EFELIA Côte d'Azur

Ecole Française de l'Intelligence Artificielle, 3IA Côte d'Azur

Formation à l'IA pour les personnels UniCA

Séance 2/5, 12 octobre 2023

UNIVERSITÉ CÔTE D'AZUR | INSTITUT INTERDISCIPLINAIRE D'INTELLIGENCE ARTIFICIELLE 3IA CÔTE D'AZUR

FRANCE 2030 cnrs EURECOM Inria Inserm skema BUSINESS SCHOOL

EFELIA Côte d'Azur : Branche formation du 3IA Côte d'Azur

- ANR CMA : 2022-2027, 8M€
- Généraliser la formation à l'IA :
  - Bac-3 à Bac+8 et FC
- Grâce à vous : 12 Nouvelles mineures de master en sep. 2023, compétences transversales L en 2024
- Ici, la FC pour vous : Vous permettre de capitaliser votre investissement pédagogique dans vos activités de recherche

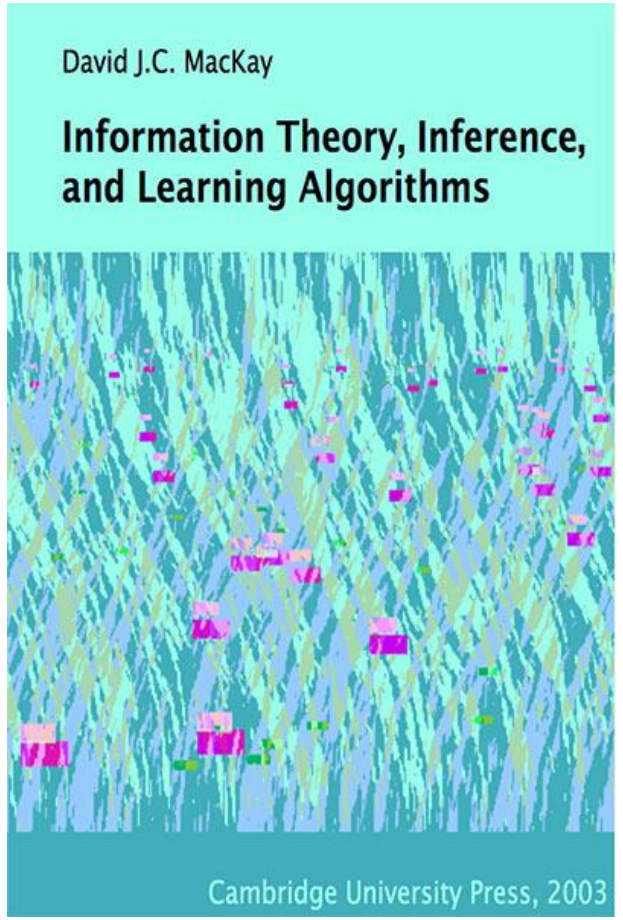
Direction du projet :
Charles Bouveyron, professeur en mathématiques appliquées, directeur du 3IA Côte d'Azur

Direction scientifique :
Lucile Sassatelli, professeure en informatique, directrice scientifique
Vincent Vandewalle, professeur en mathématiques appliquées, directeur scientifique adjoint

Pilotage opérationnel :
Violette Assati, Cheffe de Projet

D’où je viens

- Mon premier réseau de neurones artificiels : 2002 en C !
- Ingénieure en électronique, DEA en télécommunications et traitement du signal
- Ma thèse 2005-2008 : *Apport des réseaux de neurones artificiels dans l’inférence sur graphe pour les codes correcteurs d’erreur LDPC*
- David J.C. MacKay:
  - “Why unify information theory and machine learning? Because they are two sides of the same coin. In the 1960s, a single field, cybernetics, was populated by information theorist, computer scientitsts, and neuroscientists, all studying common problems. Information theory and machine learning still belong together. Brains are the ultimate compression and communication systems. And the state-of-the-art algorithms for both data compression and error-correcting codes use the same tools as machine learning.”

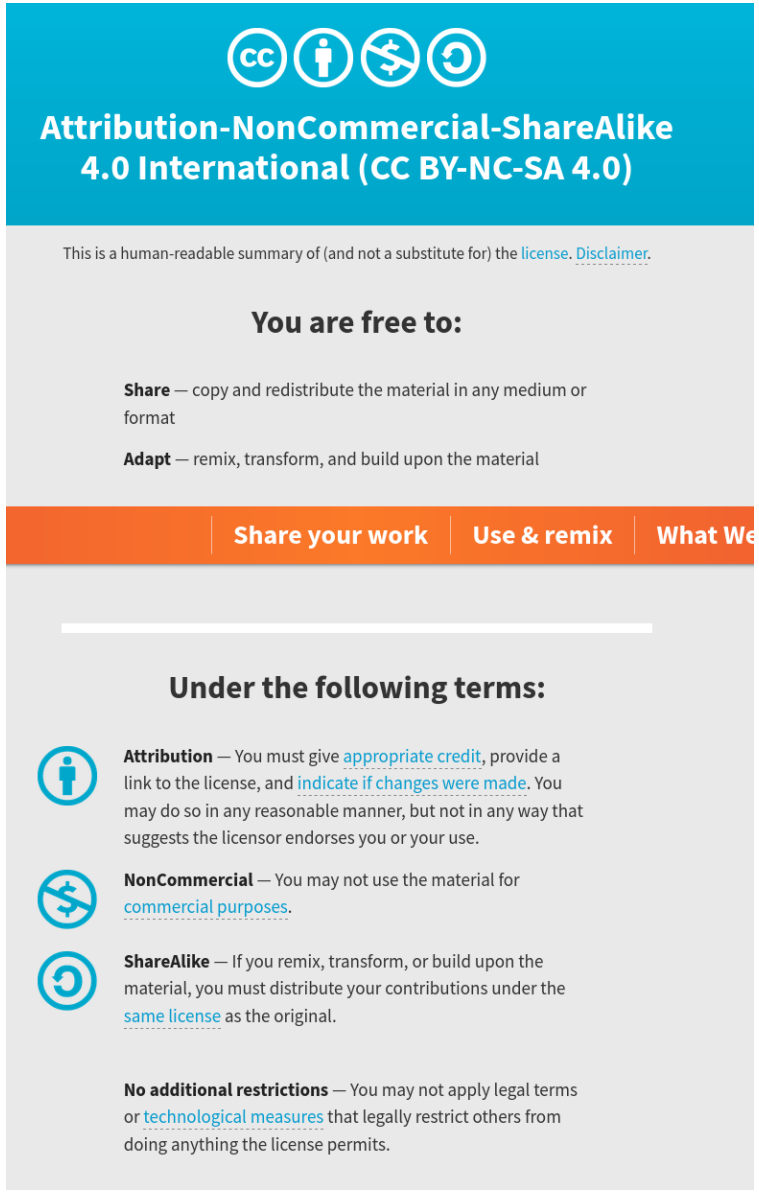


Ce que je fais

- Chaire IUF :
  - Network Streaming of Immersive Media with Machine Learning and User-centric approaches
- PI du projet ANR TRACTIVE
  - Analyse basée IA de la représentation du genre dans les films
  - 6 labos : 3 en informatique, 3 en SHS (ling. et études des médias)
- PI locale pour 3IA-UCA du projet européen AI4Media :
  - A European Excellence Centre for Media, Society and Democracy
  - 1 des 4 centres d'excellence européens en IA

- Mes activités sont centrées sur le multimédia
  - dans différents contextes : réseau, études des médias, études de genre, réalité virtuelle
  - avec des questions méthodologiques communes : analyse de données, IA/ML, optim

License de ce cours : CC BY-NC-SA



Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material

Share your work | Use & remix | What We

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for commercial purposes.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Apprentissage de représentations

Principes, modèles fondation, biais, enjeux

Lucile Sassatelli

Professeure des Universités en Informatique

Directrice scientifique de EFELIA Côte d'Azur

lucile.sassatelli@univ-cotedazur.fr , https://www.i3s.univ-cotedazur.fr/~sassatelli/

Objectifs de cette séance de formation

- Comprendre l'apprentissage de représentation des données
  - Hypothèses, avancées, questions
- Introduire les *modèles fondation*, et comment les utiliser
- Comprendre l'encodage des biais humains dans les modèles d'IA
- Mettre en perspective les enjeux de la/des disciplines

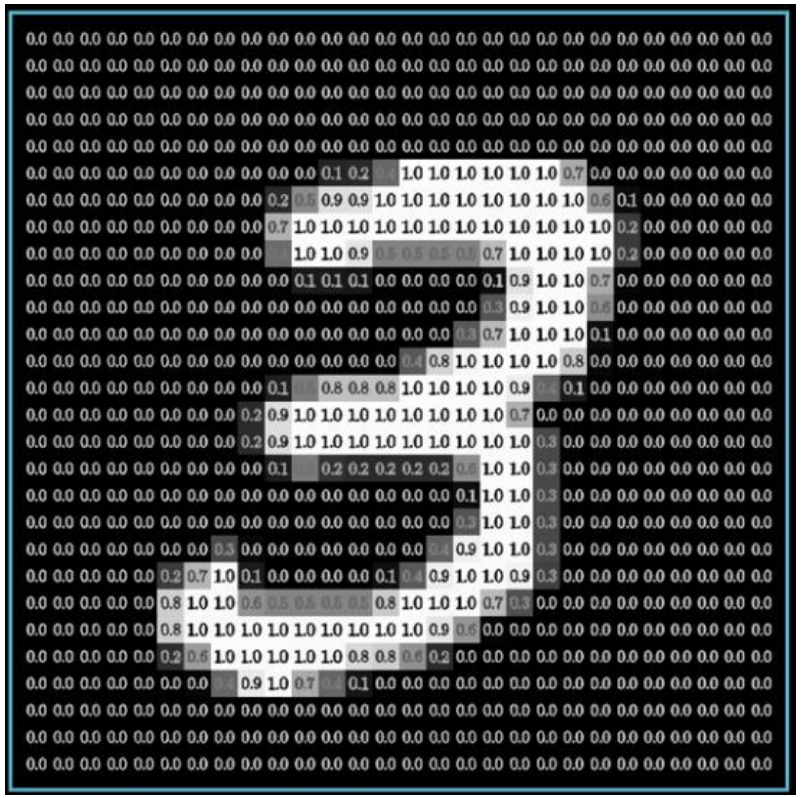
Plan de la formation

1. Apprentissage de représentation pour la reconnaissance de formes
   - Perceptron multi-couche (MLP)
   - Réseaux de neurones convolutionnels (CNN)
     - pour l'apprentissage de motifs pertinents dans les données
2. Apprentissage de représentation de mots
   - Représentations apprises par similarités de contextes
   - Représentation apprises par modèlisation du langage
     - Modèles Transformers et pré-entraînement
3. Modèles fondation : le changement de paradigme en IA
   - Emergence de capacité imprévues
   - En langage, vision, audio... et plus
   - Nouvelles méthodes pour adapter les modèles à des tâches spécifiques
4. Limites et enjeux
   - Environnement social et politique du design et du déploiement des systèmes de ML

Quels critères décrire pour détecter un chiffre manuscript ?

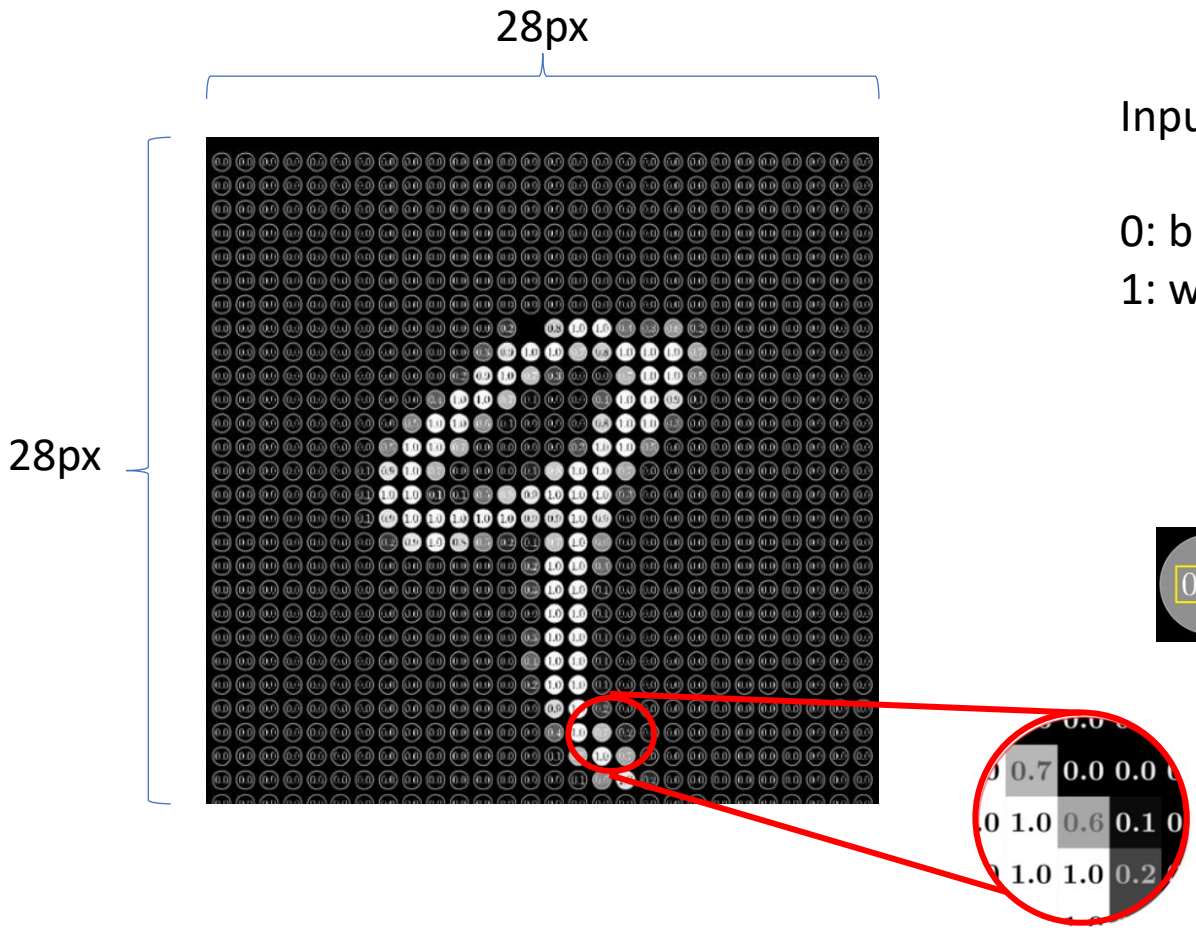
- Il n'est pas possible d'énumérer tous les motifs possibles correspondant à un seul chiffre (épaisseur, inclinaison, etc.).

→ L'IA en général et l'apprentissage automatique en particulier visent à résoudre ce type de tâche ciblée



Comment faire ? Construire une reseau de neurons artificiels ! (v0.1)

- Considérons d'abord l'entrée :

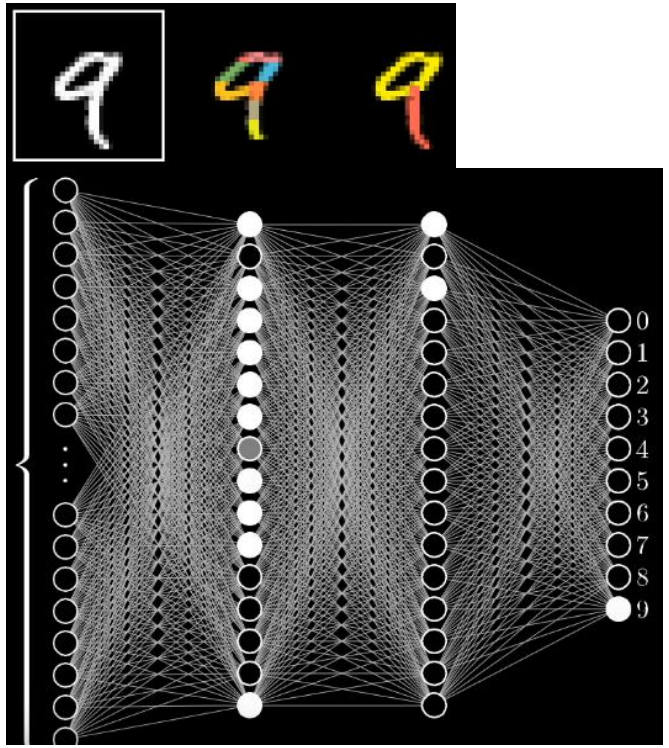


Input size: 28 x 28 = 784

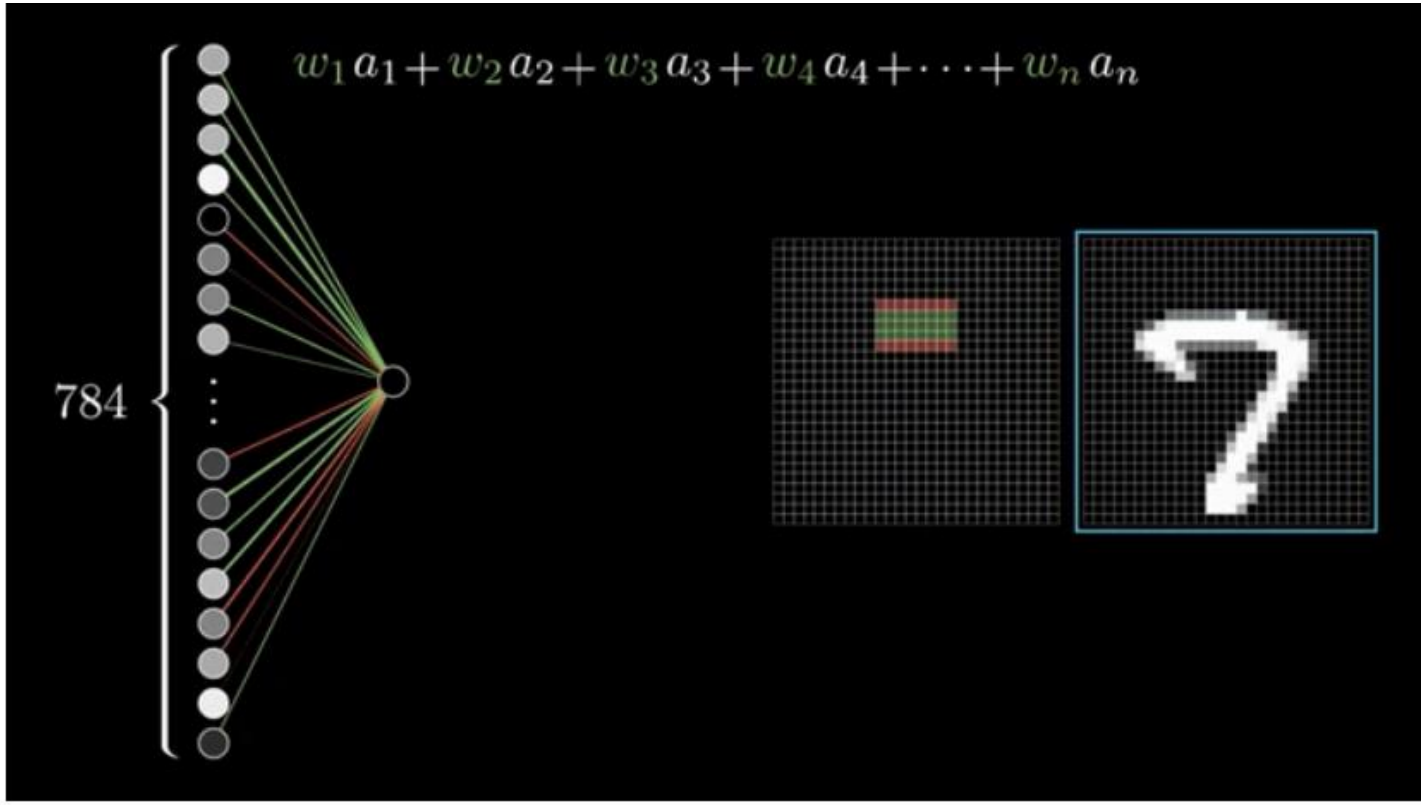
0: black
1: white

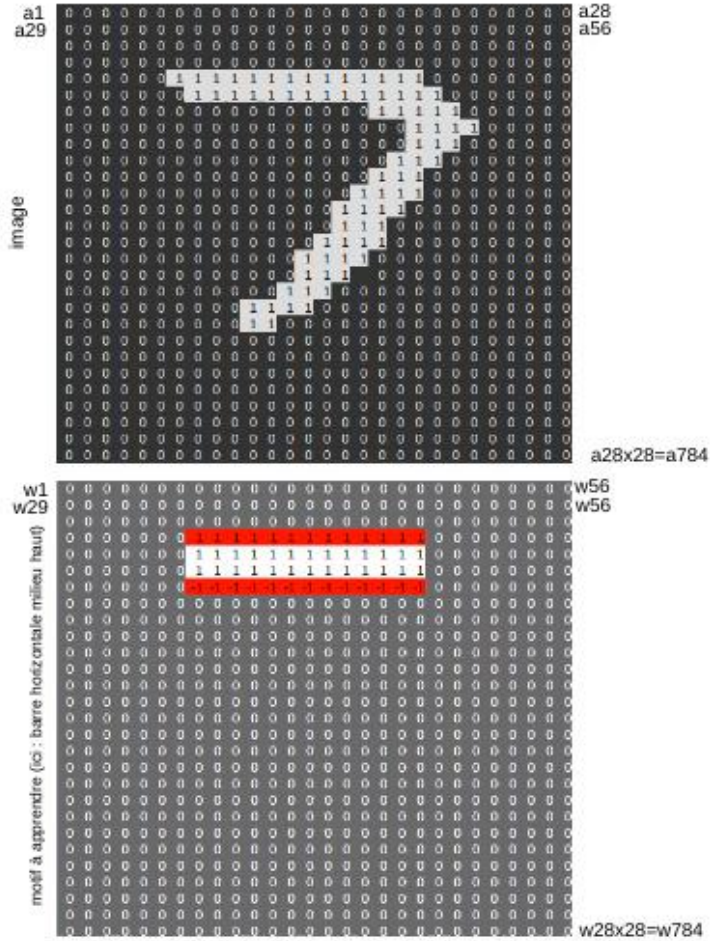
Donc on espère que les couches reconnaîtront et reconstitueront les motifs :

- A partir de l'image d'entrée/activations, chaque couche devrait reconnaître des motifs plus complexes en assemblant les activations/motifs de la couche précédente, jusqu'à activer les scores des chiffres en fonction des motifs activés ... :
- … mais peut-on faire ça avec un réseau de neurones artificiels (type MLP) ?



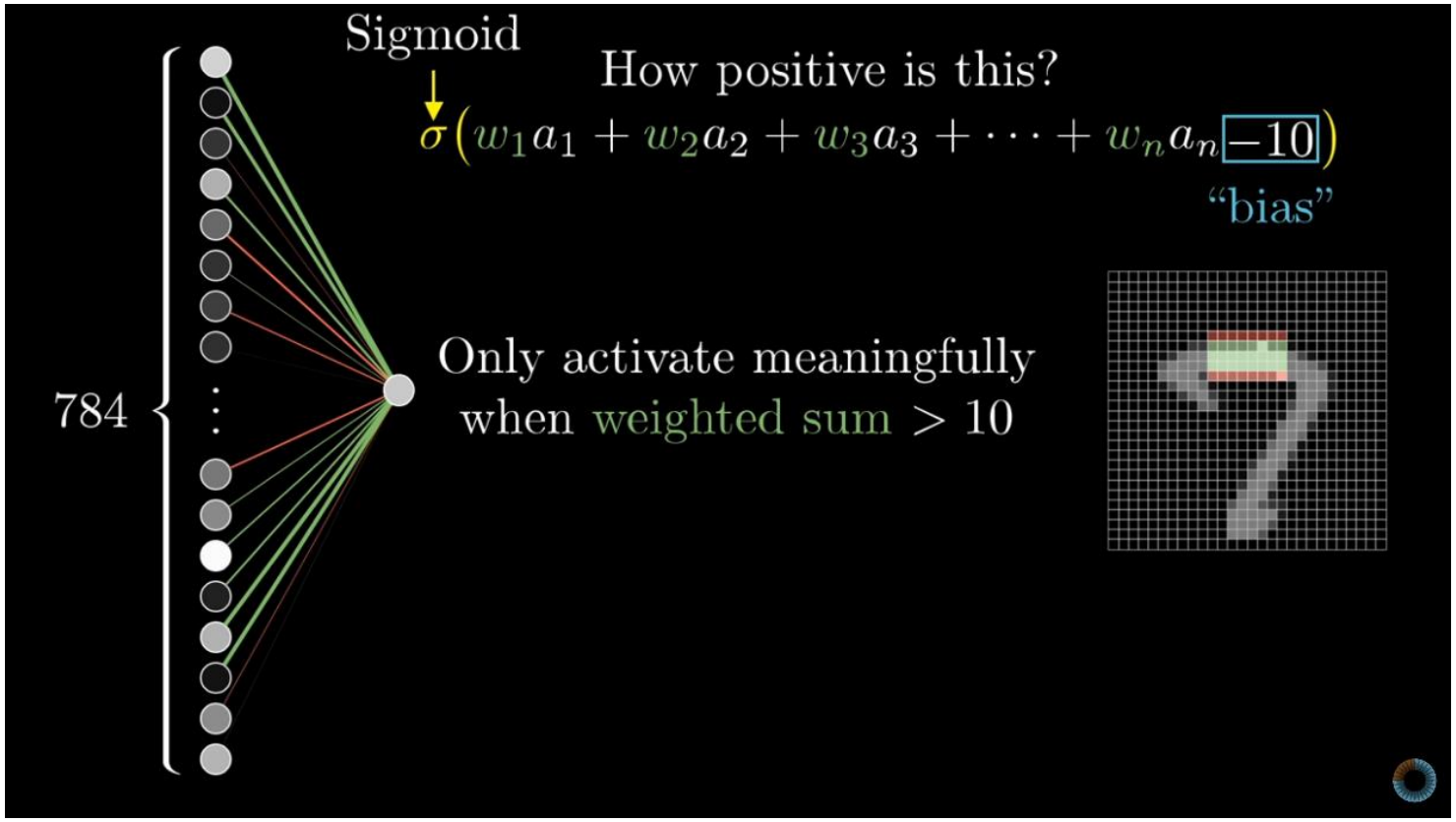
Détection des contours par multiplication de la carte des motifs et de l'image



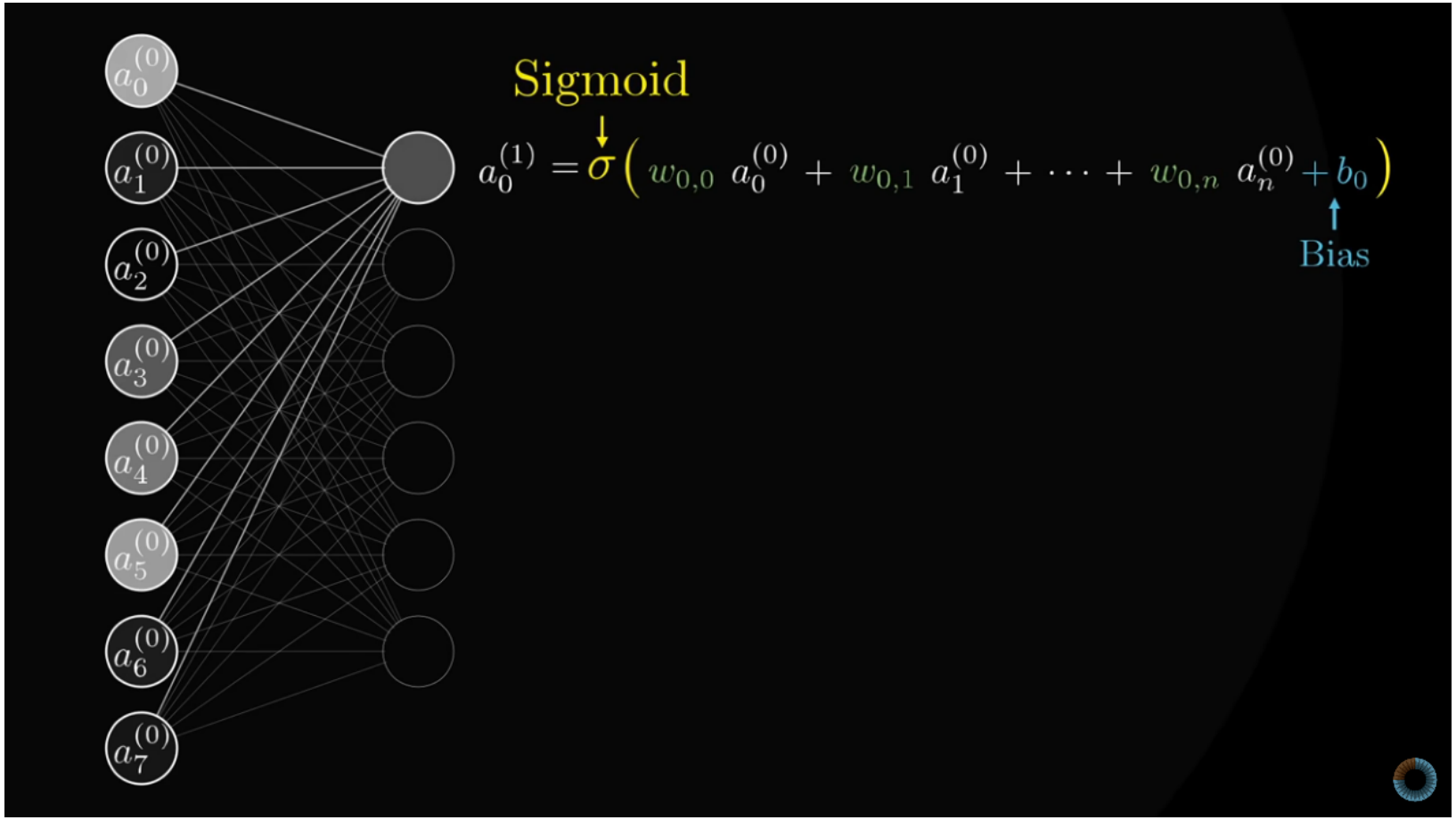


Et finalement, un “biais”/facteur additif pour chaque neurone

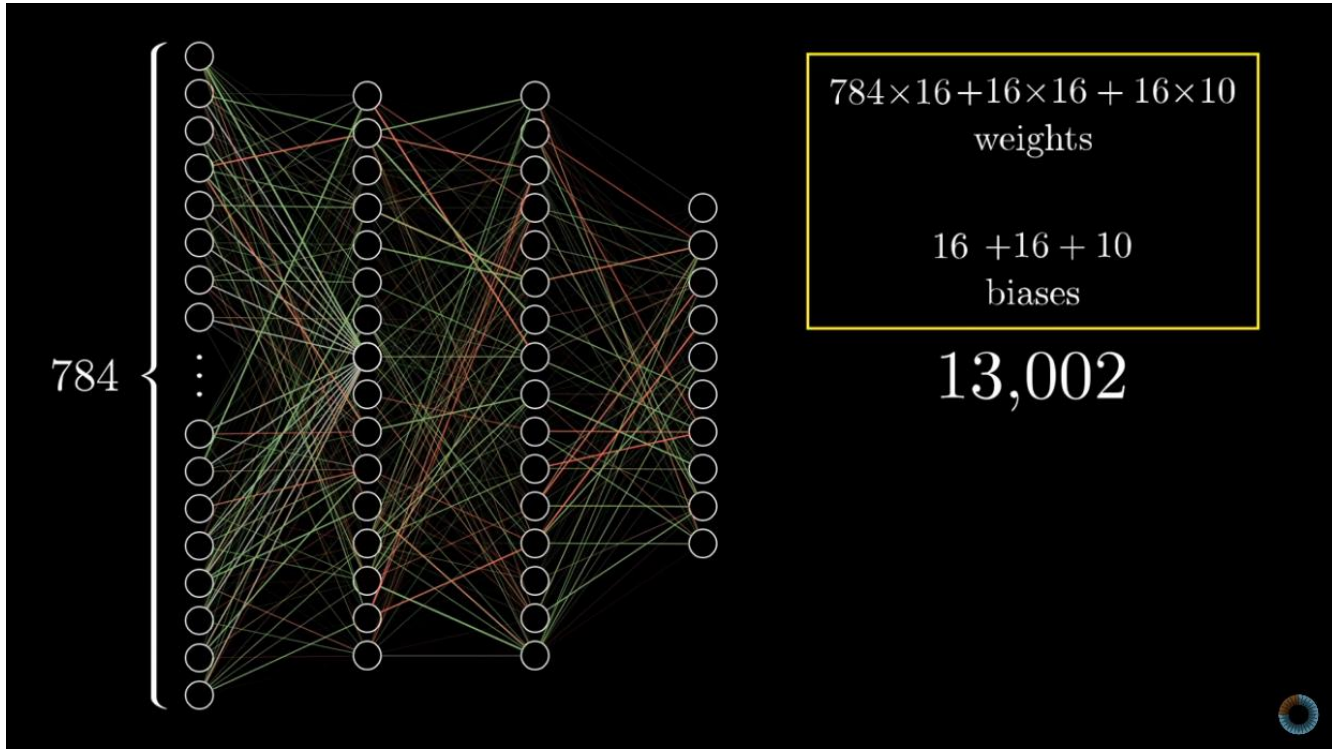
- Le ***biais*** ne permet l'activation que lorsque le bord/contour est suffisamment significatif :



Un neurone est une function paramétrée par les poids
Sa sortie est :



Et maintenant que fait-on ? On choisit les valeurs des poids/paramètres du réseau!

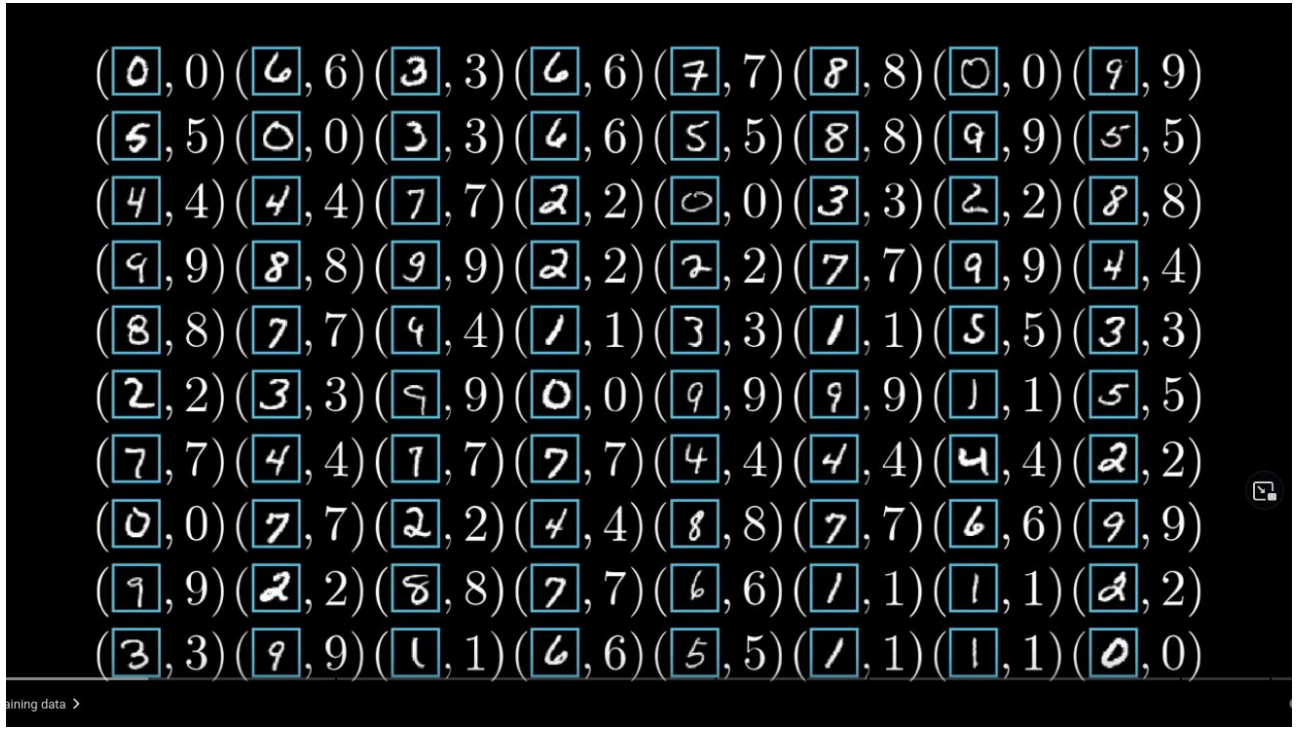


- ***Apprentissage*** **:= trouver les paramètres (w et b) de chaque neurone de l'architecture choisie**

© 3Blue1Brown

Comment entrainer : comparer les sorties de réseau à la vérité désirée

- Pour les images de chiffres à reconnaître : Un jeu de données est constitué des images et de leurs étiquettes (le chiffre correspondant), annotées par un humain.
- On partage le jeu de données en donnés pour l'entraînement (~80%), et le reste est gardé pour le test.
- Permet de tester le modèle sur des données jamais vues, et d'estimer ses capacités une fois déployédans le monde réel.

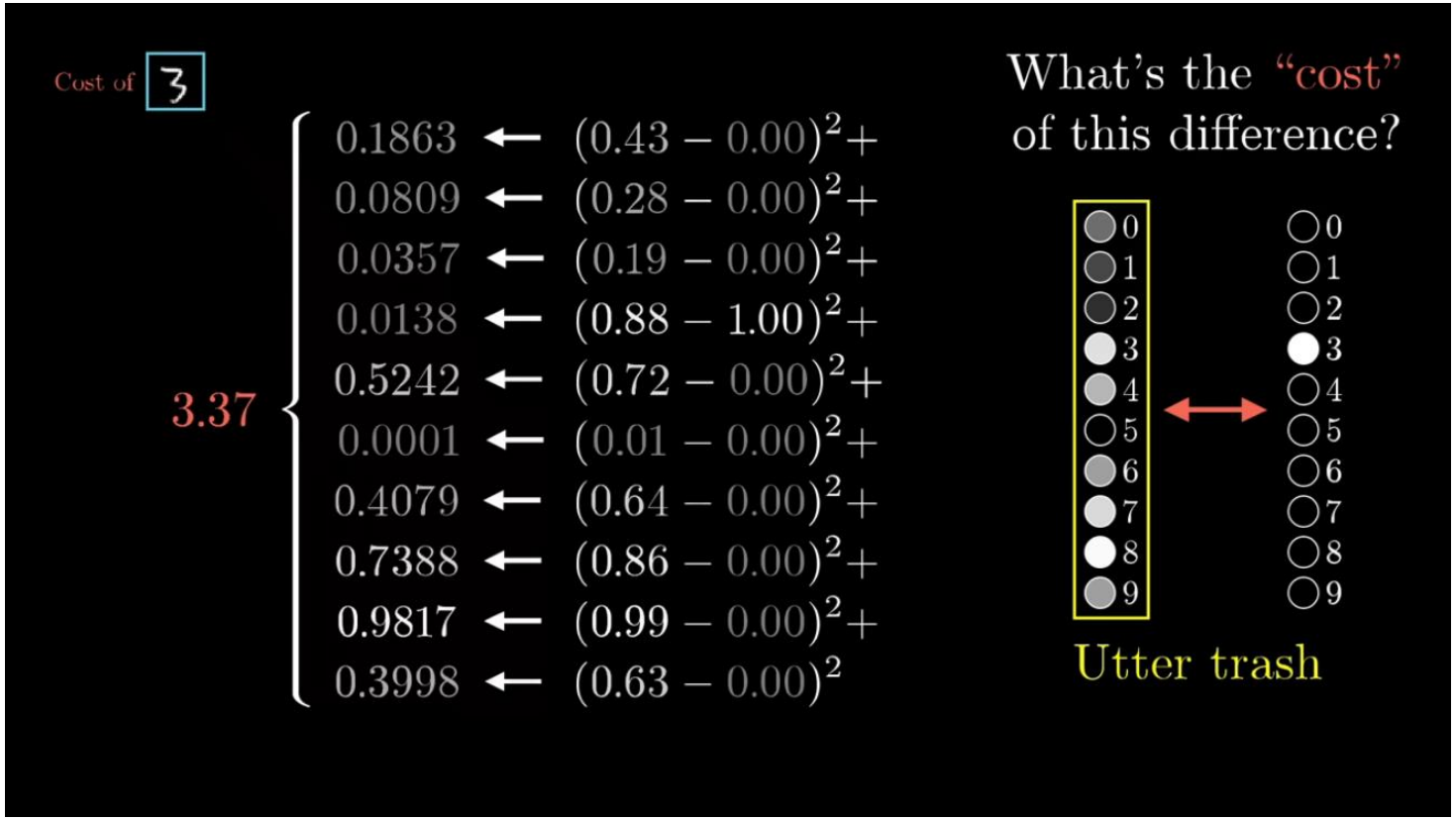


Comment entrainer : comparer les sorties de réseau à la vérité désirée

- Avec les labels de *verité terrain*

→ Calculer le coût/erreur totale pour chaque exemple d'entrainement

- Puis moyenner le coût pour tous les exemples d'entraînement



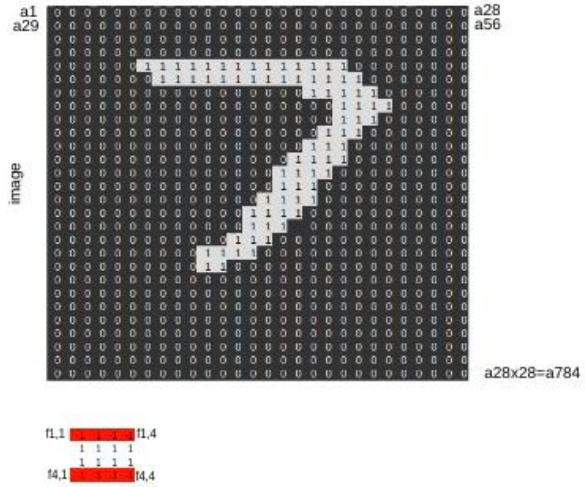
© 3Blue1Brown

Les bases à retenir pour l'approche par réseau de neurones artificiels

- Problème de classification
- Descente de gradient pour trouver les poids
- Les poids détectent des motifs et leurs combinaisons successives, et sont "appris".=optimisés
- Ces motifs appris permettent de décrire une image (une donnée) avec les niveaux de présence de chaque motif, pertinents pour la tâche
- Différent de la description explicite de règles en symbolique
- → avec les 1ères couches d'un réseau de neurones, **on apprend une représentation des données** !
- Jeu d'entraînement vs test : vérifier la généralité des motifs appris sur d'autres données que celle d'entraînement

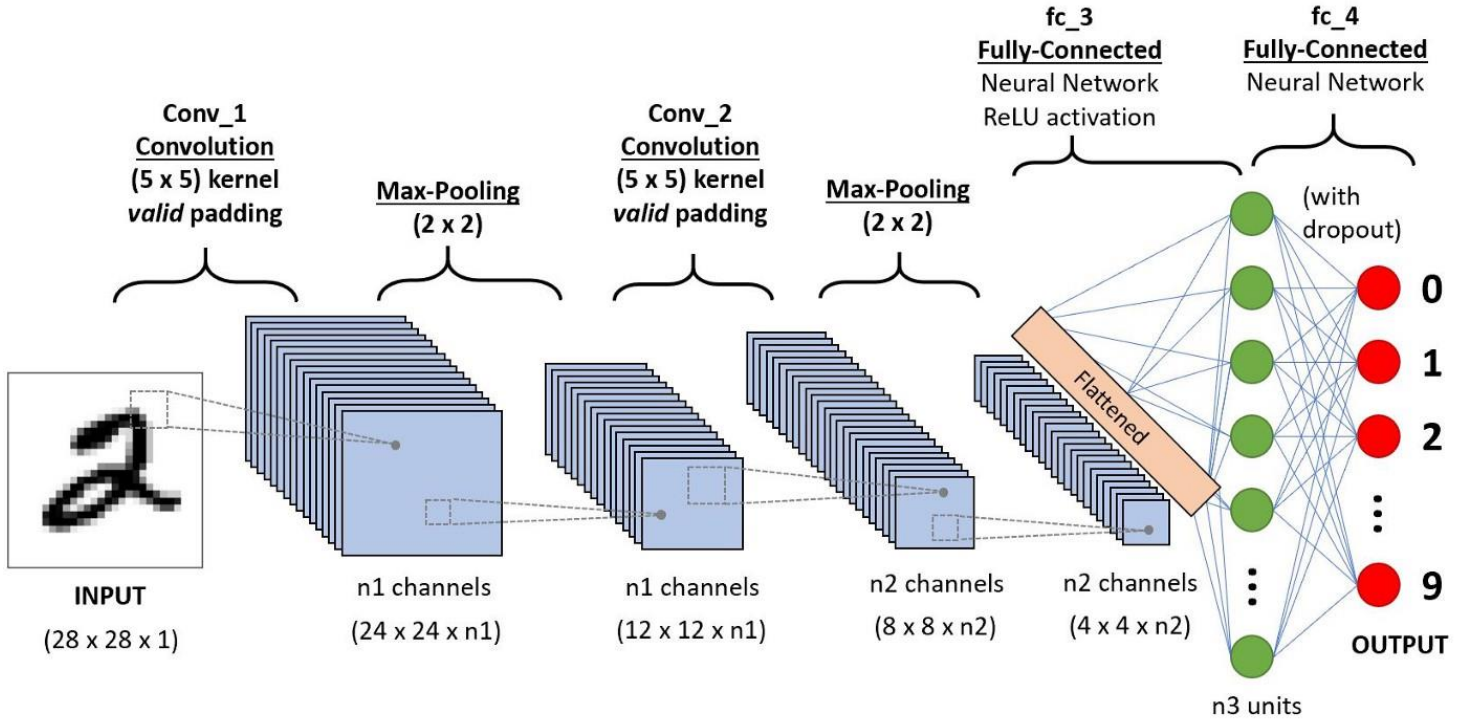
Un nouveau type de RNA : *Convolutional Neural Networks* - CNN

- Moins de paramètres pour mieux décrire les motifs visuels
  - Grâce notamment à l'invariance par translation



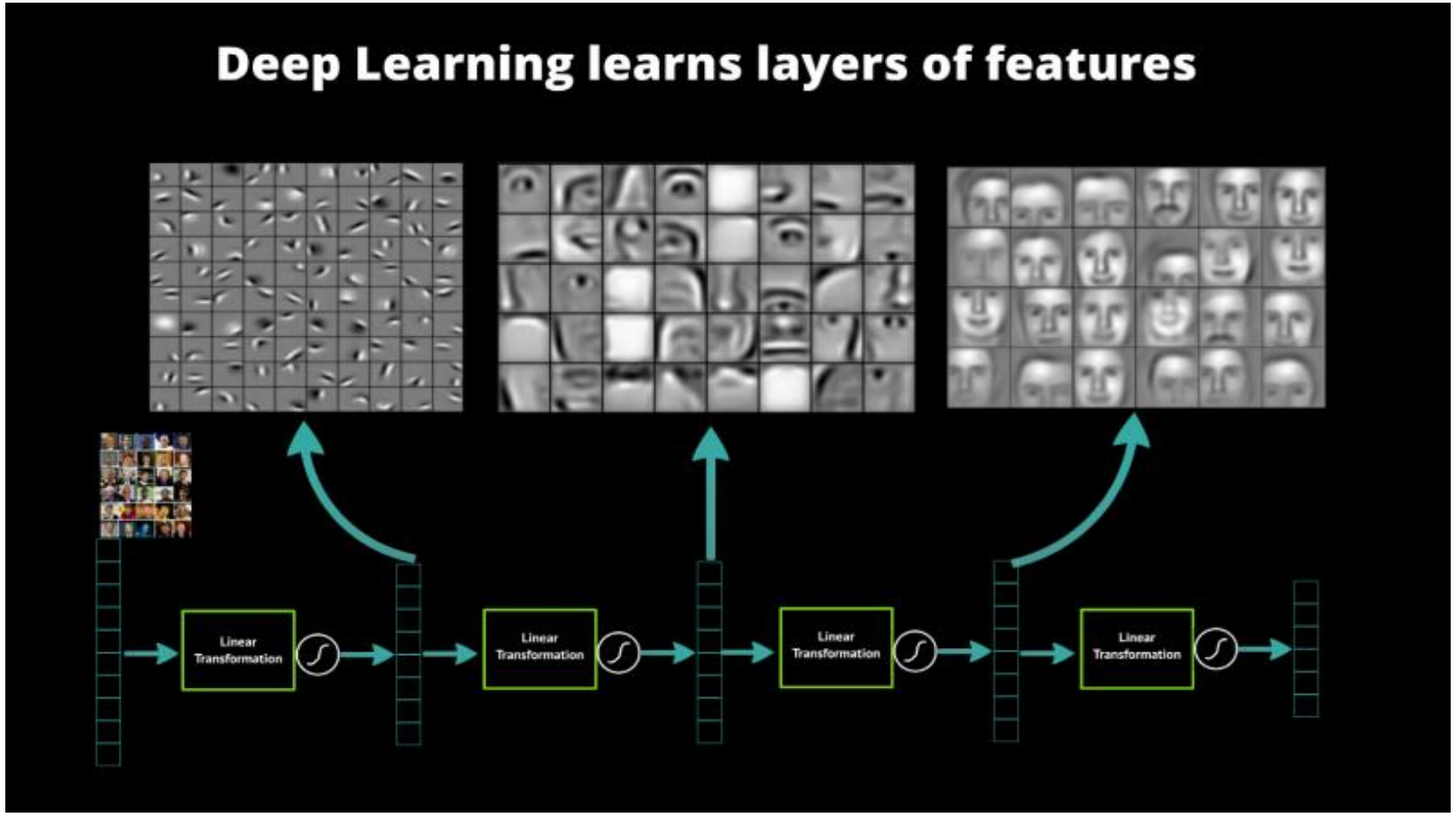
$$a'_{x,y} = \sum_{-L}^{L} \sum_{-L}^{L} f_{i,j}\, a(x+i, y+j)$$

Motifs à apprendre

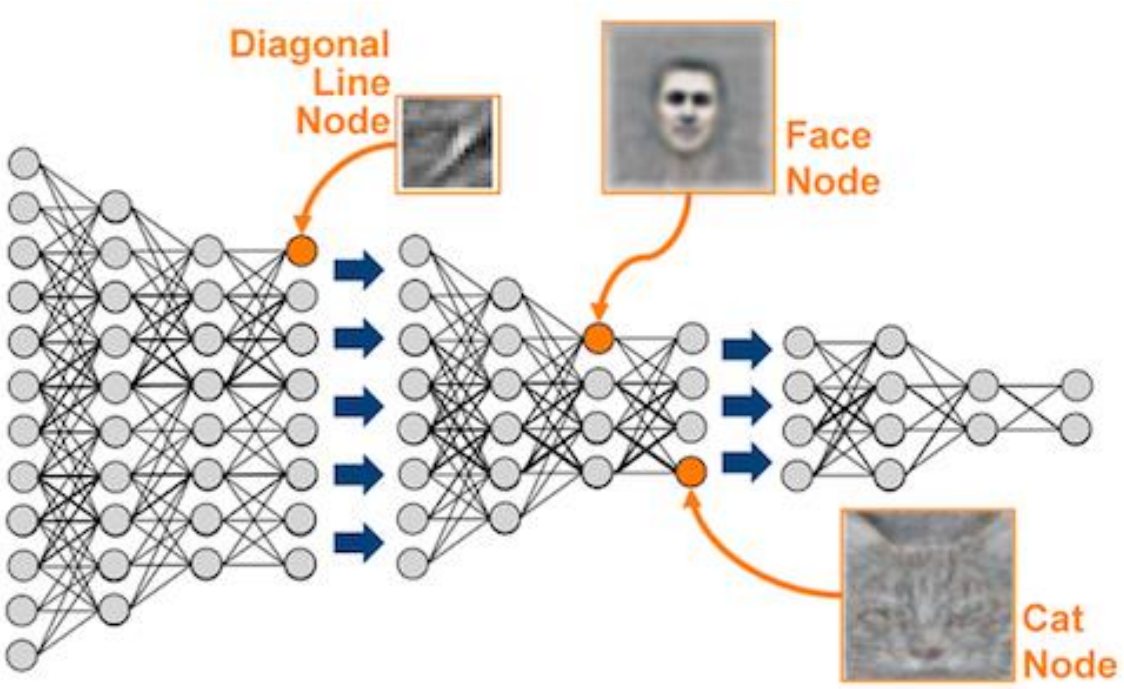


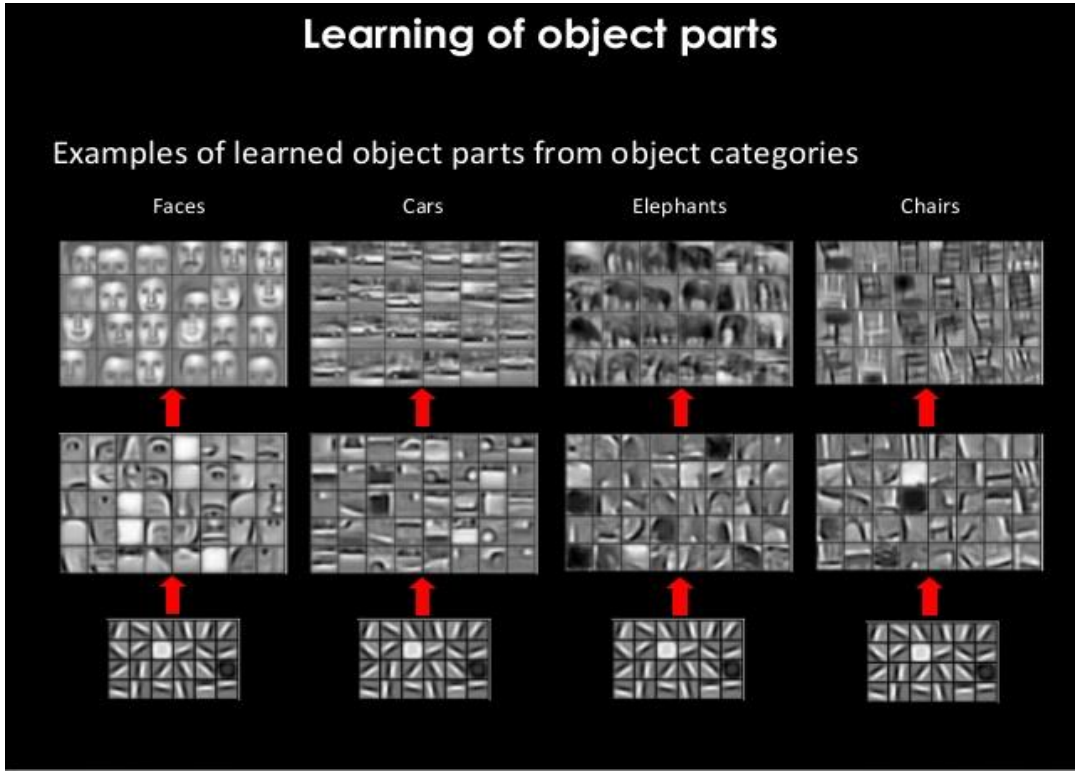
Taken from D2lAI

Les motifs appris (filtres du Réseau de neurones convolutionels)



Les motifs appris (filtres du Réseau de neurones convolutionels)





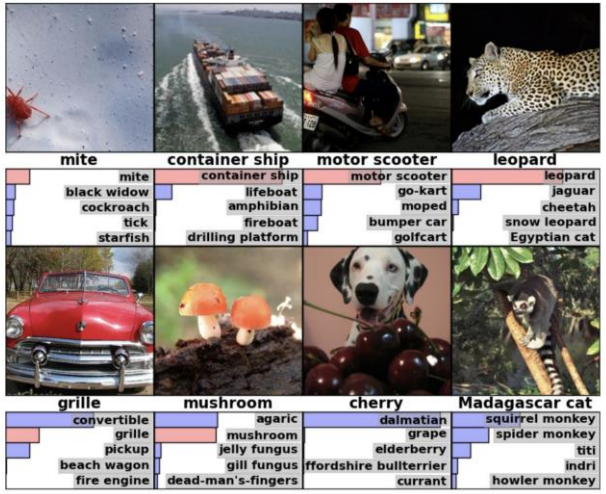
2012: Gains historique en performance

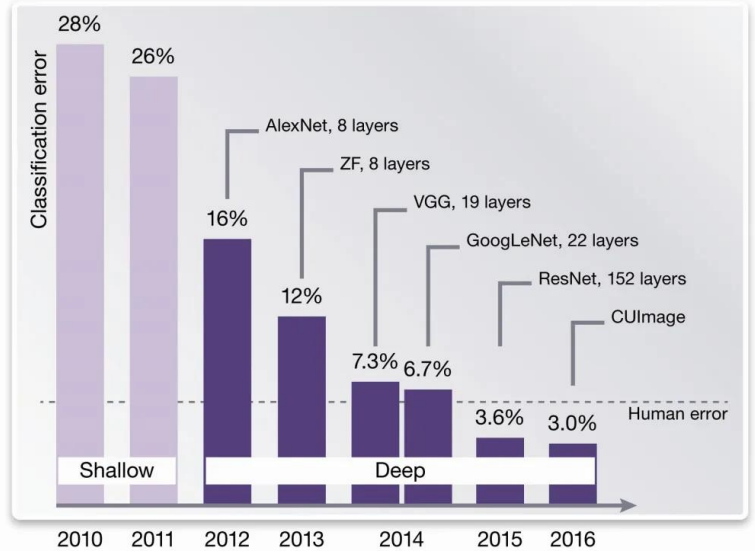
- Results on ImageNet

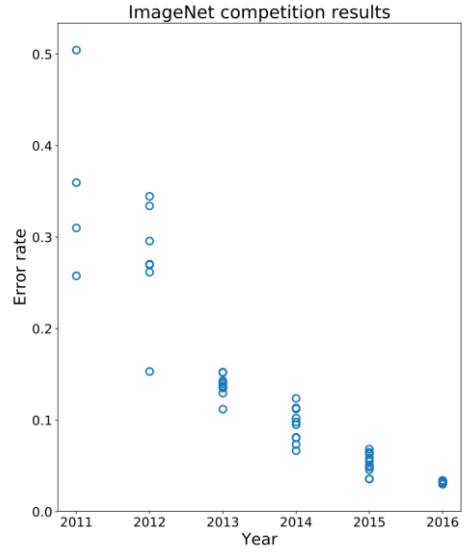
ImageNet Challenge

IMAGENET

- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.







Principe fondamental

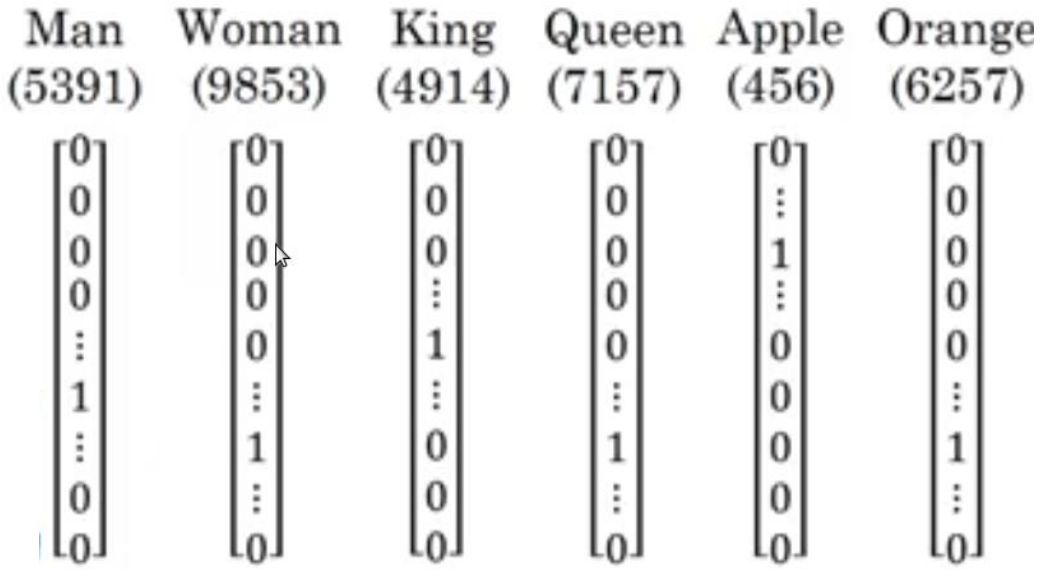
- Mise en contexte : avant on pré-déterminait les motifs qui nous paraissaient importants, et on décrivait les données ainsi, pour seulement les séparer en classes avec de l'apprentissage
- Les réseaux de neurones convolutionnels (CNN, combiné à la descente de gradient, à la puissance de calcul et à la quantité de données) permettent à présent d'apprendre les représentations pertinentes pour classer les données dans les catégories souhaitées
- → **Apprentissage de représentation**

Plan de la formation

1. Apprentissage de représentation pour la reconnaissance de formes
   - Perceptron multi-couche (MLP)
   - Réseaux de neurones convolutionnels (CNN)
     - pour l'apprentissage de motifs pertinents dans les données
2. Apprentissage de représentation de mots
   - Représentations apprises par similarités de contextes
   - Représentations apprises par modèlisation du langage
     - Modèles Transformers et pré-entraînement
3. Modèles fondation : un changement de paradigme
   - Emergence de capacité imprévues
   - En langage, visio, audio... Et plus
   - Nouvelles méthodes pour adapter les modèles à des tâches spécifiques
4. Limites et enjeux
   - Environnement social et politique du design et du déploiement des systèmes de ML

Word representation

- Vocabulary V = [a, aaron, …, zulu, <UNK>]
- 1-hot representation of words



o_{5391}

I want a glass of orange _____.

I want a glass of apple _____.

Example of use of word embedding in an NLP task: Named entity recognition

Sally	Johnson	is	an	orange	farmer.
1	1	0	0	0	0
Robert	Lin	is	an	apple	cultivator.
?=1	?=1	?=0	?=0	?=0	?=0

Trained knowing that *Sally* and *Jonhson* are proper names
→ Based on similarity *orange-apple* and *farmer-cultivator*, infer that Robert and Lin are proper names

Featurized representation: word embedding

- The model is trained to complete the last word:
  - it sees I want a glass of orange _juice_ . and is given the response "juice"
- At test time, when it sees new sentences, it must predict the last word of:
  - I want a glass of apple ___?_____.
  - To correctly guess, it must rely on some proximity between apple and orange

→ We must find a way to numerically encode this proximity of words:

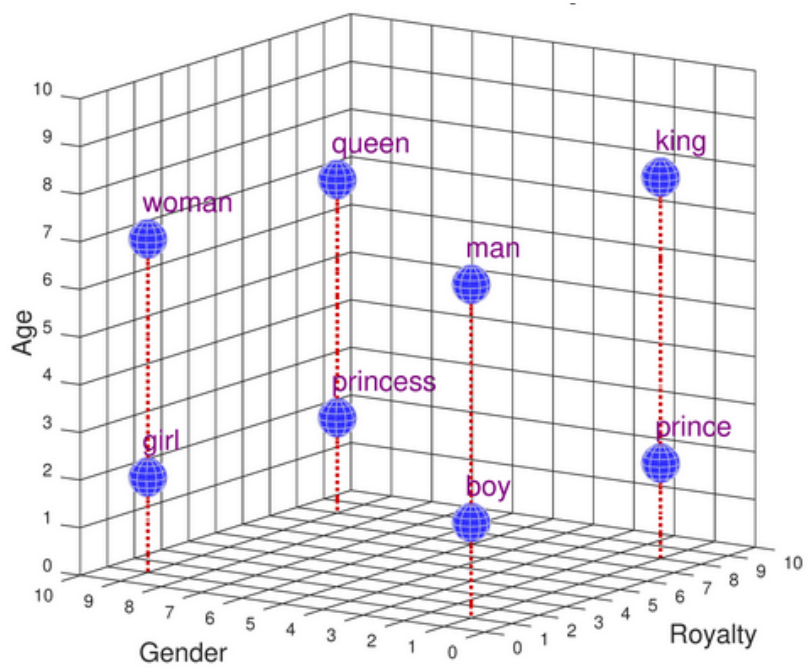
  - Rather than to represent every word with its index in the dictionary, represent it with a score on common characteristics:

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0.97	0.00	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0.02
Age	0.65	0.65	0.70	0.69	0.03	-0.02
Food	0.09	0.01	0.02	0.01	0.95	0.97

e_{5391}

Visualizing word embeddings

- ***Word embedding***: représentation numérique (par un tableau de nombre) d'un mot, habituellement dans un espace de caractéristiques sémantiques permettant d'encoder (partiellement) son sens



Word Coordinates

	Gender	Age	Royalty
man	[1,	7,	1]
woman	[9,	7,	1]
boy	[1,	2,	1]
girl	[9,	2,	1]
king	[1,	8,	8]
queen	[9,	7,	8]
prince	[1,	2,	8]
princess	[9,	2,	8]

© https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html

Questions as formulas!

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	O (
Gender	-1	1	-0.95	0.97	0.00	0.0
Royal	0.01	0.02	0.93	0.95	-0.01	0.0
Age	0.65	0.65	0.70	0.69	0.03	-0.0
Food	0.09	0.01	0.02	0.01	0.95	0.9

The difference/distance between man and king is almost the same as between woman and queen.
So we consider this distance encodes the semantic difference between these concepts, that only differ in the royal dimension, given the dimensions we have set by hand for this first word embedding.

Let us query the language model by looking for:

"King is to Man" as "? is to Woman"

$$e_{king} - e_{man} \approx \begin{matrix} 0.05 \\ 0.92 \\ 0.05 \\ -0.07 \end{matrix}$$

$$e_{queen} - e_{woman} \approx \begin{matrix} -0.03 \\ 0.93 \\ 0.04 \\ 0 \end{matrix}$$

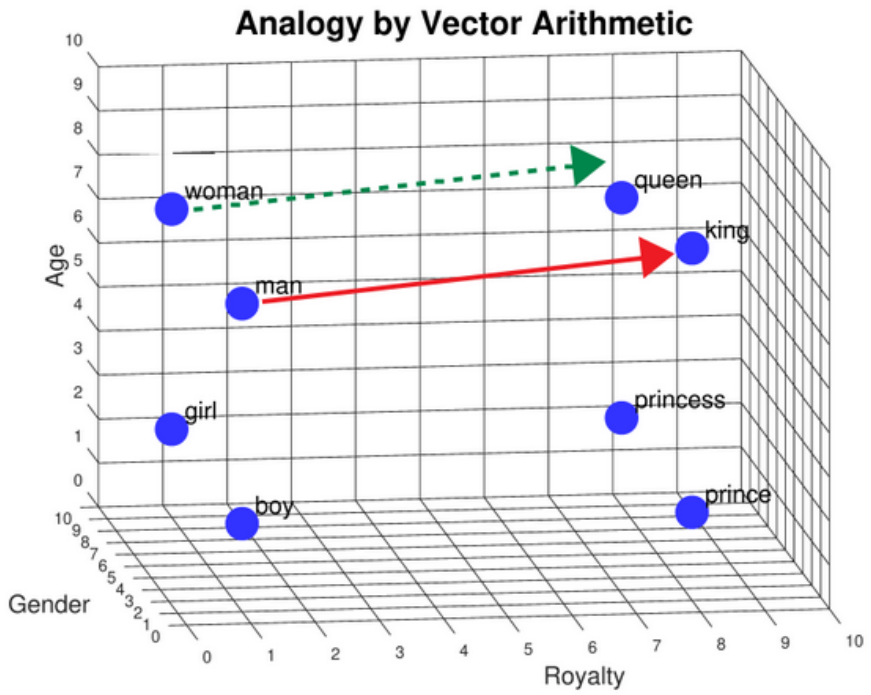
Mikolov et al., 2013, Linguistic regularities in continuous space word representations.

Analogies using word vectors

- To answer “King is to Man” as “? is to Woman”, we look for the word whose embedding is closest to:

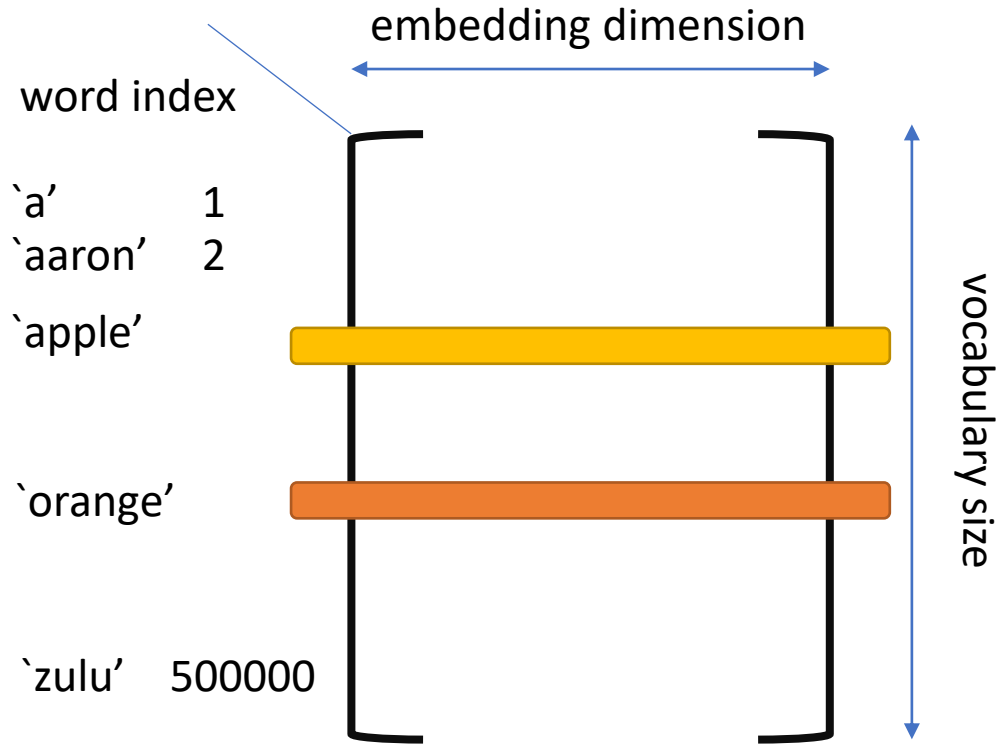
$$e_{woman} + (e_{king} - e_{man})$$

So we can build a new embedding vector by adding to e_{woman} the displacement between e_{man} and e_{king}.
Then we check what are the closest words in the neighborhood of the resulting vector.



Formalizing word embedding: the Embedding matrix E

- Table that associates a word index from the dictionary to its embedding vector, **which will be learned**:



Now: How to (automatically) learn the embedding matrix?

- Remember: Given two words w_i and w_o, we want them to be as close as possible in the projected space if they are ***similar***.
- **Deliberate choice :** *similar* in terms of the context in which they appear
  - Distributional semantics: A word's meaning is given by the words that frequently appear close-by
    - "You shall know a word by the company it keeps" (J. R. Firth 1957)

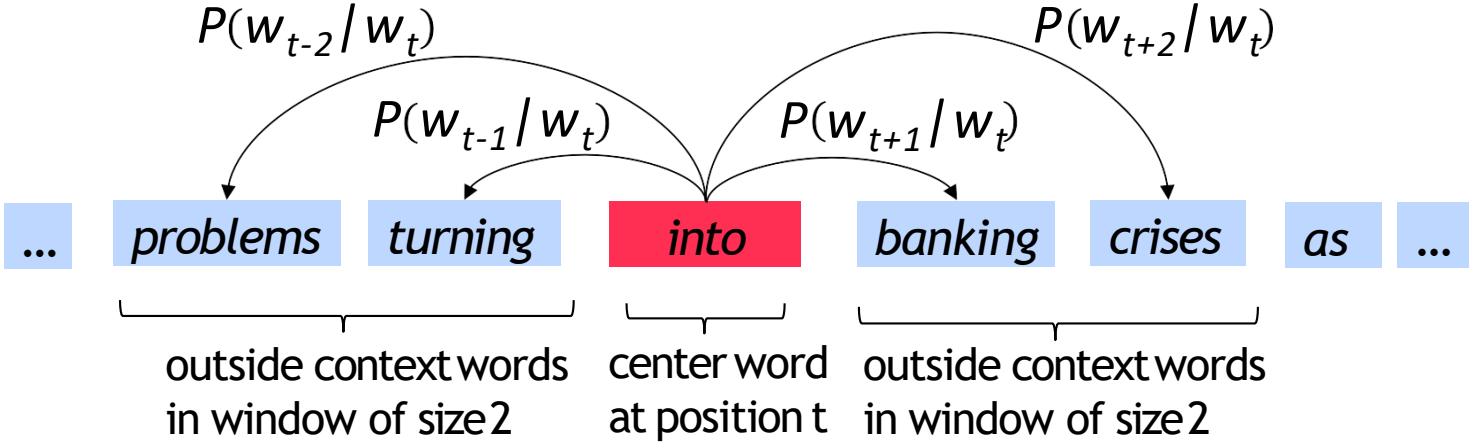
…government debt problems turning into	***banking***	*crises as happened in 2009…*
…saying that Europe needs unified	***banking***	*regulation to replace the hodgepodge…*
…India has just given its	***banking***	*system a shot in the arm…*

These context words will represent ***banking***

Adapterd from Christopher Manning

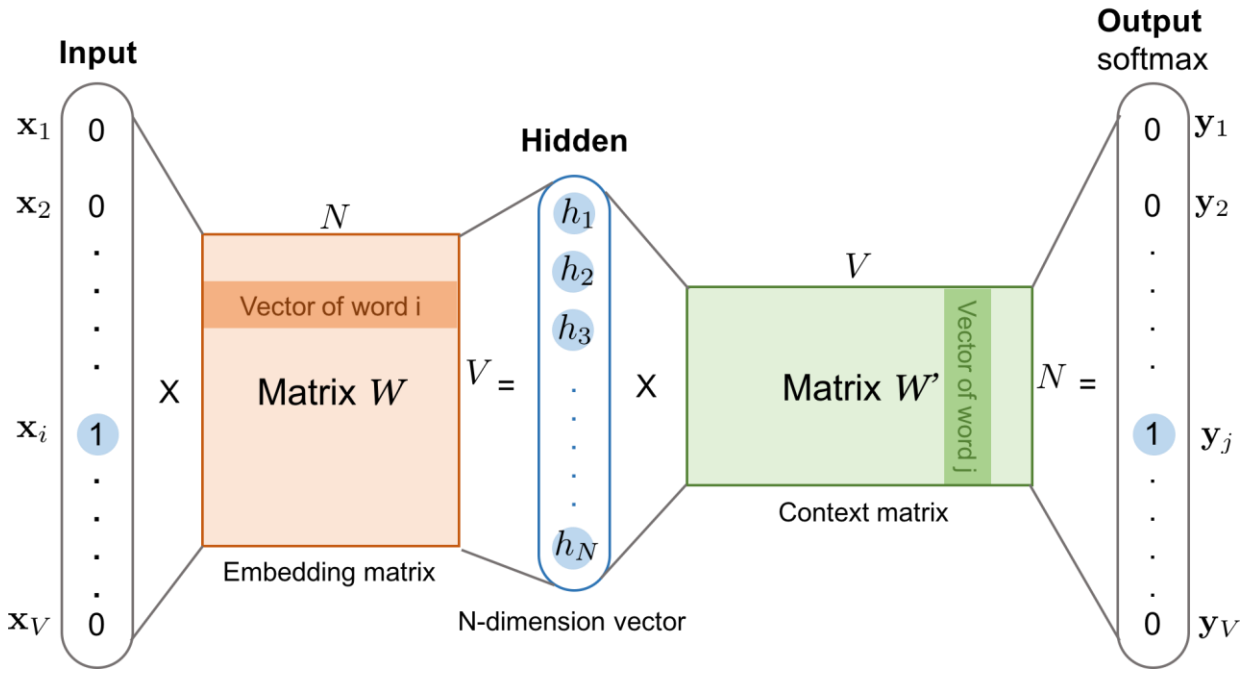
Word2Vec principle

- Idea:
  - We have a large corpus of text
  - Every word in a fixed vocabulary is represented by a vector
  - Go through each position t in the text, which has a center word c and context ("outside") words o
  - Use the similarity of the word vectors for c and o to calculate the probability of o given c (or vice versa)
  - Keep adjusting the word vectors to maximize this probability



Mikolov et al., 2013, Efficient estimation of word representations in vector space.

Context-Based: Skip-Gram Model of Word2Vec



Taken from Lilian Weng

- We want to minimize the objective function:

$$J(\theta) = -\frac{1}{T}\log L(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{-m\le j\le m\\ j\neq 0}} \log P\left(w_{t+j} \mid w_t; \theta\right)$$

- We compute $P(w_{t+j}|w_t; \theta)$ as:

$$p(w_O|w_I) = \frac{\exp({v'_{w_O}}^{\top} v_{w_I})}{\sum_{i=1}^{V}\exp({v'_{w_i}}^{\top} v_{w_I})}$$

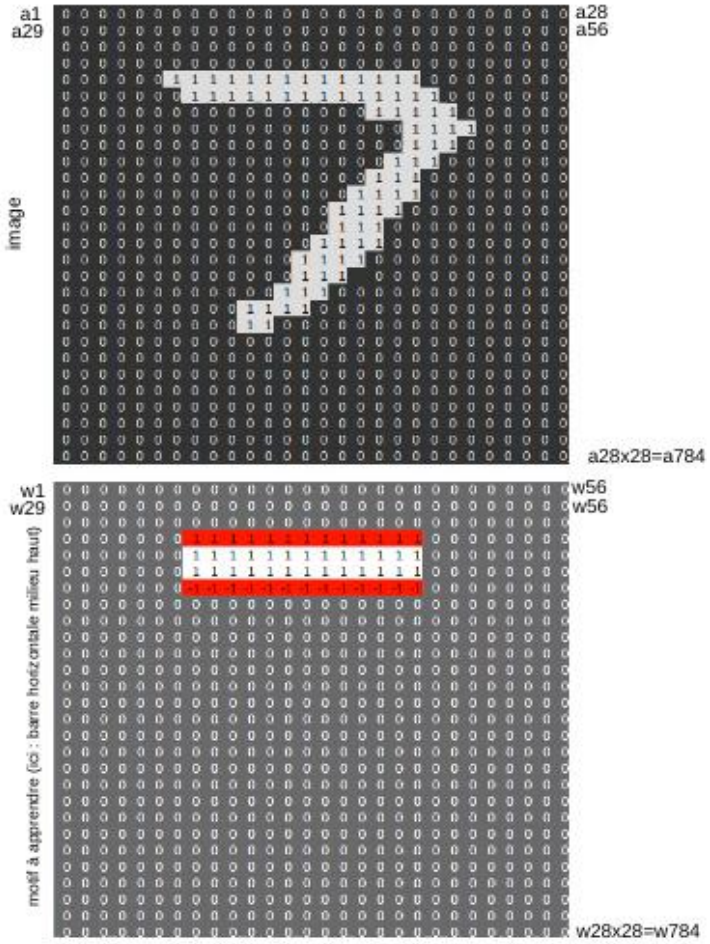
Mikolov et al., 2013, Efficient estimation of word representations in vector space.

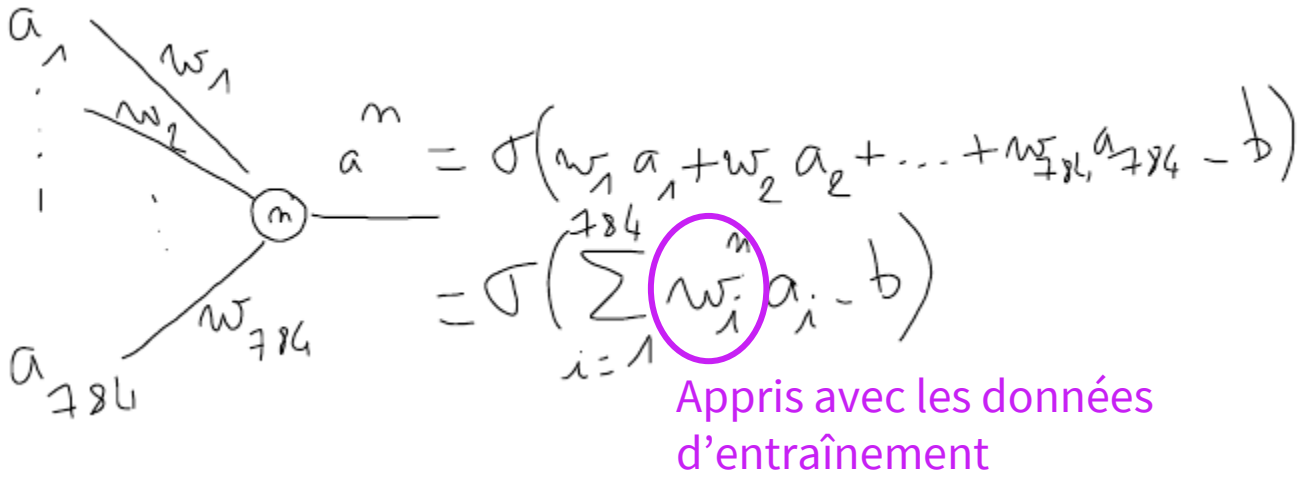
Démonstration avec des embeddings de mots Word2Vec

- http://nlp.polytechnique.fr/word2vec
- Essayez notamment :
  - France – paris + berlin
  - Paris – France + allemagne
  - Paris – France + brésil

→ Que remarquez-vous ?

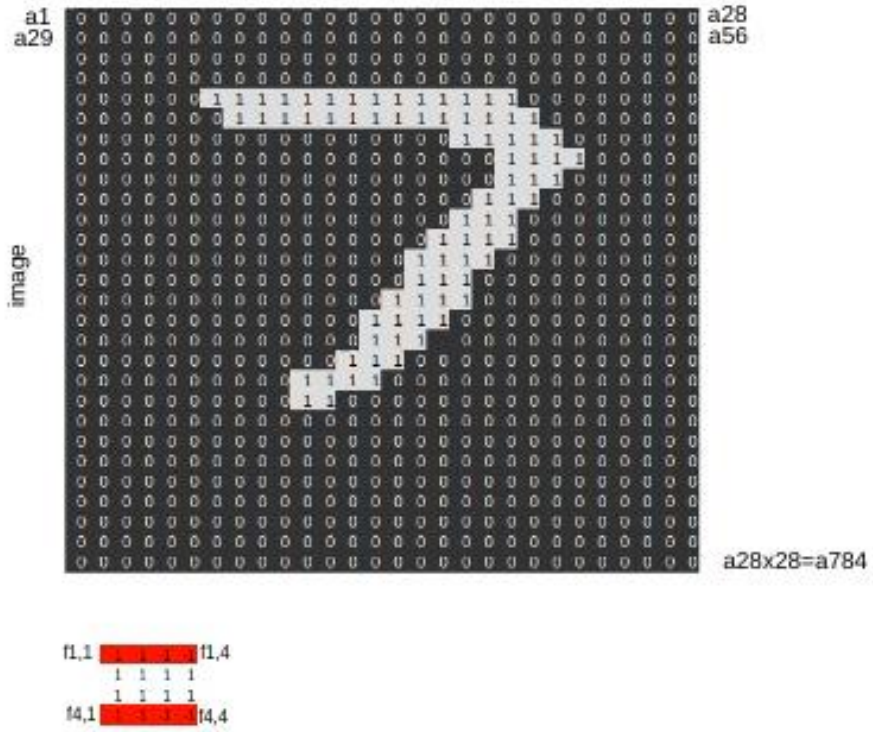
Récap : l'apprentissage de représentation jusque là : MLP





- Nouvelle représentation des données : à chaque couche, l'image est décrite par les valeurs des neurones, c'est-à-dire la présence ou l'absence des motifs appris

Récap : l'apprentissage de représentation jusque là : CNN



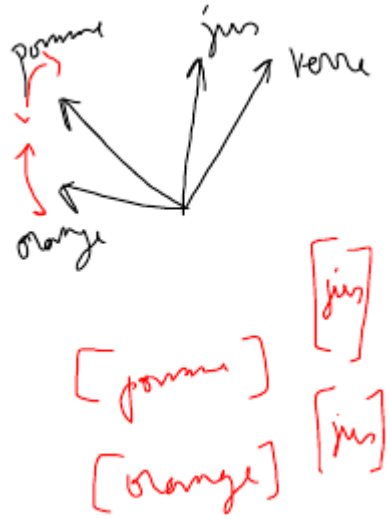
$$a^{m}_{x,y} = \sum_{-L}^{L} \sum_{-L}^{L} f_{i,j}\, a(x+i, y+j)$$

Appris avec les données d'entraînement

- Réseau de neurones convolutionnels :
  - invariance par translation → plus efficace

Récap : l'apprentissage de représentation jusque là : Word2Vec

- Aligner (maximiser le produit scalaire) les vecteurs des mots apparaissant souvent dans le même contexte :
  - verre de jus de pomme
  - verre de jus d'orange



Récap : l'apprentissage de représentation à venir : les réseaux Transformer

- On flexibilise les motifs recherchés encore plus : ils peuvent dépendre des mots ou pixels voisins !

(Multi-head self attention) permet des motifs/noyaux définis sur de **grande fenêtres** et **spécifiques aux données**

Mais lourd en calcul en test aussi, pas que en entraînement comme avant !

$$\mathbf{e}_i^{1,h} = \sum_j \text{softmax}\left(\frac{\mathbf{q}_{i,h} \cdot \mathbf{k}_{j,h}}{\sqrt{d}}\right)\mathbf{v}_{j,h}$$

$$\frac{e^{\mathbf{q}_{i,h}\cdot\mathbf{k}_{j,h}}}{\sum_l e^{\mathbf{q}_{l,h}\cdot\mathbf{k}_{l,h}}}$$

$$\mathbf{q}_{i,h} = \mathbf{W}^{q,h}\mathbf{e}_i^0 \;, \quad \mathbf{k}_{j,h} = \mathbf{W}^{k,h}\mathbf{e}_j^0 \;, \quad \mathbf{v}_{j,h} = \mathbf{W}^{v,h}\mathbf{e}_j^0$$

$$\mathbf{e}_i^1 = MLP\left(\text{Linear}\left(\mathbf{e}_i^{1,1}, \ldots, \mathbf{e}_i^{1,H}\right)\right)$$

- Le mot i est représenté par une recombinaison de (diverses représentations de) ses mots voisins, dont les facteurs varient eux-mêmes en fonction des mots voisins (pas comme avant).

A. Vaswani et al.. Attention is all you need. NeurIPS 2017.

Key idea : Attention

- Generate successive word representation as recombination of the representations of the other words:

Turn weights into probabilities

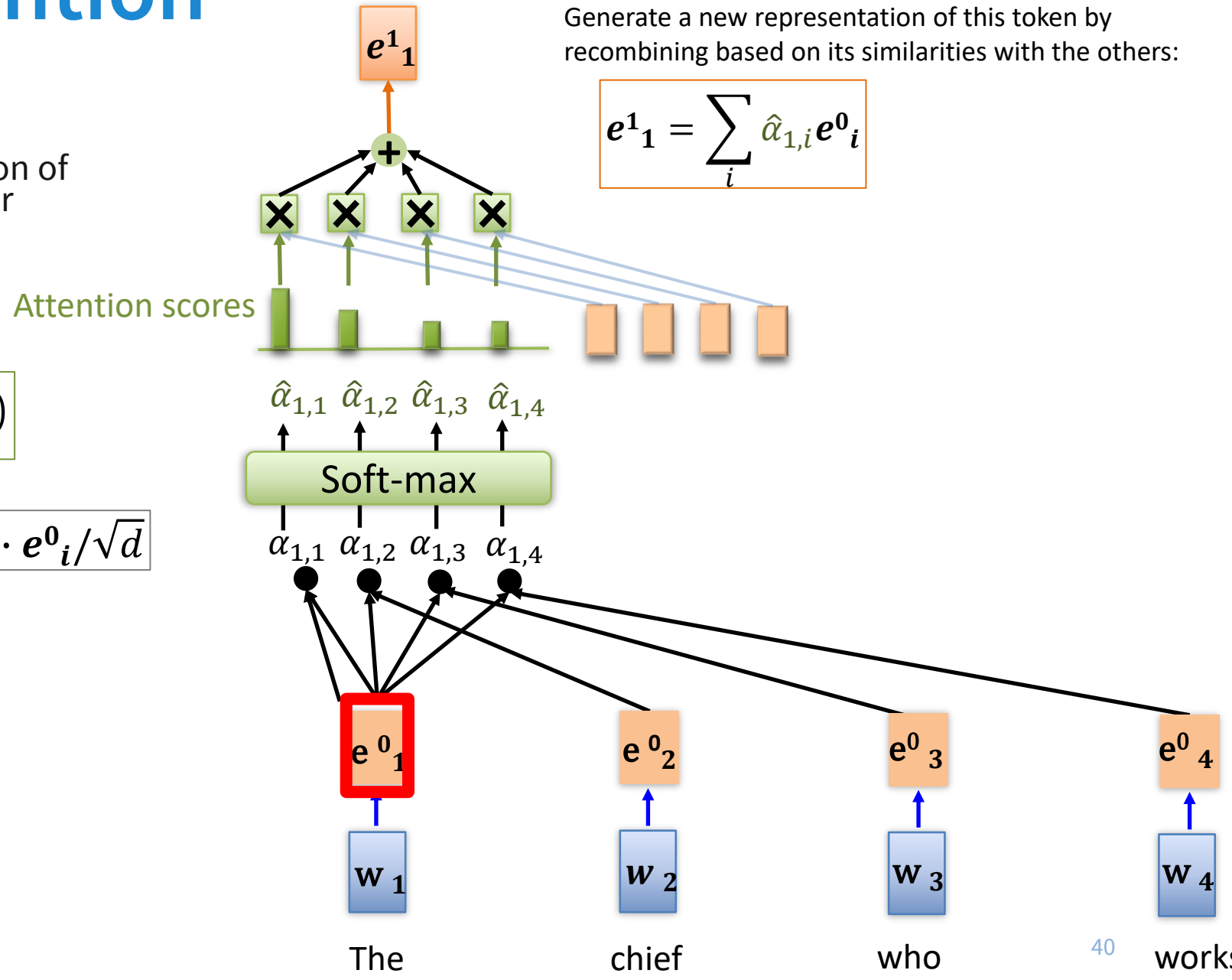
$$\hat{\alpha}_{1,i} = exp(\alpha_{1,i}) / \sum_j exp(\alpha_{1,j})$$

Compute token similarity (scaled dot-product attention):

$$\alpha_{1,i} = \boldsymbol{e^0}_1 \cdot \boldsymbol{e^0}_i / \sqrt{d}$$

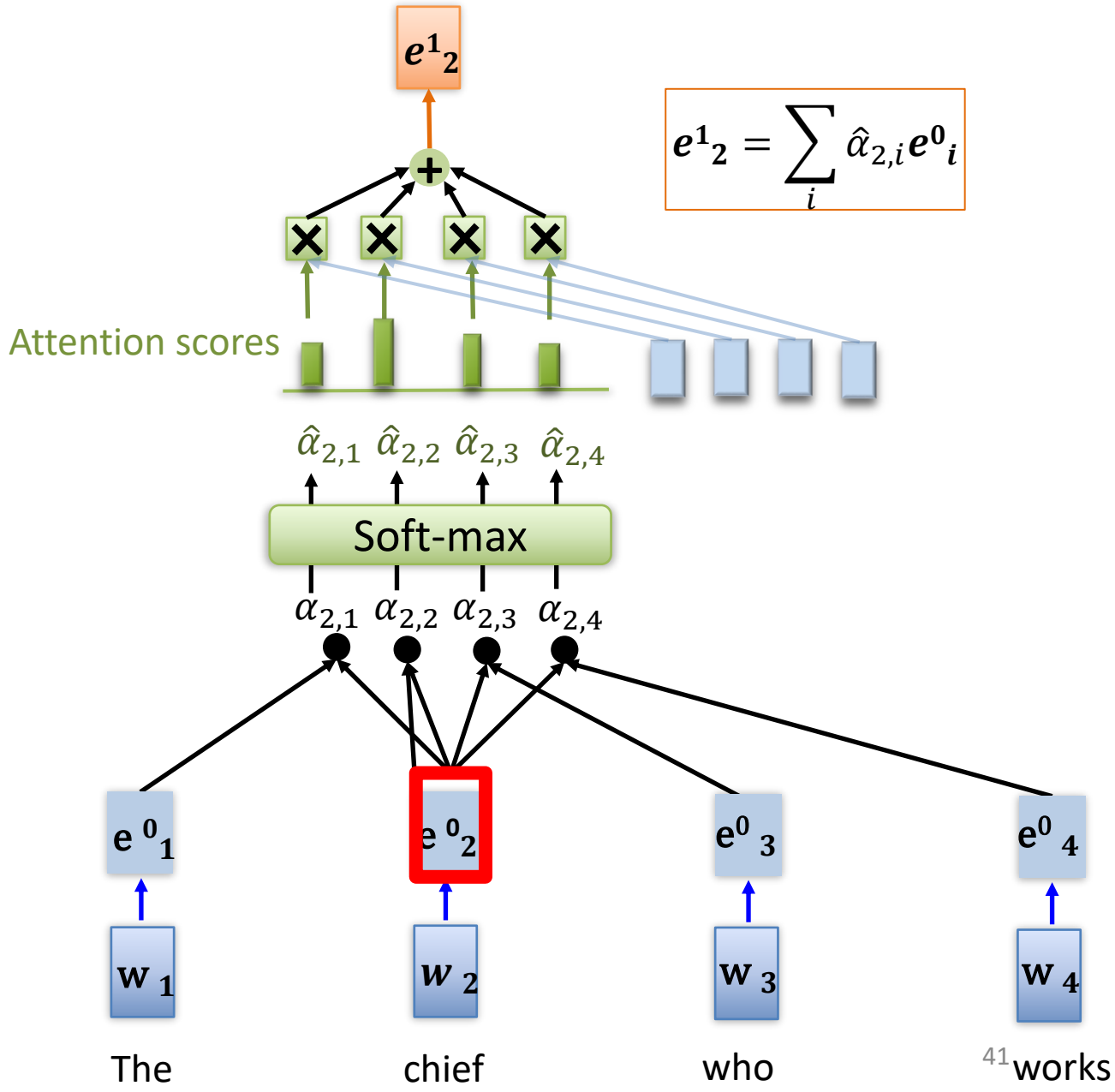
Generate a new representation of this token by recombining based on its similarities with the others:

$$\boldsymbol{e^1}_1 = \sum_i \hat{\alpha}_{1,i} \boldsymbol{e^0}_i$$



Basic idea: Self-attention

Generate a new representation of this token by recombining based on its similarities with the others:



Turn weights into probabilities

$$\hat{\alpha}_{2,i} = exp(\alpha_{2,i}) / \sum_j exp(\alpha_{2,j})$$

Compute token similarity (scaled dot-product attention):

$$\alpha_{2,i} = e^0{}_2 \cdot e^0{}_i / \sqrt{d}$$

Making Self-attention more flexible: Different representations for query, key, value roles

$\boldsymbol{q}$: query (to match others)

$$\boldsymbol{q_i} = \boldsymbol{W^q a_i}$$

$\boldsymbol{k}$: key (to be matched)

$$\boldsymbol{k_i} = \boldsymbol{W^k a_i}$$

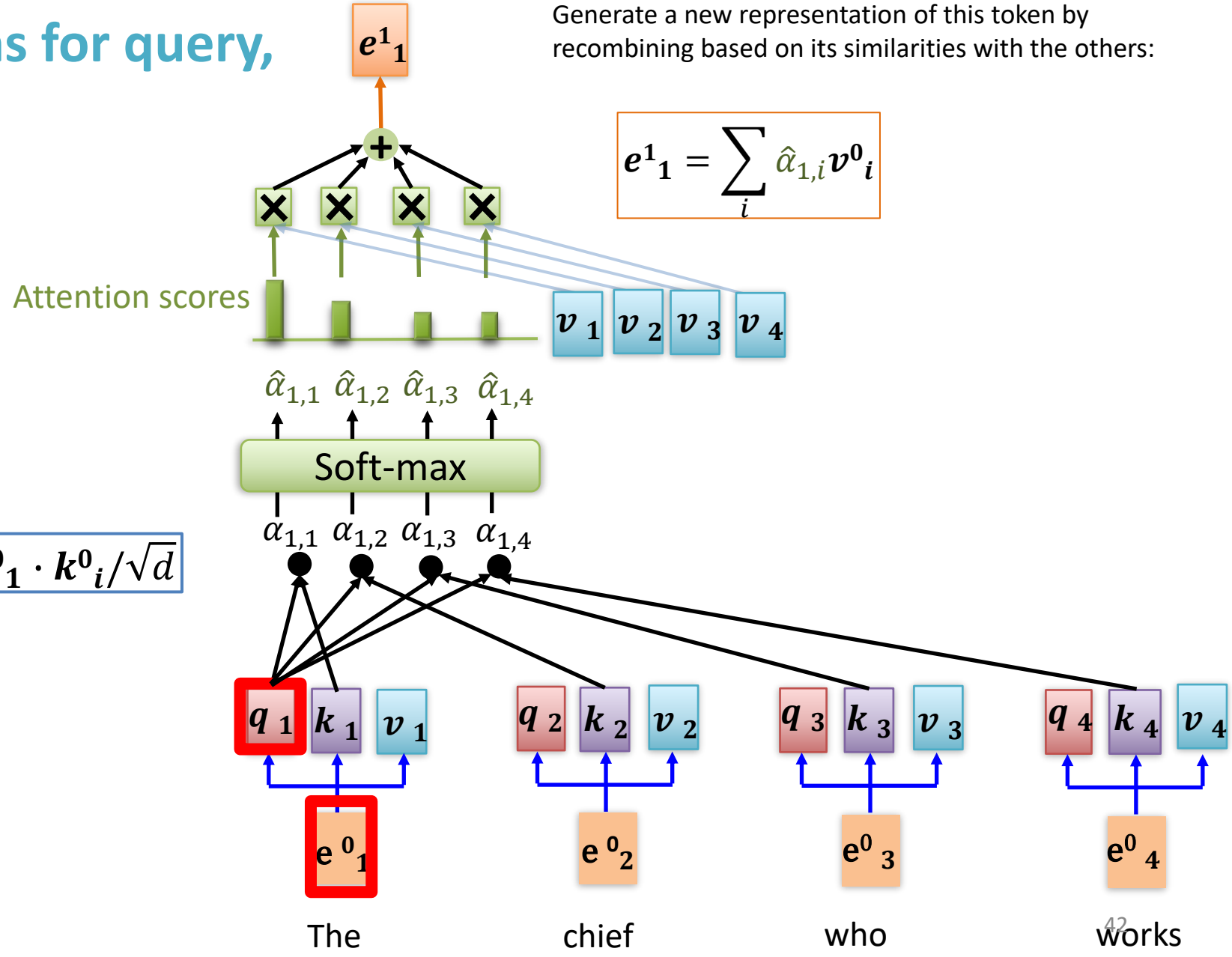
$\boldsymbol{v}$: information to be extracted

$$\boldsymbol{v_i} = \boldsymbol{W^v a_i}$$

$$\alpha_{1,i} = \boldsymbol{q^0}_1 \cdot \boldsymbol{k^0}_i / \sqrt{d}$$

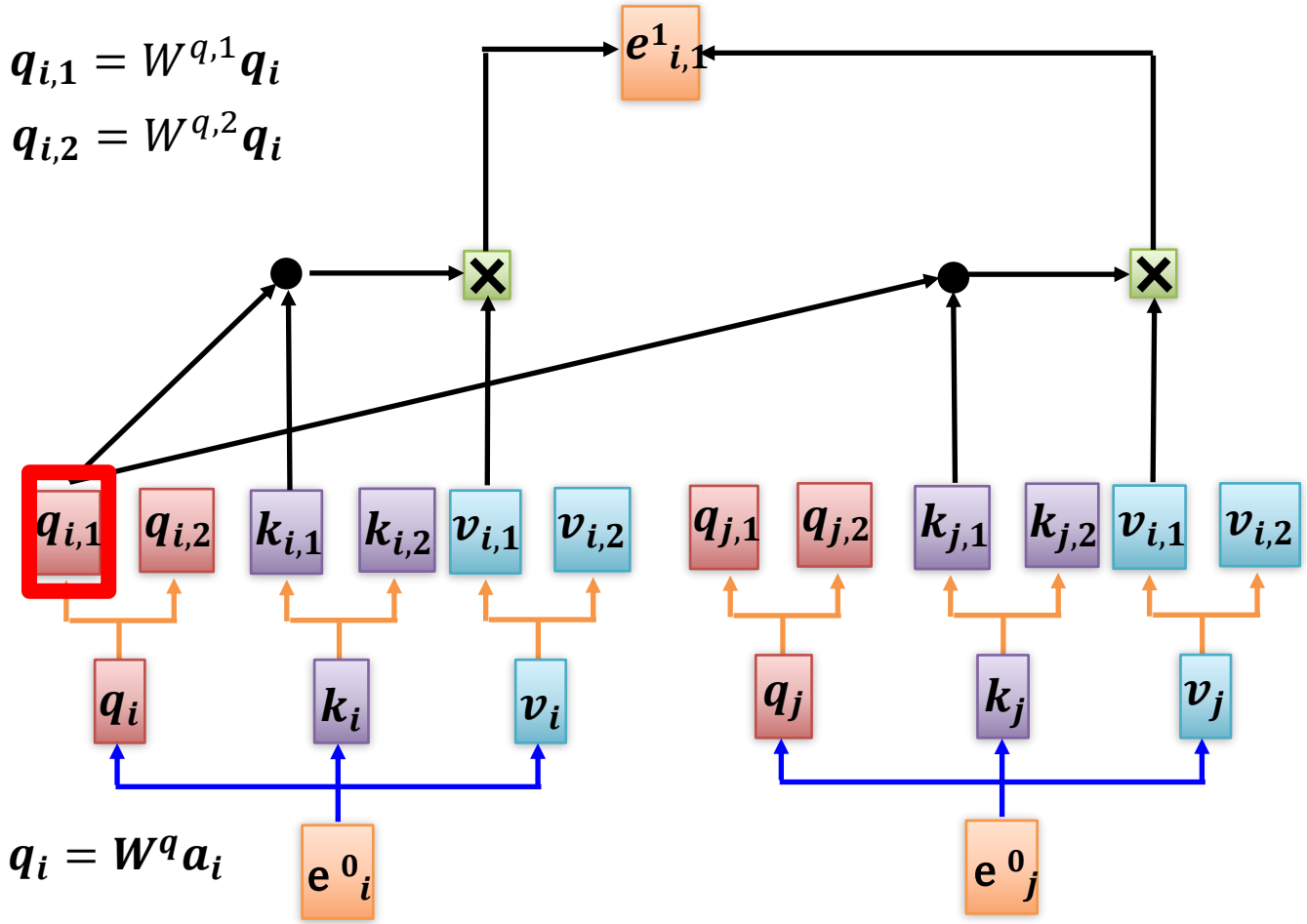
Generate a new representation of this token by recombining based on its similarities with the others:

$$\boldsymbol{e^1}_1 = \sum_i \hat{\alpha}_{1,i} \boldsymbol{v^0}_i$$



Making Self-attention more flexible: Multi-headed attention

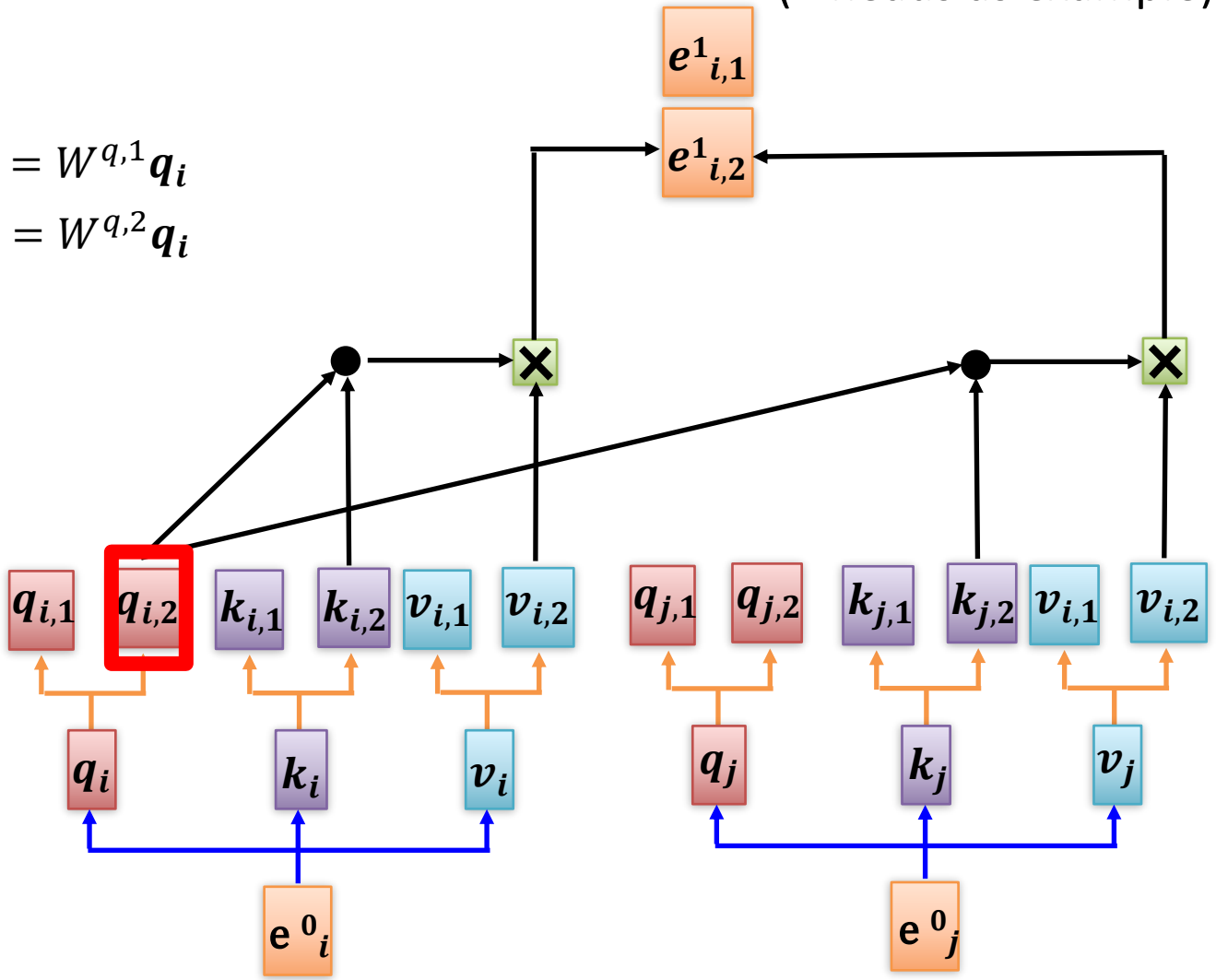
(2 heads as example)



Making Self-attention more flexible: Multi-headed attention

(2 heads as example)

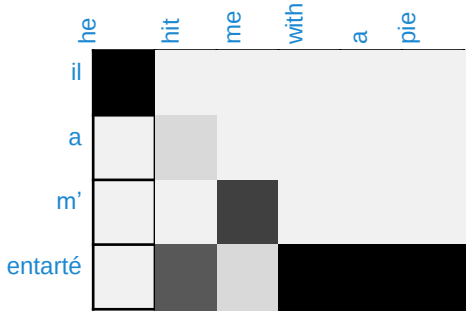
$$q_{i,1} = W^{q,1} q_i$$
$$q_{i,2} = W^{q,2} q_i$$



The principle of multi-head attention in transformers is similar to the one of multi-kernels in one convolution layer of a CNN (i.e. multi-feature maps)

Attention is great

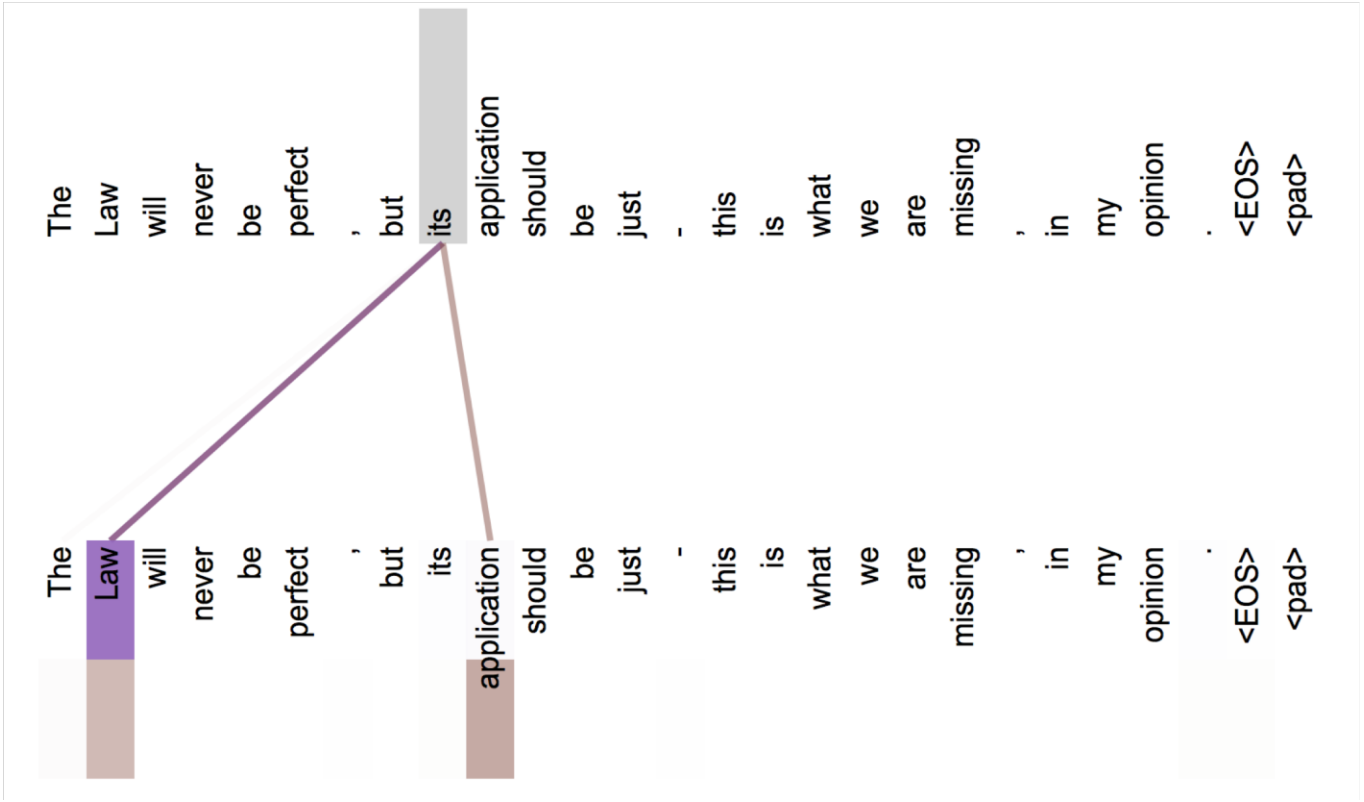
- Attention significantly improves NMT performance
  - It's very useful to allow decoder to focus on certain parts of the source
- Attention solves the bottleneck problem
  - Attention allows decoder to look directly at source; bypass bottleneck
- Attention helps with vanishing gradient problem
  - Provides shortcut to faraway states
- Attention provides some interpretability
  - By inspecting attention distribution, we can see what the decoder was focusing on
  - We get (soft) alignment for free!
  - This is cool because we never explicitly trained an alignment system
  - The network just learned alignment by itself



Abigail See

Attention visualization

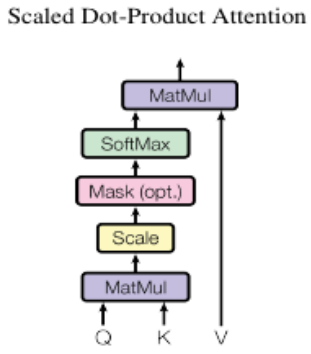
- In 5th layer:
  - Isolated attentions from just the word 'its' for attention heads 5 and 6.

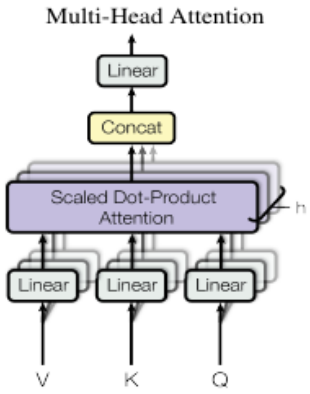


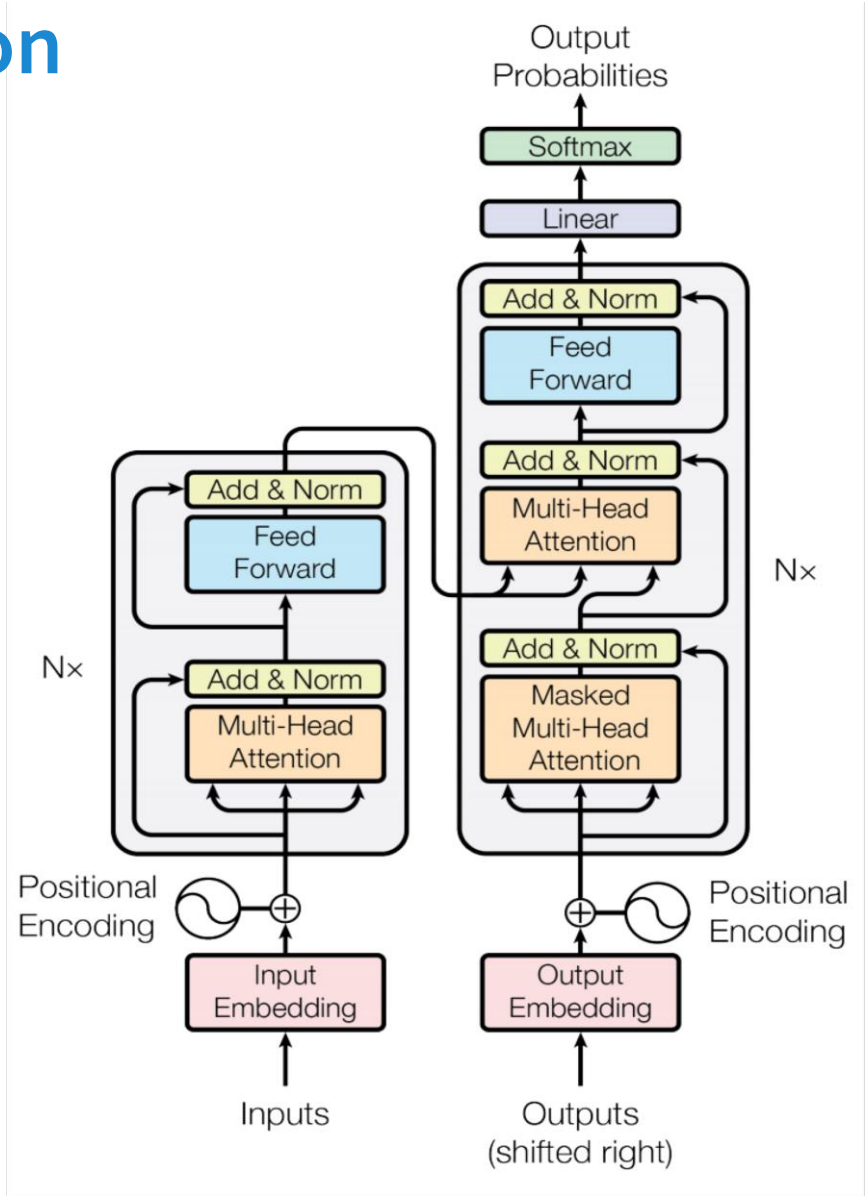
Transformer for machine translation

Attention is all you need. 2017. Aswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin
https://arxiv.org/pdf/1706.03762.pdf

- Non-recurrent sequence-to-sequence encoder-decoder model
- Task: machine translation with parallel corpus
- Predict each translated word
- Final cost/error function is standard cross-entropy error on top of a softmax classifier
  - This and related figures from paper ⇑

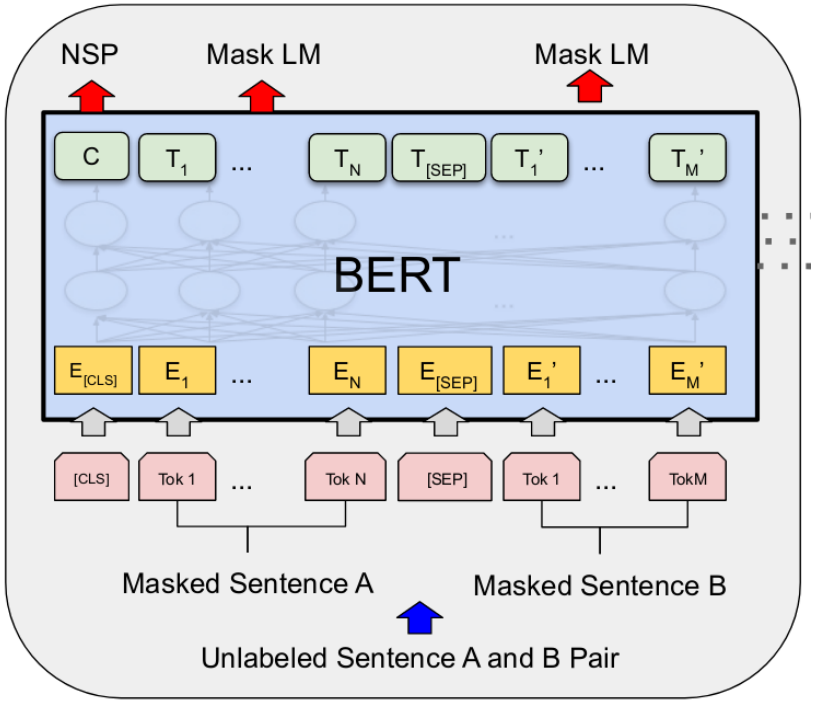






Transformer: obtenir une représentation de chaque token en fonction du contexte

- Exemple avec BERT [1]

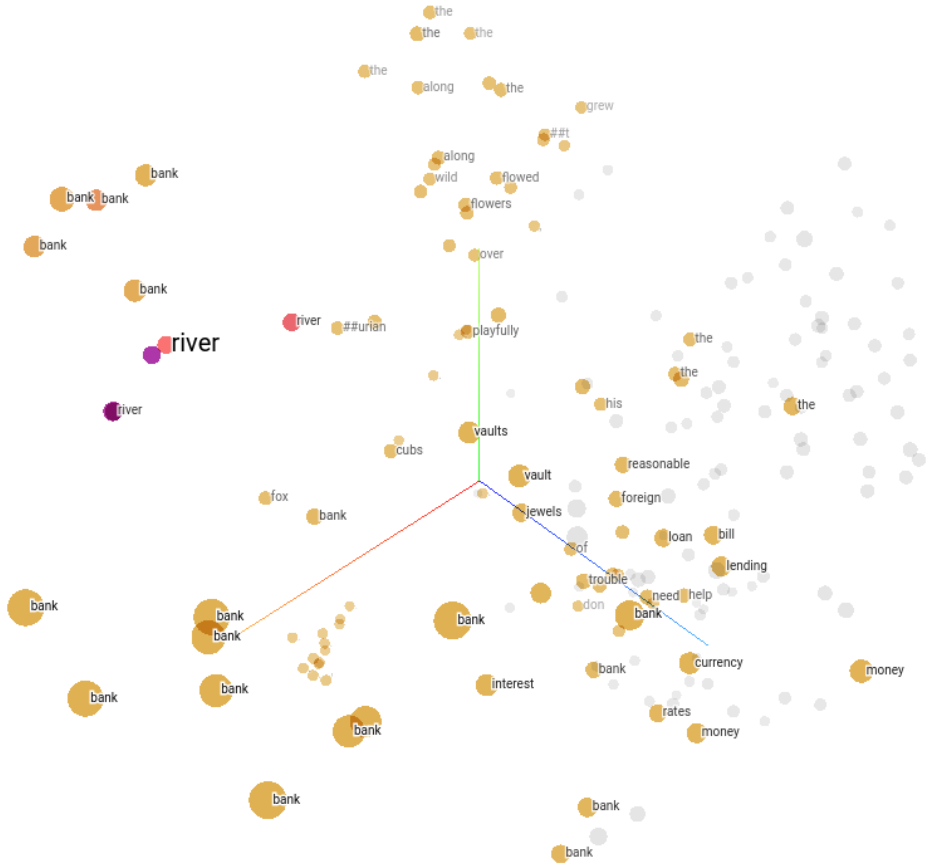


[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "**BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**". NAACL-HLT 2019.

Visualisation des embeddings par BERT

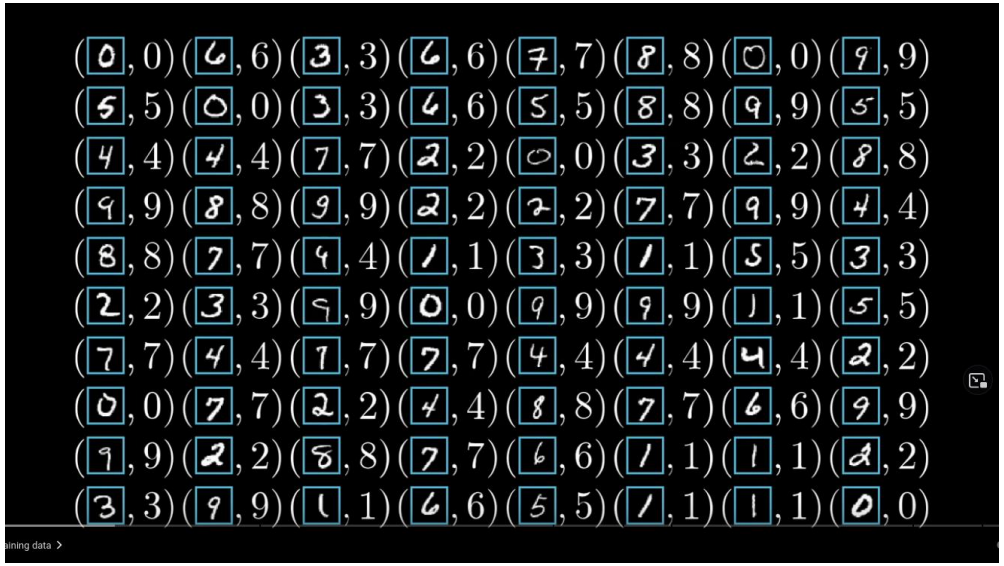
- Un vecteur pour un même mot pour chaque apparition

```
sentences = ["bank",
        "he eventually sold the shares back to the bank at a
premium.",
        "the bank strongly resisted cutting interest rates.",
        "the bank will supply and buy back foreign currency.",
        "the bank is pressing us for repayment of the loan.",
        "the bank left its lending rates unchanged.",
        "the river flowed over the bank.",
        "tall, luxuriant plants grew along the river bank.",
        "his soldiers were arrayed along the river bank.",
        "wild flowers adorned the river bank.",
        "two fox cubs romped playfully on the river bank.",
        "the jewels were kept in a bank vault.",
        "you can stow your jewellery away in the bank.",
        "most of the money was in storage in bank vaults.",
        "the diamonds are shut away in a bank vault somewhere.",
        "thieves broke into the bank vault.",
        "can I bank on your support?",
        "you can bank on him to hand you a reasonable bill for your
services.",
        "don't bank on your friends to help you out of trouble.",
        "you can bank on me when you need money.",
        "i bank on your help."
        ]
```



Au delà de l'architecture : les approches pour l'entraînement

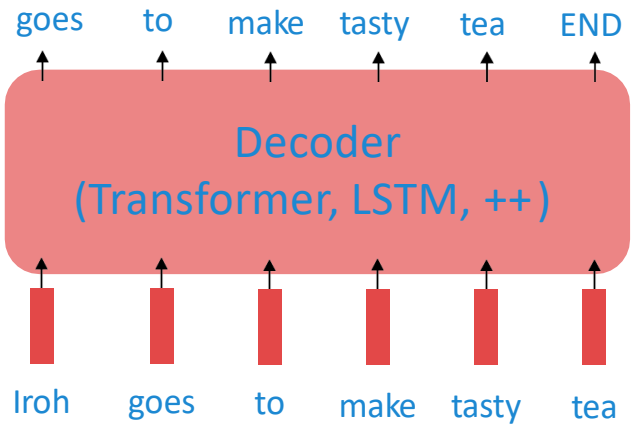
- Pour les images de chiffres à reconnaître : Un jeu de données est constitué des images et de leurs étiquettes (le chiffre correspondant), annotées par un humain.
- On partage le jeu de données en donnés pour l'entraînement (~80%), et le reste est gardé pour le test.
- Permet de tester le modèle sur des données jamais vues, et d'estimer ses capacités une fois déployédans le monde réel.



Comment entrainer un modèle Transformer pour la représentation de mots ?

- Nécessite énormément de données → difficile d'entraîner sur des datasets annotés manuellement (*apprentissage supervisé*)
- Mais on cherche à créer un modèle de langage: des représentations de mots encodant leurs distributions conditionnelles, qui puissent ensuite être utilisées pour des tâches plus spécifiques

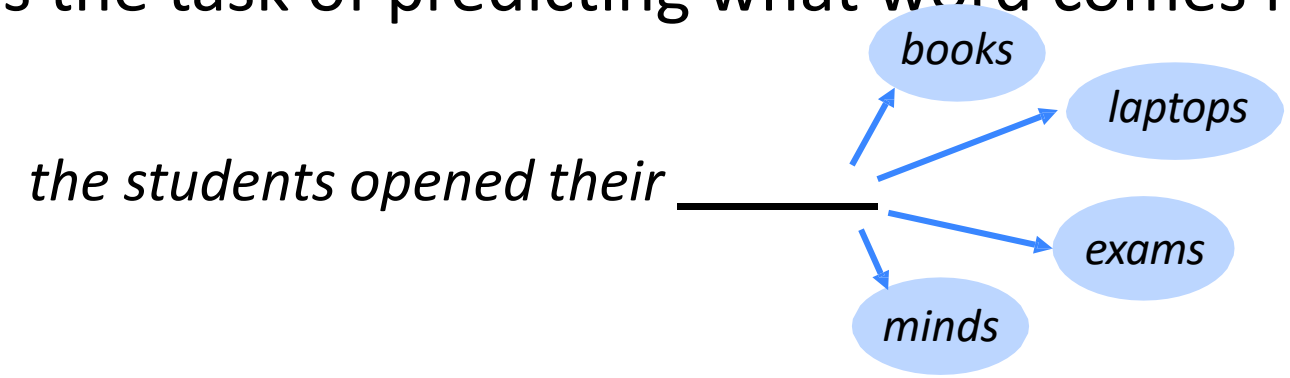
→ On entraîne en auto-supervisé (**self-supervised**) en définissant des tâches de reconstruction



[Dai and Le, 2015]

Language Modeling

- **Language Modeling** is the task of predicting what word comes next.



- More formally: given a sequence of words $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t)}$, compute the probability distribution of the next word $\boldsymbol{x}^{(t+1)}$:

$$P(\boldsymbol{x}^{(t+1)} | \boldsymbol{x}^{(t)}, \ldots, \boldsymbol{x}^{(1)})$$

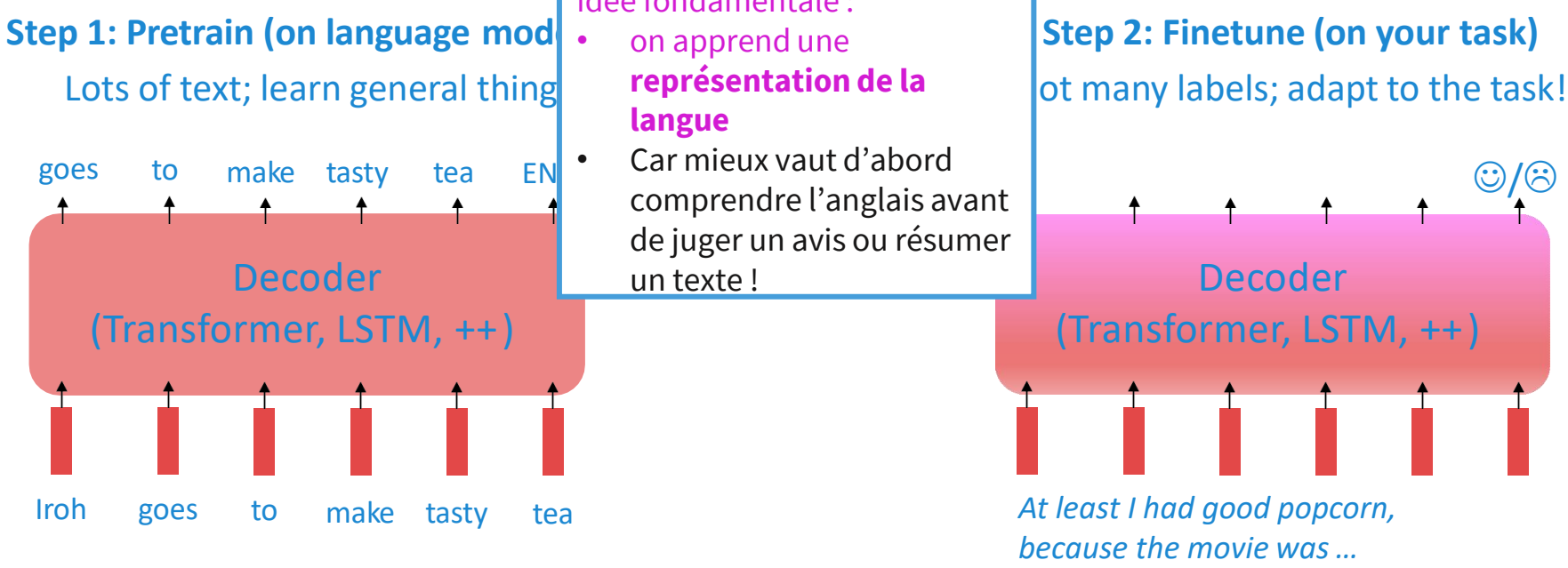
where $\boldsymbol{x}^{(t+1)}$ can be any word in the vocabulary $V = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{|V|}\}$

- A system that does this is called a **Language Model**.

Abigail See

Principe du pré-entraînement et du raffinement

- Pré-entraînement :
  - initialise les paramètres du réseau
- Raffinement :
  - adapte les paramètres du reseau à votre tâche spécifique



Various LLMs

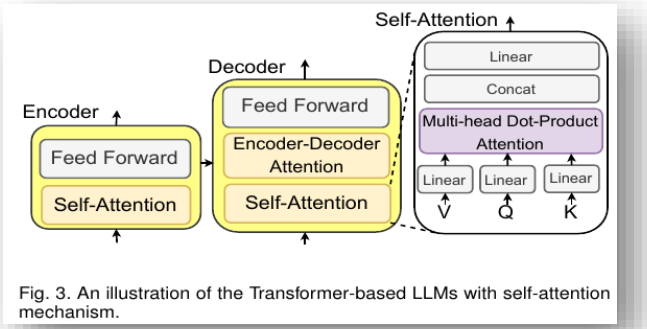


Fig. 3. An illustration of the Transformer-based LLMs with self-attention mechanism.

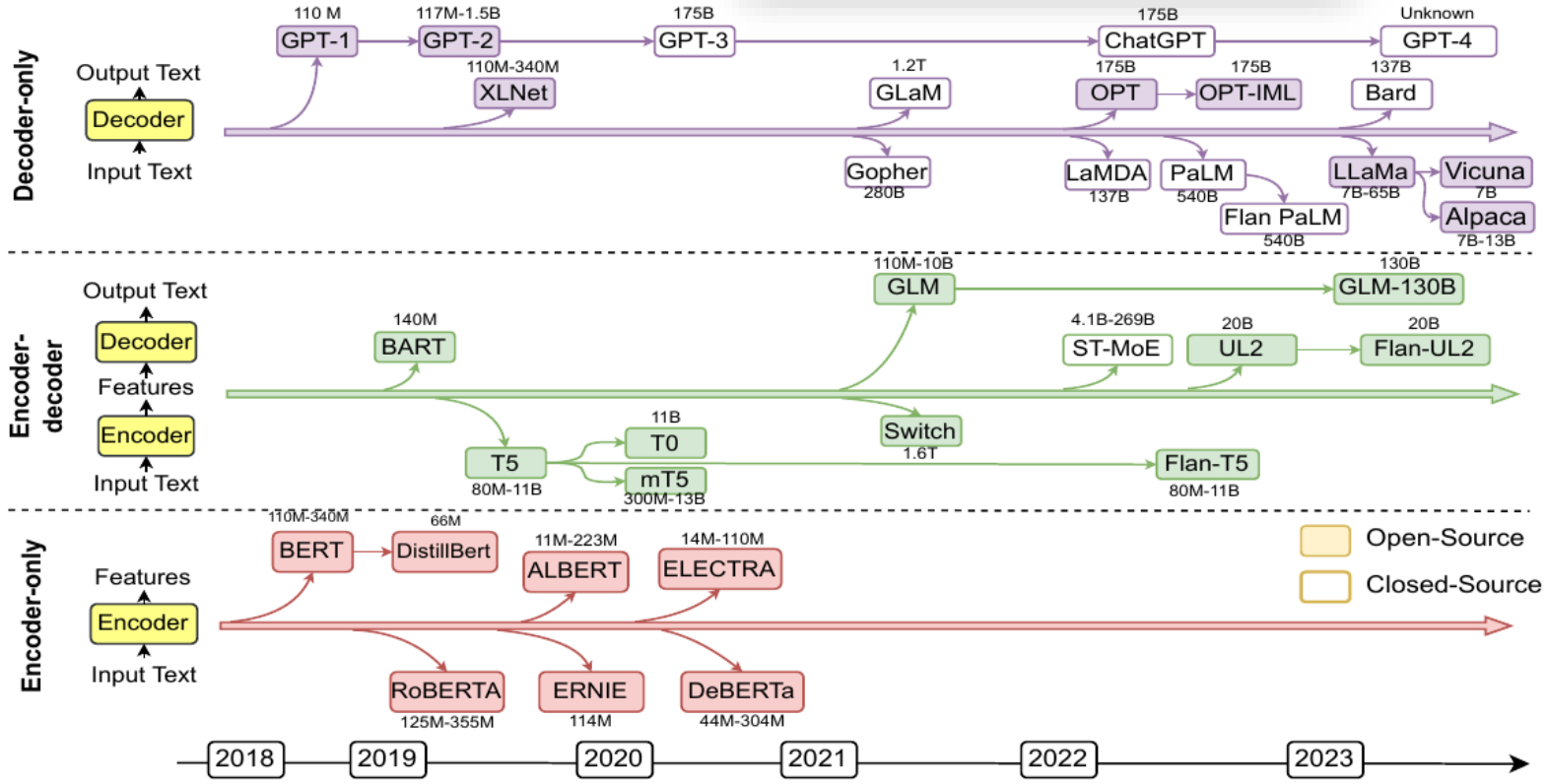


Fig. 2. Representative large language models (LLMs) in recent years. Open-source models are represented by solid squares, while closed source models are represented by hollow squares.

Plan de la formation

1. Apprentissage de représentation pour la reconnaissance de formes
   - Perceptron multi-couche (MLP)
   - Réseaux de neurones convolutionnels (CNN)
     - pour l'apprentissage de motifs pertinents dans les données
2. Apprentissage de représentation de mots
   - Représentations apprises par similarités de contextes
   - Représentation apprises par modèlisation du langage
     - Modèles Transformers et pré-entraînement
3. Modèles fondation : un changement de paradigme
   - Emergence de capacité imprévues
   - En langage, visio, audio... Et plus
   - Nouvelles méthodes pour adapter les modèles à des tâches spécifiques
4. Limites et enjeux
   - Environnement social et politique du design et du déploiement des systèmes de ML

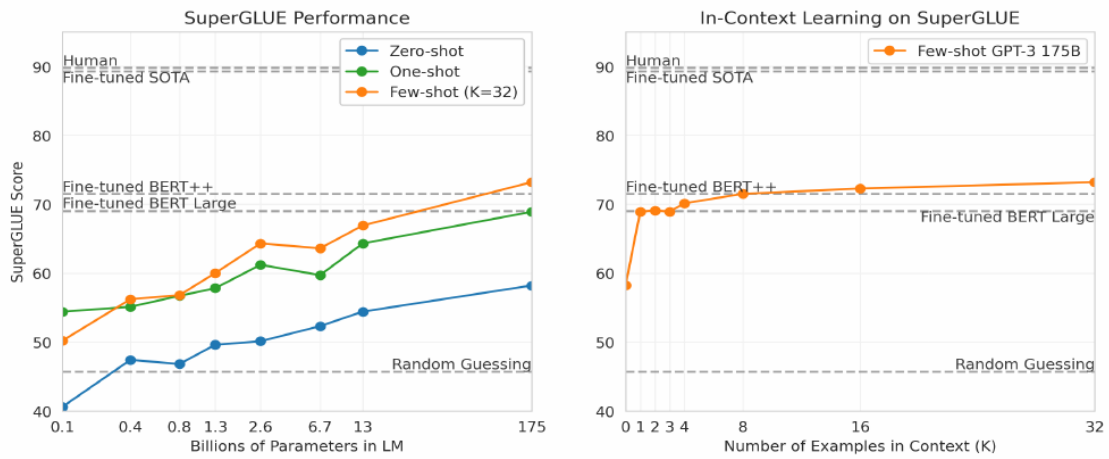
Foundation models: new paradigm

- ***Foundation model***: to describe a paradigm shift in AI by designating a model class that are distinctive in their sociological impact and how they have conferred a broad shift in practices in AI research and deployment.
- FMs led to surprising **emergence** which results from scale: GPT3 (175B params, GPT2 1.5B) permits **in-context learning**, in which the language model can be adapted to a downstream task simply by providing it with a prompt (a natural language description of the task), an emergent property that was neither specifically trained for nor anticipated to arise.

[1] R. Bommasani et al., “On the Opportunities and Risks of Foundation Models.” arXiv, Jul. 12, 2022. Accessed: Sep. 03, 2023. [Online]. Available: http://arxiv.org/abs/2108.07258

LLMs are few-shot learners

- **In-context (task) learning** by
  - Fine-Tuning (FT) - updates the weights of a pre-trained model by training on thousands of supervised labels specific to the desired task.
  - Few-Shot (FS) - the model is given a few demonstrations of the task at inference time as conditioning [RWC+19], but no weights are updated.
  - Zero-Shot (0S) - similar to few-shot but with a natural language description of the task instead of any examples.



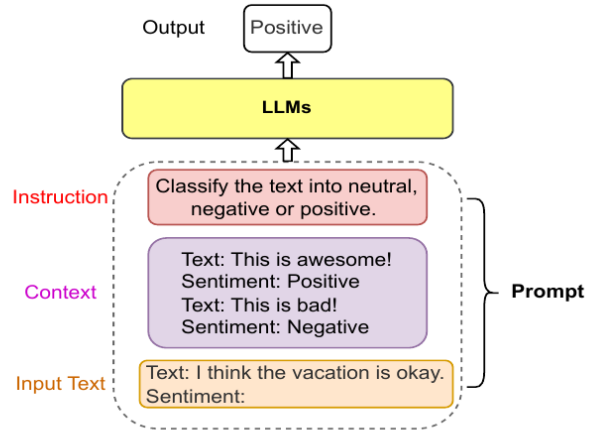


Fig. 4. An example of sentiment classification prompt.

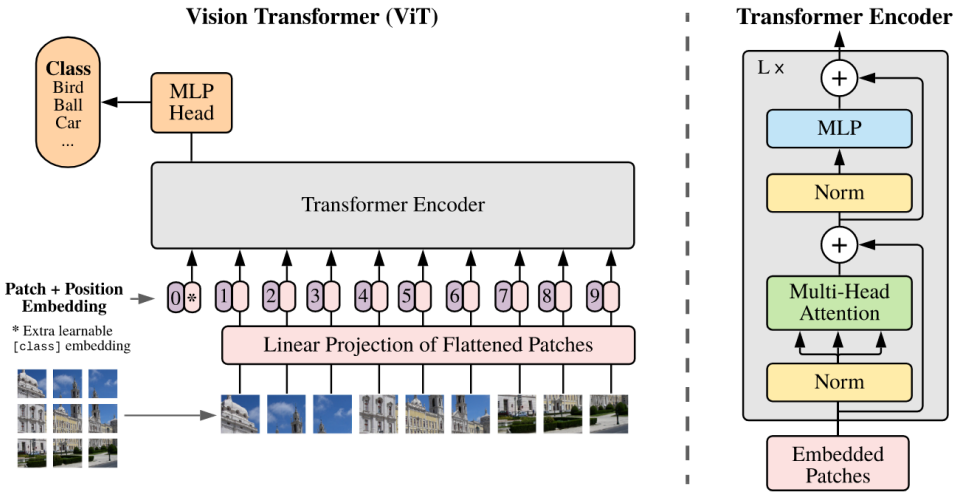
Taken from Pan et al.

[1] T. B. Brown et al., “Language Models are Few-Shot Learners”. NeurIPS 2020.

Transformers for vision

- The necessity of self-supervised learning is even greater in domains such as video analysis where annotations are more expensive
- Self-supervised pretraining of ViT: emerging features show interesting properties that do not emerge with supervised ViTs, nor with convnets

→ richer image representations from pre-training than those obtained from labeled data



[1] S. Mo, Z. Sun, and C. Li, "Multi-level Contrastive Learning for Self-Supervised Vision Transformers," WACV 2023.
[2] M. Caron et al., "Emerging Properties in Self-Supervised Vision Transformers," ICCV 2021.
[3] K. Ranasinghe, M. Naseer, S. Khan, F. S. Khan, and M. S. Ryoo, "Self-Supervised Video Transformer," CVPR 2022.

CLIP: Contrastive Language Image Pre-training

- CLIP: **to compute aligned representations of text and images**
- Dataset of 400 million (image, text) pairs collected from the internet
- Contrastive objective:
  - Predicts only which text as a whole is paired with which image and not the exact words of that text.
- Natural language is used to reference learned visual concepts, enabling zero-shot transfer of the model to downstream tasks.

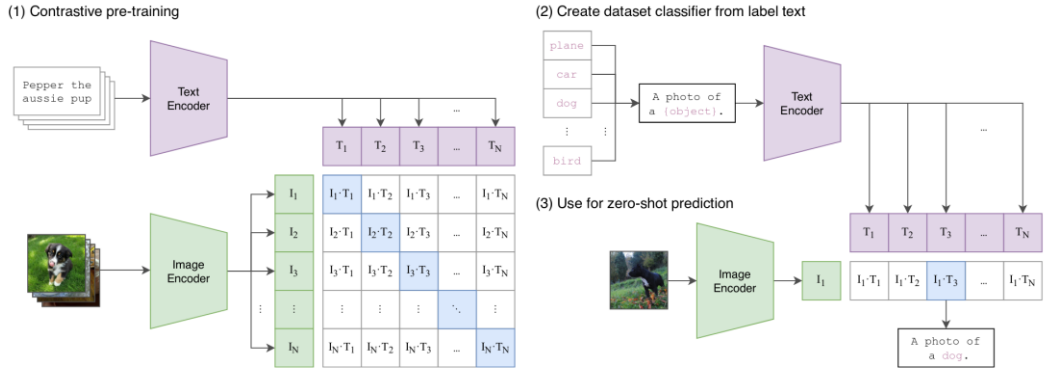


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

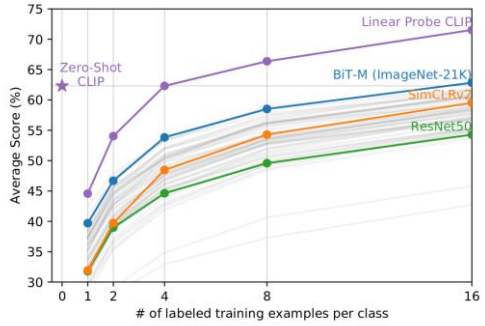


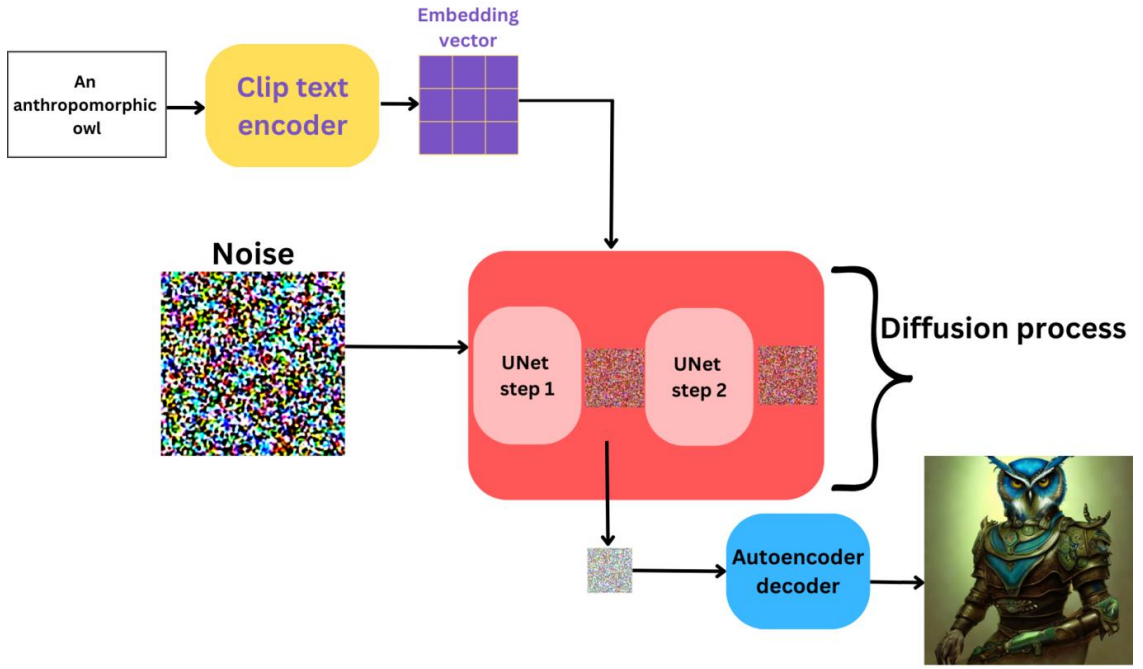
Figure 5. **Zero-shot CLIP outperforms few-shot linear probes.**

Zero-shot CLIP is much more robust to distribution shift than standard ImageNet models.

[1] A. Radford et al., "Learning Transferable Visual Models From Natural Language Supervision". ICML 2021

CLIP for Generative AI text-vision

- CLIP is used in generative models such as DALL-E 2 or Stable Diffusion:
  - CLIP embeddings are processed by the diffusion models used.



Taken from S. Rath

https://the-decoder.com/new-clip-model-aims-to-make-stable-diffusion-even-better/

Foundation models: new paradigm

- ***Foundation model***: to describe a paradigm shift in AI by designating a model class that are distinctive in their sociological impact and how they have conferred a broad shift in practices in AI research and deployment.
- FMs led to surprising **emergence** which results from scale: GPT3 (175B params, GPT2 1.5B) permits **in-context learning**, in which the language model can be adapted to a downstream task simply by providing it with a prompt (a natural language description of the task), an emergent property that was neither specifically trained for nor anticipated to arise.
- **Homogenization**: similar Transformer-based sequence modeling approaches for text, images, speech, code, protein sequences, organic molecules, reinforcement learning
  → towards a unified set of tools as FMs across a wide range of modalities
- **Homogenization** of actual models across research communities in the form of multimodal models
  → multimodal foundation models to fuse all the relevant information about a domain and adapt to tasks that also span multiple modes

[1] R. Bommasani et al., "On the Opportunities and Risks of Foundation Models." arXiv, Jul. 12, 2022. Accessed: Sep. 03, 2023. [Online]. Available: http://arxiv.org/abs/2108.07258

Foundation models

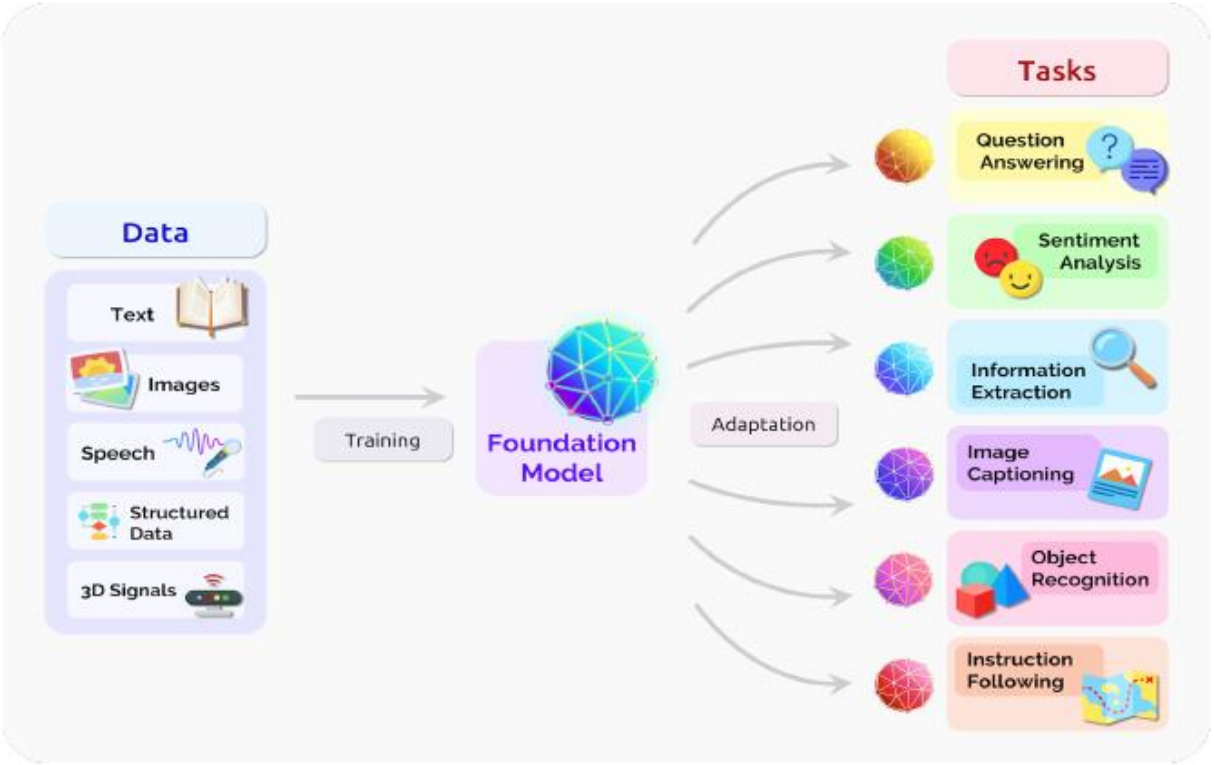
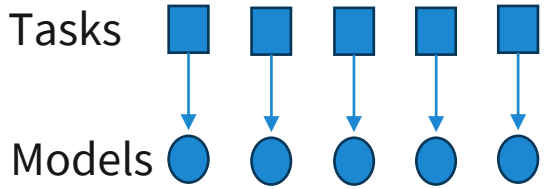
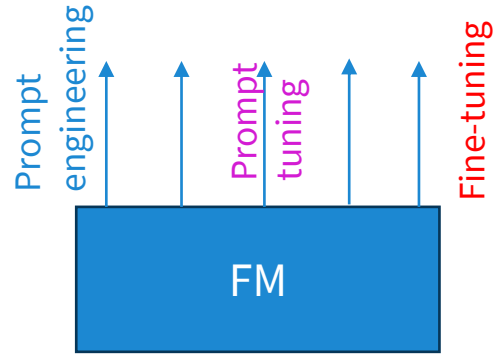


Fig. 2. A foundation model can centralize the information from all the data from various modalities. This one model can then be adapted to a wide range of downstream tasks.

Before:



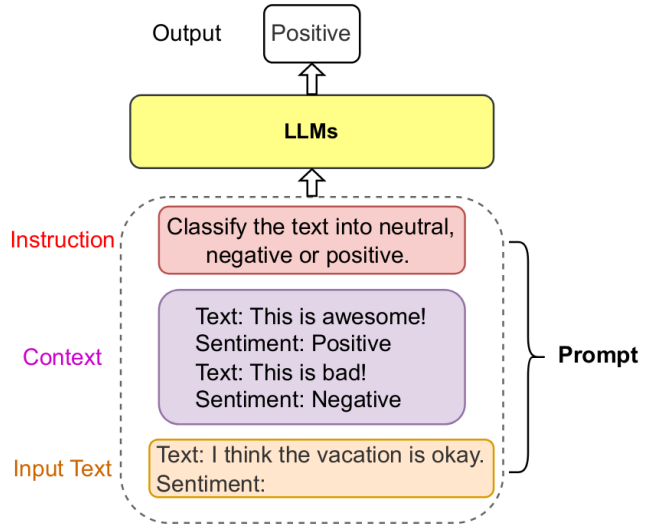
Now:



[1] R. Bommasani et al., “On the Opportunities and Risks of Foundation Models.” arXiv, Jul. 12, 2022. Accessed: Sep. 03, 2023. [Online]. Available: http://arxiv.org/abs/2108.07258

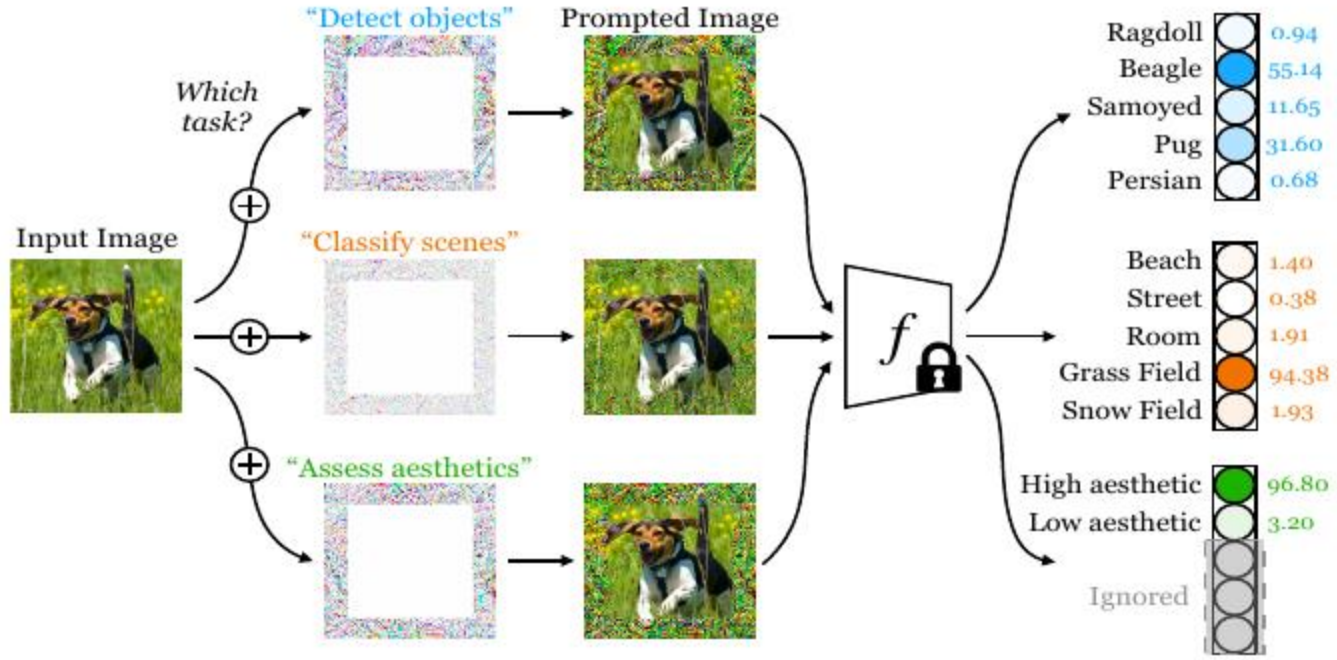
How to adapt a foundation model to build a new task-specific model?

- Fine-tuning of the FM:
  - Adapts the weights of the FM
  - Requires lots of data
  - Risk of over-fitting
- Prompting and prompt engineering:
  - Adapt input w/o access to the model
  - Practices and computational approaches: https://www.promptingguide.ai/
- Prompt tuning:
  - Learning to adapt target data in input space



Prompt-tuning: example for vision

- Prompt Learning in Pixel Space
- Prompt = a continuous task-specific vector



H Bahng, A Jahanian, S Sankaranarayanan, P Isola. Exploring visual prompts for adapting large-scale models. arXiv preprint arXiv:2203.17274 1 (3), 4, 2022.

Example to use scarce additional info

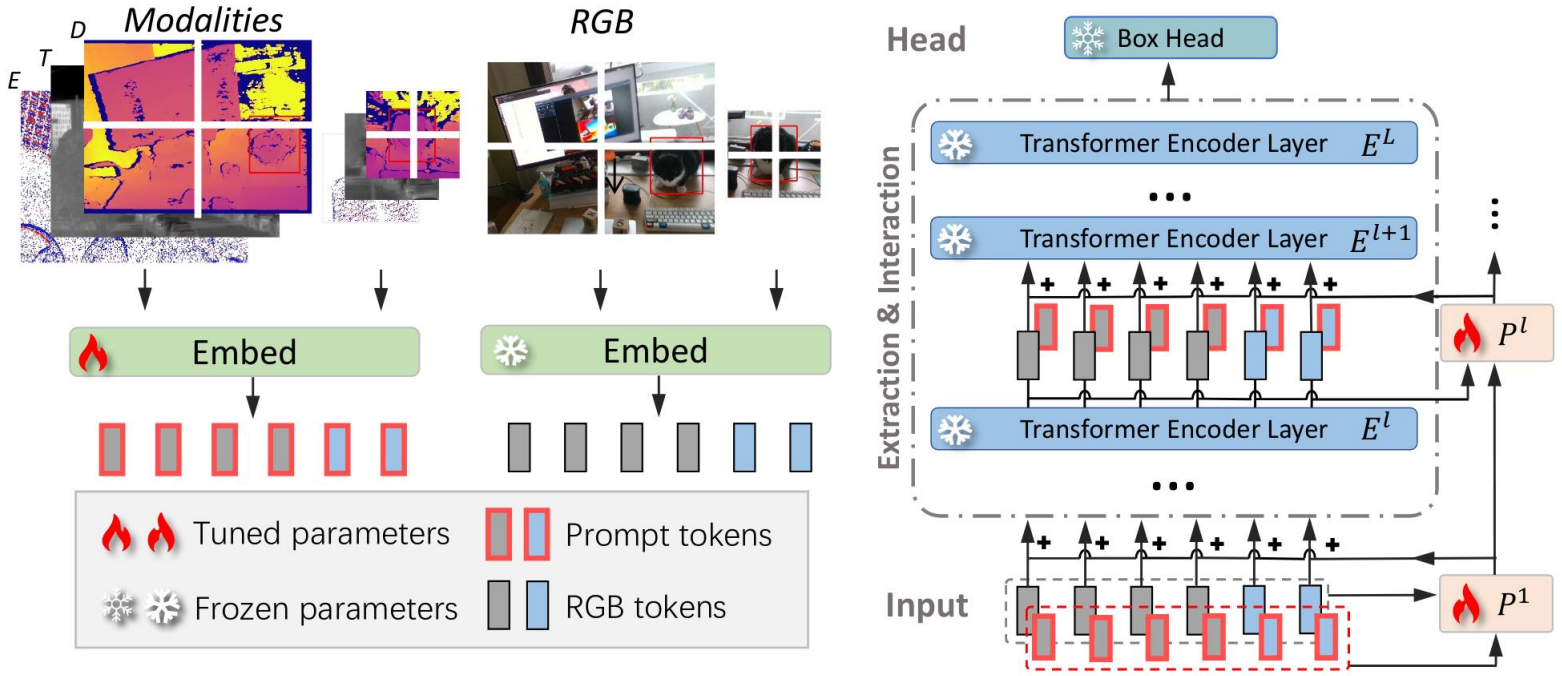
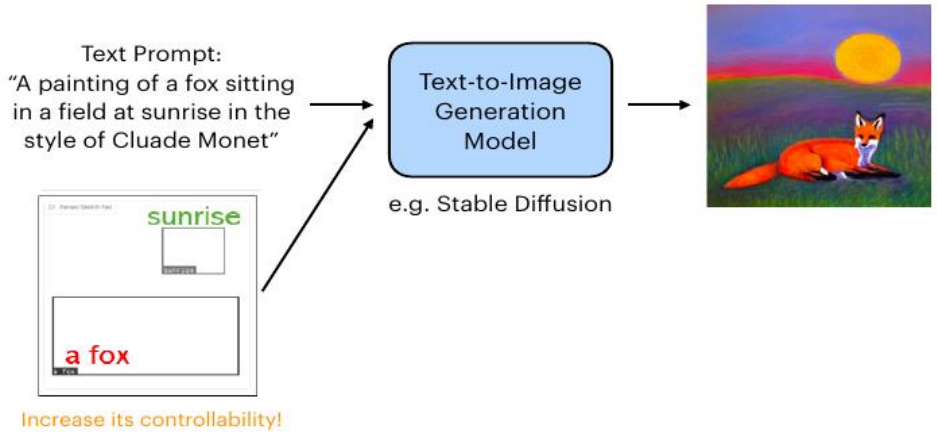


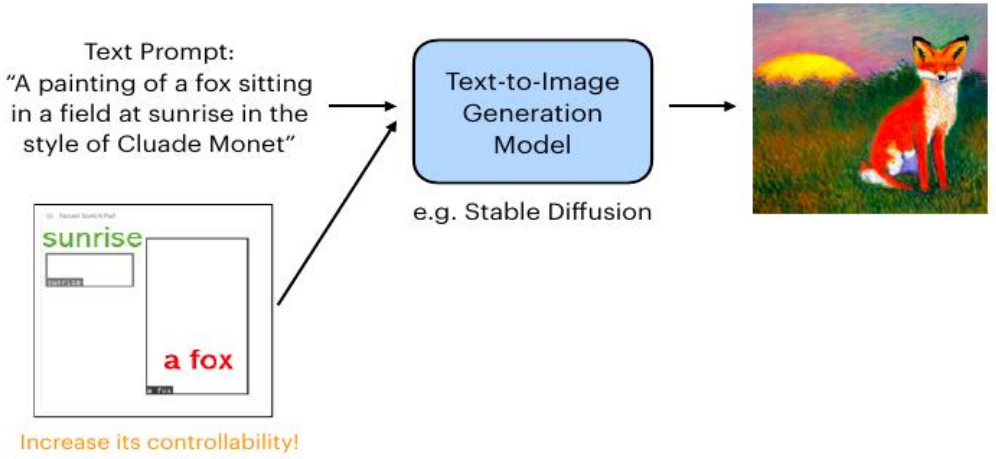
Figure 2. **Overview architecture of our ViPT.** The RGB- and auxiliary- modal inputs are first fed to the patch embed to generate the corresponding RGB and prompt tokens. L-layer stacked vision transformer (ViT) backbone is used for feature extraction and interaction. Modality-complementary prompters $\{P^l\}_{l=1}^{L}$ are inserted into the foundation model to learn the effective visual prompts.

[1] J. Zhu, S. Lai, X. Chen, D. Wang, and H. Lu, “Visual Prompt Multi-Modal Tracking”.

Example for controllable generation

- Promptable image generation





Y. Li et al. GLIGEN: Open-Set Grounded Text-to-Image Generation. CVPR 2023.

Wrapping-up (1/2)

- Top-down/symbolic AI and bottom-up AI/Machine learning
  - Knowledge-driven, data-driven
- Clé du Machine Learning moderne (deep learning) : l'apprentissage de représentations
- Image :
  - Neurone artificiel comme classifieur linéaire, réseau de neurones artificiels (MLP) pour apprentissage de motifs de plus en plus complexes par combinaisons de couches en couches
  - Permet de représenter les données par la présence ou l'absence de ces motifs
  - Permettre plus de motifs plus complexes à apprendre : CNN pour invariance par translation et motifs appris sous forme de filtres
- Texte :
  - Représentation numérique unique d'un mot (embedding) en alignant sur les représentations des mots ayant le même contexte (Word2Vec)
  - Transformers: représentation du mot dépend de son contexte, obtenue par recombinaisons des autres mots et le motif (la façon de recombiner) dépend lui-même des voisins, et dépendance longue distance
- Puissance de l'apprentissage auto-supervisé sur grand corpus
- Emergence de capacité d'adaptation à de nouvelles tâches
- Modèles fondation : modèles pré-entraînés pour un domaine, adaptables à d'autres tâches (avec prompt engineering et prompt tuning en plus de fine-tuning)

Risques et enjeux de ces systèmes socio-techniques

- Semblant de “compréhension” par cohérence grammaticale, mais:
  - Les langues sont des systèmes de signes, des paires de formes et de significations, les LLM ne font qu’enchaîner des formes en fonction d'informations probabilistes sans accès au sens.
  - Les humains partagent un socle commun de communication, en sont mutuellement conscients, ont des intentions, utilisent le langage pour les transmettre, et modélisent l'état mental de l'autre.
- Energie consommée : en entraînement (modèle 2019: vol altantique), à présent en inférence
- Représentativité des données et biais +/- subtils dans le texte produit (vue hégémonique)
- Défauts des FM hérités par les modèles avals
- Erreurs factuelles
- Exploitation des travailleur·euses du clic
- Modèle économique et souveraineté

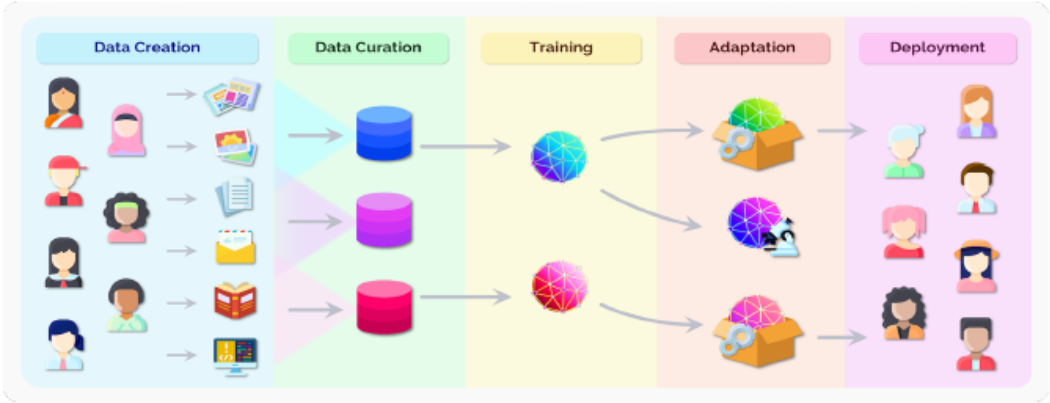


Fig. 3. Before reasoning about the social impact of foundation models, it is important to understand that they are part of a broader ecosystem that stretches from data creation to deployment. At both ends, we highlight the role of people as the ultimate source of data into training of a foundation model, but also as the downstream recipients of any benefits and harms. Thoughtful data curation and adaptation should be part of the responsible development of any AI system. Finally, note that the deployment of adapted foundation models is a decision separate from their construction, which could be for research.

Taken from Bommasani et al.

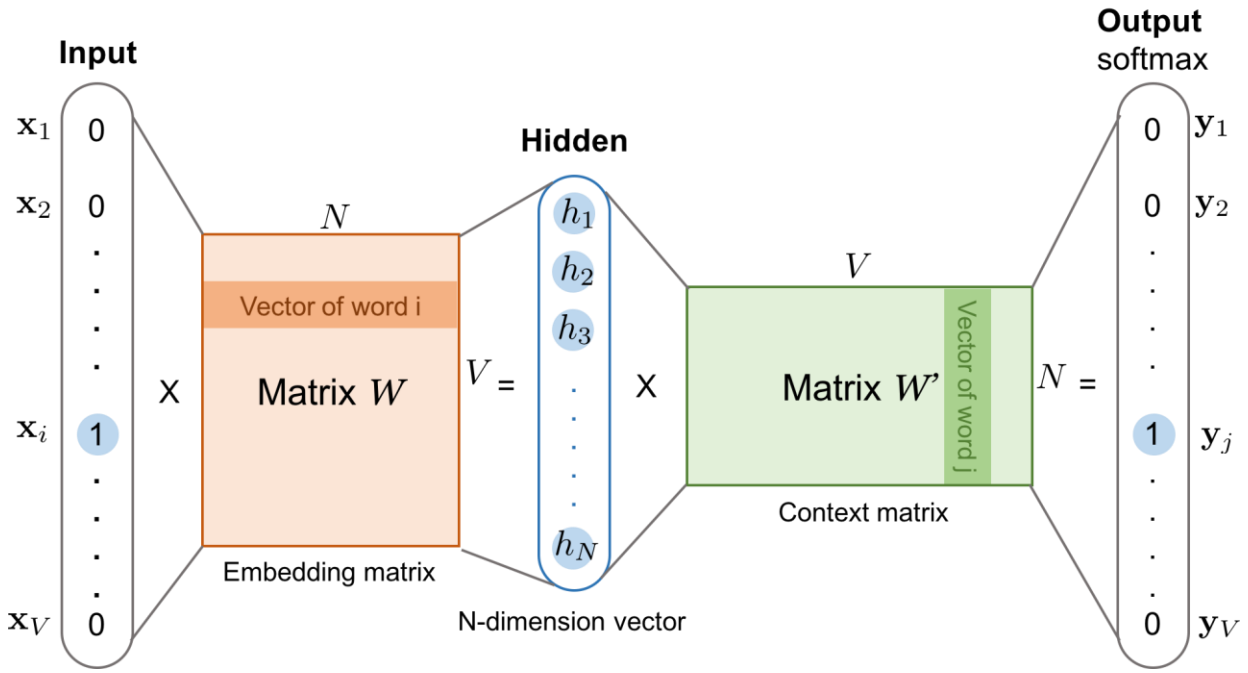
[1] TIME. *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic*. Avril 2022 , https://time.com/6247678/openai-chatgpt-kenya-workers/
[2] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Mar. 2021.
[3] L. Sassatelli. Is data fixable? On the need of socially-informed practices in ML research and education - Part 2: A more holistic perspective on data creation and expectations, Medium, 3IA Côte d'Azur, Jan. 2023.

Plan de la formation

1. Apprentissage de représentation pour la reconnaissance de formes
   - Perceptron multi-couche (MLP)
   - Réseaux de neurones convolutionnels (CNN)
     - pour l'apprentissage de motifs pertinents dans les données
2. Apprentissage de représentation de mots
   - Représentations apprises par similarités de contextes
   - Représentation apprises par modèlisation du langage
     - Modèles Transformers et pré-entraînement
3. Modèles fondation : un changement de paradigme
   - Emergence de capacité imprévues
   - En langage, visio, audio... Et plus
   - Nouvelles méthodes pour adapter les modèles à des tâches spécifiques
4. Limites et enjeux
   - Environnement social et politique du design et du déploiement des systèmes de ML

Biais dans les représentations de mots et d’images

Context-Based: Skip-Gram Model of Word2Vec



Taken from Lilian Weng

- We want to minimize the objective function:

$$J(\theta) = -\frac{1}{T}\log L(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{-m\le j\le m \\ j\neq 0}} \log P(w_{t+j} \mid w_t; \theta)$$

- Mots avec contextes proches auront des représentatiosn proches
- We compute $P(w_{t+j}|w_t; \theta)$ as:

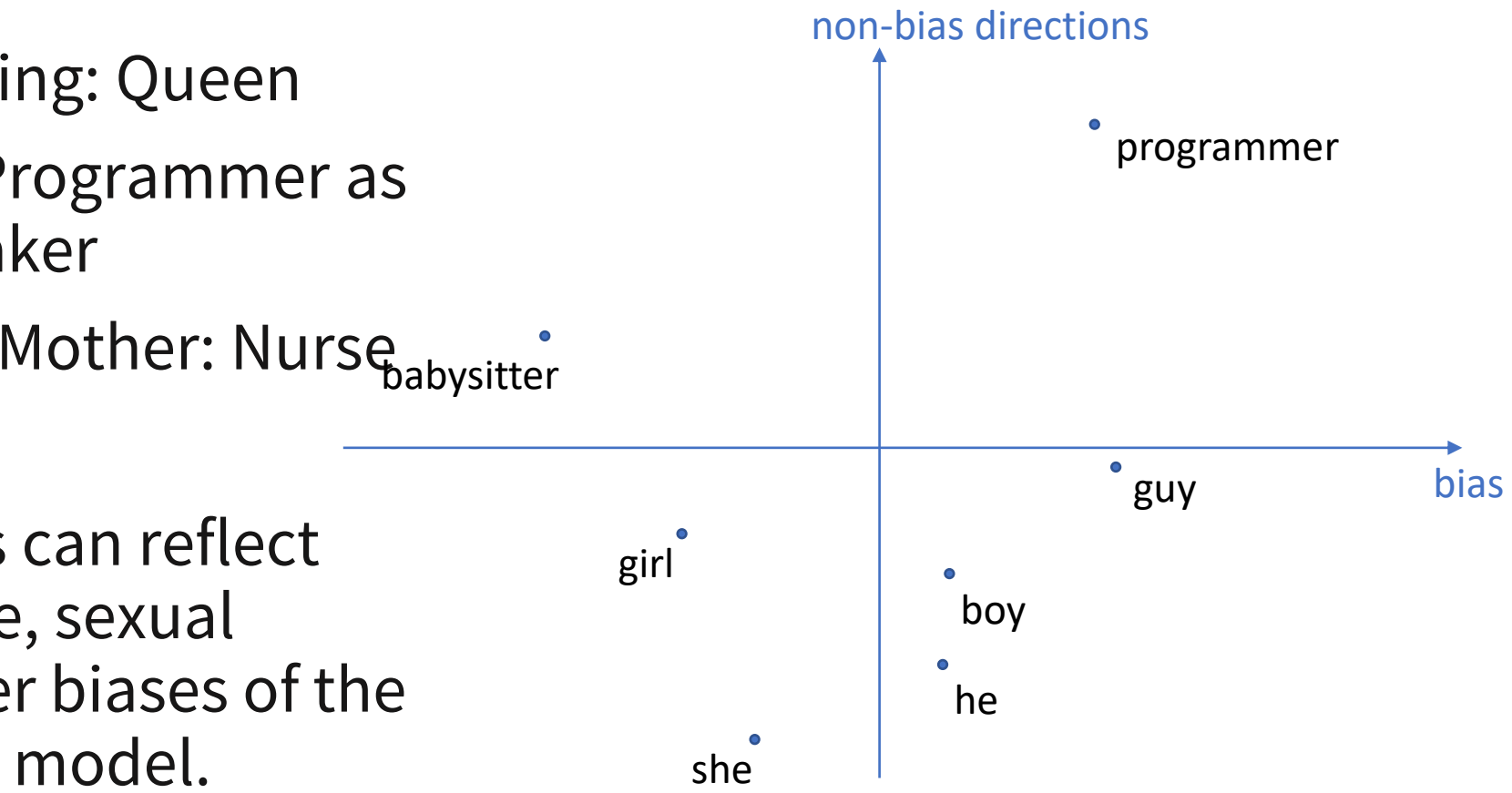
$$p(w_O|w_I) = \frac{\exp({v'_{w_O}}^\top v_{w_I})}{\sum_{i=1}^{V}\exp({v'_{w_i}}^\top v_{w_I})}$$

Mikolov et al., 2013, Efficient estimation of word representations in vector space.

The problem of bias in word embeddings

- Man: Woman as King: Queen
- Man: Computer_Programmer as Woman: Homemaker
- Father: Doctor as Mother: Nurse

→ Word embeddings can reflect gender, ethnicity, age, sexual orientation, and other biases of the text used to train the model.



Bolukbasi et al., 2016, Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. NeurIPS 2016.

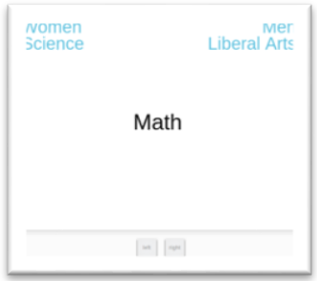
Définition de biais : humain, dans les données, appris par le système d'IA

- **Un choix préalable, arbitraire, non technique** : représenter le sens d'un mot par son contexte
- Conséquence : la représentation numérique obtenue (par le modèle d'IA) pour chaque mot reproduit des associations de co-occurrences entre concepts, dont les problématiques
- Biais (statistique) : caractéristique avec différentes fréquences d'apparition entre les groupes, dans l'absolu ou par rapport à d'autres caractéristiques
- Biais (éthique pour l'IA) : les déviations statistiques d'un modèle pour des groupes de personnes sur la base d'attributs protégés tels que le genre, la race, l'âge, etc.

[1] S. Fabbrizzi, S. Papadopoulos, E. Ntoutsi, and I. Kompatsiaris, "A survey on bias in visual datasets," Computer Vision and Image Understanding,2022.

Biais humains : mesurés par les scores IAT

- Formulé en psychologie cognitive et sociale
- Dans notre cerveau, le réseau de notre mémoire sémantique fonctionne par des associations entre concepts.
  - Biais quand ces associations sont problématiques :
    - Onpense plus vote à la savanne quand on nous parle lions, mais aussi généralement plus vite aux hommes quand on nous parle science
- Et on peut mesurer la force de ces automatismes : **Implicit Association Test (IAT)**

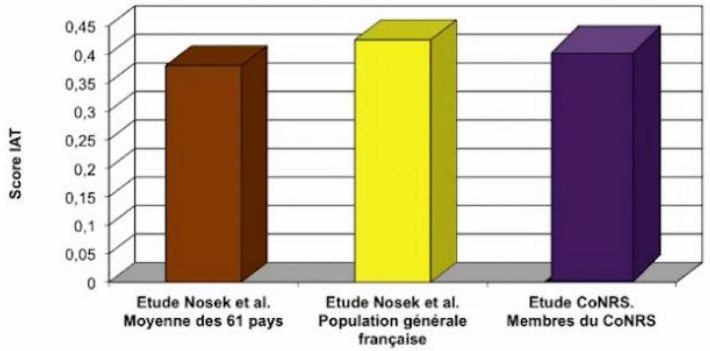


$$Score\ IAT = \frac{tps\ rép.pour\ assoc.incomp.avec.stéréo - tps\ rép.pour\ assoc.comp.avec.stéréo}{écart\ type\ des\ tps\ de\ rép.}$$

- Pop. générale : individus rapides sur les essais compatibles avec le stéréotype
- Score positif statistiquement significatif : stéréotype est bien installé dans les réseaux de la mémoire sémantique

→ tester la rapidité d'exécution teste la force des connexions entre concepts

→ IAT très fiable pour tester la stéréotypie implicite



A. G. Greenwald, D. E. McGhee, and J. L. Schwartz, "Measuring individual differences in implicit cognition: the implicit association test.," Journal of personality and social psychology, June 1998.

Quantifier les biais d'un modèle de langage: l'IAT étendu aux représentations apprises de mots

- Word Embedding Association Test (WEAT):
  - A and B are target groups, w are attributes (like occupation)

$$s(w, A, B) = \frac{\text{mean}_{a \in A} \cos(\vec{w}, \vec{a}) - \text{mean}_{b \in B} \cos(\vec{w}, \vec{b})}{\text{std-dev}_{x \in A \cup B} \cos(\vec{w}, \vec{x})}$$

- Several associations can be probed:
  - age and pleasantness, sexuality (gay or straight) and pleasantness, Arab-Muslim and pleasantness, gender and science, gender and career
- Language models trained on large-scale data learn similar biased associations between concepts:
  - the distances between the vectorized latent representations of words related to the same pairs of concepts (WEAT) reflect the IAT scores of tested populations.

[1] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics Derived Automatically from Language Corpora Contain Human-like Biases. Technical Report 6334. Science.
[2] W. Guo and A. Caliskan, "Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases," in Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event USA: ACM, Jul. 2021, pp. 122–133. doi: 10.1145/3461702.3462536.
[3] K. Kurita, N. Vyas, A. Pareek, A. W. Black, and Y. Tsvetkov, "Measuring Bias in Contextualized Word Representations," in Proceedings of the First Workshop on Gender Bias in Natural Language Processing, Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 166–172. doi: 10.18653/v1/W19-3823.

Semantics Derived Automatically from Language Corpora Contain Human-like Biases

Category	Targets	Templates
Pleasant/Unpleasant (Insects/Flowers)	flowers,insects,flower,insect	T are A, the T is A
Pleasant/Unpleasant (EA/AA)	black, white	T people are A, the T person is A
Career/Family (Male/Female)	he,she,boys,girls,men,women	T likes A, T like A, T is interested in A
Math/Arts (Male/Female)	he,she,boys,girls,men,women	T likes A, T like A, T is interested in A
Science/Arts (Male/Female)	he,she,boys,girls,men,women	T likes A, T like A, T is interested in A

Table 2: Template sentences used and target words for the grammatically correct sentences (T: target, A: attribute)

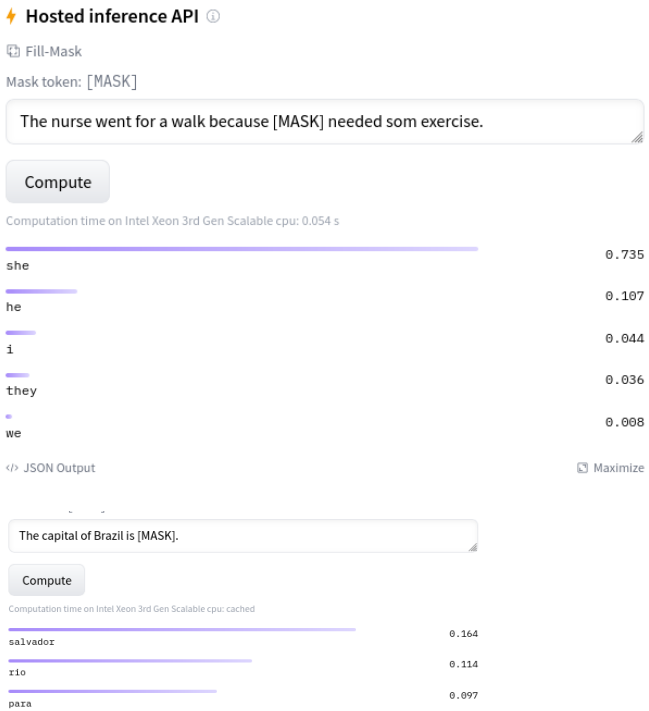
Category	WEAT on GloVe	WEAT on BERT	Ours on BERT *Log Probability Bias Score*
Pleasant/Unpleasant (Insects/Flowers)	1.543*	0.6688	0.8744*
Pleasant/Unpleasant (EA/AA)	1.012	1.003	0.8864*
Career/Family (Male/Female)	1.814*	0.5047	1.126*
Math/Arts (Male/Female)	1.061	0.6755	0.8495*
Science/Arts (Male/Female)	1.246*	0.8815	0.9572*

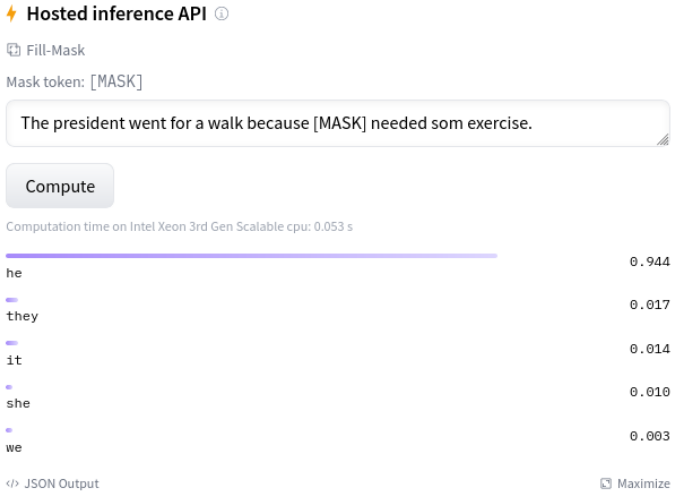
Table 3: Effect sizes of bias measurements on WEAT Stimuli. (* indicates significant at $p < 0.01$)

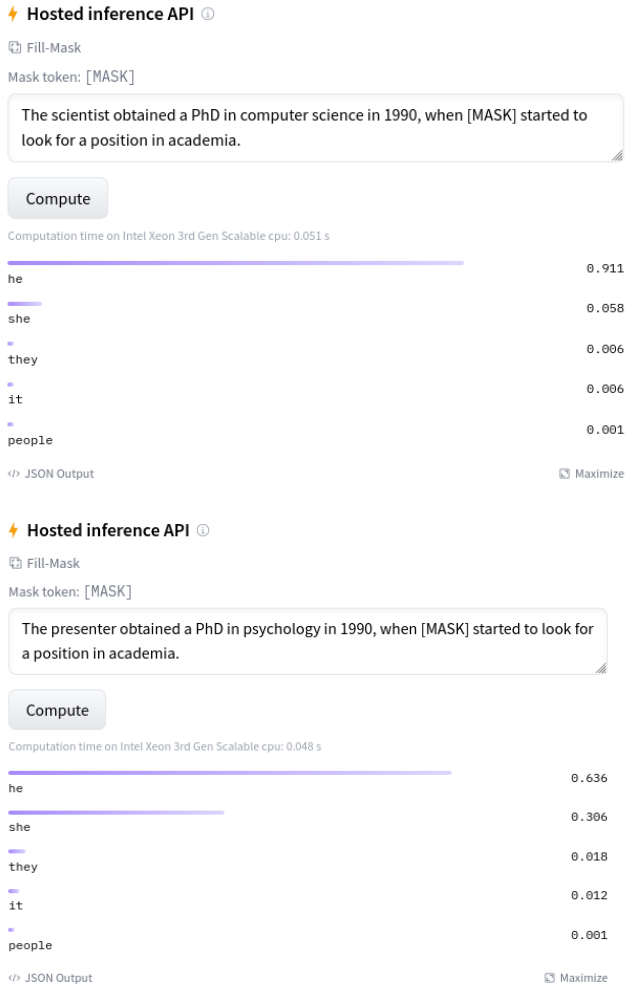
[1] K. Kurita, N. Vyas, A. Pareek, A. W. Black, and Y. Tsvetkov, "Measuring Bias in Contextualized Word Representations," in Proceedings of the First Workshop on Gender Bias in Natural Language Processing, Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 166–172. doi: 10.18653/v1/W19-3823.

Démo des biais dans BERT

- Depuis l'interface de Hugging Face: [lien]







https://huggingface.co/bert-base-uncased
https://dmccreary.medium.com/showing-bias-in-bert-475e98dabf51

Impact of biased word representation on generated text

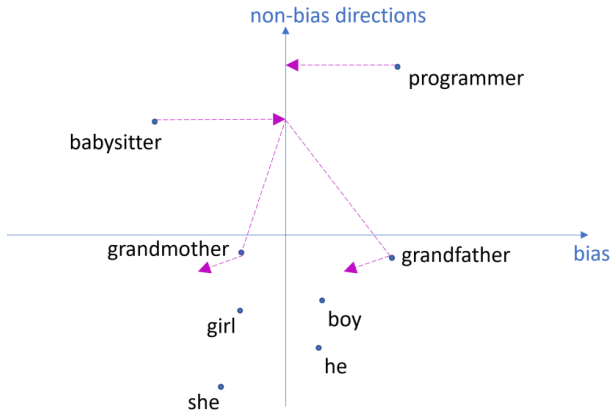
- Open Google translate
- Try English → Persian → English
  for the sentence “Her name is Sandra. She obtained her PhD in computer science in 1989.”
- Try English → Persian → English
  for the sentence “Her name is Sandra. She is a nurse working with babies in hospital.”
- Do it again this time with “His name is Alexander. He obtained his PhD in computer science in 1989.”
- Then “His name is Alexander. He is a nurse working with babies in hospital.”

→ Can you see a problem there?

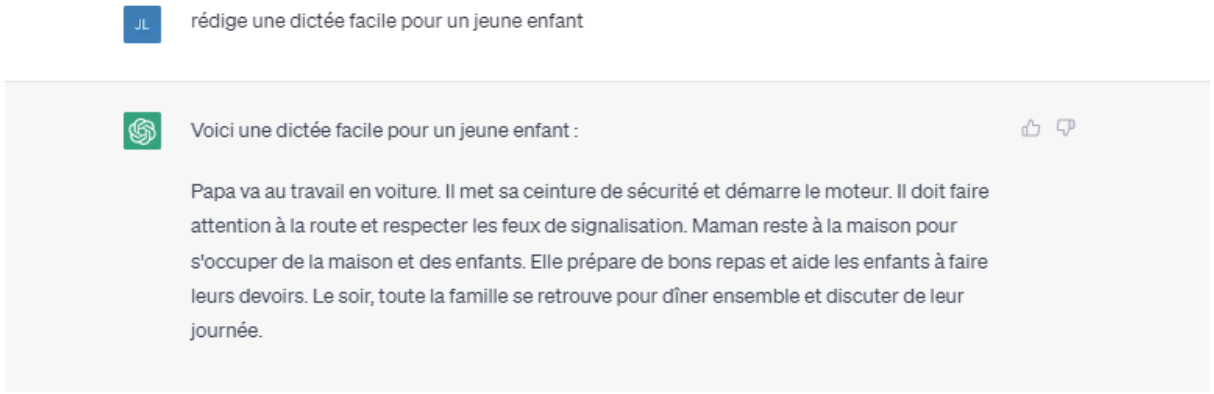
Bias mitigation techniques

Data Manipulation			
Data Augmentation	**Gender Tagging**	**Balanced Fine-Tuning**	**Adding Context**
Madaan et al. [2018]; Park et al. [2018] Hall Maudslay et al. [2019]; Zhao et al. [2018a] Emami et al. [2019]; Zmigrod et al. [2019] Bartl et al. [2020]; Zhao et al. [2019] de Vassimon Manela et al. [2021]; Sen et al. [2021]	Moryossef et al. [2019]; Vanmassenhove et al. [2018] Habash et al. [2019]; Stafanovičs et al. [2020] Saunders et al. [2020]	Park et al. [2018]; Saunders and Byrne [2020] Costa-jussà and de Jorge [2020]	Basta et al. [2020]
Methodological Adjustment			
Projection-Based Debiasing	**Adversarial Learning**	**Constraining Output**	**Other**
Bolukbasi et al. [2016]; Schmidt [2015] Bordia and Bowman [2019]; Park et al. [2018] Ethayarajh et al. [2019]; Sahlgren and Olsson [2019] Karve et al. [2019]; Sedoc and Ungar [2019] Liang et al. [2020]; Prost et al. [2019] Dev et al. [2020]; Kaneko and Bollegala [2021a]	Li et al. [2018]; Zhang et al. [2018]	Ma et al. [2020]; Zhao et al. [2017]	Qian et al. [2019]; Zhao et al. [2018b] Jin et al. [2021]; Kaneko and Bollegala [2019]

Table 3. Classification of gender bias mitigation methods with respective publications.



- Beaucoup de méthodes computationnelles, mais de nombreuses limites subsistent
  - Factuel (*La présidente de la France*, *Les hommes ont manifesté pour une justice reproductive qui leur a été déniée jusque là*,...)
  - Biais raciaux, validistes, etc.
  - Points de vue, styles, ...



[1] K. Stanczak and I. Augenstein, “A Survey on Gender Bias in Natural Language Processing.” arXiv, Dec. 28, 2021. Available: http://arxiv.org/abs/2112.14168

Et dans les images ?

- In the MS-COCO dataset:
  - women and persons with darker skin tones are less likely to take a large area of the image.
  - Persons with darker skin tones tend to appear more in outdoor transportation scenes, while women tend to appear more in indoor scenes like shopping and dining, and men appear in more outdoor scenes related to sports and vehicles categories.
- The less people appear in gender-traditional roles, the less their gender is correctly classified.

Figure 2: Taken from Wang et al. [28]. Errors on gender classification depending on the appearance of each binary gender. The classifier is an SVM using FC7 features from pretrained AlexNet. The figure shows " what the Linear SVM has learned on OpenImages for the sports uniform and flower categories. For sports uniform, males tend to be represented as playing outdoor sports like baseball, while females tend to be portrayed as playing an indoor sport like basketball or in a swimsuit. For flower, we see another drastic difference in how males and females are portrayed, where males pictured with a flower are in formal, official settings, whereas females are in staged settings or paintings." This shows the difficulty of addressing bias in data, beyond the simple count of occurrences.

Taken from [1]

[1] Angelina Wang, Alexander Liu, Ryan Zhang, Anat Kleiman, Leslie Kim, Dora Zhao, Iroha Shirai, Arvind Narayanan, and Olga Russakovsky, "Revise: A tool for measuring and mitigating bias in visual datasets," International Journal of Computer Vision, 2022.

Quantifier les biais d'un modèle de vision : l'IAT étendu aux représentations d'images

- An image transformer model like iGPT generates image representations to solve the next pixel prediction task.
  - 2 computer vision models published in summer 2020, iGPT and SimCLRv2. We extract representations of image stimuli with these two pre-trained, unsupervised image representation models
- Introduce the iEAT test

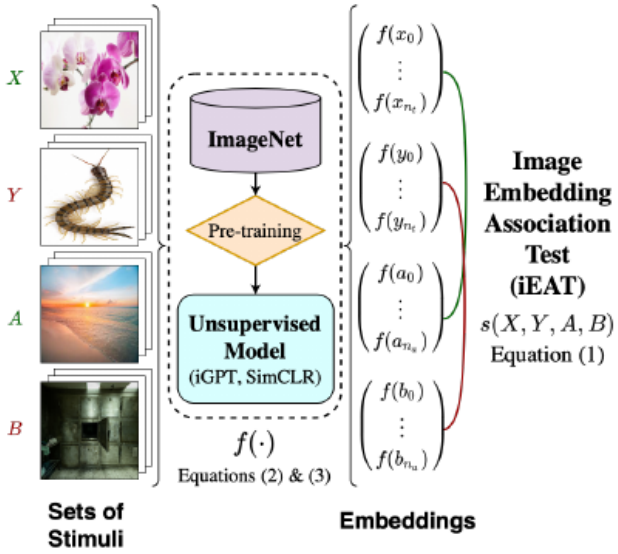


Figure 2: Example iEAT replication of the Insect-Flower IAT [31], which measures the differential association between flowers vs. insects and pleasantness vs. unpleasantness.

Taken from [1]

[1] R. Steed and A. Caliskan, "Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Canada, 2021.

Quantifier les biais d'un modèle de vision : l'IAT étendu aux représentations d'images

$$s(w, A, B) = \frac{\text{mean}_{a \in A} \cos(\vec{w}, \vec{a}) - \text{mean}_{b \in B} \cos(\vec{w}, \vec{b})}{\text{std-dev}_{x \in A \cup B} \cos(\vec{w}, \vec{x})}$$

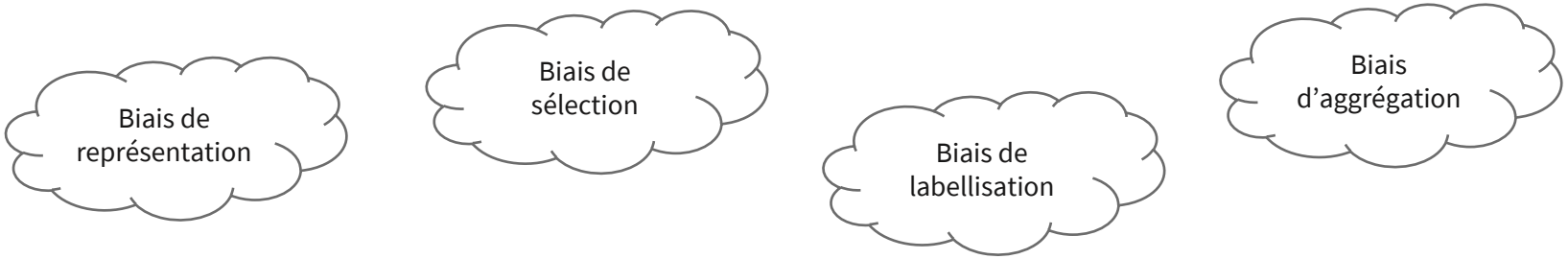
Table 1: iEAT tests for the association between target concepts X vs. Y (represented by n_t images each) and attributes A vs. B (represented by n_a images each) in embeddings generated by an unsupervised model. Effect sizes d represent the magnitude of bias, colored by conventional small (0.2), medium (0.5), and large (0.8). Permutation p-values indicate significance. Reproduced from Nosek et al. [56], the original human IAT effect sizes are all statistically significant with $p < 10^{-8}$; they can be compared to our effect sizes in sign but not in magnitude.

	X	Y	A	B	n_t	n_a	Model	iEAT d	iEAT p	IAT d
Age†	Young	Old	Pleasant	Unpleasant	6	55	iGPT	0.42	0.24	1.23
							SimCLR	0.59	0.16	1.23
Arab-Muslim	Other	Arab-Muslim	Pleasant	Unpleasant	10	55	iGPT	0.86	0.02	0.33
							SimCLR	1.06	$< 10^{-2}$	0.33
Asian§	European American	Asian American	American	Foreign	6	6	iGPT	0.25	0.34	0.62
							SimCLR	0.47	0.21	0.62
Disability†	Disabled	Abled	Pleasant	Unpleasant	4	55	iGPT	-0.02	0.53	1.05
							SimCLR	0.38	0.34	1.05
Gender-Career	Male	Female	Career	Family	40	21	iGPT	0.62	$< 10^{-2}$	1.1
							SimCLR	0.74	$< 10^{-3}$	1.1
Gender-Science	Male	Female	Science	Liberal Arts	40	21	iGPT	0.44	0.02	0.93
							SimCLR	-0.10	0.67	0.93
Insect-Flower	Flower	Insect	Pleasant	Unpleasant	35	55	iGPT	0.34	0.07	1.35
							SimCLR	1.69	$< 10^{-3}$	1.35
Native§	European American	Native American	U.S.	World	8	5	iGPT	-0.33	0.73	0.46
							SimCLR	-0.19	0.65	0.46
Race†	European American	African American	Pleasant	Unpleasant	6	55	iGPT	-0.62	0.85	0.86
							SimCLR	-0.57	0.83	0.86
Religion	Christianity	Judaism	Pleasant	Unpleasant	7	55	iGPT	0.37	0.25	-0.34
							SimCLR	0.36	0.26	-0.34
Sexuality	Gay	Straight	Pleasant	Unpleasant	9	55	iGPT	-0.03	0.52	0.74
							SimCLR	0.04	0.47	0.74
Skin-Tone†	Light	Dark	Pleasant	Unpleasant	7	55	iGPT	1.26	$< 10^{-2}$	0.73
							SimCLR	-0.19	0.71	0.73
Weapon§	White	Black	Tool	Weapon	6	7	iGPT	0.86	0.07	1.0
							SimCLR	1.38	$< 10^{-2}$	1.0
Weapon (Modern)	White	Black	Tool	Weapon	6	9	iGPT	0.88	0.06	N/A
							SimCLR	1.28	0.01	N/A
Weight†	Thin	Fat	Pleasant	Unpleasant	10	55	iGPT	1.67	$< 10^{-3}$	1.83
							SimCLR	-0.30	0.74	1.83

§ Originally a picture-IAT (image-only stimuli). † Originally a mixed-mode IAT (image and verbal stimuli).

Taken from [1]

[1] R. Steed and A. Caliskan, "Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Canada, 2021.

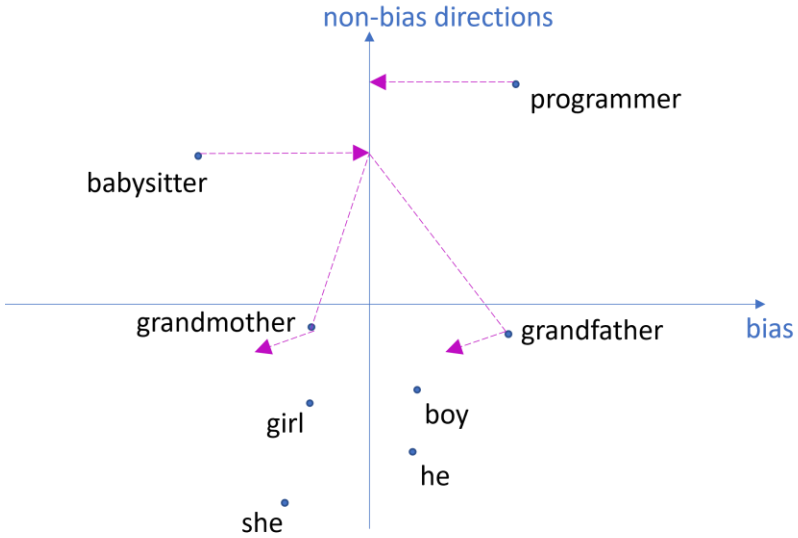


Prenons un peu de recul : la faute aux données ?

Spoiler : oui mais pas que

Est-ce que le problème n'est pas dans les données ?

- Dans les données....
- ET dans nos choix de méthodes !
  - On fait le choix en ML de ne prendre aucun a priori sur la représentation pertinente à avoir des données (pas de règles ou motifs pré-établis à partir de connaissance humaine), seulement (et c'est pas rien !) :
    - Données
    - Fonction de coût/grandeur à modéliser (probas conditionnelles de mots, similarité de contexte, etc.)
- Mais alors peut-on corriger les données ?



Une impossible correction ?

- La taille n'est pas garante de diversité : Qui écrit les textes sur Internet mis dans les datasets ?
  - Sur-représentation des jeunes utilisateurs et des pays développés
  - Encore plus dans les données collectées :
    - Ex: données pour GPT-2's issues de liens depuis Reddit : 67% d'hommes utilisateurs aux US, 64% 18-29
    - Wikipedia: 8.8–15% de femmes

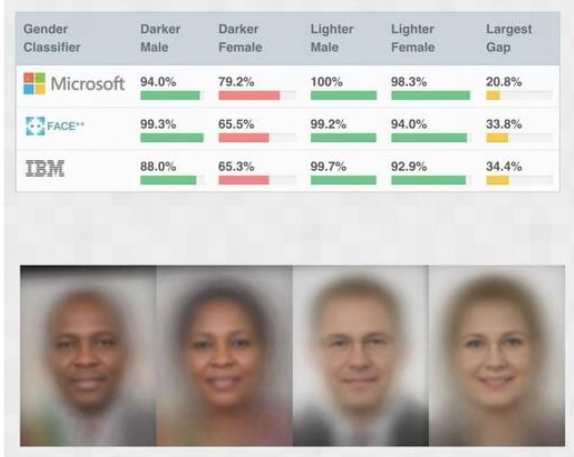
  → Une vue hégémonique est véhiculée dans les textes utilisés pour l'entraînement
- LLMs présentent de multiples biais (dont associations stéréotypiques) :
  - Intersectionalité : BERT, GPT-2 encodent plus de biais contre les identités marginalisées dans plusieurs dimensions
  - BERT : phrases avec personnes avec handicap ont plus de mots négatifs, ...
  - GPT-2 : 272K documents de sites non-fiables et 63K de subreddits interdits
  - GPT-3 : phrases générées fortement toxiques même pour des prompts non-toxiques
- Spoiler: problèmes similaire pour les modèles de vision

E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜," *ACM Conference on Fairness, Accountability, and Transparency*, Mar. 2021.

Une impossible correction ?

Raji et al. explicitent les **tensions éthiques** quand on tente de diversifier les datasets pour entraîner les modèles des reco faciale :

- Définir plusieurs catégories de groupes pour analyser l'équité peut ne pas tenir compte de l'intersectionnalité et nuire à l'équité.
- Représentativité vs vie privée :
  - Accroître le nombre d'échantillons des groupes sous-représentés sur Internet augmente de façon disproportionnée le risque lié à la vie privée
  - Les datasets majeurs crées avec des pratiques prédatrices en matière de collecte de données(ImageNet, CelebA, MS-Celeb)
  - EU, autorités de protection des données ont sanctionné Clearview AI pour de telle pratiques prédatices et illégales au regard du RGPD (France, Italy, UK)



Source: *Buolamwini & Gebru (2018)*

(c) J. Buolamwini

Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton, "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing,", arxiv, 2020.

Est-ce que accroître les données est possible et suffisant ?

- IBM Diversity in Face
  - Collected from Yahoo! Flickr CC
  - Avec catégories de genre, de couleur de peau, d'âge et de «structure faciale» :
    - Symmétrie faciale et forme du crâne !
    - Rappellent les pseudosciences de craniométrie et phrénologie pour tenter d'établir un déterminisme biologique de l'intelligence, toujours selon le genre et la race
- Crawford et Paglen : ces tentatives de corriger les datasets révèlent les actes politiques restant souvent implicites quand on construit un dataset de ML :
  - choisir les quelques catégories dans lesquelles diviser un monde continu
  - décider qui doit étiqueter chaque échantillon de données dans chaque catégorie, qui doit superviser les processus d'annotation
  - tenter de quantifier la diversité et choisir une certaine formule d'équité

Kate Crawford and Trevor Paglen, "Excavating ai: the politics of images in machine learning training sets," AI & SOCIETY, 06 2021.

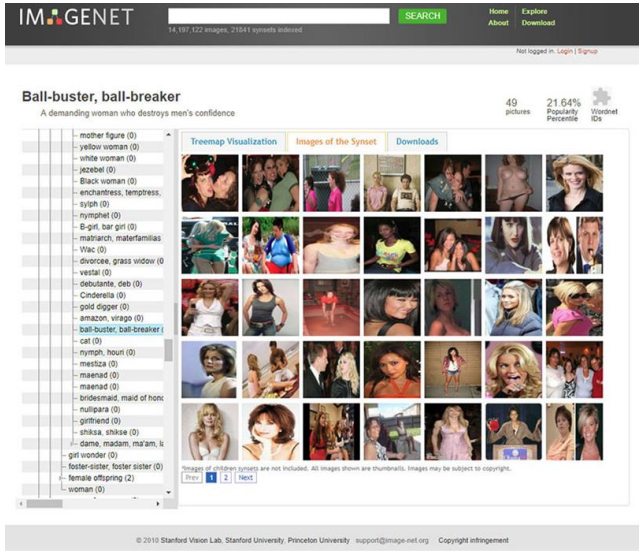
Les relations de pouvoir sous-tendent la création des datasets

Ces actes politiques exemplifiés dans 2 datasets majeurs :

- CelebA : 40 attributs binaires annotés par des travailleurs du clic d' Amazon Mechanical Turk. Catégories telles que "double chin", "pointy nose", "narrow eyes", "big lips", or "attractive". Problématiques car
  - intrinsèquement subjectives
  - historiquement utilisées pour des classifications racistes, antisémites ou sexistes
  - implicitement définies en référence à une certaine norme : masculine, hétérosexuelle, mince et caucasienne.

→ Norme pervasive au processus de création, reflétant les relations de pouvoir

- Denton et al. font une généalogie de ImageNet :
  - Initialement des sous-catégories alarmantes ("salope, alcoolique, buveur, casse-couilles, mulâtre, plouc").
  - Certaines étiquettes reflètent "une vision qui associe les bikinis aux femmes, les sports aux hommes", mais aussi "les truites aux trophées de pêche et les homards aux dîners".
  - ces vêtements, activités et animaux pourraient être décrits différemment selon d'autres points de vue sociaux



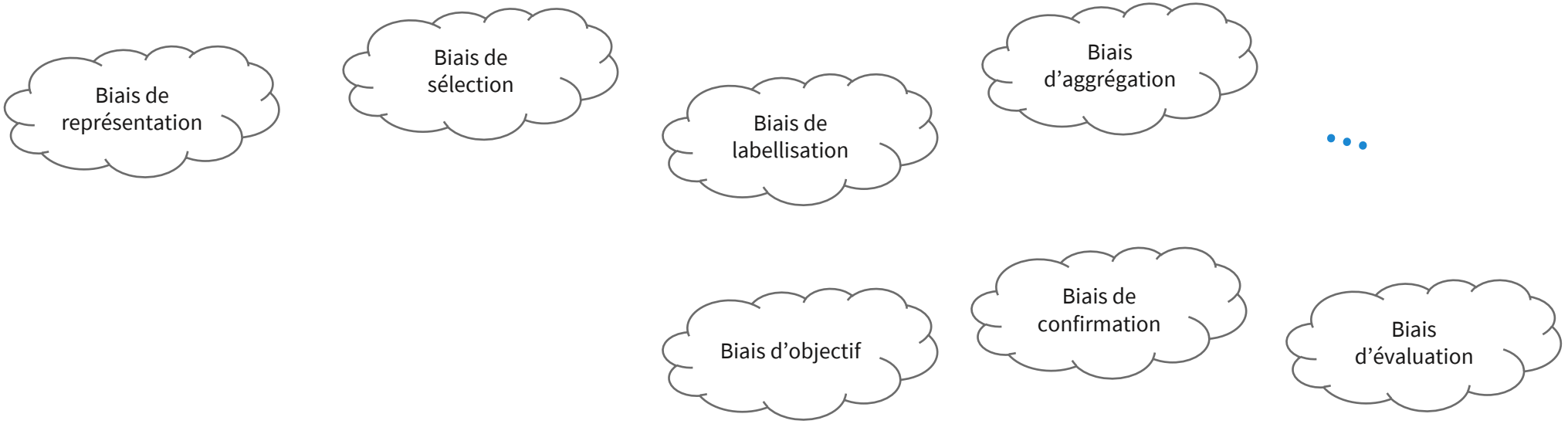
→ Les points de vue présents dans le dataset reflètent un "regard d'homme blanc occidental"

[1] Kate Crawford and Trevor Paglen, "Excavating ai: the politics of images in machine learning training sets," AI & SOCIETY, 06 2021.
[2] Inioluwa Deborah Raji and Genevieve Fried, "About Face: A Survey of Facial Recognition Evaluation," arxiv, 2021.
[3] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole, "On the genealogy of machine learning datasets: A critical history of ImageNet," Big Data & Society, 2021.
[4] Catherine D'Ignazio. The Urgency of Moving from Bias to Power. Medium post of Data + Feminism Lab, MIT, May 2023.

Dataset collection, annotation, compensation

- Sambasivan et al.: “high-stakes domains lacked pre-existing datasets, so practitioners were necessitated to collect data from scratch. ML data collection practices were reported to conflict with existing workflows and practices of domain experts and data collectors. Limited budgets for data collection often meant that data creation was added as extraneous work to on-the-ground partners (e.g., nurses, patrollers, farmers) who already had several responsibilities, and were not adequately compensated for these new tasks.”
- Data practices chosen to create instrumental datasets such as ImageNet failed to recognize the data work performed by click-workers, making their work invisible, overlooking the interpretive work of these humans, wrongly considered as a homogeneous pool .
- Goyal et al.: different pools of raters annotate speech toxicity differently depending on their multiple identities (self-declared racial group and sexual orientation)

N. Sambasivan et al., ““Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI,” ACM CHI 2021.
N. Goyal et al., “Is Your Toxicity My Toxicity? Exploring the Impact of Rater Identity on Toxicity Annotation,” Tech. Rep., arXiv, 2022.



Prenons encore un peu plus de recul : quelles tâches visons-nous ?

Est-que tant qu'on a de bonnes données, on va y arriver ?...

Spoiler : Pas nécessairement

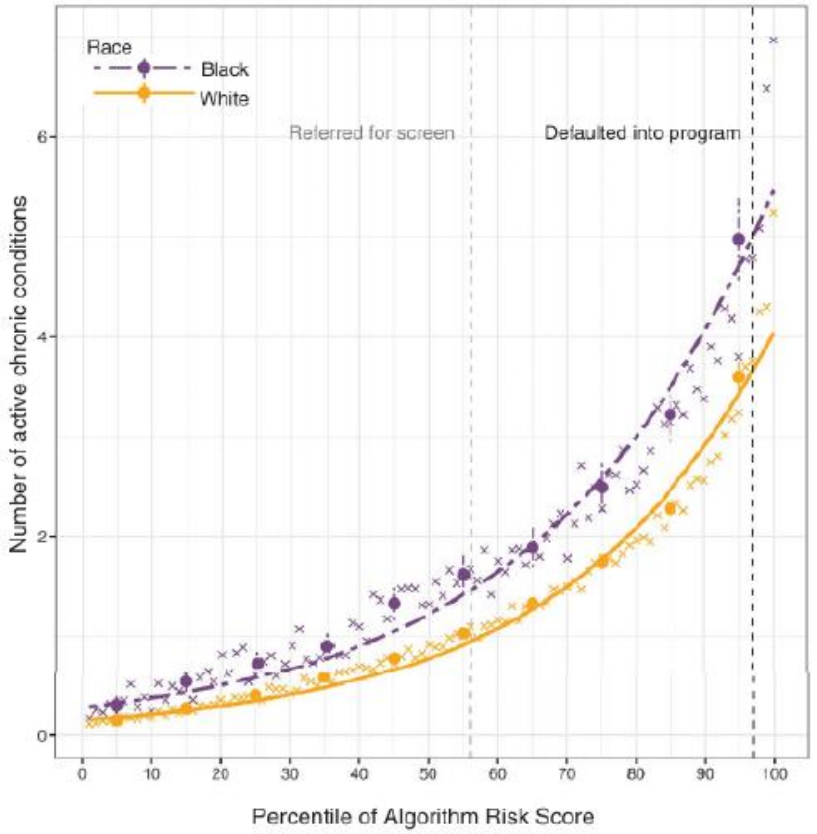
Échecs de déploiement de systèmes de ML dans des scénarios à fort enjeu

- Des systèmes d'IA ont déjà été déployés dans des scénarios à fort enjeu et ont échoué avec de terribles conséquences :
  - Des détecteurs de fraude aux allocations de chômage basés sur l'IA ont laissé des personnes (innocentes) sans revenus, des personnes paralysées ont vu leur aide à domicile réduite de moitié, ...
- Et ces échecs discriminent de façon disproportionnée les groupes sociodémographiques défavorisés. La communauté afro-américaine a été excessivement ciblée par les défaillances de systèmes :
  - utilisés pour identifier les criminels et prédire le taux de récidive
  - les personnes à faibles revenus ont été identifiées à tort comme ayant moins besoin d'assistance médicale
  - plus susceptibles de commettre des abus sur les enfants
  - Les femmes ont été identifiée comme moins intéressantes à recruter

[1] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst, "The Fallacy of AI Functionality," in 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea, June 2022, pp. 959–972, ACM.
[2] Thomas Liao, Rohan Taori, Inioluwa Deborah Raji, and Ludwig Schmidt, "Are We Learning Yet? A Meta Review of Evaluation Failures Across Machine Learning," in Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2), 2021.

A tort ou à raison ?

- Soyons claires : ces décisions étaient fausses
- Prévention de la maltraitance des enfants :
  - Fort sur-échantillonnage des familles ouvrières et de couleur, soumettant les parents et les enfants pauvres à des enquêtes plus fréquentes
- Allocation de lits d'hospital :
  - Les besoins de soin quantifiés par les dépenses de santé des individus !
- Recrutement des femmes :
  - La verité terrain (cible du système) faîte d'un historique de decisions humaines biaisées



Taken from Obermeyer et al. [7]. The figure shows that at a given risk score produced by the algorithm, Black patients are considerably sicker than White patients.

[1] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst, "The Fallacy of AI Functionality," in 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea, June 2022, pp. 959–972, ACM.

Reconnaissance faciale

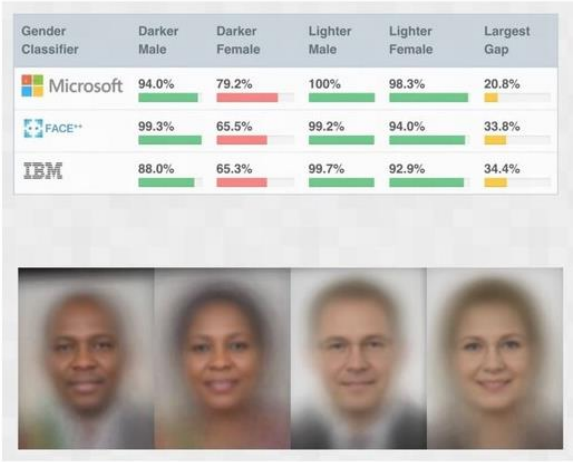
- Buolamwini and Gebru 2018 : inéquité de reconnaissance suivant une intersection de facteurs : genre + couleur de peau
- La pression persiste malgré des échecs systématiques en déploiement
  - Municipalités françaises sous pression
  - 2022, EDPB : appel certaines interdiction de reco. faciale pour police
  - 2024, JO : lobbying de la France dans le AI Act
- Et pas seulement :
  - Sur-représentation de visage à peau claire dans les datasets de visages pour la reconnaissance faciale
  - D'objets du monde occidental pour la reco d'objets
  - De pronoms masculins et noms masculins pour la reconnaissance d'entités nommées

→ Réalisation que les datasets reflètent la domination de groupes intersectionnels, et que ça impacte les modèles.

2015: Misclassification of minorities

2020: Biased super-resolution

2018: Intersectional bias in face tech.



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Source: *Buolamwini & Gebru (2018)*

[1] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst, "The Fallacy of AI Functionality," in 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul Republic of Korea, June 2022, pp. 959–972, ACM.[1]
[2] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns*, Nov. 2021.
[3] J. Hourdeaux. JO 2024 : l'expérimentation de la vidéosurveillance algorithmique inquiète. Mediapart, janvier 2023.

Mais alors

- Quelles hypothèses faire pour réduire la complexité du monde pour modéliser le problème ? Quelles données utiliser ?
- Doit-on essayer de les corriger ? Peut-on ?
- Doit-on questionner l'automatisation-même, quelque soit la méthode, de ces décisions sensibles ?

Humans involved

- Every dataset therefore involves humans: those who decide the target task, those who decide how to collect data samples, those who decide the annotation guidelines, those who decide who annotates, those who are assigned the annotation work, those whose personal data is used. This sheer fact yields limitations and biases in every data-driven approach, however massive it may be.
- To understand the connection between our ML research practices and the social and structural problems pervading datasets, let us introduce the guidelines of Paullada et al. for ML practitioners when we:
  - (1) define a problem to be tackled with ML,
  - (2) create or choose existing data to use
  - (3) analyze the model performance and envision real-world deployment.

A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, “Data and its (dis)contents: A survey of dataset development and use in machine learning research,” *Patterns*, Nov. 2021

Define a problem to be tackled with ML

- Tasks can be defined abstractly (“intensionally”) as a problem statement (e.g., object recognition, speech-to-text translation) or “extensionally”, that is instantiated by a learning problem made of a dataset of (input, output) pairs and an evaluation metric (e.g., top-1 accuracy)
- One must first analyze the intensional definition of the task and the mapping we can foresee between input and output.

Quels tâches attaquer avec du ML ?

Quelle correspondance envisage t-on entre l'entrée et la sortie ?

- Visage → orientation sexuelle ou employabilité
  - Tâche pseudo-scientifique reposant sur des affirmations d'essentialisme des traits humains
- Réponses textuelles courtes d'étudiant·es → score de QI
  - La responsabilité réside dans (i) la légitimation du score de QI en tant que quantité raisonnable, (ii) la prédiction du QI avec une approche ML, et (iii) la supposition que le QI peut être prédit à partir de réponses textuelles courtes.
- Prediction de la récidive avec le système COMPAS dans la justice US :
  - Récidivistes violents blancs 63% plus susceptibles d'être mal classé bas risque que les récidivistes noirs
  - Comment ? La race n'est pas une entrée, mais 137 questions comme "Was one of your parents ever sent to jail or prison?" "How many of your friends/acquaintances are taking drugs illegally?" and "How often did you get in fights while at school?"

→ Un·e scientifique doit oser demander : faut-il prédire la récidive ? Faut-il prédire la récidive pour informer les décisions de justice à l'égard des individus ?

→ Si nous voulons donner à toute personne des chances égales quelle que soit son origine sociale, existe-t-il des caractéristiques acceptables sur lesquelles fonder la prédiction de la récidive ?

- Le déterminisme social sous-tend des tâches abordées avec le ML (exemple : prédire la réussite des étudiants ? Pour éclairer les décisions de Parcoursup ?)

[1] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, "Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.," Propublica, 2016.
[2] Timnit Gebru and Emily Denton, "Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision at CVPR 2020," https://sites.google.com/view/fatecv-tutorial/home, 2020.
[3] Cathy O'Neil. Weapons of math destruction. 2016.

Essentialisme, biais, menace du stéréotype… et ML ?

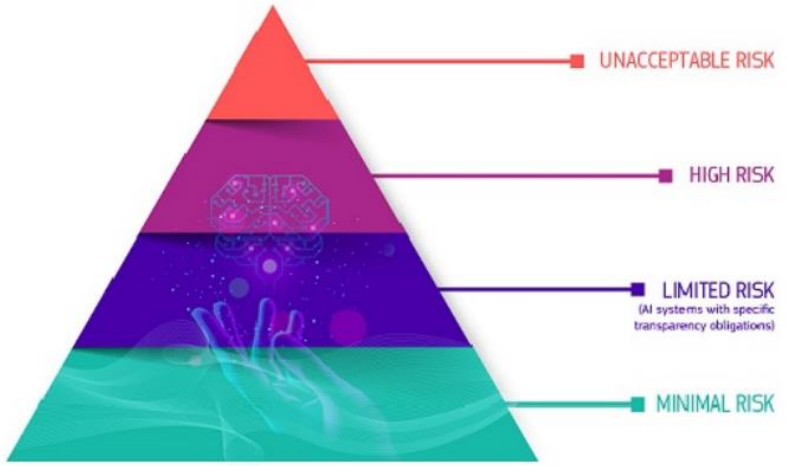
- Evaluation pré-recrutement :
  - Réponses, jeux, courte vidéo du ou de la candidate → employabilité
  - Au delà de mapping entrée-sortie questionable, un autre problème crucial dans les discriminations :

"Les évaluations cognitives ont eu des effets négatifs sur les populations minoritaires depuis leur introduction dans l'usage courant. Les critiques ont longtemps soutenu que les différences observées entre les groupes dans les résultats des tests indiquaient des défauts dans les tests eux-mêmes, et un consensus croissant s'est formé autour de l'idée que si les évaluations ont une certaine validité prédictive, elles désavantagent souvent les minorités en dépit du fait que les candidats minoritaires ont des performances professionnelles similaires à celles de leurs homologues blancs dans le monde réel. L'American Psychological Association (APA) reconnaît ces préoccupations comme des exemples de "biais prédictifs" (lorsqu'une évaluation sur- ou sous-prédit systématiquement les résultats d'un groupe particulier) [...] Les disparités dans les résultats des évaluations pour les populations minoritaires ne se limitent pas aux évaluations préalables à l'emploi. Dans la littérature sur l'éducation, l'impact négatif des évaluations sur les minorités est bien documenté. Cela a donné lieu à une série d'ouvrages qui, depuis des décennies, cherchent à mesurer et à atténuer les disparités observées"

Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," Conference on Fairness, Accountability, and Transparency (ACM FACCT), 2020.

Notre responsabilité dans la chaîne

- Catherine Tessier Référente intégrité scientifique et éthique de la recherche de l'ONERA, membre du Comité national pilote d'éthique du numérique
  - « Il ne peut pas y avoir d'algorithme éthique, mais il y a besoin d'une éthique de l'autonomie »
  - **La machine morale est un leurre nous masquant les vrais choix que les scientifiques et la société doivent pouvoir considérer honnêtement, en dehors de l'algorithme**.
- Jacobsen: emphasizes that "When assessing whether a task is solvable, we first need to ask: should it be solved? And if so, should it be solved by AI?"
- Pour ces raisons : EU AI act
  - Les systèmes d'IA utilisés dans l'administration de la justice, pour contrôler l'accès à l'éducation et à l'emploi sont désormais classés comme des systèmes à haut risque dans le dernier règlement de l'UE sur l'IA.



L'approche fondée sur le risque définit quatre niveaux de risque. Les systèmes d'IA à haut risque comprennent ceux "qui peuvent déterminer l'accès à l'éducation et le parcours professionnel d'une personne", "utilisés dans l'emploi, la gestion des travailleurs et l'accès à l'activité indépendante", "utilisés dans l'administration de la justice et les processus démocratiques".

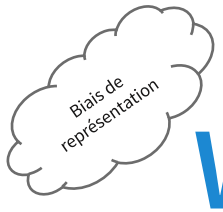
[1] Jörn-Henrik Jacobsen, Robert Geirhos, and Claudio Michaelis, "Shortcuts: Neural networks love to cheat," The Gradient, 2020.
[2] Catherine Tessier. Il n'y a pas de « décision autonome éthique » mais nécessité d'une éthique de l'« autonomie ». La revue de la société savante de l'Aéronautique et de l'Espace, Fév. 2021.

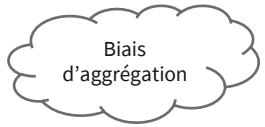
Plan de la formation

1. Apprentissage de représentation pour la reconnaissance de formes
   - Perceptron multi-couche (MLP)
   - Réseaux de neurones convolutionnels (CNN)
     - pour l'apprentissage de motifs pertinents dans les données
2. Apprentissage de représentation de mots
   - Représentations apprises par similarités de contextes
   - Représentation apprises par modèlisation du langage
     - Modèles Transformers et pré-entraînement
3. Modèles fondation : un changement de paradigme
   - Emergence de capacité imprévues
   - En langage, visio, audio... Et plus
   - Nouvelles méthodes pour adapter les modèles à des tâches spécifiques
4. Limites et enjeux
   - Environnement social et politique du design et du déploiement des systèmes de ML

Wrapping-up (1/2)

- Top-down/symbolic AI and botto-up AI/Machine learning
  - Knowledge-driven, data-driven
- Clé du Machine Learning moderne (deep learning) : l'apprentissage de représentations
- Image :
  - Neurone artificiel comme classifieur linéaire, réseau de neurones artificiels (MLP) pour apprentissage de motifs de plus en plus complexes par combinaisons de couches en couches
  - Permet de représenter les données par la présence ou l'absence de ces motifs
  - Permettre plus de motifs plus complexes à apprendre : CNN pour invariance par translation et motifs appris sous forme de filtres
- Texte :
  - Représentation numérique unique d'un mot (embedding) en alignant les représentations des mots voisins (Word2Vec)
  - Transformers: représentation du mot dépend de son contexte, obtenue par recombinaisons des autres mots et le motif (la façon de recombiner) dépend lui-même des voisins, et dépendance longue distance
- Puissance de l'apprentissage auto-supervisé sur grand corpus
- Emergence de capacité d'adaptation à de nouvelles tâches
- Modèles fondation : modèles pré-entraînés pour un domaine, adaptables à d'autres tâches (avec prompt engineering et prompt tuning en plus de fine-tuning)






Wrapping-up (2/2)

- Image, texte, IA générative, systèmes sociaux-techniques avec risques et enjeux (de recherche, de droit – EU AI Act, économiques, pédagogiques, etc.)
- Les systèmes de ML déployés dans des scénarios à fort enjeu et ayant un impact direct sur des vies humaines ne fonctionnent souvent pas. Leurs échecs pénalisent de manière disproportionnée les personnes appartenant à des groupes socialement défavorisés.
- Nous ne pouvons pas espérer que les jeux de données soient exempts de biais, nous devons donc reconnaître les problèmes sociaux et structurels omniprésents dans les données et les prendre en compte dans nos pratiques de ML :
  - Réflechir à la correspondence entrée-sortie envisage au moment de définir une tâche
  - Expliciter toute implication humaine dans le processus de creation de jeu de données et le documenter
  - Etre consciente de la faiblesse des benchmarks d'évaluation (inclure des analyses déaggrégées sur les groups, coût environnemental, etc.), et ne pas survendre les capacité de généralisation d'un modèle.
  - Re-visiter les hypothèses techniques pour concevoir des systèmes performants alignés sur nos valeurs.
- Reconnaître que pour les problèmes complexes du monde réel, des solutions pluridisciplinaires sont nécessaires et que nous ne devons pas ignorer les limites de nos approches techniques. Cela peut être enseigné dans les cours de ML et d'éthique afin de savoir quand nous, en ML, ne pouvons pas savoir.

→ **Beaucoup d'espaces d'agentivité à distinguer dans le discours dominant et à ré-investir dans nos et vos travaux !**

[1] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," *Patterns*, Nov. 2021
[2] Timnit Gebru and Emily Denton, "Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision at CVPR 2020," https://sites.google.com/view/fatecv-tutorial/home, 2020.

Ressources

Posts Medium 3IA – EFELIA :

- Is data fixable? On the need of socially-informed practices in ML research and education - Part 1: Deployment failures and approaches to data, Medium, 3IA Côte d'Azur, Jan. 2023.
- Is data fixable? On the need of socially-informed practices in ML research and education - Part 2: A more holistic perspective on data creation and expectations, Medium, 3IA Côte d'Azur, Jan. 2023.
- Is data fixable? On the need of socially-informed practices in ML research and education - Part 3: AI ethics and our ML education practices, Medium, 3IA Côte d'Azur, Jan. 2023.

EFELIA Côte d’Azur

Merci de votre attention

https://univ-cotedazur.fr/efelia-cote-dazur
Twitter : 3IAcotedazur
LinkedIn : 3IA Côte d’Azur

UNIVERSITÉ CÔTE D’AZUR | INSTITUT INTERDISCIPLINAIRE D’INTELLIGENCE ARTIFICIELLE 3IA CÔTE D’AZUR