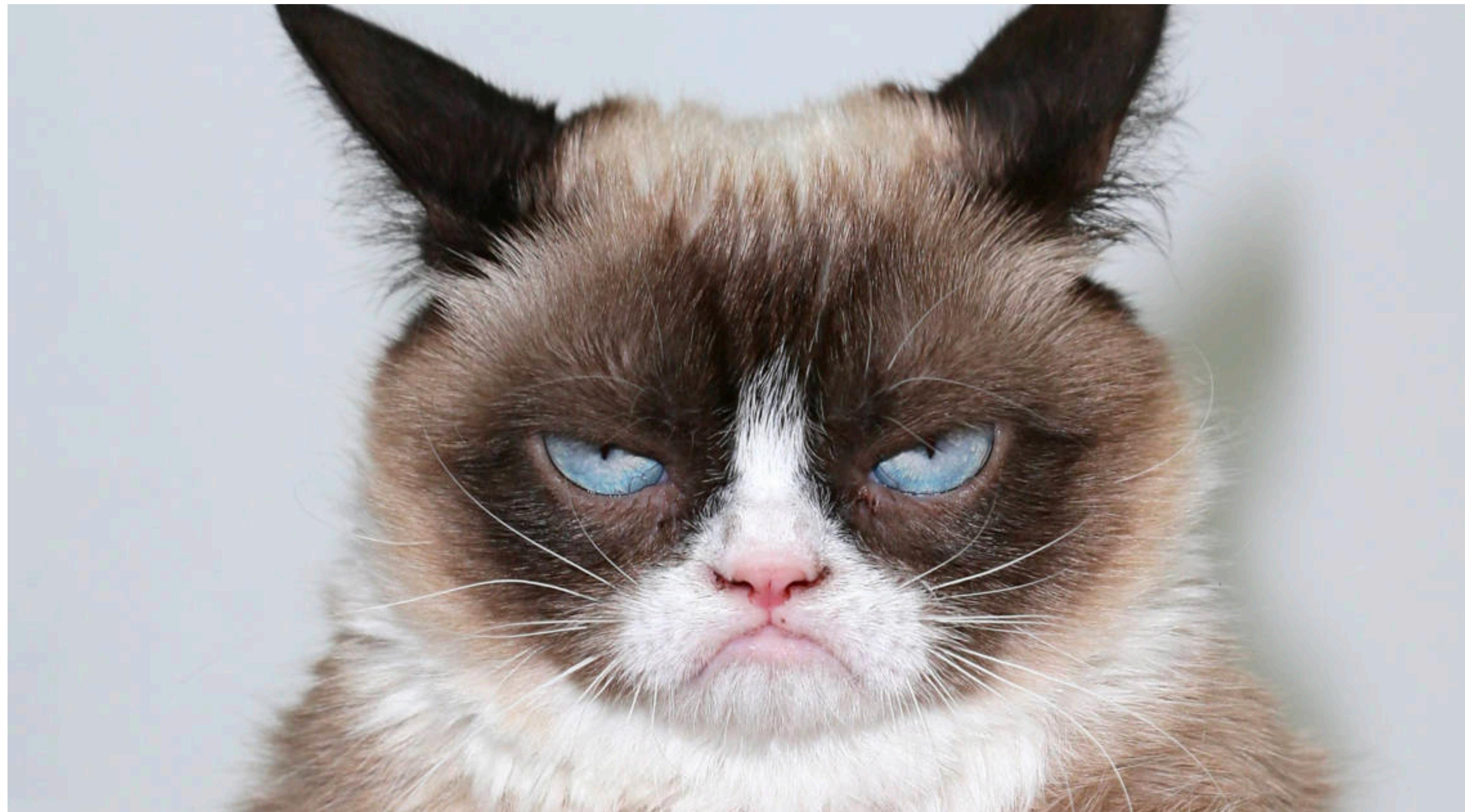
Test 1



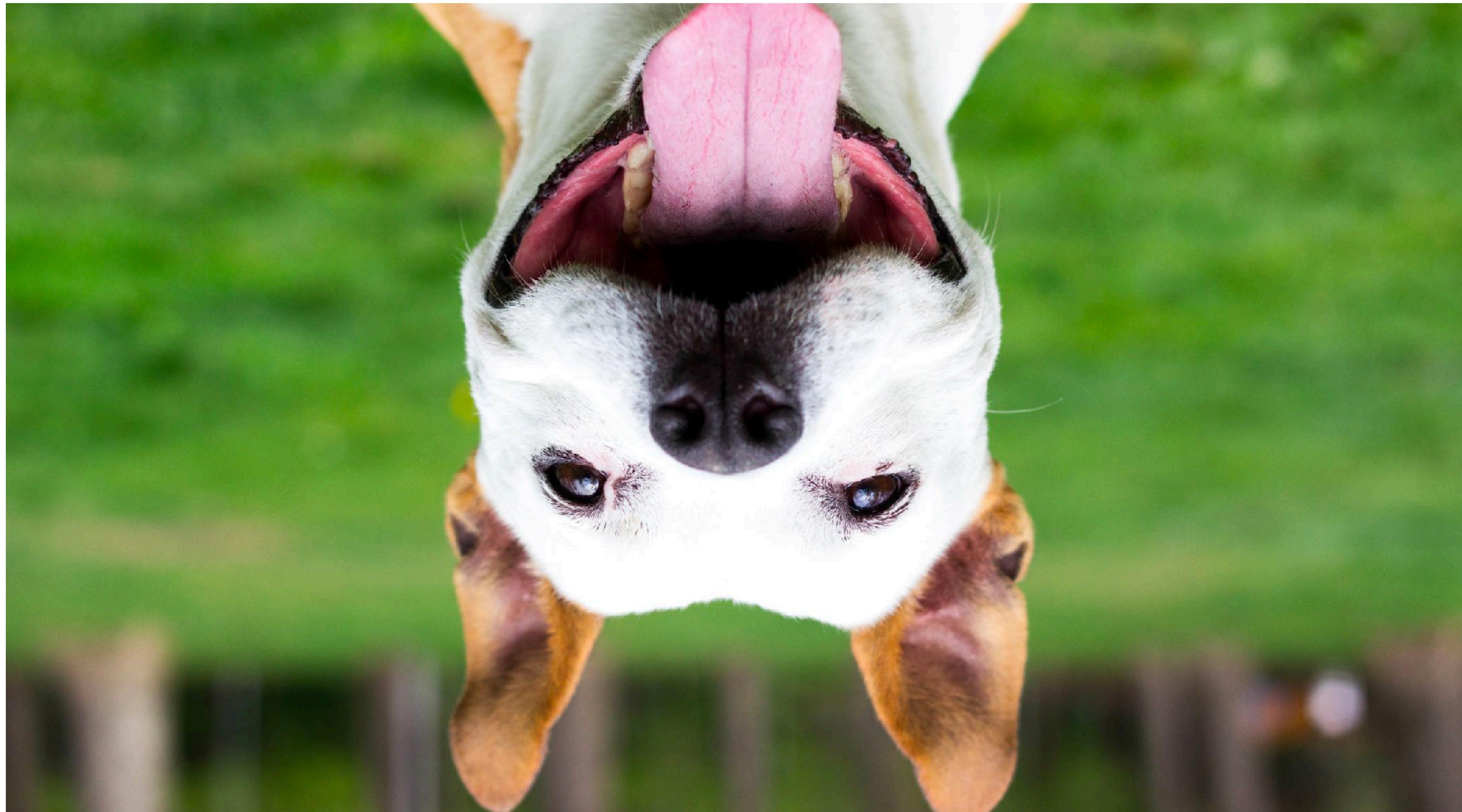
Test 2



Test 3



Test 4



Test 5



Test 6



Test 7



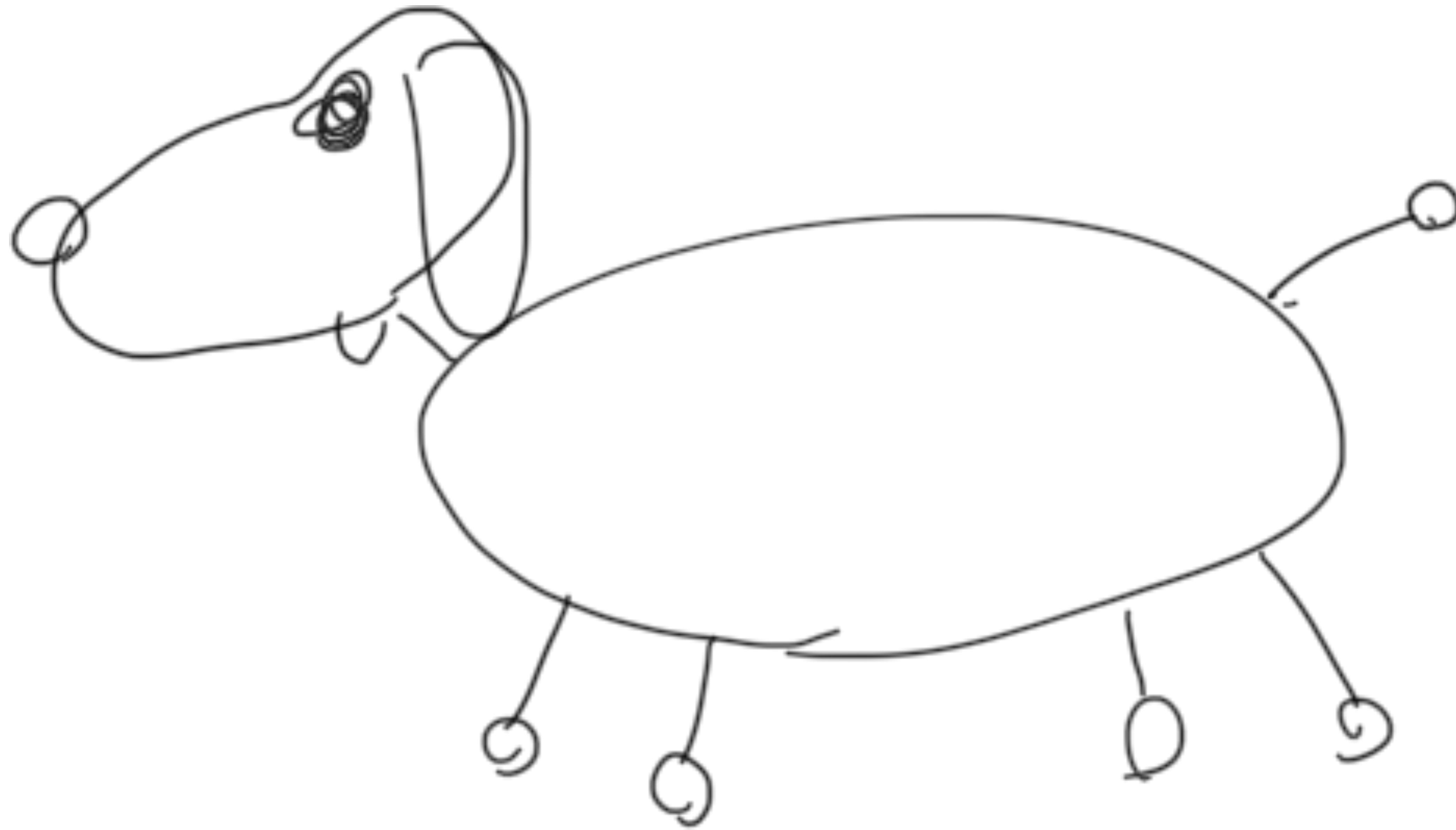
Test 8



Test 9



Test 10



Test 11



Test 12



Let's Classify! (Part 2)

Quel est le prix du bien immobilier ?

Test 1



Test 2



Test 3



Let's Classify! (Part 3)

Test 1



Test 2



Test 3



Test 4



Test 5



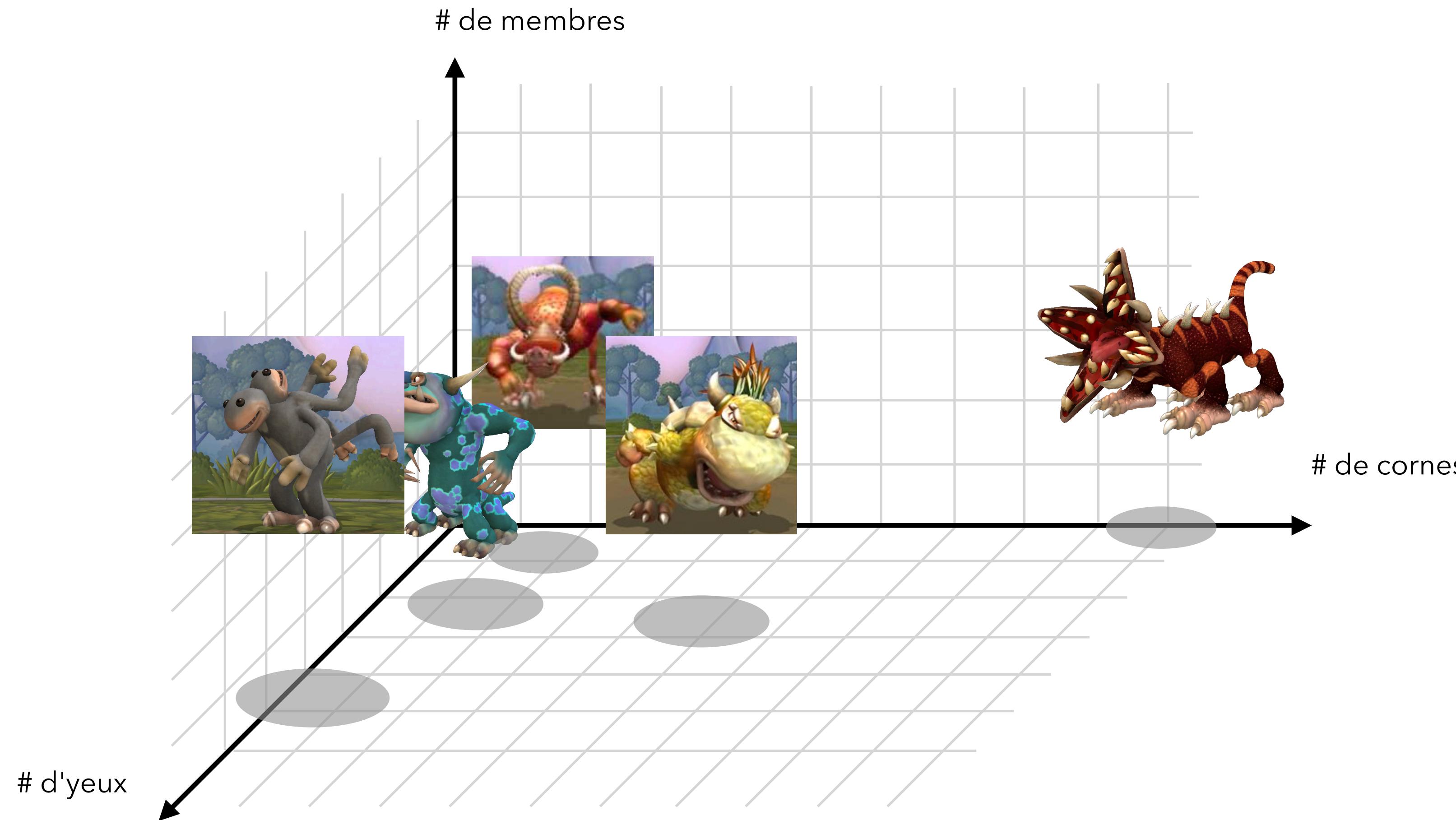
Classer sans label



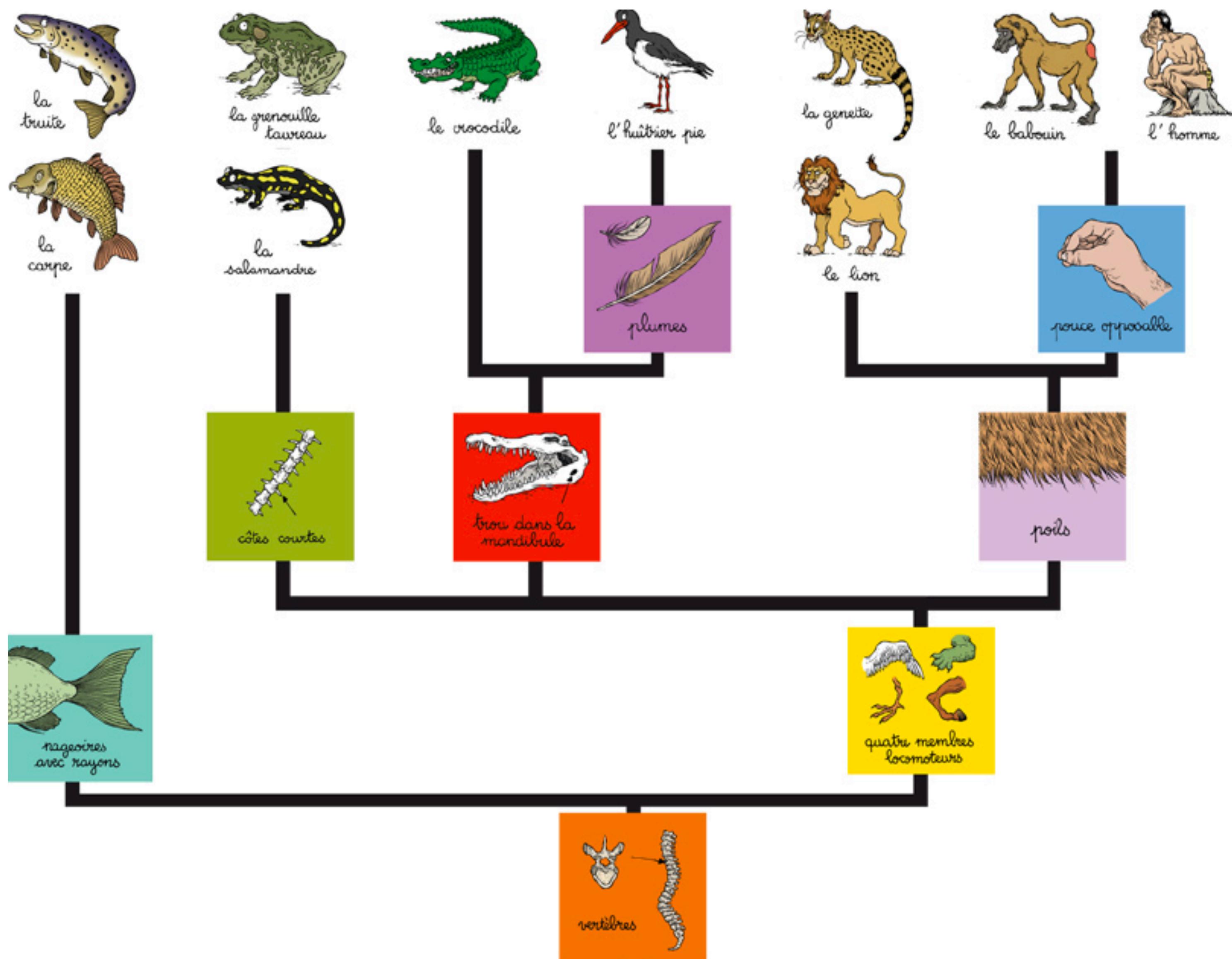
?



Classer sans label



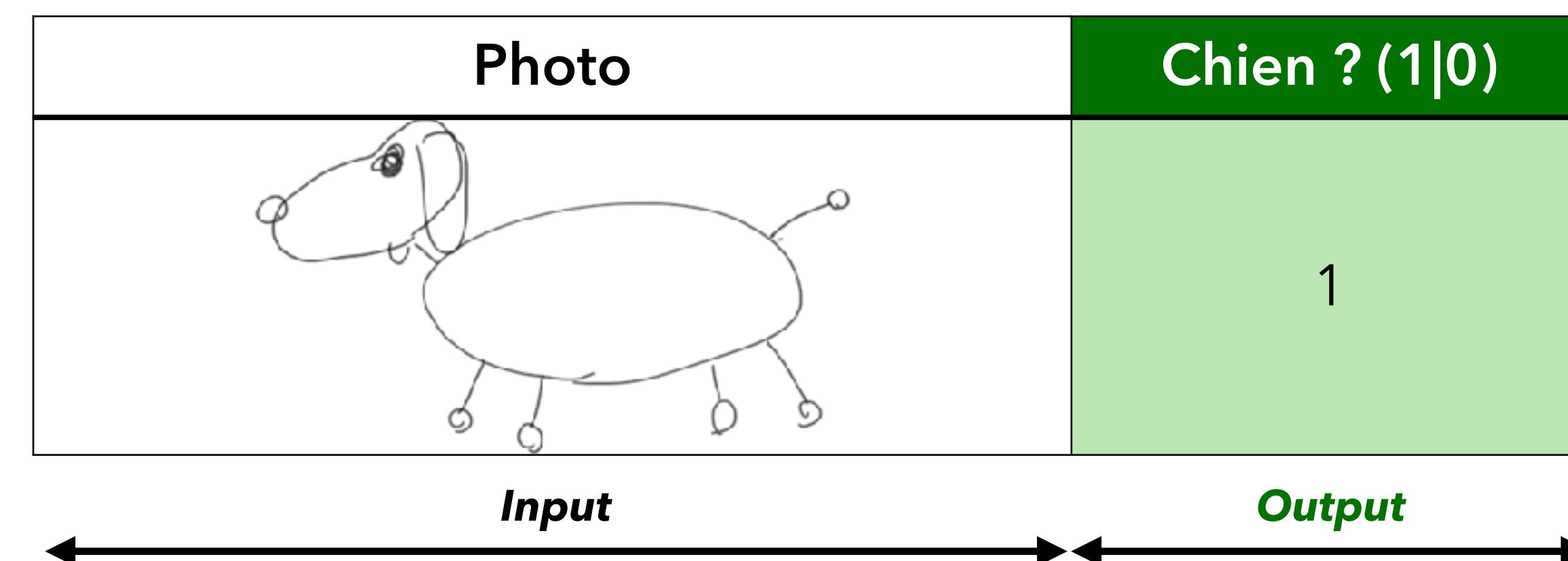
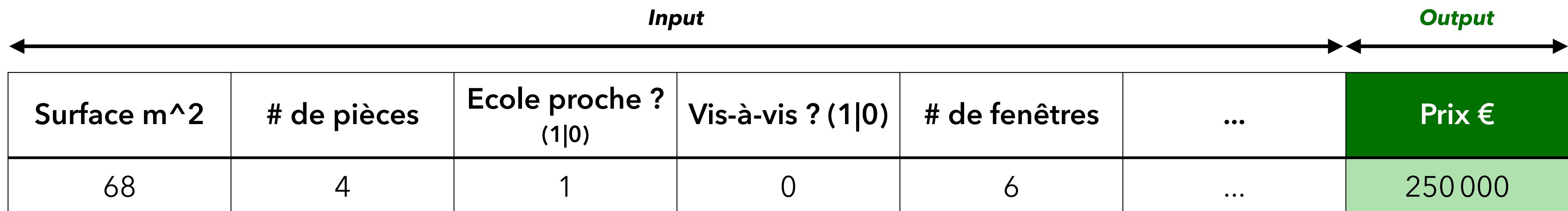
Classer sans label: Darwin



Apprentissage Supervisé

Input et Output

- Nous avons des données et nous savons à quoi correspond une donnée "correcte" (**label**).
 - Il y a un label **quantifiable** symbolisant la relation entre *l'input* et *l'output*



Classification et Régression

- **Régression** : Prédire des résultats avec des outputs *continus*.
- **Classification** : Prédire des résultats avec des outputs *discrets*.

Régression

Surface m ²	# de pièces	Ecole proche ? (1 0)	Vis-à-vis ? (1 0)	# de fenêtres	...	Prix €
68	4	1	0	6	...	250 000

Photo	Chien ? (1 0)
	1

Classification

Apprentissage Non-Supervisé

Une donnée sans label

- Approcher un problème avec peu, ou pas, d'idée précise d'à quoi ressemble le résultat final.
- Une analyse de la *structure* des données, sans savoir le vrai impact de chaque *variable / caractéristique*.

<i>Variable/ Caractéristique</i>	Transfert €	Type de transfert	Date d'envoi	RIB de transfert	Date d'arrivée	RIB d'arrivée	...
Donnée	200	PayPal	01/12/1998	CE42 500 6000 3200	08/12/1998	BE55 420 69000 1000	...

Clustering et Détection d'anomalies

- **Clustering** : Grouper les données similaires en catégories distinctes.
- **Anomalies** : Détection d'anomalies dans un jeu de données.

Grouper ces symboles par similarités.



Est ce que ce transfert est une fraude ou non ?

Transfert €	Type de transfert	Date d'envoi	RIB de transfert	Date d'arrivée	RIB d'arrivée	...
200	PayPal	01/12/1998	CE42 500 6000 3200	08/12/1998	BE55 420 69000 1000	...