

# GUIDE À L'INTENTION DES DIRIGEANTS D'ENTREPRISE

*À lire sans modération par tout utilisateur de l'IA*

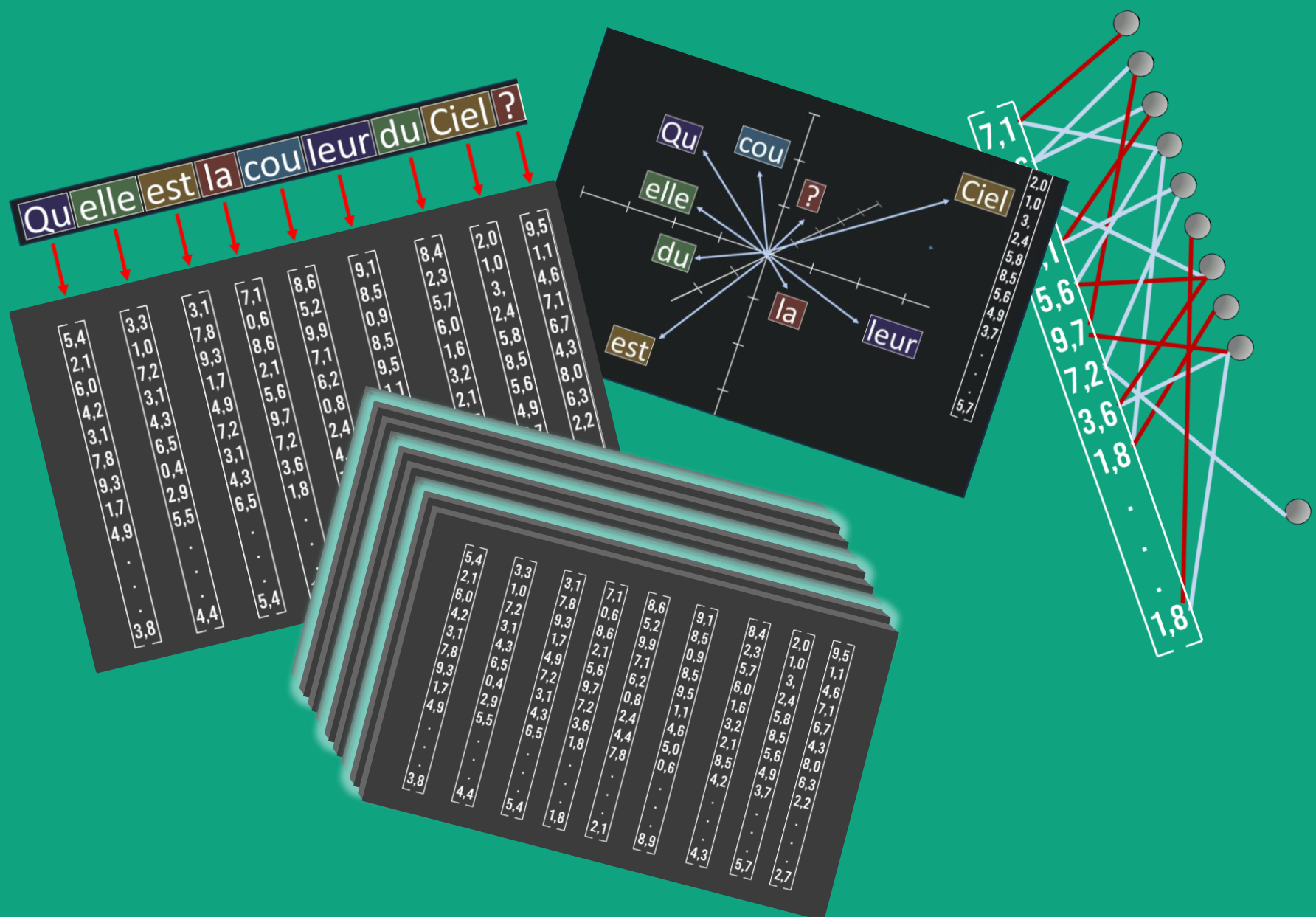
## INTELLIGENCE ARTIFICIELLE

Guide complet du fonctionnement l'IA générative

expliquée par



**CLOUD BUSINESS**  
School



## L'IA va plus changer le monde en 4 ans, que l'arrivée conjuguée du PC, du Web et des mobiles en 40.

L'IA est dans toutes les bouches, et presque tout le monde a son avis. Certains pour, certains contre, certains ne jurent que par elle, d'autres y voient tous les dangers pour les ados, les démocraties et l'humanité.

Pourtant, très peu savent comment fonctionne réellement l'IA. **Sans doute le sujet le moins compris** des sujets dont tout le monde parle le plus.

**Cette ignorance ouvre la porte à beaucoup de fantasmes**, à des **craintes** infondées, à une **surestimation** de ses capacités, et à une **mauvaise utilisation** de l'outil.

Signe qu'il est sans doute temps de prendre 30 min pour comprendre ce qu'est réellement l'IA, cette technologie qui n'a rien de commun avec l'intelligence contrairement au nom dont on la qualifie.



## L'IA est le sujet sur lequel se disent le plus d'absurdités sur le web :

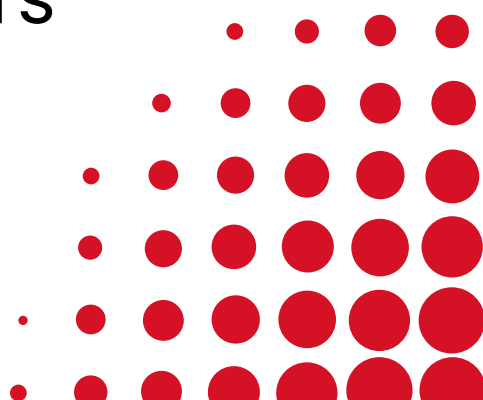
- "L'IA apprend de vos interactions."
- "Il faut attribuer à l'IA un rôle de spécialiste du sujet sur lequel on l'interroge pour améliorer ses réponses."
- "Il faut tutoyer l'IA, c'est bien."
- "Il faut être poli avec l'IA."
- Etc.

Ceux qui colportent ces inepties, et se font parfois payer pour le faire, le font parce qu'ils n'ont aucune idée du fonctionnement de l'IA.

La plupart du temps, ils ne font que copier-coller ce qu'ils ont lu ailleurs sans conscience de l'absurdité du propos. Parce qu'ils n'ont pas fait le minimum pour se former et comprendre l'IA, trop pressés d'en tirer profit.

Et ceux qui les croient le font pour la même raison.

Nous allons ici expliquer en détail le fonctionnement de l'IA, et les implications pour permettre à vos collaborateurs d'en tirer le meilleur parti.



# FAKE NEWS

# CHATGPT

## Fonctionnement de ChatGPT

1. ChatGPT apprend seul des interactions avec les utilisateurs
2. ChatGPT et les IA utilisent des réseaux de neurones
3. Les résultats de ChatGPT et des IA sont fiables
4. ChatGPT ne fournit que des réponses banales
5. OpenAI aurait bridé ChatGPT avec des limites hackables
6. ChatGPT n'est pas adapté au Français
7. Les tokens sont des syllabes ou mots
8. Les tokens, ça ne sert à rien
9. L'IA sait déterminer le process pour répondre juste à nos questions
10. Idées reçues sur le fonctionnement ChatGPT : 1. Tokenization, 2. embedding, 3. transformer
11. ChatGPT c'est de l'intelligence artificielle
12. OpenAI a inventé la technologie de ChatGPT
13. ChatGPT est trop gentil, il va toujours dans votre sens (idée de business = chaussettes pour chien)
14. ChatGPT est mieux (ou moins bien) que Claude/Mistral, etc.



**La bêtise,  
c'est de la paresse.**  
Jacques Brel

## Prompt engineering

15. Il faut attribuer un rôle à ChatGPT
16. Le prompt ultime : le prompt à faire le prompt idéal
17. Utiliser des formules de politesse est préférable
18. On peut tutoyer ChatGPT
19. Il faut donner à ChatGPT le temps de respirer

## Risques de ChatGPT et des IA

20. ChatGPT pose des problèmes de sécurité des données
21. Le risque que l'IA remplace l'humain
22. L'IA va amenuiser la capacité des jeunes à réfléchir

## SEO et ChatGPT

23. ChatGPT peut créer du contenu optimisé SEO
24. ChatGPT est nul pour le SEO
25. Le SEO ne fonctionne plus à cause de l'IA
26. L'IA va tuer Google
27. SearchGPT d'OpenAI va détrôner Google

## Idées reçues sur les fonctionnalités de ChatGPT

28. Ce que sont vraiment les Custom Instructions
29. On peut entraîner les GPT sur une tâche spécifique
30. Avec sa fonction mémoire, ChatGPT apprend



# Sommaire

## 1

### 6 Introduction

- 9 Principe Général
- 10 Différents modèles

## 2

### 12 Fonctionnement IA

- 15 Tokenization
- 17 Vectorisation
- 21 Mécanisme d'attention
- 25 Perceptron multicouche
- 28 Répétition
- 32 Calcul de probabilité
- 37 Chainage de probabilités
- 42 Interface IA

## 3

### 51 Implication sur l'IA

- 52 Mauvaises pratiques
- 61 Failles de l'IA

## 4

### 65 Prompt Engineering

- 66 Les 7 couches d'infos
- 74 Les erreurs à éviter
- 78 Les marqueurs
- 81 Les paramètres

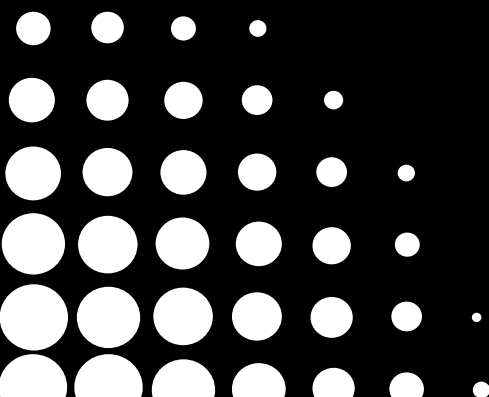
## 5

### 82 Conclusion



# 1

# Introduction



Quand on parle d'IA depuis 2 ans, on parle en réalité d'une sous-sous-sous-famille de l'IA qui est l'IA générative.

- **IA (Intelligence Artificielle)** : Le (pour l'instant) fantasme d'un système capable de reproduire des tâches effectuées par l'intelligence humaine, comme la perception, la logique ou la prise de décision.
- **Machine Learning (Apprentissage Automatique)** : Sous-domaine de l'IA où les machines apprennent à partir de données pour effectuer des prédictions ou des décisions sans être explicitement programmées.
- **Deep Learning (Apprentissage Profond)** : Branche du Machine Learning traitant de grandes quantités de données complexes, comme les images, les vidéos ou le langage naturel.
- **IA Generative IA** : Type d'IA qui génère du contenu original (texte, images, audio, etc.) en s'appuyant sur des modèles entraînés à reproduire des structures issues de données apprises.



L'IA générative a été révélée au grand public en novembre 2022 par OpenAI. C'est la technologie utilisée par Gemini (Google), LLaMA (Meta anciennement Facebook), Claude (Anthropic), Mistral (le petit français de la bande), ChatGPT (OpenAI), etc.

C'est une technologie basée sur les **transformers**, plus connus sous le terme de GPT (Generative Pretrained Transformer).

**G**enerated : signifie qu'il s'agit de logiciels qui génèrent du nouveau texte, ou sons, ou images, ou code...

**P**retrained : indique que le modèle a été soumis à un processus d'apprentissage en l'exposant à de grandes quantités de textes.

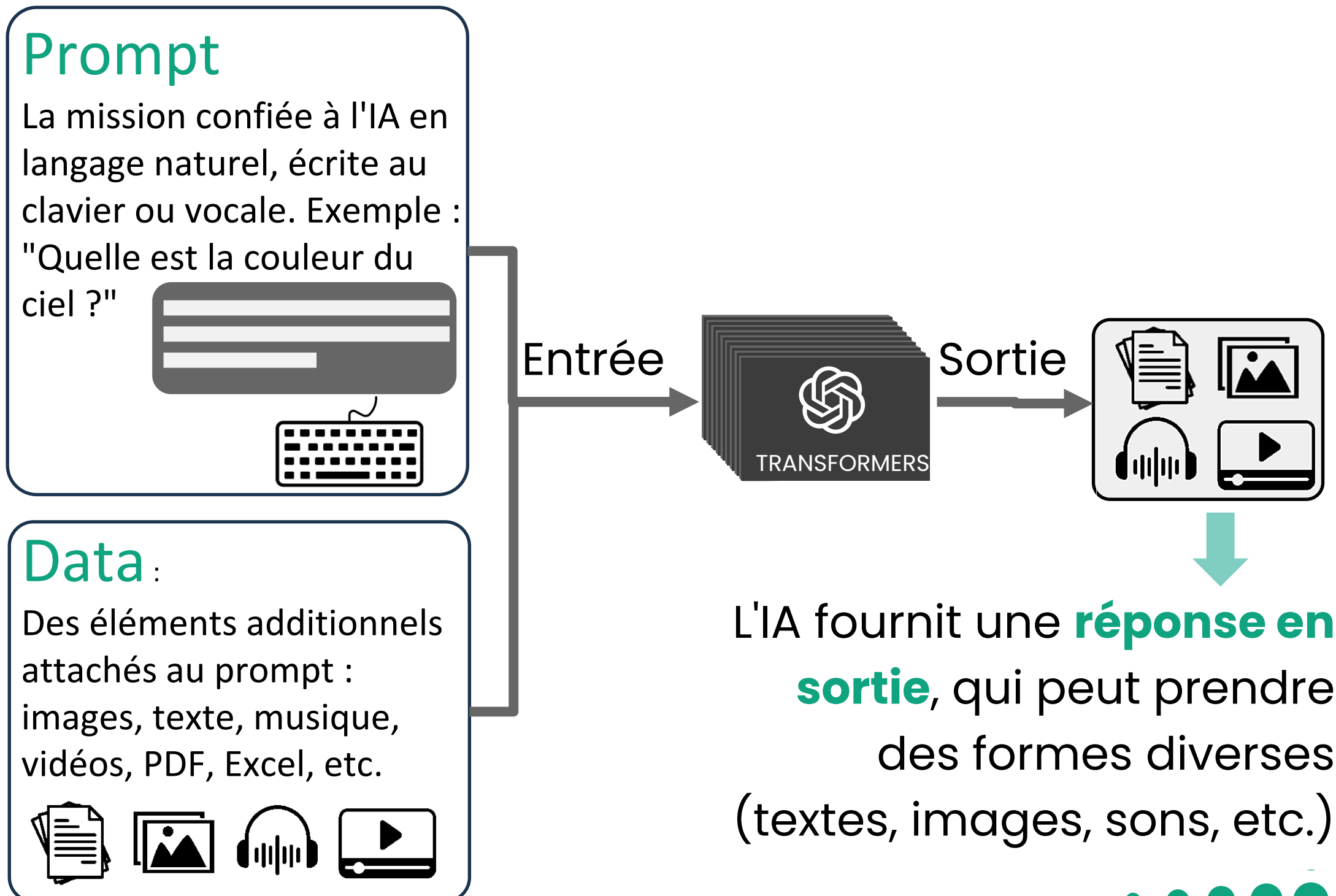
**T**ransformers, qualifie le processus qui génère du texte, des sons, des images.

**Les Transformers sont l'invention centrale derrière l'essor actuel de l'IA Générative.**

# A. Principe général de l'IA Gen

L'utilisateur fournit en entrée à l'IA une mission (une tâche à faire, ou une question) : le **prompt**.

Il peut attacher au prompt **des données** (un document à traduire, des consignes à respecter, un document à consulter...)



## B. Différents modèles d'IA.

- **Voice-to-text** : L'IA transforme des données audio en une transcription écrite.



- **Text to voice** : l'IA fait exactement l'inverse, elle transforme du texte en audio.



- **Text to Image** : l'IA crée une image selon les instructions écrites dans le prompt.



- **Machine translation** : Le 1<sup>er</sup> transformer introduit par Google en 2017 permettait la traduction d'un texte d'une langue à une autre.

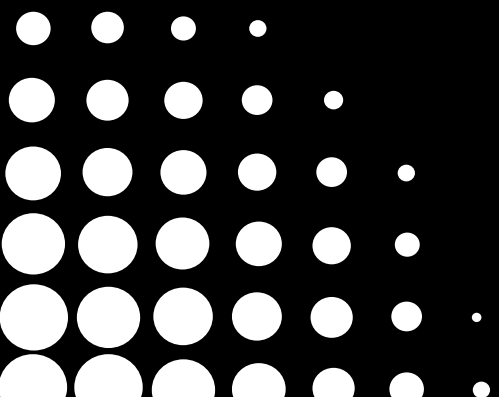


Nous allons dans ce document expliquer le fonctionnement de l'IA générative pour le cas le plus général : le modèle sous-jacent à Gemini (Google), LLaMA (Meta ), Claude (Anthropic), Mistral (le français), ChatGPT (OpenAI), et 80% des applications IA basées sur cette technologie comme Articul8 (spin-up) d'Intel, et des centaines d'autres.

C'est-à-dire des modèles entraînés à transformer un morceau de texte, accompagné ou pas de fichiers (images, sons, PDF, DOC, etc.) et à produire une prédiction de ce qui vient ensuite.

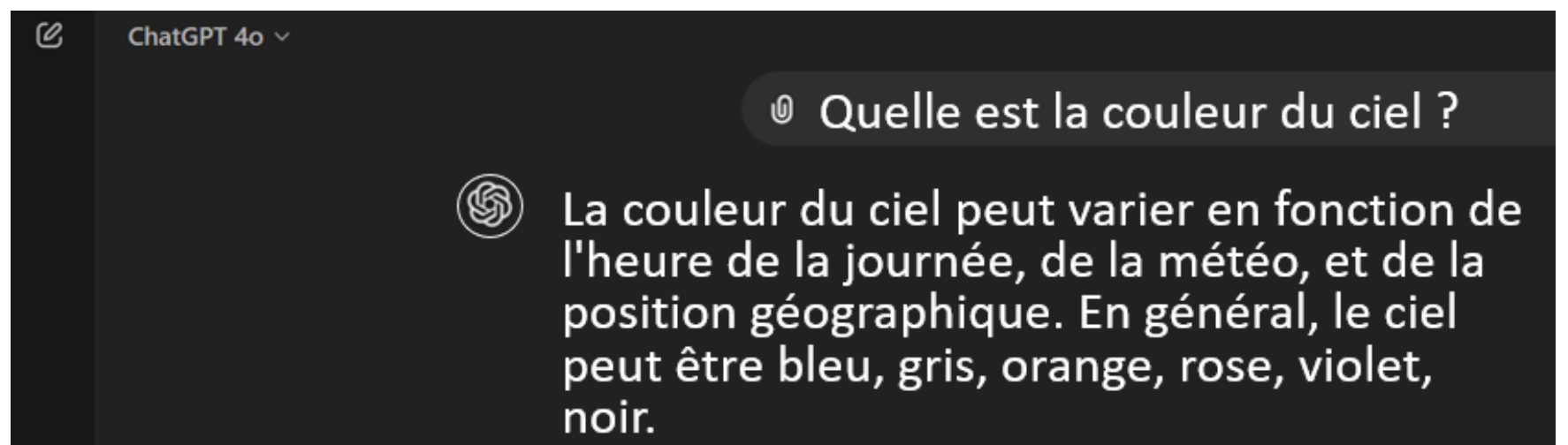


# 2 **Fonctionnement de l'IA**



## Fonctionnement de l'IA

Nous allons regarder sous le capot de l'IA, et voir le fonctionnement qui rend possible de passer par exemple de la question posée ici "quelle est la couleur du ciel" à la réponse de l'IA.



Si vous êtes utilisateur de l'IA, vous avez vu "la magie opérer" : on pose une question, et l'IA *donne l'impression et l'illusion de comprendre* notre question, et produit une réponse adéquate en quelques secondes.

Mais peu d'utilisateurs savent réellement comment la chose est possible. Et vous allez voir que cela n'a rien de magique, ni rien à voir avec aucune forme d'intelligence.



# L'IA : un process en 7 étapes

Plus exactement, un process que nous allons résumer en 7 étapes à des fins pédagogiques :

- 1. Tokenization**
- 2. Vectorisation**
- 3. Mécanisme d'attention**
- 4. Perceptron multicouche**
- 5. Répétition**
- 6. Calcul de probabilité**
- 7. Chainage de probabilités**

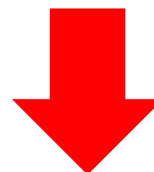


# 1. Tokenization

Si vous pensiez que l'IA comprend ce que vous dites... oubliez cette idée absurde.

L'IA ne comprend pas votre question : la 1<sup>ère</sup> chose que fait l'IA quand vous lui donnez un prompt, c'est le décomposer en tokens. Un token est un mot ou une fraction de mot, composé en moyenne de 3 ou 4 lettres.

Quelle est la couleur du ciel ?



Quelle est la couleur du ciel ?

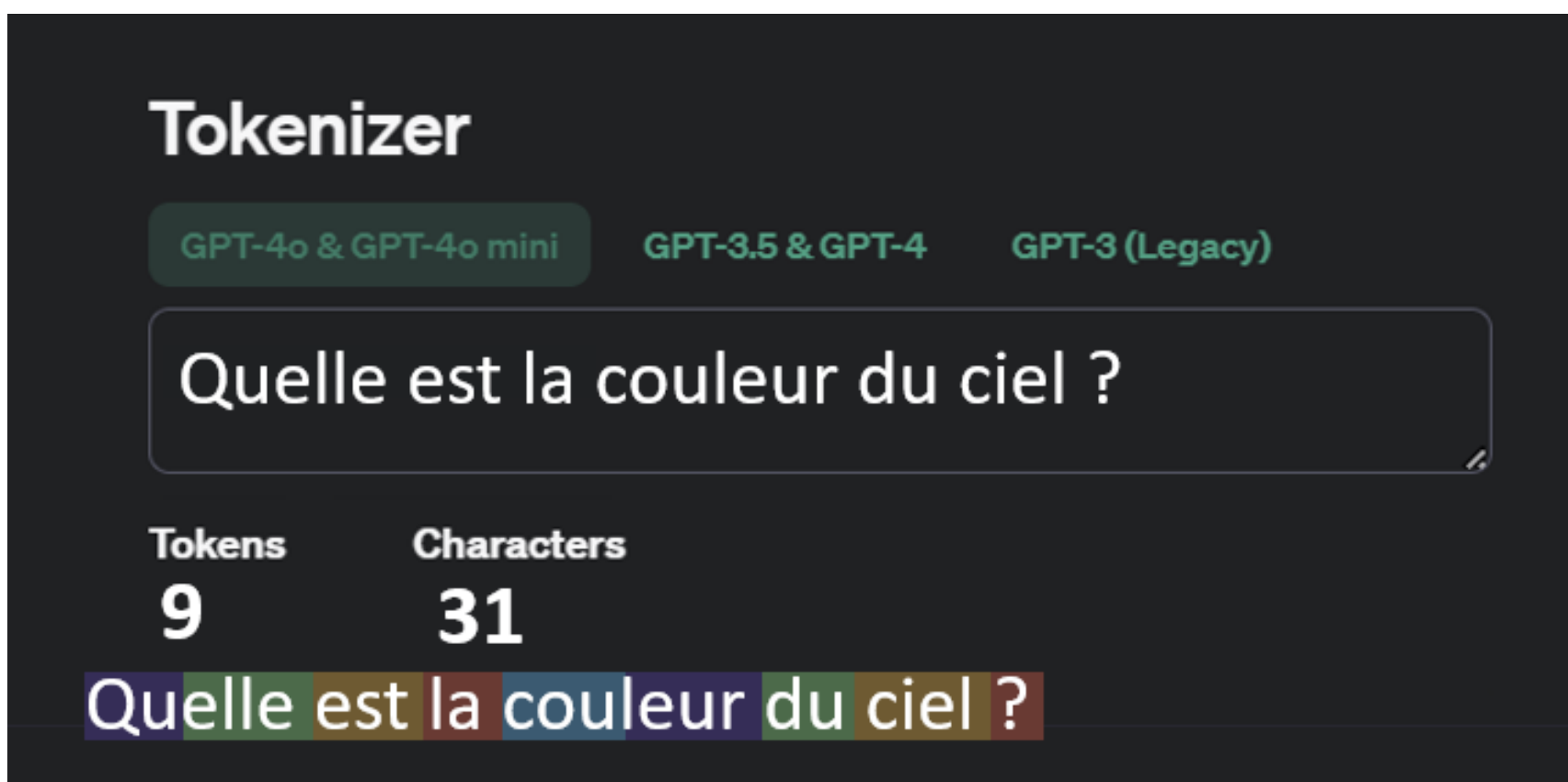
Une langue comme le français est composée de 600 000 mots (plus d'un million si l'on rajoute les jargons). Mais ces mots ne sont composés que d'environ 50 000 tokens, communs à toutes les langues.



L'IA ne parle donc pas français, allemand ou anglais : elle parle "le token" : 50 000 tokens sont bien plus rapides à opérer qu'un million de mots, et ils lui permettent de reconstituer non seulement le million de mots de la langue française, mais aussi toutes les langues du monde.

Vous pouvez tester la manière dont ChatGPT réalise cette décomposition, dans le tokenizer mis à disposition gratuitement par OpenAI ici :

<https://platform.openai.com/tokenizer>



**Tokenizer**

GPT-4o & GPT-4o mini    GPT-3.5 & GPT-4    GPT-3 (Legacy)

Quelle est la couleur du ciel ?

Tokens	Characters
9	31

Quelle est la couleur du ciel ?

## 2. Vectorisation

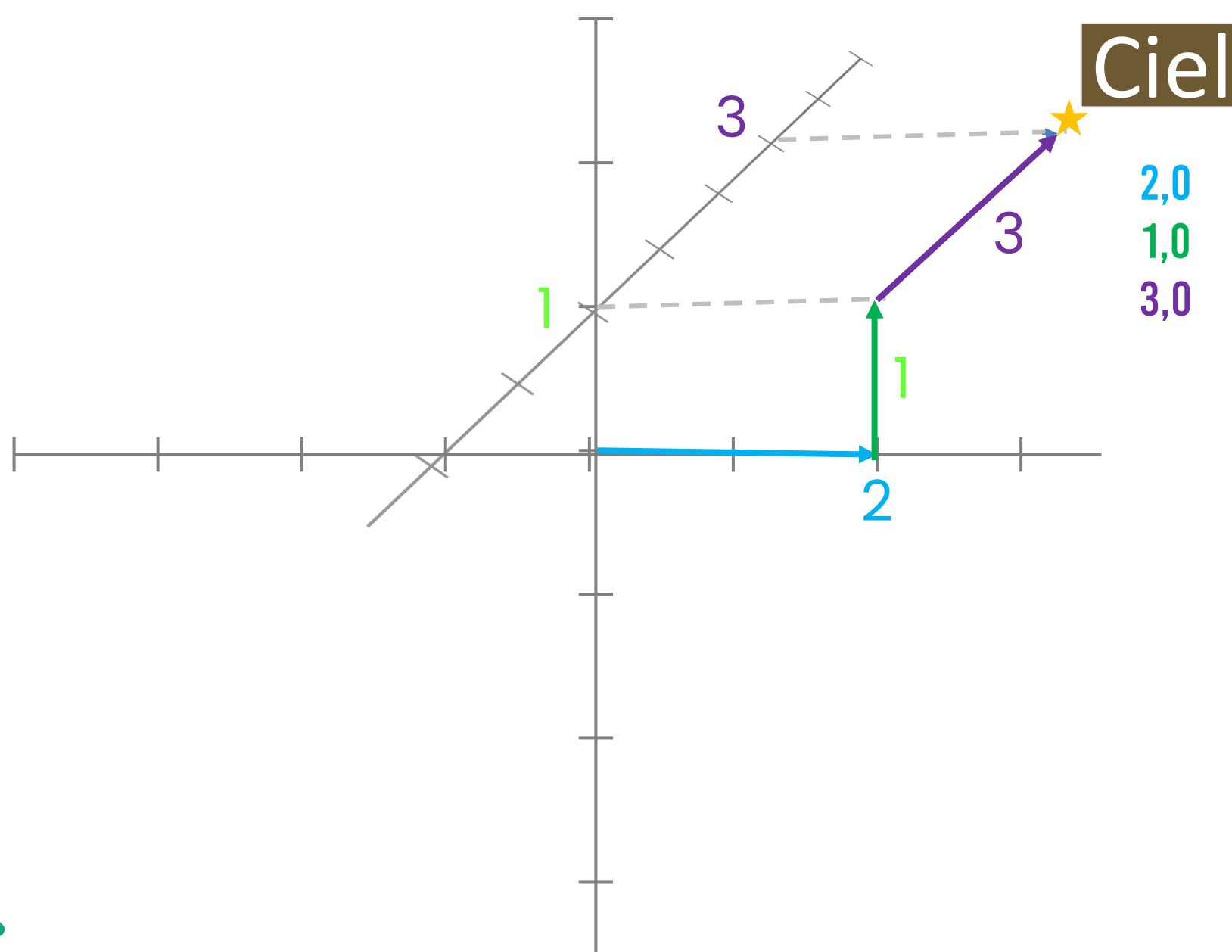
Chacun de ces tokens a été préalablement associé à un vecteur dans une immense matrice appelée "matrice d'intégration" (embedding matrix). Ces 50 000 vecteurs se présentent sous forme d'une liste de nombres, qui sont les coordonnées de chaque token dans l'espace.

Quelle est la couleur du ciel ? ...

↓	↓	↓	↓	↓	↓	↓	↓	↓	???
5,4	3,3	3,1	7,1	8,6	9,1	8,4	2,0	9,5	
2,1	1,0	7,8	0,6	5,2	8,5	2,3	1,0	1,1	
6,0	7,2	9,3	8,6	9,9	0,9	5,7	3,	4,6	
4,2	3,1	1,7	2,1	7,1	8,5	6,0	2,4	7,1	
3,1	4,3	4,9	5,6	6,2	9,5	1,6	5,8	6,7	
7,8	6,5	7,2	9,7	0,8	1,1	3,2	8,5	4,3	
9,3	0,4	3,1	7,2	2,4	4,6	2,1	5,6	8,0	
1,7	2,9	4,3	3,6	4,4	5,0	8,5	4,9	6,3	
4,9	5,5	6,5	1,8	7,8	0,6	4,2	3,7	2,2	
.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	
3,8	4,4	5,4	1,8	2,1	8,9	4,3	5,7	2,7	

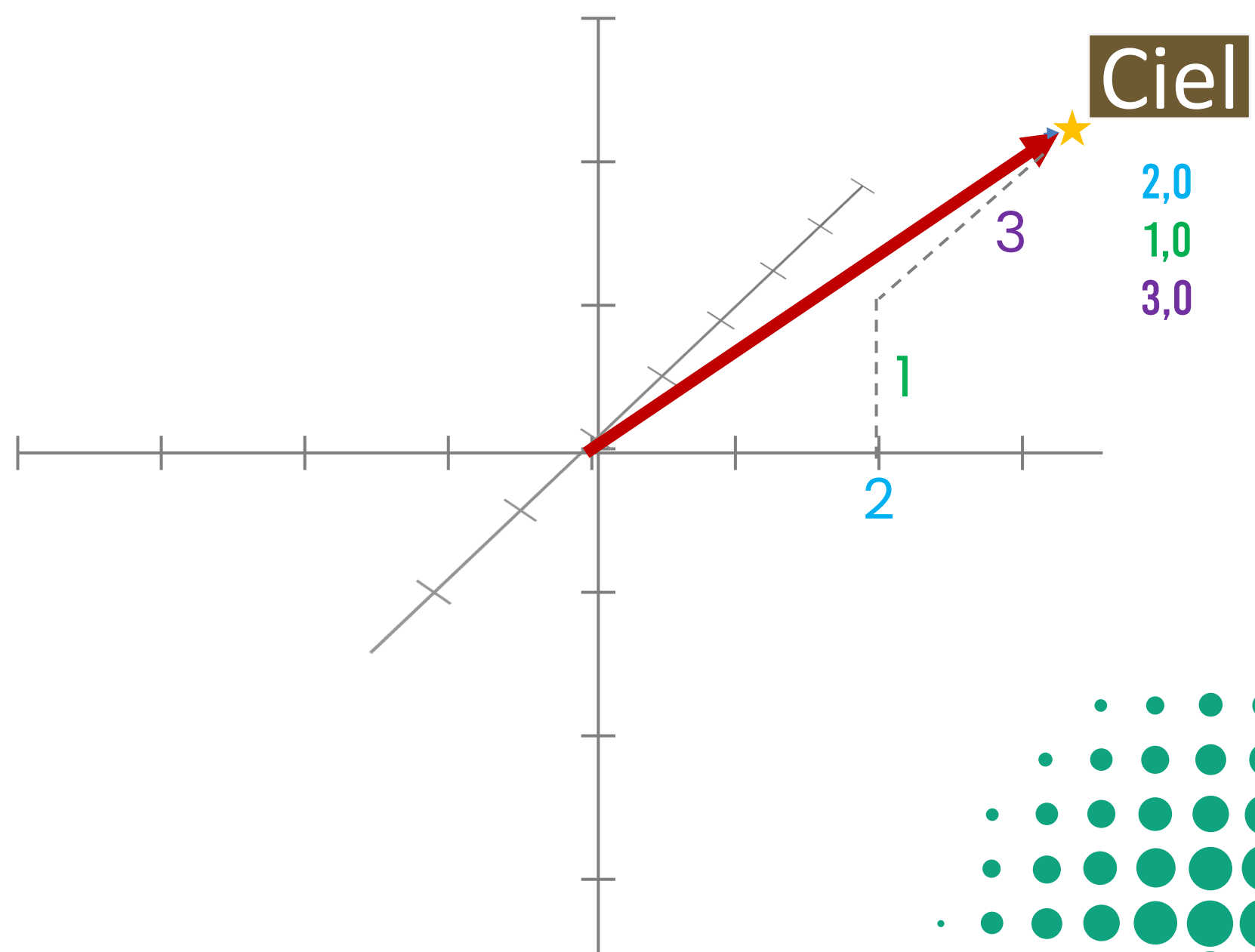
Pour faciliter la compréhension, imaginons que ces nombres soient les coordonnées de chaque token dans un **espace à 3 dimensions**. Ici le token "ciel".

Ses coordonnées (2, 1, 3) définissent sa position dans l'espace : 2 sur l'axe horizontal, 1 sur l'axe vertical, 3 sur l'axe profondeur.



La position du token "ciel" dans l'espace peut alors être définie par un **vecteur** : rappelez-vous vos cours de mathématiques de 3<sup>e</sup>, un vecteur c'est 3 valeurs : un point, une longueur (appelée "norme", et une direction.

Ici le vecteur rouge a son point d'origine au croisement des axes, sa longueur et sa direction situent parfaitement la position du token "ciel" dans l'espace.

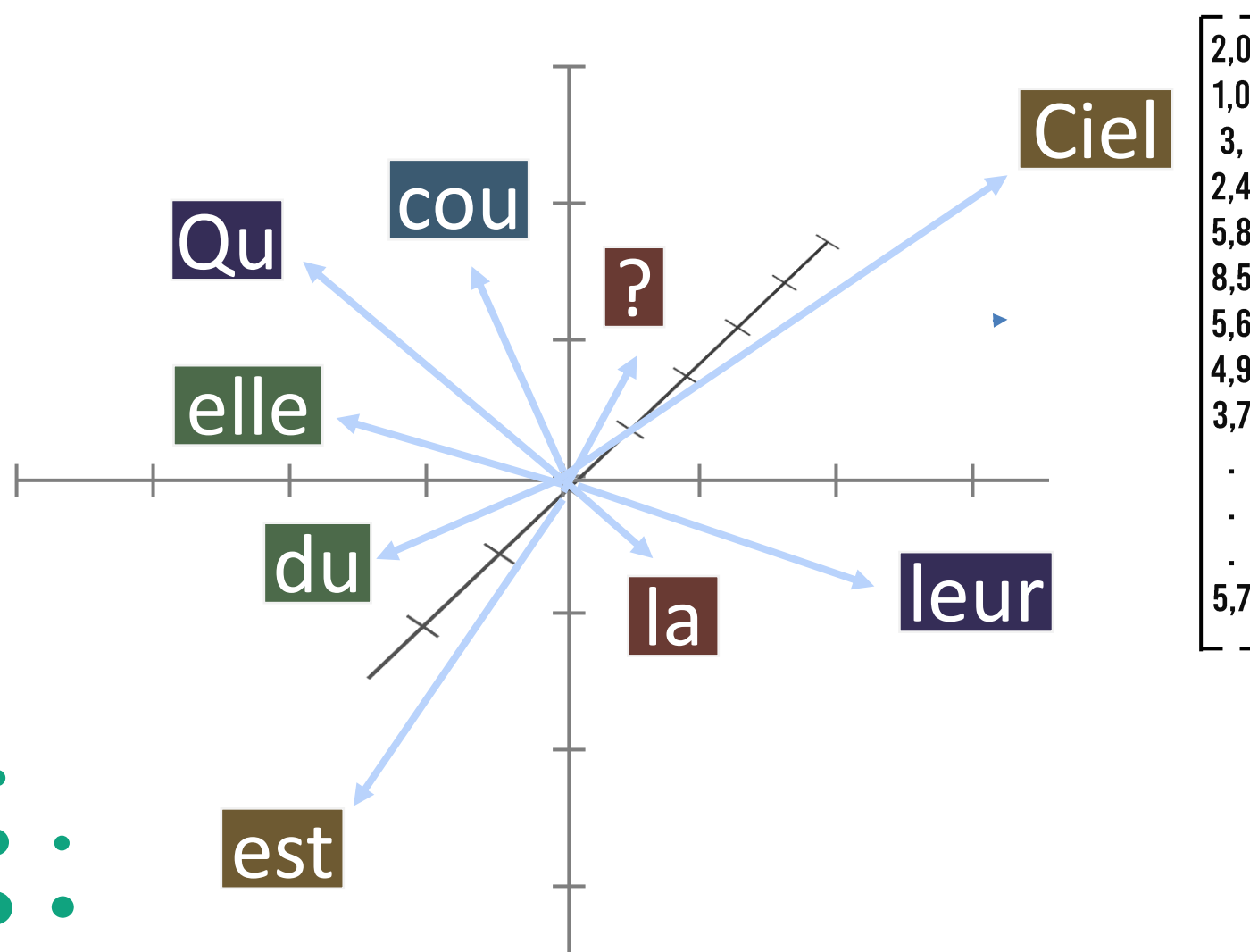


La matrice d'intégration définit donc la position dans l'espace de chacun des 50 000 tokens par ses coordonnées vectorielles. Les tokens ayant des significations similaires sont proches les uns des autres dans l'espace.

**Sauf que ce n'est pas un espace à 3 dimensions :**

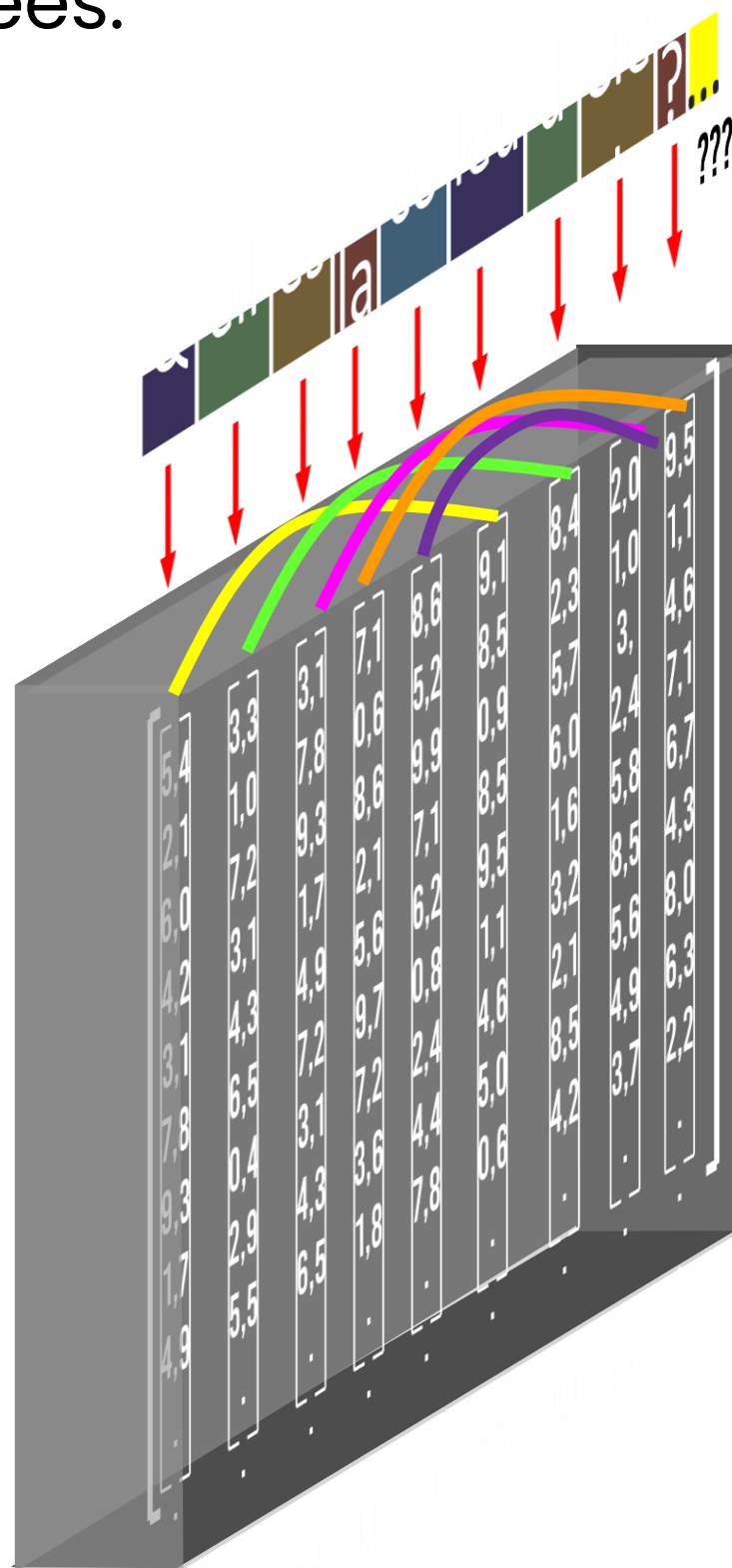
ChatGPT-3 était basé sur un espace à 12 888 dimensions. On estime (en l'absence de publication officielle d'OpenAI) que celui de ChatGPT-4 a été porté à 16 384 dimensions... Plus que votre cerveau ne peut imaginer.

Chaque token a donc **16 384 coordonnées.**



### 3. Mécanisme d'attention

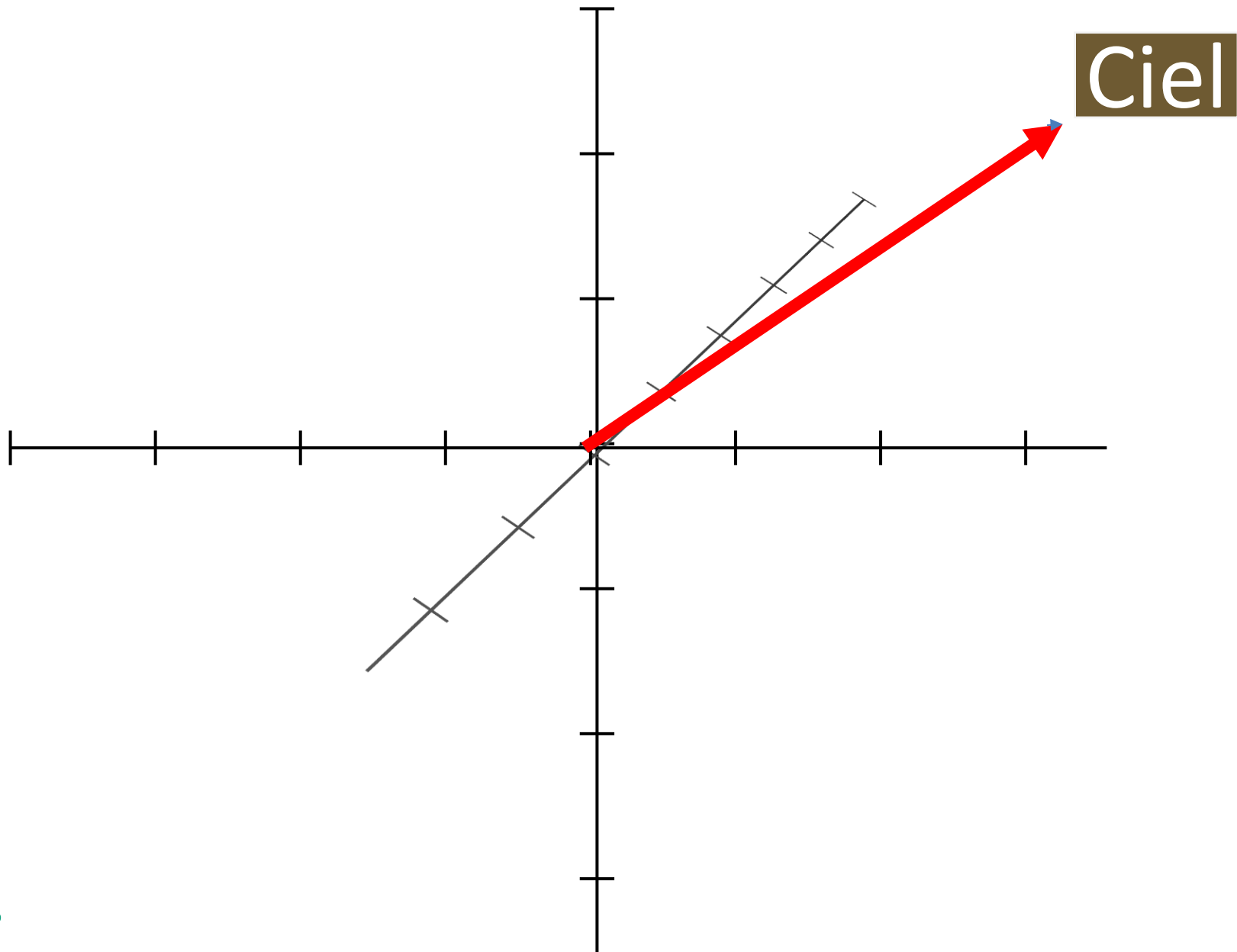
Les vecteurs passent ensuite par une opération connue sous le nom de bloc d'attention, qui permet aux vecteurs de communiquer entre eux et de transmettre des informations pour mettre à jour leurs coordonnées.



# Pourquoi mettre à jour les coordonnées des vecteurs ?

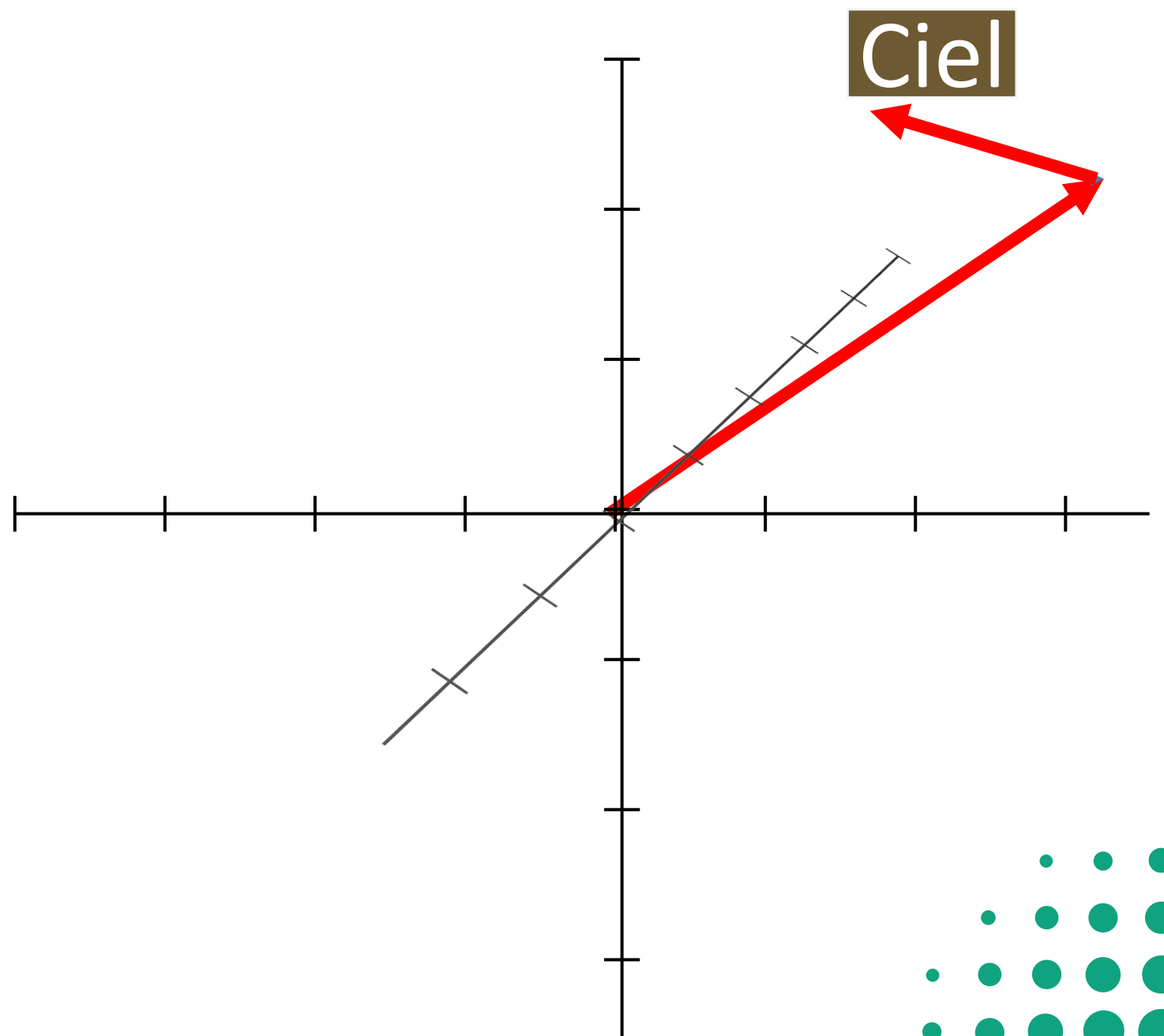
Prenons l'exemple de la position dans l'espace du mot "ciel" dans la question :

Quelle est la couleur du Ciel ? ...



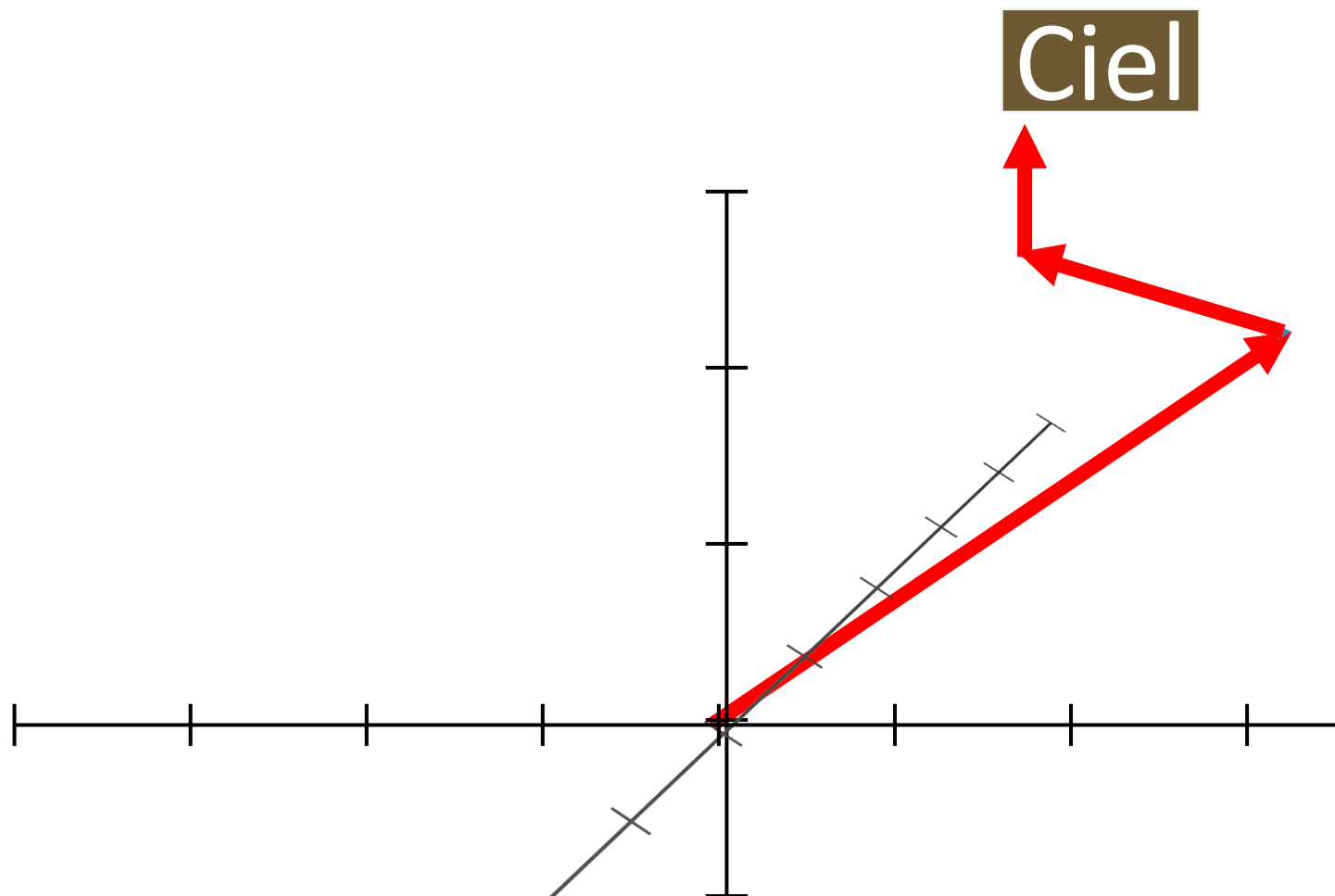
Le sens du mot "ciel" dans la question initiale, n'est pas exactement le même si la question porte sur le ciel de la France. La position du token ciel dans la matrice d'intégration était une position "standard", ne rendant pas compte de cette nuance contextuelle.

Quelle est la couleur du Ciel en France? ...



Et l'écart se creuse avec la position de la matrice d'intégration, si le terme ciel est inséré dans une question sur le ciel de France en hiver...

Quelle est la couleur du Ciel en France l'hiver ? ...



Le bloc d'attention détermine quels mots du contexte sont pertinents pour mettre à jour la signification de quels autres mots, et comment exactement ces significations doivent être mises à jour.

## 4. Perceptron multicouche

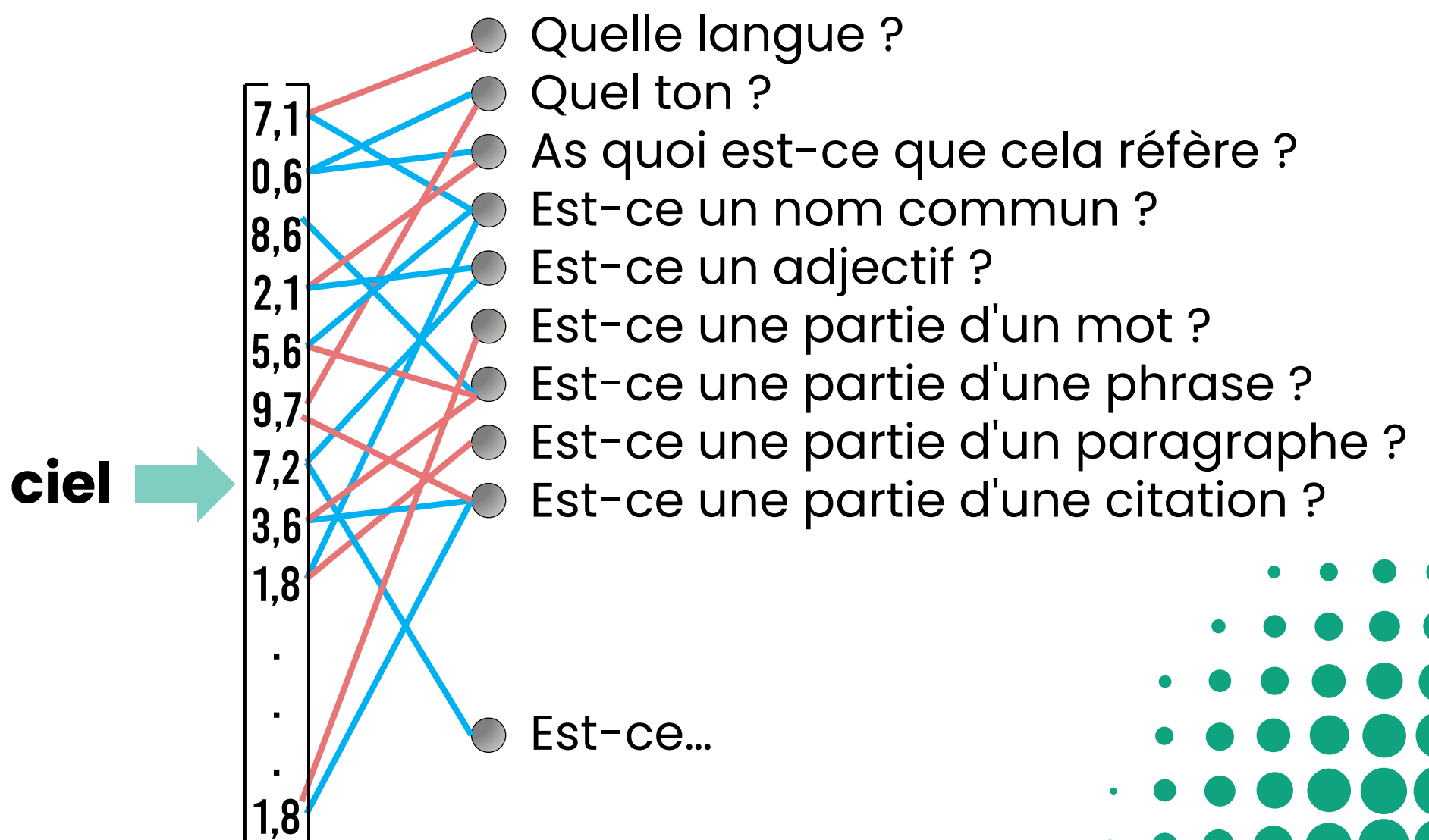
$$\hat{y} = \varphi_{\text{out}} \left( \sum_i w_i^{(2)} h_i^{(2)} + b^{(2)} \right)$$

$$\forall i, h_i^{(2)} = \varphi \left( \sum_j w_{ij}^{(1)} h_j^{(1)} + b_i^{(1)} \right)$$

$$\forall i, h_i^{(1)} = \varphi \left( \sum_j w_{ij}^{(0)} x_j + b_i^{(0)} \right)$$

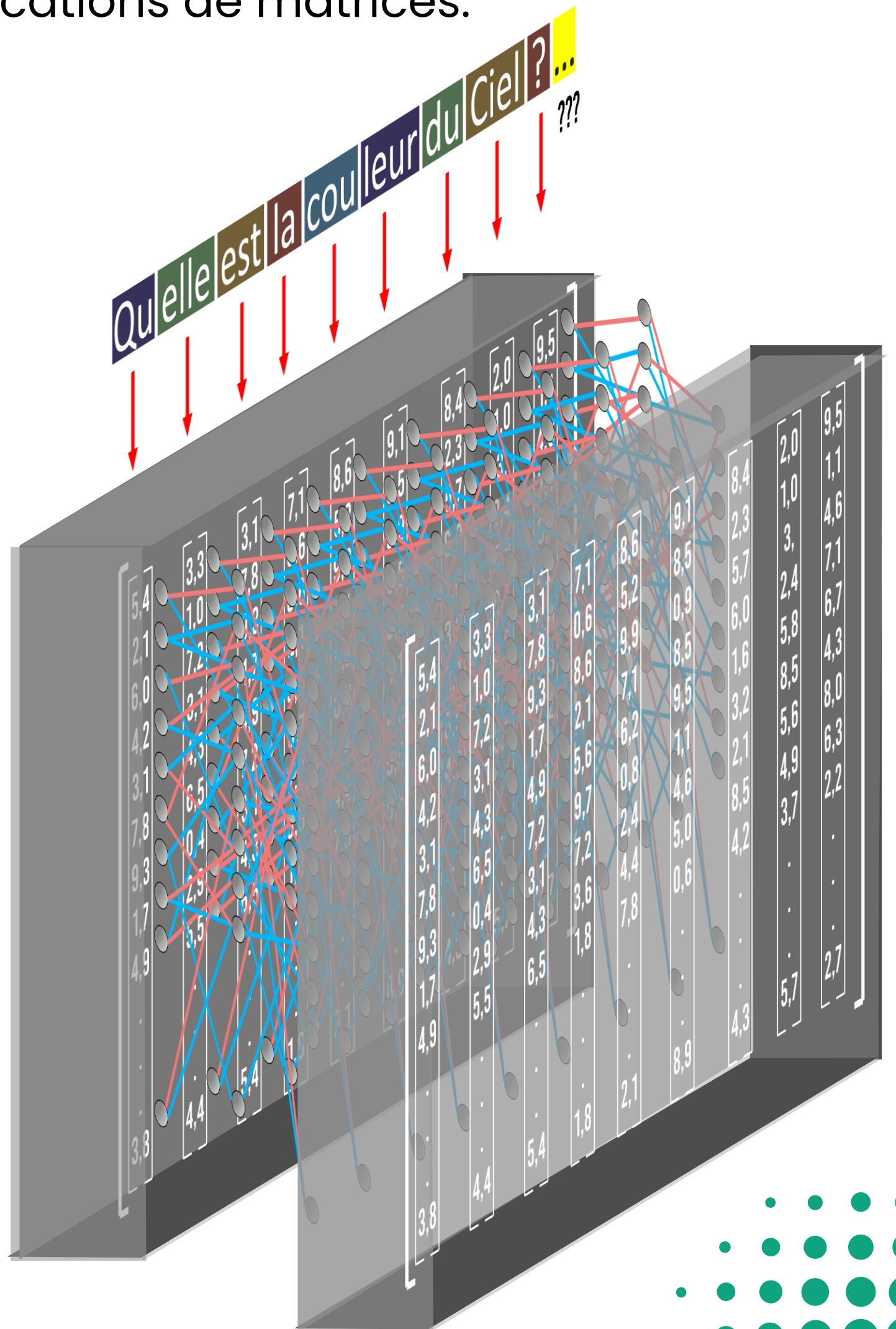
Cette étape (en anglais : **Multilayer Perceptron** ou **MLP**) est une étape de recalcul des coordonnées vectorielles, modélisée par ce genre d'équations mathématiques.

En clair, elle consiste à poser une longue liste de questions, indépendamment à chaque vecteur, pour préciser son sens dans le contexte. Puis à mettre à jour les coordonnées vectorielles du token en fonction des réponses à ces questions.





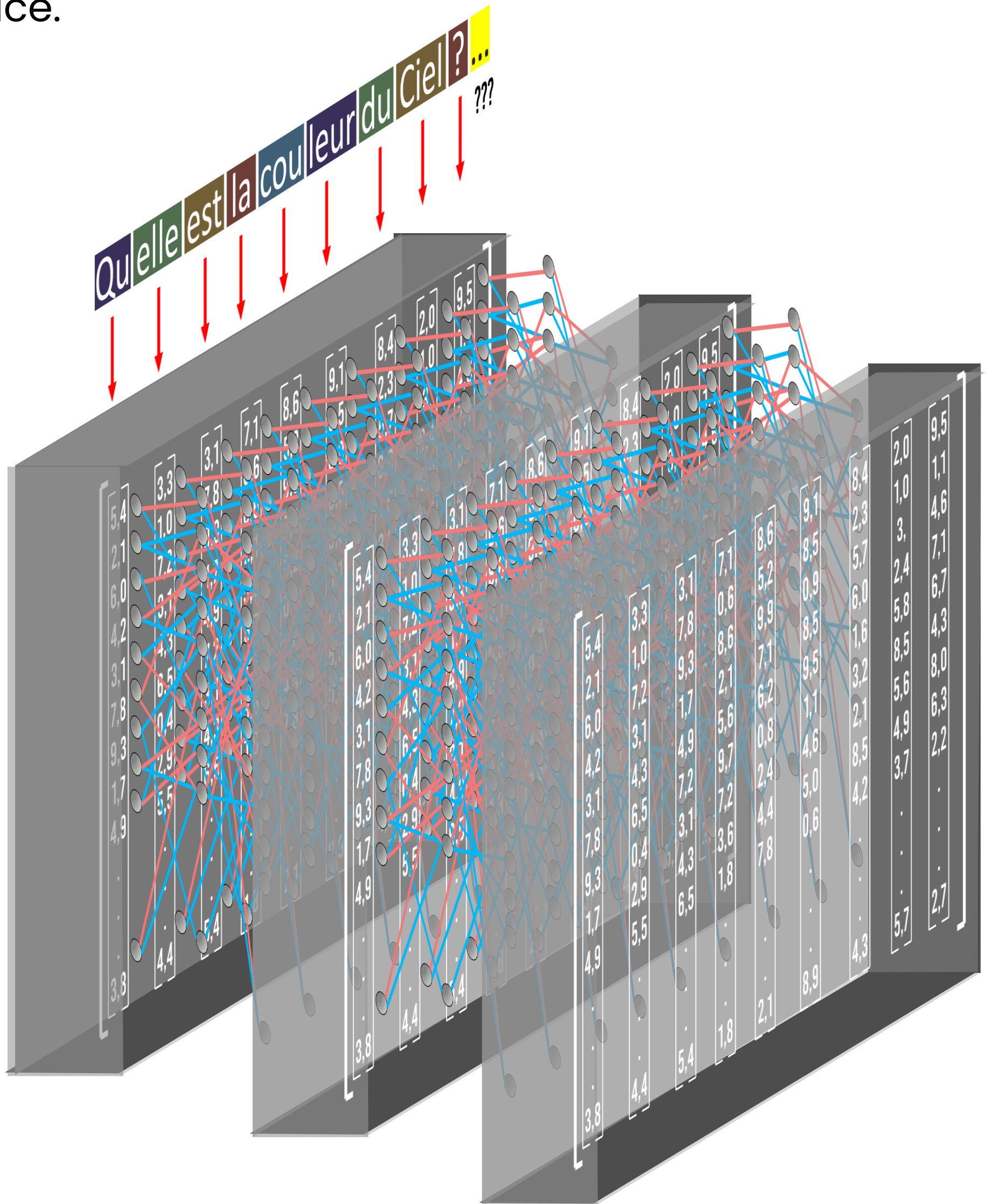
Toutes les opérations dans ces deux blocs ressemblent à une pile géante de multiplications de matrices.



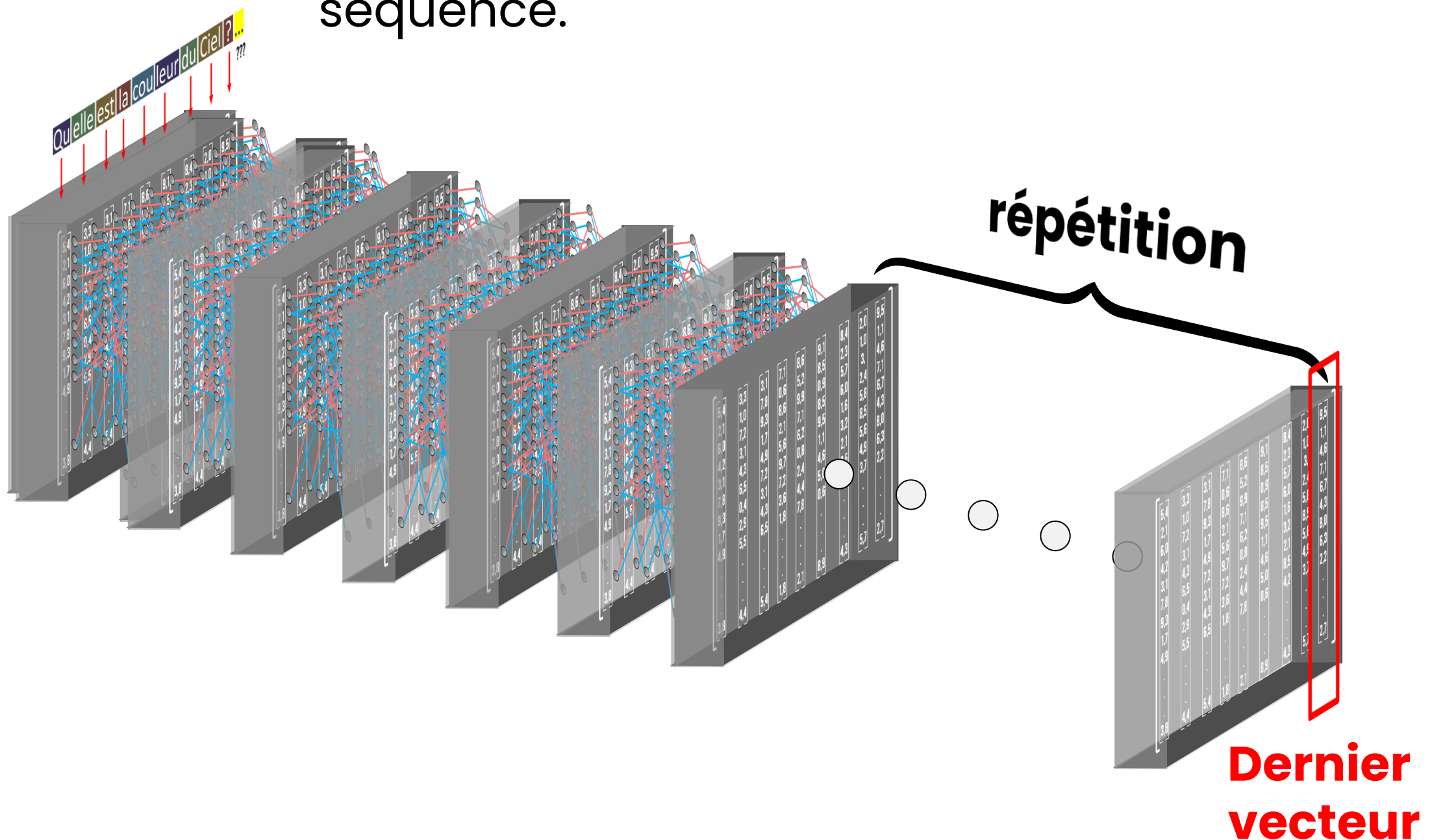




Ce qui aboutit à une nouvelle mise à jour de la matrice.



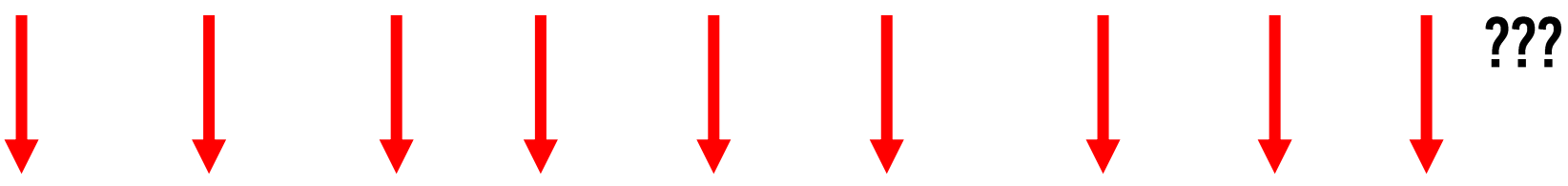
Et ainsi de suite : le processus répète en faisant des allers-retours entre les blocs d'attention et les blocs de perceptron multicouche. Jusqu'à ce qu'à la toute fin, tout le sens essentiel du prompt (on l'espère) a été incorporé d'une manière ou d'une autre dans le dernier vecteur de la séquence.



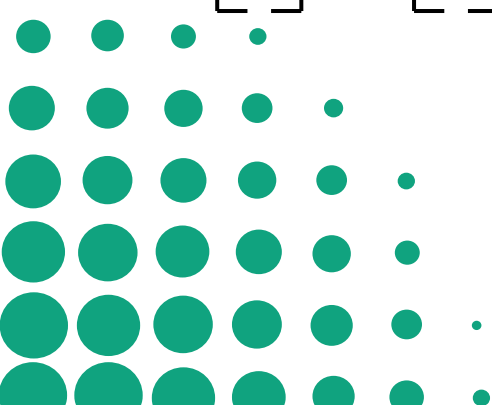
# 6. Calcul de probabilité

À ce stade, l'IA va effectuer un calcul de probabilité du token suivant le dernier vecteur, maintenant que celui-ci est représentatif du sens essentiel du prompt.

Quelle est la couleur du Ciel ? ...



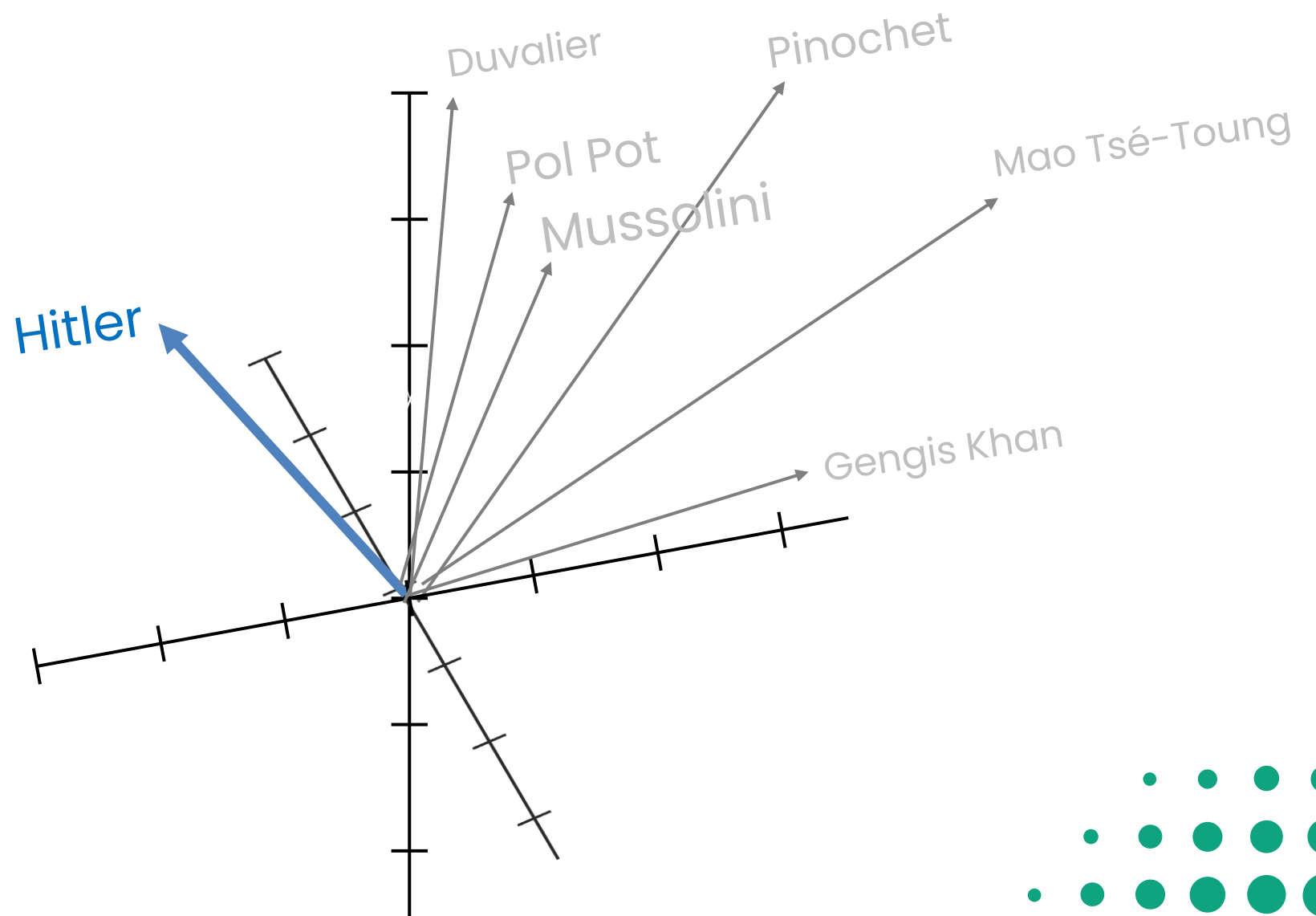
5,4	3,3	3,1	7,1	8,6	9,1	8,4	2,0	9,5
2,1	1,0	7,8	0,6	5,2	8,5	2,3	1,0	1,1
6,0	7,2	9,3	8,6	9,9	0,9	5,7	3,	4,6
4,2	3,1	1,7	2,1	7,1	8,5	6,0	2,4	7,1
3,1	4,3	4,9	5,6	6,2	9,5	1,6	5,8	6,7
7,8	6,5	7,2	9,7	0,8	1,1	3,2	8,5	4,3
9,3	0,4	3,1	7,2	2,4	4,6	2,1	5,6	8,0
1,7	2,9	4,3	3,6	4,4	5,0	8,5	4,9	6,3
4,9	5,5	6,5	1,8	7,8	0,6	4,2	3,7	2,2
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
3,8	4,4	5,4	1,8	2,1	8,9	4,3	5,7	2,7



## Calcul du token suivant

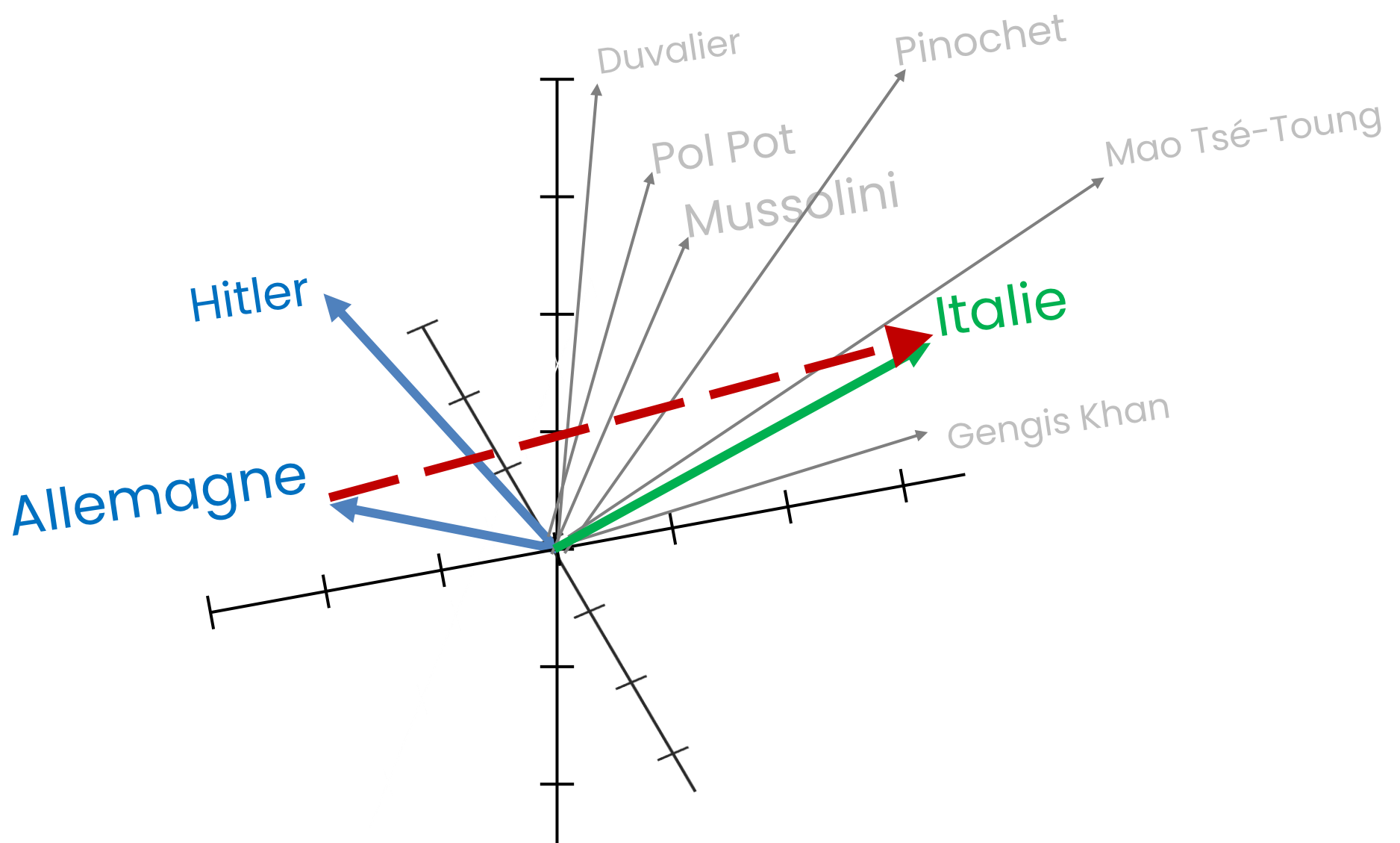
La vectorisation a positionné proches les uns des autres dans l'espace, les tokens similaires sémantiquement.

Prenons un exemple différent ici, avec un dictateur : Hitler. On trouve au voisinage d'Hitler dans l'espace multidimension, tous les autres dictateurs, de Gengis Khan à Mao Tsé Toung, en passant par Mussolini.



Imaginons que notre prompt parlait non seulement d'Hitler, mais aussi d'Allemagne et d'Italie. Imaginons que le prompt demande à l'IA d'identifier à un moment donné quel dictateur était à l'Italie ce qu'Hitler a été pour l'Allemagne.

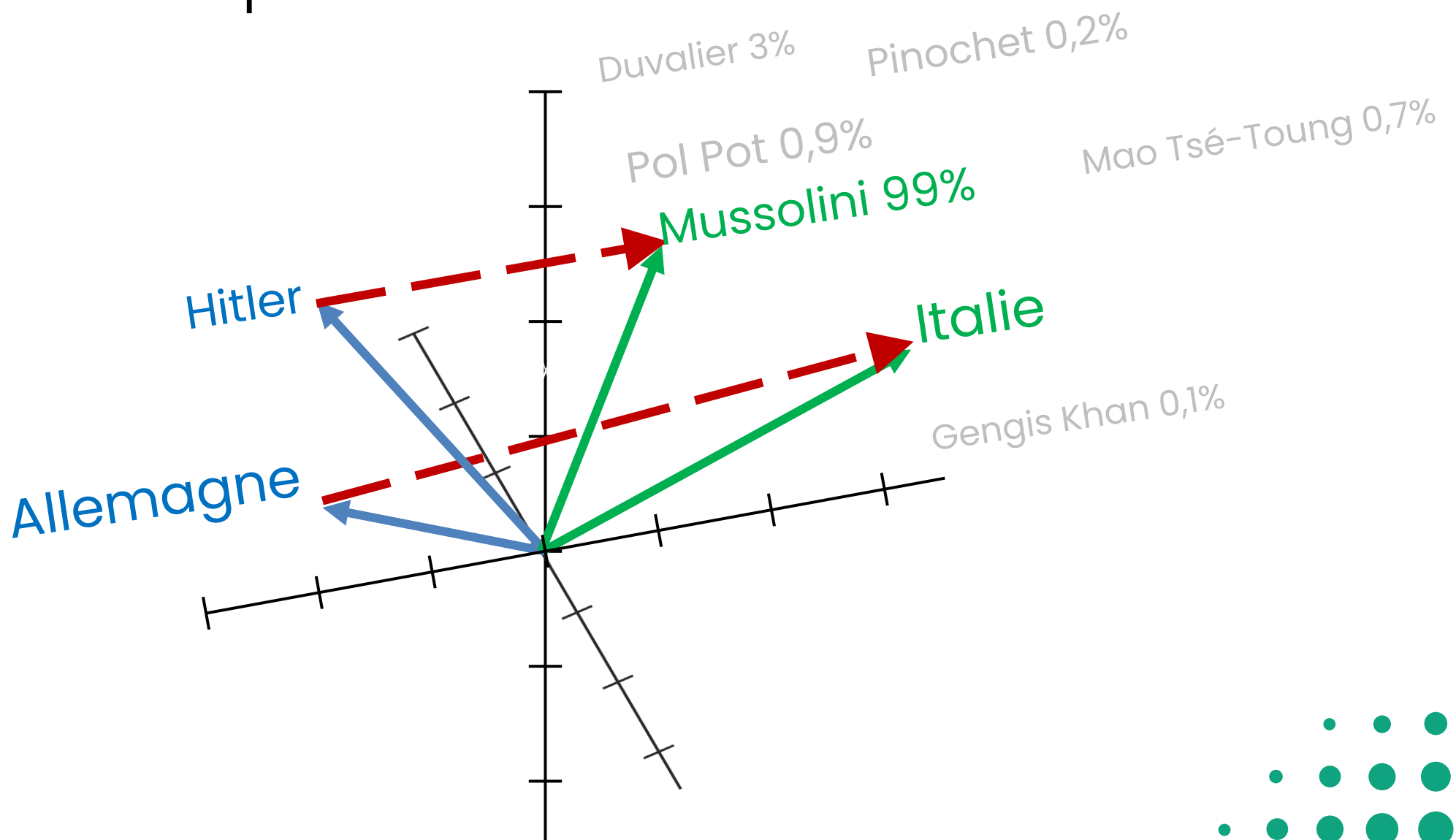
L'IA va effectuer un calcul vectoriel pour mesurer l'écart séparant Allemagne d'Italie.



En appliquant la distance entre Allemagne et Italie à Hitler, le vecteur pointe sur Mussolini comme token le plus proche. Et il est possible de calculer la distance entre "Hitler" et tous les autres dictateurs.

L'IA ramène ces distances à une valeur totale de 1, ce qui revient à traduire les distances en pourcentage.

Et les pourcentages obtenus permettent d'établir que le dictateur italien le plus probable est Mussolini.













## Le playground d'OpenAI :

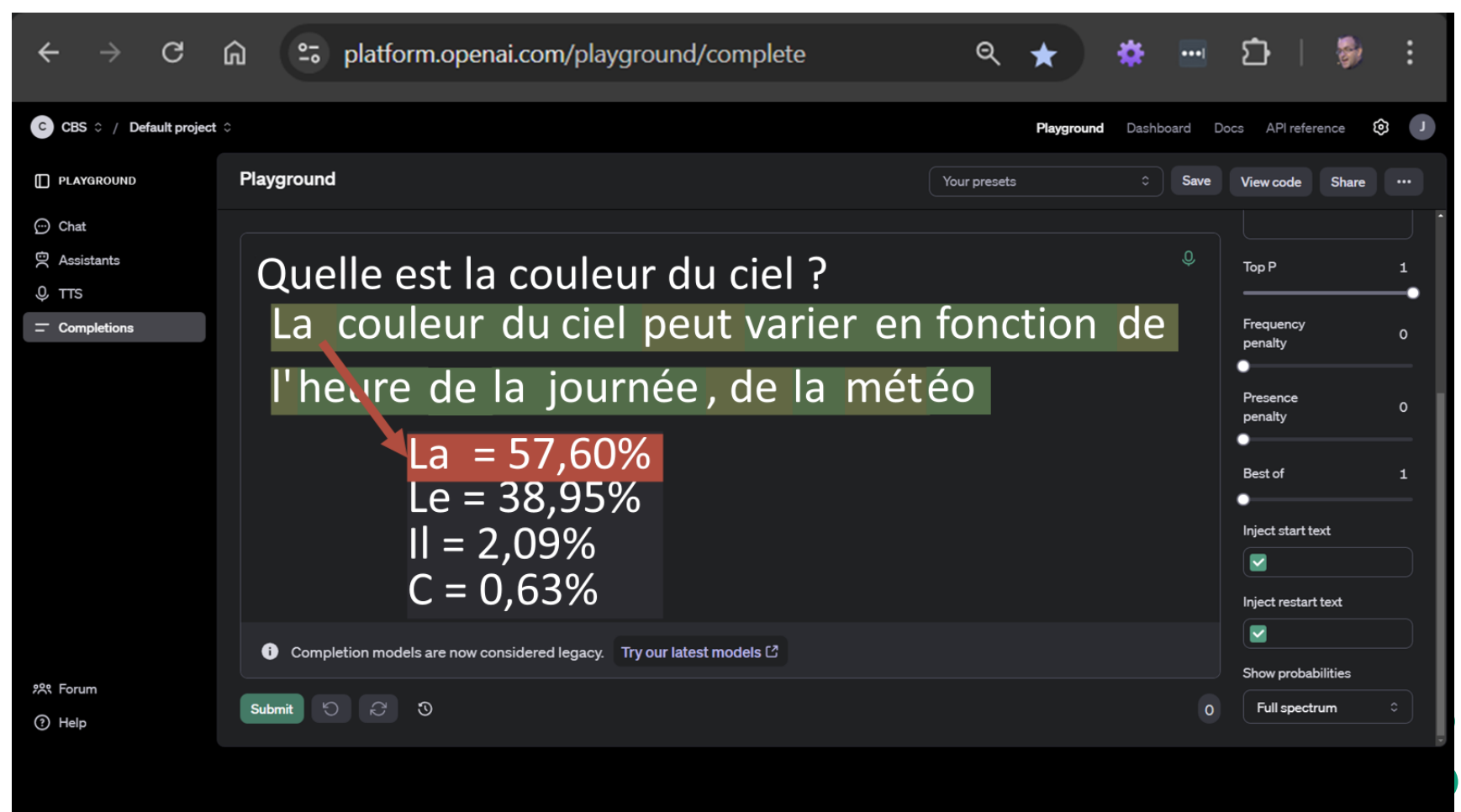
OpenAI met à disposition un outil très utile pour visualiser et comprendre ce process : le Playground. Il est accessible ici :

<https://platform.openai.com/playground/>

### Procédez à 4 réglages :

- Option "completion"
- Température = 0
- Top P = 0
- Show Probabilities : Full Spectrum

Dès lors, saisissez un prompt, cliquez sur "Submit", l'IA répond, et les probabilités de chaque token s'affichent lors de son survol.



The screenshot shows the OpenAI Playground interface. The prompt is "Quelle est la couleur du ciel ?". The completion is "La couleur du ciel peut varier en fonction de l'heure de la journée, de la météo". A red arrow points to the completion text, which is highlighted in green. Below the completion, a tooltip shows the probability breakdown for the next token:

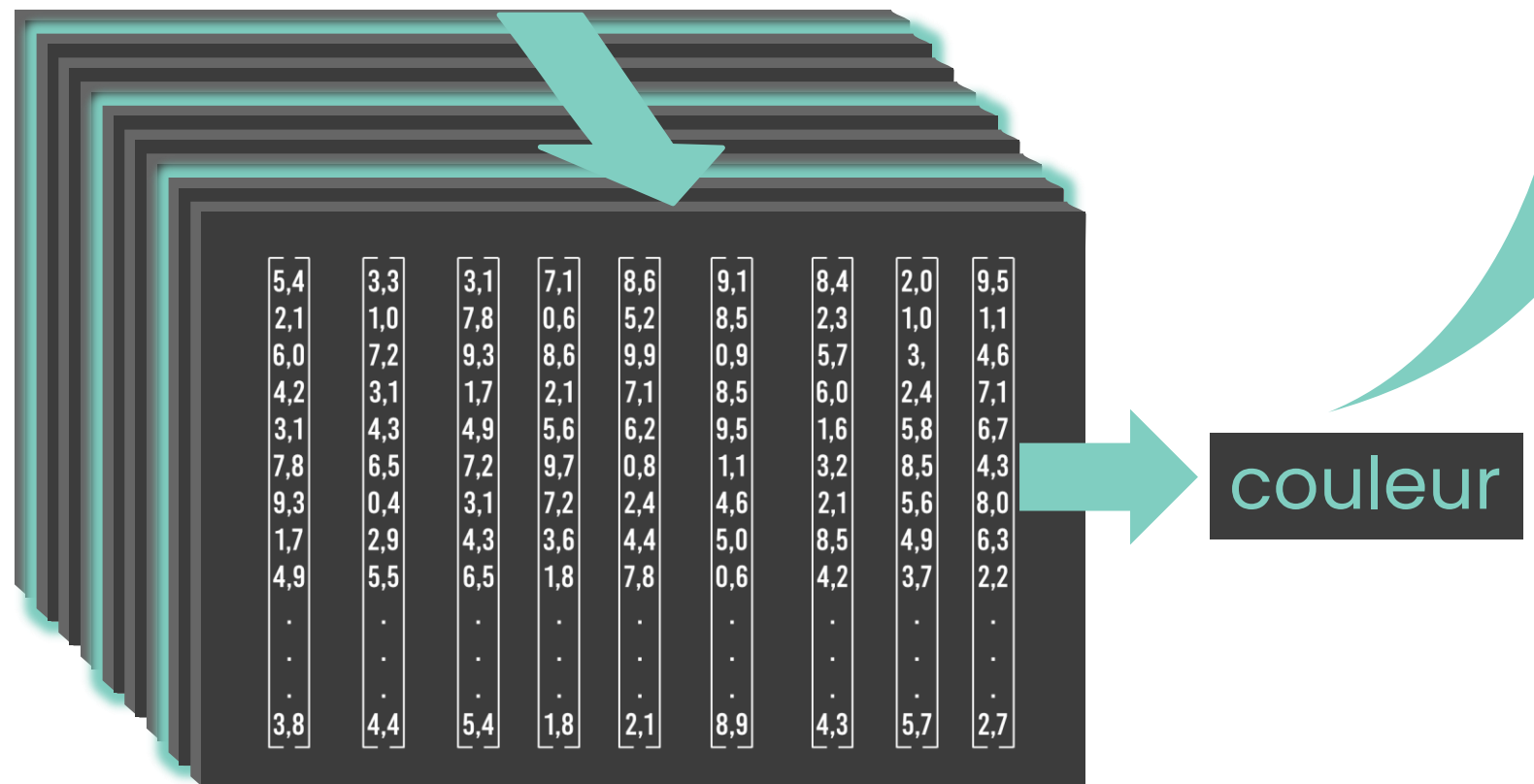
La	= 57,60%
Le	= 38,95%
Il	= 2,09%
C	= 0,63%

The interface also shows various settings on the right side, including Top P (1), Frequency penalty (0), Presence penalty (0), Best of (1), Inject start text (checked), and Inject restart text (checked). The "Show probabilities" dropdown is set to "Full spectrum".



Utilisateur : Quelle est la couleur du ciel ?

assistant : La



Le 1<sup>er</sup> token ayant la plus forte probabilité de suivre le prompt apparaît alors dans la zone de réponse. Cela crée l'illusion que l'IA a compris votre question et commence à vous répondre. C'est le 2<sup>ème</sup> rôle en IA, **le rôle de l'assistant** : fournir la réponse en sortie.

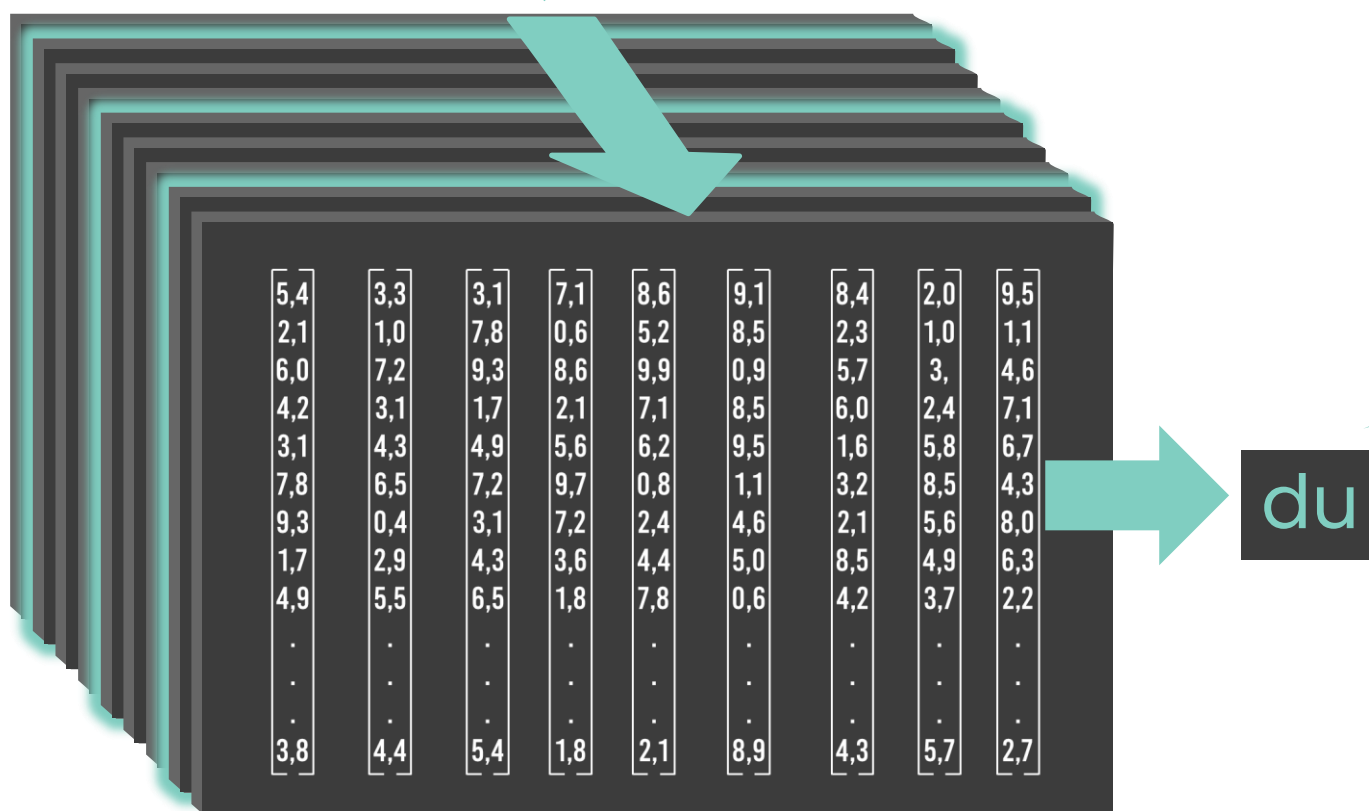
Le process s'applique à nouveau au prompt + le nouveau token, pour calculer le token le plus probable à suivre.

Ici le mot "couleur".



Utilisateur : Quelle est la couleur du ciel ?

assistant : La couleur

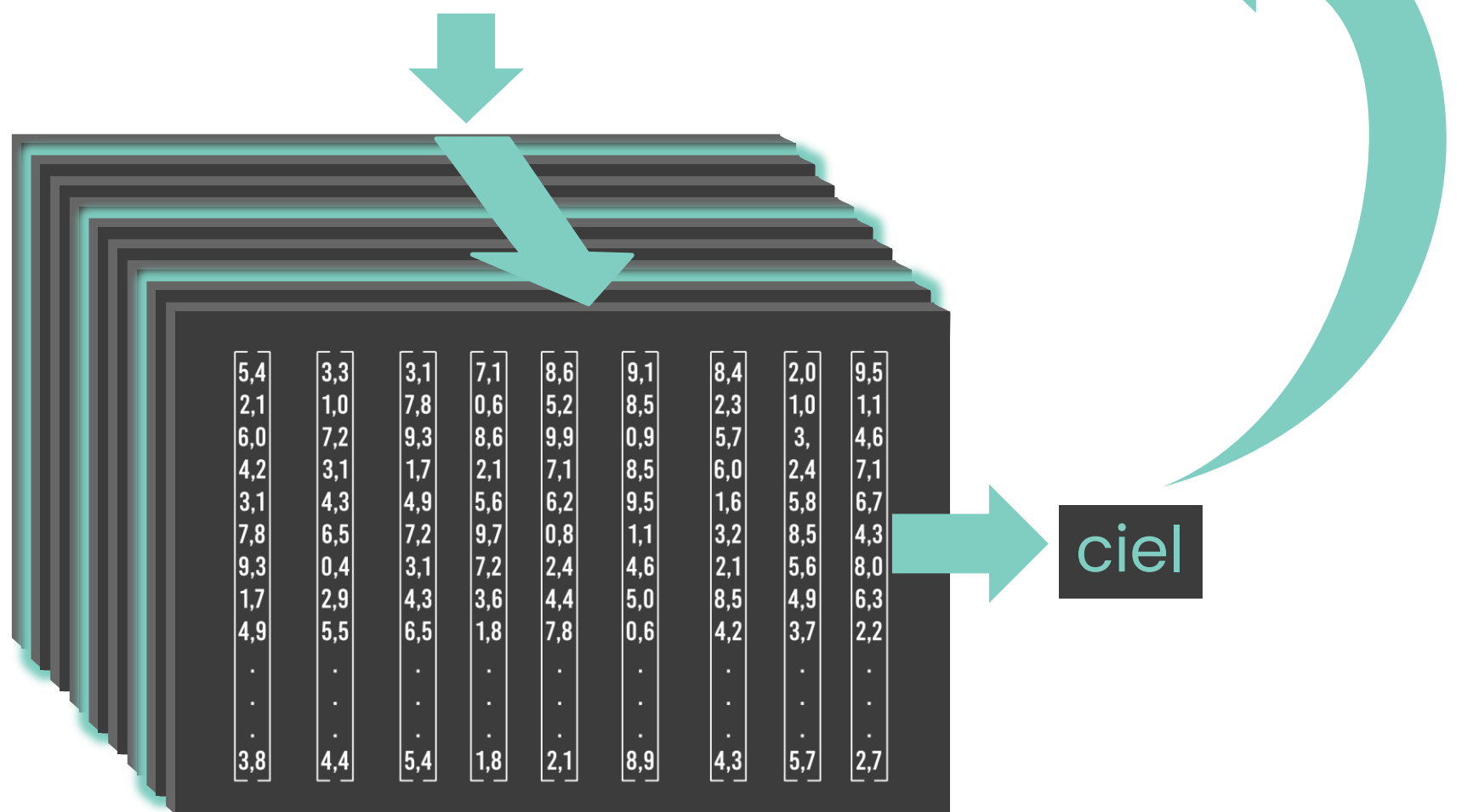


Le token "couleur" qui avait la plus forte probabilité de suivre le prompt + "la" est ajouté par l'assistant dans la zone de réponse.

Le process s'applique à nouveau au prompt + les 2 nouveaux tokens, pour calculer le token le plus probable à suivre. Ici le mot "du".

Utilisateur : Quelle est la couleur du ciel ?

assistant : La couleur du



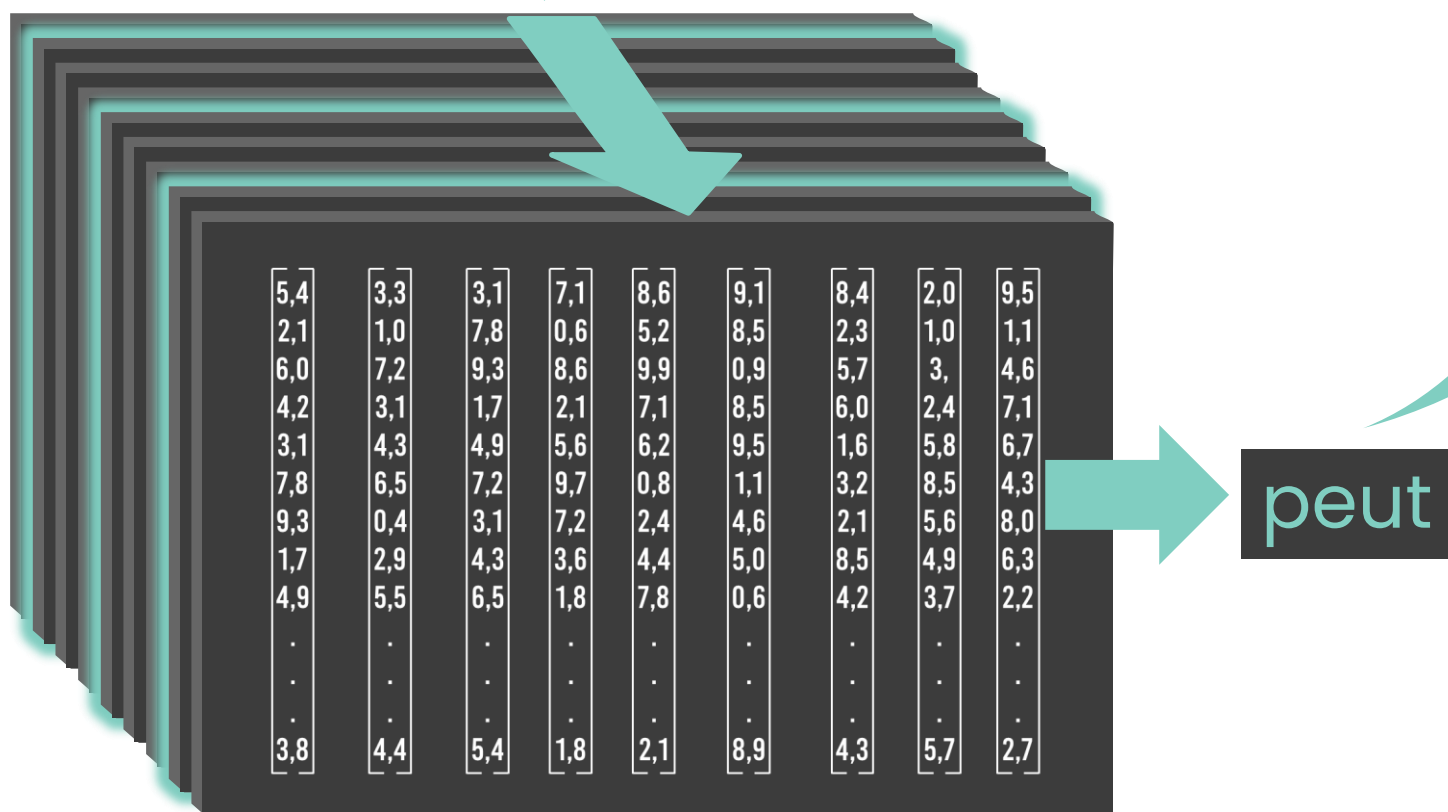
Le token "du" qui avait la plus forte probabilité de suivre le prompt + "la couleur" est ajouté par l'assistant dans la zone de réponse.

Le process s'applique à nouveau au prompt + les 3 nouveaux tokens, pour calculer le token le plus probable à suivre. Ici le token "ciel".



Utilisateur : Quelle est la couleur du ciel ?

assistant : La couleur du ciel



Le token "ciel" qui avait la plus forte probabilité de suivre le prompt + "la couleur du" est ajouté par l'assistant à la réponse.

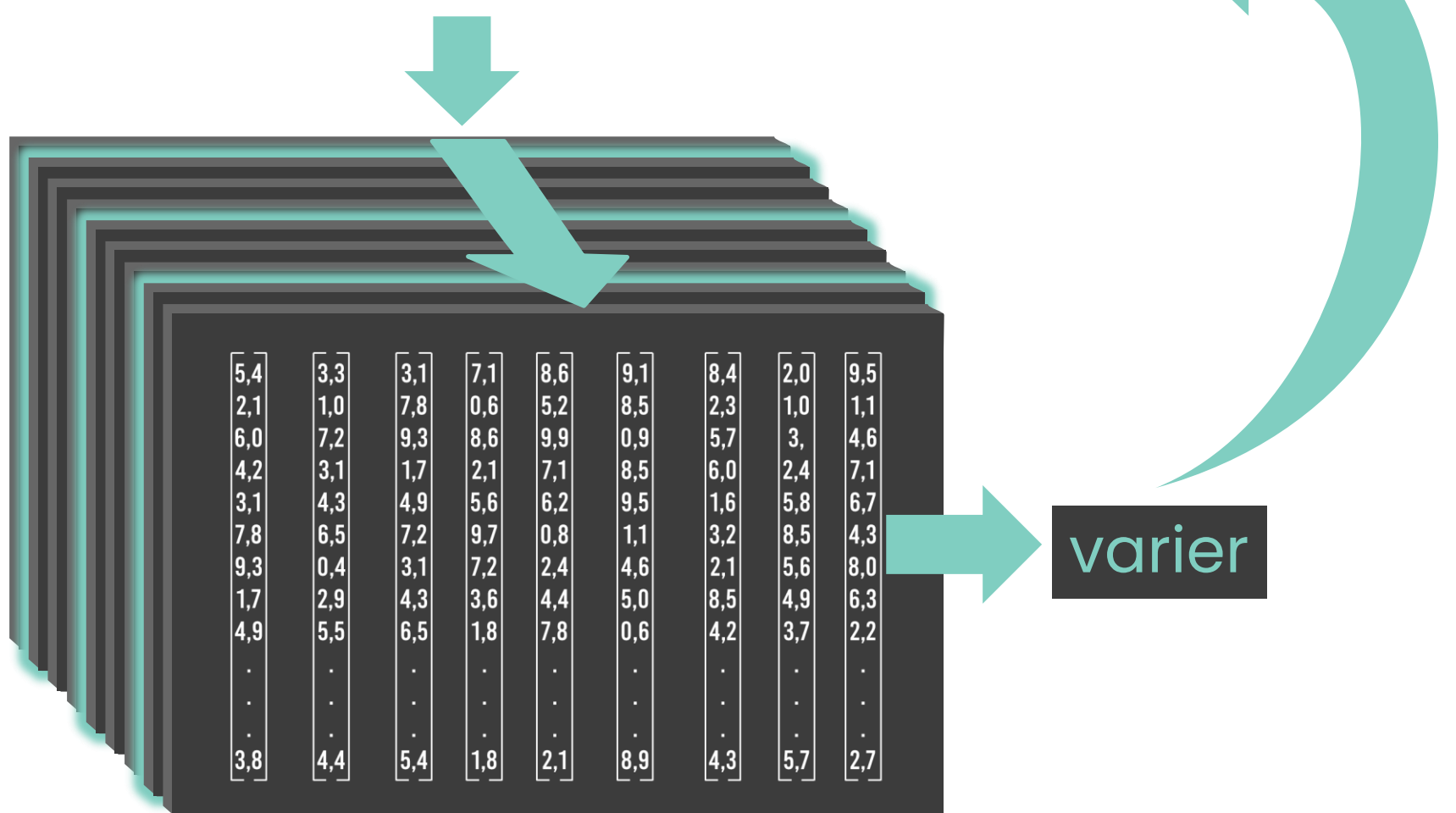
Le process s'applique à nouveau au prompt + les 4 nouveaux tokens, pour calculer le token le plus probable à suivre. Ici : "peut"

Utilisateur :

Quelle est la couleur du ciel ?

assistant :

La couleur du ciel peut



Le token "peut" qui avait la plus forte probabilité de suivre le prompt + "la couleur du ciel " est ajouté à la réponse.

Le process s'applique à nouveau au prompt + les 5 nouveaux tokens, pour calculer le token le plus probable à suivre. Ici : "varier"





Nous avons annoncé 3 rôles : Le 3<sup>ème</sup>, **le rôle "system"**, est la partie du prompt donnant des instructions pour que la mission ou tâche demandée soit effectuée correctement.

Il n'y en avait pas dans l'exemple simpliste précédent. Elle se matérialise par des instructions ajoutées dans le prompt. Par exemple :

- **Soit très simples :**

Procéder en 2 étapes, faire une recherche internet avant de répondre, puis répondre.

- **Soit plus complexes :**

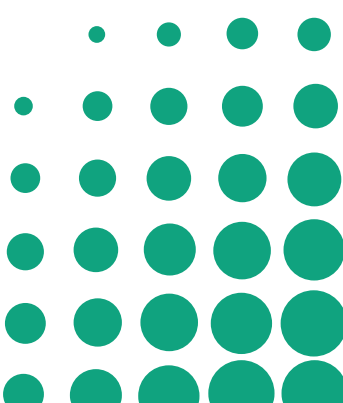
Optimisation du processus : {Utiliser une méthode claire et directe pour minimiser les efforts cognitifs tout en maximisant la précision, et réduire le risque de confusion ou d'interprétation erronée. Simplifier le processus pour éviter les étapes superflues et accélérer la résolution.},

*NB : Ce dernier exemple est l'une des 9 instructions ajoutées en amont de votre prompt par o1-preview pour fiabiliser les réponses. (Les autres sont : méthodologie systématique, pensée critique, mobilisation de ressources. Etc. )*



**Truc de pro :** Ajouter ces 9 instructions aux prompts hisse la qualité des réponses vers celle de o1-preview.

**Demandez les en commentaire, nous vous donneront accès...**



## En synthèse

Le principe de ces modèles de langage (LLM) qualifiés d'autorégressifs est de :

- Calculer la probabilité du 1<sup>er</sup> token suivant une série,
- Puis celle du 2<sup>e</sup> token après ajout du 1<sup>er</sup> à la série
- Et ainsi de suite.

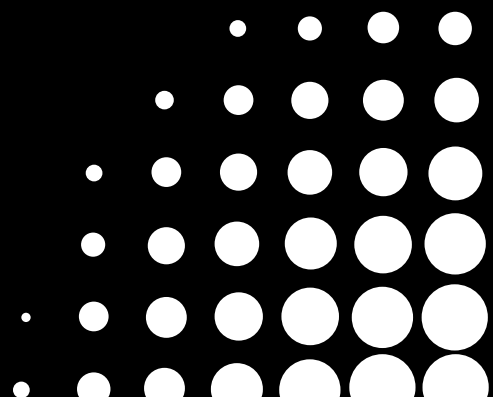
**Il n'y a donc pas d'intelligence** à proprement parler dans l'IA, mais une chaîne de calculs de probabilités par calculs vectoriels matriciels.

**Il n'y a pas d'approximation** non plus. Il s'agit d'une chaîne de calculs de probabilités.

**L'IA ne fait jamais d'erreurs** : ses calculs statistiques sont vérifiables. Ce que nous percevons comme des erreurs (appelées "hallucinations" de l'IA), sont simplement des probabilités qui ne se réalisent pas, car "probable" ne signifie pas "certain".



# 3 Implications pour l'usage de l'IA



# 1. Les mauvaises pratiques

Maintenant que vous connaissez le fonctionnement réel de l'IA, il vous sera aisé de comprendre pourquoi la majorité des conseils et pratiques que vous avez pu trouver sur Internet sont du grand n'importe quoi.

- Tutoyer l'IA
- Attribuer un rôle à l'IA
- Les prompts simplistes
- Le prompt ultime... prompt à faire des prompts
- Etc.





**L'anthropomorphisme** : Il y a d'autres inconvénients à tutoyer l'IA : le principal s'appelle l'anthropomorphisme, et c'est la santé mentale de ceux qui tutoient l'IA qu'il met en danger. Pour en savoir plus sur le risque d'anthropomorphisme à tutoyer l'IA, [cliquez ici >>](#)

**Dans tous les cas, retenez qu'il faut parler à l'infinitif à l'IA pour éviter tous ces pièges.**

**Un test significatif** à mener dans votre entreprise : combien de collaborateurs tutoient l'IA dans leurs prompts ? **Plus il est important, plus vous devez rapidement organiser une VRAIE formation à l'IA.**

**Vous-même** si vous êtes utilisateur, vous utilisez l'infinitif ?





**À l'origine de ce conseil absurde, une erreur de compréhension et/ou de traduction**, faite de la documentation d'OpenAI en anglais, qui explique l'utilisation des 3 rôles quand on passe par l'API. (Assistant, User, System)

Puis ils sont des milliers à avoir repris bêtement cette idée absurde "d'attribuer un rôle". Bêtement, car cela sous-entend que :

- Donner un rôle à l'IA (pompiers par exemple) pourrait lui greffer par incantation les connaissances correspondantes !!! (pour établir le plan de sécurité incendie d'un bâtiment par ex.)
- Où que l'IA va parcourir en une fraction de seconde les milliards de données de son corpus d'entraînement ? Si c'était possible (ça ne l'est pas), ce dernier n'est constitué que d'extraits et de résumés d'ouvrages (jamais les ouvrages entiers pour des raisons de droits d'auteur).

Si vous avez encore un doute, sur l'absurdité du conseil, [cliquez ici >>](#)

Connaissant le fonctionnement de l'IA, il est aisé de comprendre que rajouter une séquence de mots tels que "**agis comme un expert en ressources humaines au courant de toutes les techniques RH**"... ne va pas greffer les compétences qu'aurait un tel expert. Qui pourrait croire à telle inéptie ?

Cette série de tokens va juste faire reclasser les positions dans l'espace des autres tokens en prenant en compte les termes "ressources humaines" et "techniques RH".

Mais si vous vouliez que lesdites techniques soit prises en compte, **il fallait les écrire explicitement dans le prompt**, ou dans un document attaché au prompt, ou demander dans le prompt à aller les lire sur une URL qui sera précisée à ce moment.

**Oubliez donc ce conseil idiot  
d'attribuer un rôle à l'IA.**

**Et fuyez les pseudo-experts  
qui le propagent.**





## D. Le prompt à faire des prompts

Avec ce que vous avez appris au chapitre précédent, vous pouvez désormais comprendre l'utilisation que l'IA va faire de ce type de prompt : "En tant que spécialiste du prompt engineering, élabore le meilleur prompt pour résoudre le problème x..."

- L'IA le décompose en 19 tokens :

En tant que spécialiste du prompt engineering, élabore le meilleur prompt pour résoudre le problème x...

- 7, soit presque 40% n'apporteront rien dans le processus : En, tant, que, du, le, pour, prompt (quand il est cité pour la 2<sup>o</sup> fois)
- **spécialiste et engineering** vont modifier la position vectorielle de certains tokens, mais ne va pas greffer de compétences de spécialiste en prompt engineering.
- Les tokens utiles étaient : prompt, élabore, résoudre, problème x. Oui l'IA va élaborer un prompt, mais pauvre et banal.

**il aurait fallu joindre des instructions pertinentes pour qu'il le soit!**





## 2. Les failles

Maintenant que vous connaissez le fonctionnement réel de l'IA, il vous sera aisé de comprendre les 3 failles de l'IA Générative, et donc la nécessité de former vos collaborateurs à faire de vrais prompts capables de contourner ces 3 pièges.

### 1. Les erreurs (hallucinations)

**L'IA ne se trompe jamais**, elle effectue des calculs de probabilités de tokens. Comme une prévision météo, il arrive régulièrement que la réalité soit différente de la probabilité, et que la probabilité ne se vérifie pas dans les faits. Ce qui n'empêche pas que le calcul statistique effectué ait été juste, et que la probabilité existait réellement.

En moyenne, 20% des prévisions de l'IA ne se vérifient pas, et seront perçues comme des erreurs.



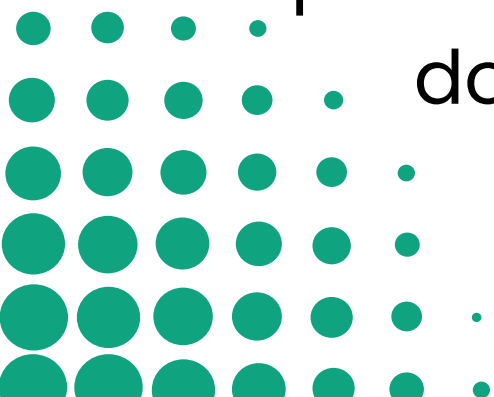
## **Une bonne maîtrise des techniques de prompt engineering permet de se prévenir de ces 20% d'erreurs.**

Raison pour laquelle vous devez former vos salariés : Une vraie formation basée sur une connaissance solide de l'IA.

Fuyez les formations qui tutoient l'IA et préconisent d'attribuer un rôle par exemple : c'est la certitude que le formateur et/ou l'organisme ne connaît pas le sujet.

Faute de quoi, vos salariés génèrent systématiquement les 20% d'erreurs, puis utilisent et divulguent ces informations erronées en interne et externe comme des vérités absolues.

Une catastrophe, dont l'outil n'est pas responsable : c'est l'utilisation par des gens non formés. Comme une voiture conduite par quelqu'un qui n'a pas le permis, comme un avion par un utilisateur qui n'a pas de brevet de pilote... comme n'importe quelle technologie dont on ignore le fonctionnement.




## 2. Les réponses banales

Vous savez que l'IA calcule des probabilités de tokens formant les mots les plus probables. Et "contenu le plus probable" est synonyme de "contenus le plus banal" **pour 100% des réponses de l'IA sans instructions spécifiques le rôle "System"**.

À nouveau il est possible de contourner cette faiblesse de l'IA par une bonne maîtrise du prompt engineering. C'est une autre raison pour laquelle vous devez former vos salariés.

## 3. Les réponses incomplètes

Vous l'avez vu dans le fonctionnement de l'IA : à aucun moment l'IA n'est programmée pour fournir des réponses complètes et exhaustives sur un sujet. **Sans instructions spécifiques pour le rôle "System", les prompts simplistes sont la certitude absolue d'obtenir 100% de réponses incomplètes.**







# 1. 7 couches d'info indispensables

De même que le process se décompose en 7 étapes, il nécessite des prompts incluant 7 couches distinctes d'information



La **tâche** que vous allez confier au prompt.

Le **but** que la tâche sert à atteindre.

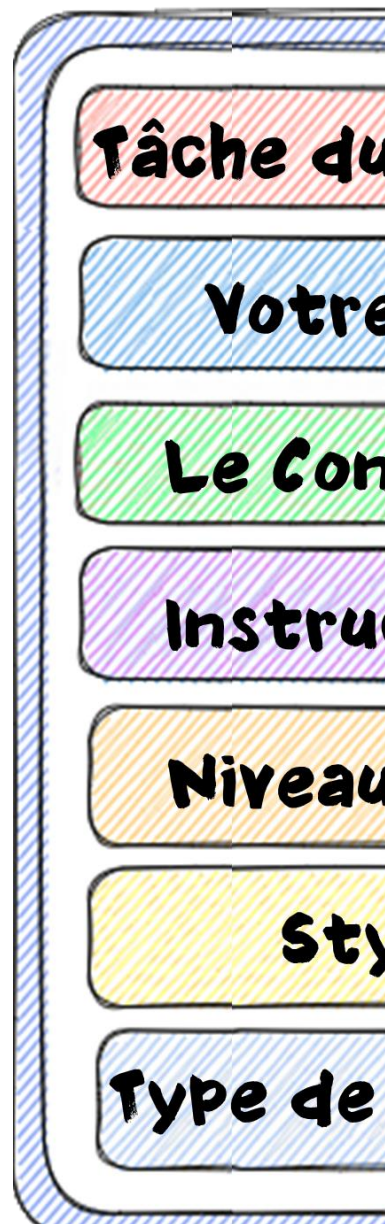
Le **contexte** dans lequel se situe votre demande.

Les **instructions** que le prompt doit exécuter.

Le **niveau** de maîtrise de la **cible**

Le **style** souhaité pour la réponse

Le type de **réponse** attendue.







# 3

## Le Contexte

C'est la couche essentielle à personnaliser, pour obtenir une réponse adaptée à votre entreprise, votre offre, votre positionnement, votre cible. Elle fait souvent à elle seule 2 ou 3 pages.

### Exemple

**Contexte :** **Je suis** Directeur Commercial et Marketing de **l'entreprise** Bon Café, **notre site** <https://fr.combier.com/fr>, **localisé en** France, **notre offre** : la machine à café Trio, **notre cible** sont les Adultes amateurs de café, **nos avantages** : 1. : Fait du café à partir de grains fraîchement moulus, pas en capsule, 2. : trois machines en une grâce à 3 broyeurs de précision, **les problèmes** que nous résolvons : 1. : Le café en capsule n'est pas écologique., 2. : Les machines à grain ne peuvent faire qu'un seul de type de café alors qu'une même personne peut en consommer des différentes [...], **nos concurrents** sont, etc.

Prompt

But

Contexte

Instructions

Cible

Style

Réponse



## 5

## Niveau cible

On ne répond pas de la même manière si la réponse est destinée à un enfant de 10 ans ou à un BAC +12 !

Préciser dans cette couche le public cible de la réponse :

- **Fonction** du public cible
- **Niveau** : Expert, Spécialiste, Expérimenté, Moyen, Débutant
- **Compétences** : Lister les compétences maîtrisées.

## Exemple

Niveau de la cible : je suis Directeur marketing, ce sujet est mon métier, je suis familier de toutes les techniques et du vocabulaire de la discipline.

prompt

But

texte

ctions

cible

Le

réponse





## 2. Les erreurs à éviter

**Il n'y a pas de 8<sup>e</sup> couche consistant à attribuer un rôle à ChatGPT.**



C'est absurde d'imaginer qu'attribuer à ChatGPT un rôle d'expert d'un sujet, va lui greffer comme par magie les compétences correspondantes !

Ce n'est pas le cas, qui relèverait de la pensée magique. Jamais cette idée saugrenue ne figure dans la documentation d'OpenAI.

Si vous aviez un doute, consultez le volume 3 de notre série sur les 3 erreurs les plus courantes dans ChatGPT. Ou notre article «attribuer un rôle d'écureuil à ChatGPT».



# 4

## Un prompt ne peut pas être une question...

→ comme Microsoft vous y incite via Bing et Copilot



→ ... comme la majorité des novices le font.

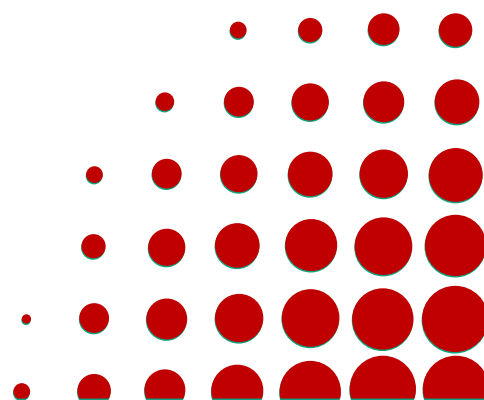
→ ... et comme les bases de prompts offertes ou vendues par des pseudo-experts autoproclamés sur le web le font.



 Rédige un message promotionnel sur les médias sociaux à partir du contenu ci-dessus.

 Rédige X idées d'accroche pour un message sur les médias sociaux concernant [sujet]

 Aide-moi à structurer une histoire captivante pour mon produit/service type en utilisant le storytelling

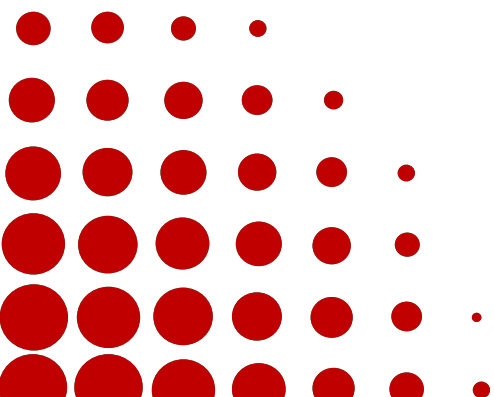




## Pas 2, pas 4, pas 6... Mais 7 couches.

Les 7 couches d'informations listées ici ont toutes un rôle distinct, et sont toutes indispensables.

Fuyez les méthodes qui préconisent 4, 5 ou 6 couches. Encore plus si une d'elles est inutile comme peut l'être l'attribution d'un rôle.



# 6

Avec 7 couches d'instructions, un prompt peut nécessiter 1 300 mots, soit 1 800 tokens. Bien plus si on a attaché des documents...

**1 800 tokens**



Entrée



Sortie

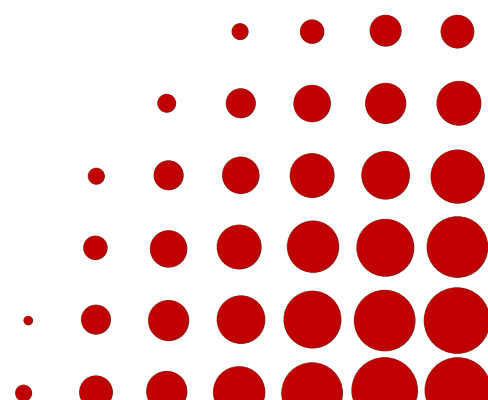
Réponse



**2 300 tokens ➡ 1 600 mots**

En raison de la limite du modèle, il n'en reste que 2 300 max. pour la réponse = 1 600 mots.

Rédiger des articles de 2000 mots, tutoriels, contenus de formation, plans stratégiques, etc. **demande des techniques avancées** comme les itérations.



# 3. Les Marqueurs







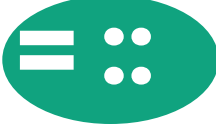

Avec 7 couches les prompts sont longs. Or le process vise à compiler toutes ces informations dans un seul vecteur final. Plus le prompt est long, plus le process risque de regrouper des choses qui ne doivent pas l'être, et en éliminer d'importantes.



Vous devez donc **impérativement utiliser** des **marqueurs** (appelés aussi "balises") : ce sont des caractères spéciaux qui encadrent des éléments et définissent des structures spécifiques dans le prompt, pour qu'il puisse les interpréter correctement.

1. Accolades simples `{ }` pour définir des ensembles d'informations.
2. Guillemets `" "` doubles ou simples pour séparer des éléments de texte au sein d'un ensemble délimité par accolades.
3. Crochets `[ ]` pour indiquer des options ou des choix possibles. ("`[Oui/Non]`" par exemple).



4. Parenthèses  pour regrouper des éléments et indiquer une hiérarchie.
  5. Barres obliques  pour séparer des éléments ou des options.
  6. Virgules  pour séparer des éléments dans une liste ou pour délimiter des clauses dans une phrase.
  7. Triples guillemets  *file.pdf*  pour mentionner des noms de fichiers.
  8. Accolades doubles  pour définir des paramètres (voir chapitre suivant).
  9. Egal et deux fois deux points  pour définir un paramètre
  10. Deux fois deux points  pour définir un paramètre
- etc.



## Exemple

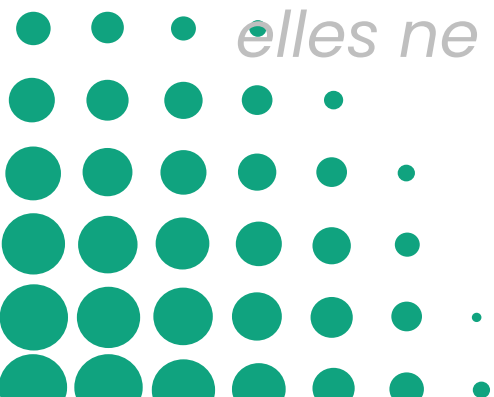
Contexte : { Je suis "Directeur Commercial et Marketing", de l'entreprise "Bon Café", localisée en France,  
notre **{{offre}}** = "Machine à café Trio Espresso.", notre **{{cible}}** = "Adultes amateurs de café.",  
nos **{{avantages}}** = {"1." : {Fait du café à partir de grains fraîchement moulus, pas en capsule.},  
"2." : {deux machines en une grâce à 2 broyeurs de précision : peut faire du café avec ou sans caféine, ou bien une torréfaction foncée pour un espresso et un mélange doux pour un lungo : Capable de traiter deux variétés de café différentes, elle offre une multitude d'options pour tous les goûts.},  
"3." : {peut couronner votre boisson d'une mousse [...]

**{{actualités}}** = {le lancement de notre nouvelle machine J8 twin.},

*Tous les caractères en couleur sont des marqueurs.*

*En **bleu** les paramètres : voir chapitre suivant.*

*Les interfaces d'IA ne gèrent pas les couleurs, elles ne sont indiquées ici qu'à titre pédagogique.*



## 4. Les paramètres

Un "**paramètre**" est une variable passée aux différentes couches du prompt pour leur transmettre des données ou des informations afin qu'elles puissent les utiliser

**sans avoir à les répéter.**

Il existe plusieurs manières d'indiquer ces paramètres dans le prompt.

1. Double accolade :

**{{paramètre}} = xxx.**

2. une combinaison de deux-points suivis d'un mot-clé et du paramètre après le signe "=" **::paramètre =xxx**



### 3

Truc de pro : ce codage est compris par l'API, et très utile dans son usage.



# Les paramètres

## Exemple

Contexte : { Je suis "Directeur Commercial et Marketing", de l'entreprise "Bon Café", localisée en France, notre **{{offre}}** "Machine à café Tri-Expresso.", notre **{{cible}}** "Adultes amateurs de café", nos **{{avantages}}** {"1." : {Fait du café à partir de grains fraîchement moulus, pas en capsule.}, "2." : { [...]

"Process et instructions" : suivre scrupuleusement les 23 étapes suivantes : {"1.- Enrichissement des données" : "Faire 4 [...]

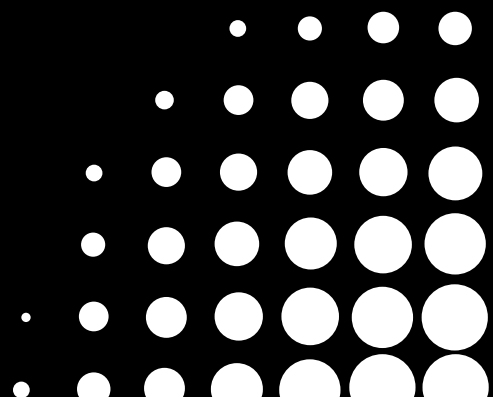
{« 5.- Collecte de thématiques sur le Web " : "Rechercher sur le web les sujets d'actualité récents pour les **{{cibles}}** sur les thèmes **{{offre}}**, **{{Problématiques}}**, **{{avantages}}**, [...]

• • • Tous les caractères en couleur sont des marqueurs.

En **bleu**, les paramètres..

• ChatGPT ne gère pas les couleurs, elles ne sont indiquées ici qu'à titre pédagogique.

# CONCLUSION

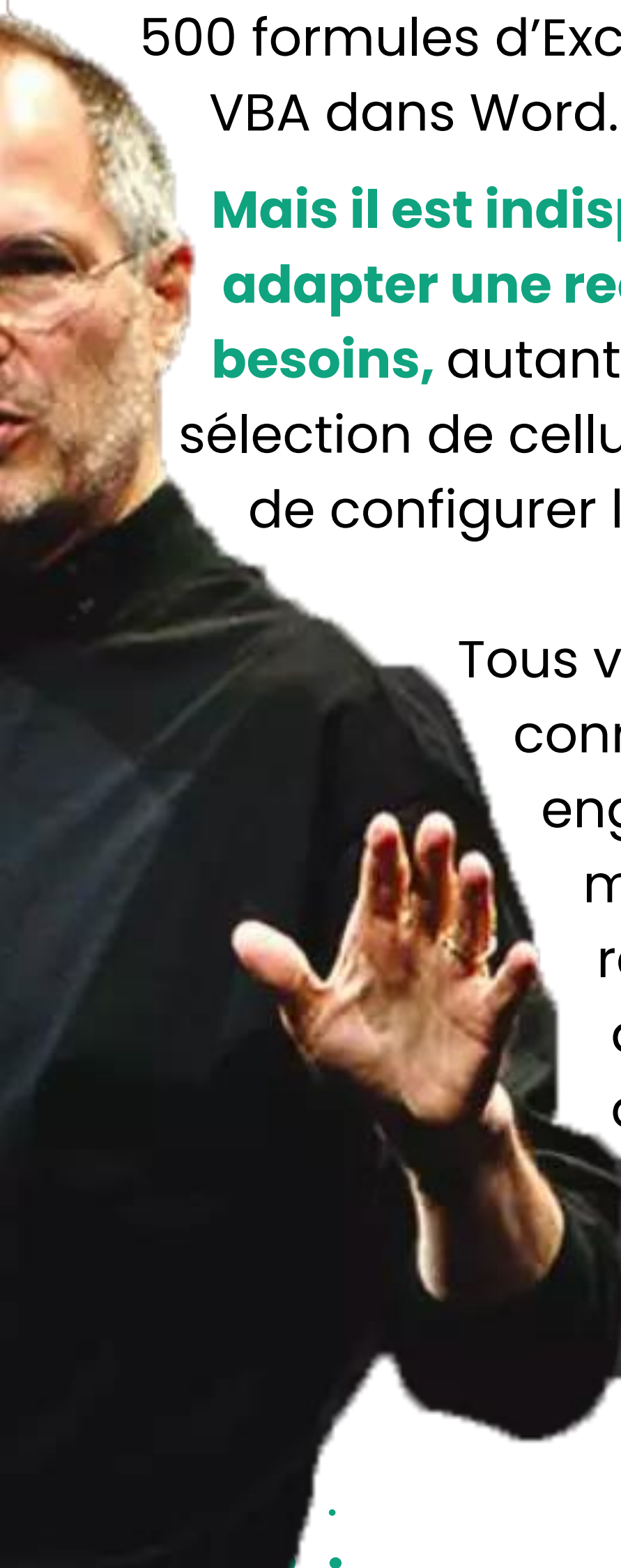




Il n'est pas nécessaire de former chaque employé aux techniques avancées de l'IA, tout comme il n'est pas essentiel pour tous de connaître les 500 formules d'Excel ou de maîtriser les macros en VBA dans Word.

**Mais il est indispensable que chacun sache adapter une requête (prompt) à ses besoins,** autant que de savoir ajuster la sélection de cellules pour une formule Excel ou de configurer la table des matières dans Word.

Tous vos collaborateurs doivent connaître les bases du prompt engineering. Cette maîtrise minimum de l'IA deviendra rapidement une compétence aussi cruciale pour l'entreprise d'un côté, et pour l'employabilité des salariés de l'autre, que l'est aujourd'hui la maîtrise minimum de Word ou d'Excel.



C'est pourquoi l'IA doit être une compétence partagée aussi largement que possible dans l'entreprise. Ne commettez pas l'erreur de désigner une "Madame IA" ou un "Monsieur IA" :

- Cela reviendrait à faire de l'IA le « problème » d'une seule personne, détournant le reste des collaborateurs du sujet, comme on a pu le voir dans les démarches ISO ou plus récemment dans la RSE.
- Auriez-vous imaginé, dans les années 90, nommer un "Monsieur Excel" ou une "Madame Word" ?

Pour que l'entreprise tire pleinement profit de ces technologies, c'est bien l'ensemble des collaborateurs qui doit en maîtriser les bases.





ChatGPT-4o langage natif de la formation

-50% black Friday jusqu'au 29/11/2024

~~479 € HT~~  
**239,50 € HT**

<https://bit.ly/3Qyz62f>

**Cliquer pour  
vous inscrire**

## FORMATION ONLINE CHATGPT et IA

**1** La méthode exclusive **DOMPTER**  
de prompt engineering  
de Cloud-Business-School.

**3** Automatiser ChatGPT  
avec l'automatisation  
la plus puissante du marché

**5** La banque de +700 prompts **DOMPTER**,  
compatible ChatGPT, Mistral,  
Copilot, Gemini, Claude etc.

**6** **Photoprompt**, générateur  
de prompts d'images pour Dall-E,  
Midjourney, et +50 IA.

**7** Support de formation  
avec tous les tutoriels  
et des prompts prêts à coller

**2** La seule formation SEO et IA  
pour apprendre vraiment à faire  
des contenus qui référencent bien.

**4** S'affranchir des limites,  
faire des contenus de  
1000, 2000 mots et plus.

Jusqu'au 29/11 coupon remise 50 %  
**OFFRE-BLACK-FRIDAY-HJ7NMB**  
Spécial Black Friday 2024 à saisir à la commande



# 4 AVANTAGES MAJEURS AVEC CLOUD BUSINESS SCHOOL

- ↳ Une garantie de professionnalisme avec Cloud Business School **certifié Qualiopi.**
- ↳ **La formation IA la plus avancée du marché.** Pourquoi ? Parce que nous avons réalisé des milliers de tests pour construire et sélectionner les meilleures techniques d'utilisation de l'IA.
- ↳ **Méthode DOMPTER l'IA :** Une révolution dans l'apprentissage du prompt engineering. Bénéficiez de notre méthode exclusive « DOMPTER », une synthèse de connaissances acquises à travers des milliers de tests.
- ↳ **Une connaissance avancée de l'IA** et de son usage en entreprise. Beaucoup se présentent comme experts de l'IA, mais reproduisent toutes les erreurs démontrant leur méconnaissance de l'IA (attribuer un rôle, tutoyer l'IA, etc.). Nous faisons partie des quelques rares spécialistes à VRAIMENT bien connaître le sujet.



