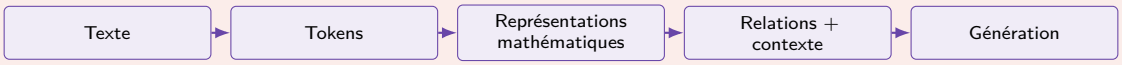
Comment une IA comprend une requête ?

IA générative ▪ Concepts fondamentaux ▪ Niveau débutant

L'IDÉE ESSENTIELLE

Une IA moderne ne « comprend » pas une requête exactement comme un humain. Elle transforme progressivement le texte en **représentations mathématiques**, calcule des relations entre les éléments, utilise le contexte puis génère une réponse.



1. LE TEXTE DEVIENT DES TOKENS

L'IA ne manipule pas directement les mots comme nous les percevons. Le texte est d'abord découpé en **tokens** : morceaux de mots, mots ou signes.

Chaque token est ensuite associé à une représentation numérique.

2. DES REPRÉSENTATIONS MATHÉMATIQUES

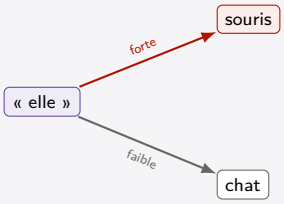
On peut représenter conceptuellement :

$$\begin{aligned} \texttt{feuilles} &\rightarrow [0.12, -0.43, 0.87, \ldots] \\ \texttt{arbres} &\rightarrow [0.15, -0.39, 0.82, \ldots] \\ \texttt{automne} &\rightarrow [0.68, 0.19, -0.11, \ldots] \end{aligned}$$

Ces nombres ne sont pas des définitions. Ils constituent une **représentation vectorielle** permettant au modèle de calculer des proximités et des relations.

3. L'ATTENTION RELIE LES ÉLÉMENTS

Le mécanisme d'**attention** permet au modèle de pondérer les relations entre les différents éléments du contexte.



Le modèle peut ainsi prendre en compte les relations entre les éléments plutôt que de traiter chaque mot isolément.

4. CONTEXTE ET INTENTION

Le même concept peut prendre une signification différente selon la requête.

« Paris »
Concept isolé

« Quelle est la population de Paris ? »
Paris + population + recherche d'information

Le contexte conversationnel joue également un rôle :

`Comment fonctionne XPath ?` → `Et avec XML ?` → **XPath + XML**

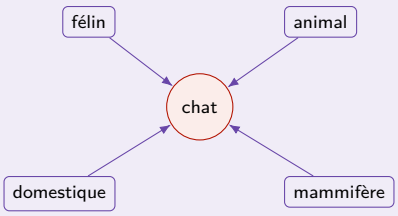
La seconde question peut donc être interprétée grâce à la première.

5. LES CONCEPTS SONT DISTRIBUÉS

Il n'existe généralement pas une simple case mémoire :

```
CONCEPT = CHAT
```

Les concepts sont représentés à travers de nombreuses dimensions et relations apprises pendant l'entraînement.



C'est notamment ce qui permet au modèle de rapprocher *« mon petit félin domestique »* de *« mon chat »*.

6. DE LA REPRÉSENTATION À LA RÉPONSE

Après traitement de la requête et du contexte, le modèle génère progressivement la réponse. À chaque étape, il choisit des tokens en fonction des représentations et des probabilités calculées.

Requête → représentation → contexte → probabilités → tokens successifs → réponse

LE MÉMO

Texte → tokens → vecteurs → relations → contexte → génération

À retenir : une IA ne possède pas nécessairement un « concept » comme un humain conscient de ce qu'il connaît. Elle manipule des représentations mathématiques distribuées des régularités et relations apprises dans les données.

Point de vigilance : cette représentation est volontairement simplifiée. Les modèles réels utilisent des architectures et des mécanismes beaucoup plus complexes.