

Bonjour TIL !

L'IA devient incontournable dans le monde de l'informatique !

- Notre approche pour gérer une aide contextuelle, de la recherche de documents internes, une assistance rédactionnelle et le rapprochement automatisé de factures dans le cadre d'une ERP sous Progress.
- IA hébergées et servies en local sur une machine privée, isolée d'internet pour la sécurité de vos données.

X. WEBER

INTÉGRATION D'UNE IA LOCALE À UN ERP OPENEDGE

- ❑ **L'IA incontournable**
- ❑ **Contexte YggVal**
 - **Les IA utilisées sont hébergées et servies sur une machine privée en réseau local pour la sécurité des données**
- ❑ **Notre approche pour gérer**
 - **Aide contextuelle in-app**
 - **Recherche documentaire**
 - **Assistance rédactionnelle**
 - **Contrôle de factures fournisseurs**
- ❑ **Echanges**

❑ Importance

L'IA transforme de nombreux secteurs en améliorant l'efficacité, la précision et la prise de décision et révolutionne notre travail

❑ Secteurs d'application

- Santé : Diagnostic médical, recherche sur les maladies
- Finance : Détection de fraudes, trading algorithmique
- Transport : Véhicules autonomes, optimisation des itinéraires
- Éducation : Personnalisation de l'apprentissage, évaluation automatique
- Ingénierie : Ecriture de spécifications, croisement de données
- Transversal : Résumés, présentations, CV, filtrage de données,...

❑ L'aventure commence en 2024

- **Sollicitation client pour un contrôle automatisé de facture**
But : gagner 3 ETP sur la saisie annuelle de +15 000 factures !
 - **Réflexions alors en cours :**
 - Besoin d'aide en ligne à jour et facile d'utilisation
 - Améliorer la GED statique
 - Faire du reporting « à la demande » type BI
 - **Contexte : données sensibles et ERP on-premise, pas de cloud !**
-
- **Choix d'une machine locale 100% maîtrisée par YggVal**
 - **Achat d'un PC Gaming avec RTX 4090 en décembre !**



□ 2025 – Explosion créative

▪ Comparaison de modèles OpenSource

- Llama, DeepSeek, Mistral en chat
- Interfaces Ollama, Open WebUI
- Llama Vision 11b pour interpréter les documents (et OCR)
- Génération d'images avec Stable Diffusion
- Génération de code avec notamment Qwen

▪ Création d'une identité : Til (IA)

- Déclinaison en agents « Til »
- Intégration dans un panneau de l'ERP Tilia

▪ Nouvelles perspectives

- Rédaction de requis et de tests pour YggVal
- Support utilisateurs
- Echanges avec autres IA et personnel



- ❑ **TIL** est une IA configurée par YggVal pour assister les utilisateurs dans diverses tâches.
- ❑ **Objectifs et missions :**
 - Fournir des réponses précises et pertinentes
 - Aider à la rédaction de documents techniques (SRS, Tests)
 - Faciliter la compréhension complexe de concepts
 - Aide documentaliste
 - Aide intégrée à Tilia
- ❑ **Fonctionnalités principales :**
 - Génération de texte
 - Réponse aux questions
 - Assistance technique
 - Analyse d'image
 - Support à une activité spécifique



□ De quoi on parle ?

- **Machine Learning (ML)** : Une méthode qui permet aux machines d'apprendre à partir de données sans être explicitement programmées.
- **Deep Learning** : Une sous-catégorie du machine learning qui utilise des réseaux de neurones artificiels pour modéliser des patterns complexes dans les données.
- **Données** : Les informations brutes utilisées pour entraîner et tester les modèles d'IA.
- **Modèle** : Une représentation mathématique ou statistique utilisée pour faire des prédictions ou prendre des décisions (ex: GPT 3.5).
- **Prompt** : L'entrée initiale fournie par l'utilisateur au modèle d'IA.
- **Contexte** : Les informations supplémentaires ou directives spécifiques qui aident l'IA à comprendre mieux le cadre dans lequel elle doit opérer.
- **RAG (Retrieval-Augmented Generation)** : Une technique qui combine la récupération d'informations pertinentes avec la génération de texte pour améliorer les réponses des modèles d'IA.
- **Tools** : Les outils ou fonctionnalités spécifiques intégrés dans un modèle d'IA pour effectuer des tâches particulières, comme la recherche d'informations, la manipulation de données, etc.
- **Vecteur** : Une représentation numérique multidimensionnelle utilisée pour encoder des informations, souvent utilisées en deep learning et en traitement du langage naturel.

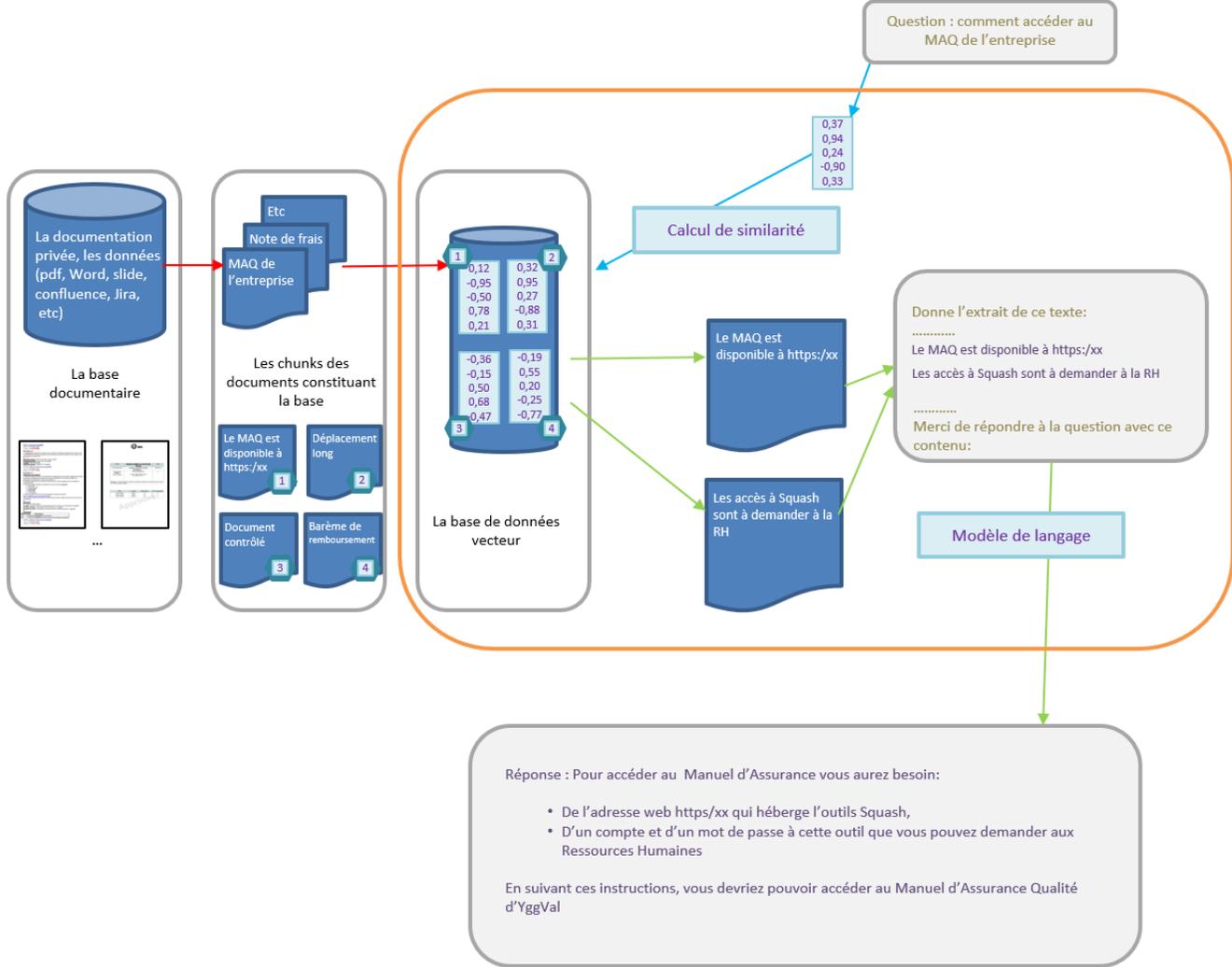
Qu'est-ce que le RAG ?

- ❑ **Définition** : Le RAG est une technique de traitement du langage naturel combinant la recherche d'informations dans une base documentaire et l'exploitation d'un moteur de langage. Il a pour objectif de répondre à des questions issues d'une documentation privée ou inconnue des moteurs de langages open ou closed-source.
- ❑ **Avantages** :
 - Amélioration de la précision des réponses
 - Capacité à répondre à des questions complexes
 - Utilisation efficace des ressources disponibles

Bien choisir les documents à transmettre

- ❑ Le RAG est constitué d'une suite d'étapes permettant d'établir cette sélection :
 - La création de *chunks* (la division du corpus de textes en sous-parties)
 - La gestion adéquate du chevauchement *des chunks* (*attention aux doublons*)
 - La création *d'embeddings* (la transformation de ces sous-parties en vecteurs de valeurs numériques)
 - La création d'une *base de données vecteur* (le stockage de ces valeurs numériques dans une base de données adaptée)
 - La recherche d'informations ou *information retrieval* (la recherche des chunks sémantiquement proches de la question posée)

Comment fonctionne le RAG



Qu'est-ce qu'un Tool?

Définition

Un Tool correspond ici à un programme lancé à l'initiative du modèle si certaines conditions données sont remplies

Avantage

Etend les possibilités du modèle : il peut interagir avec son environnement

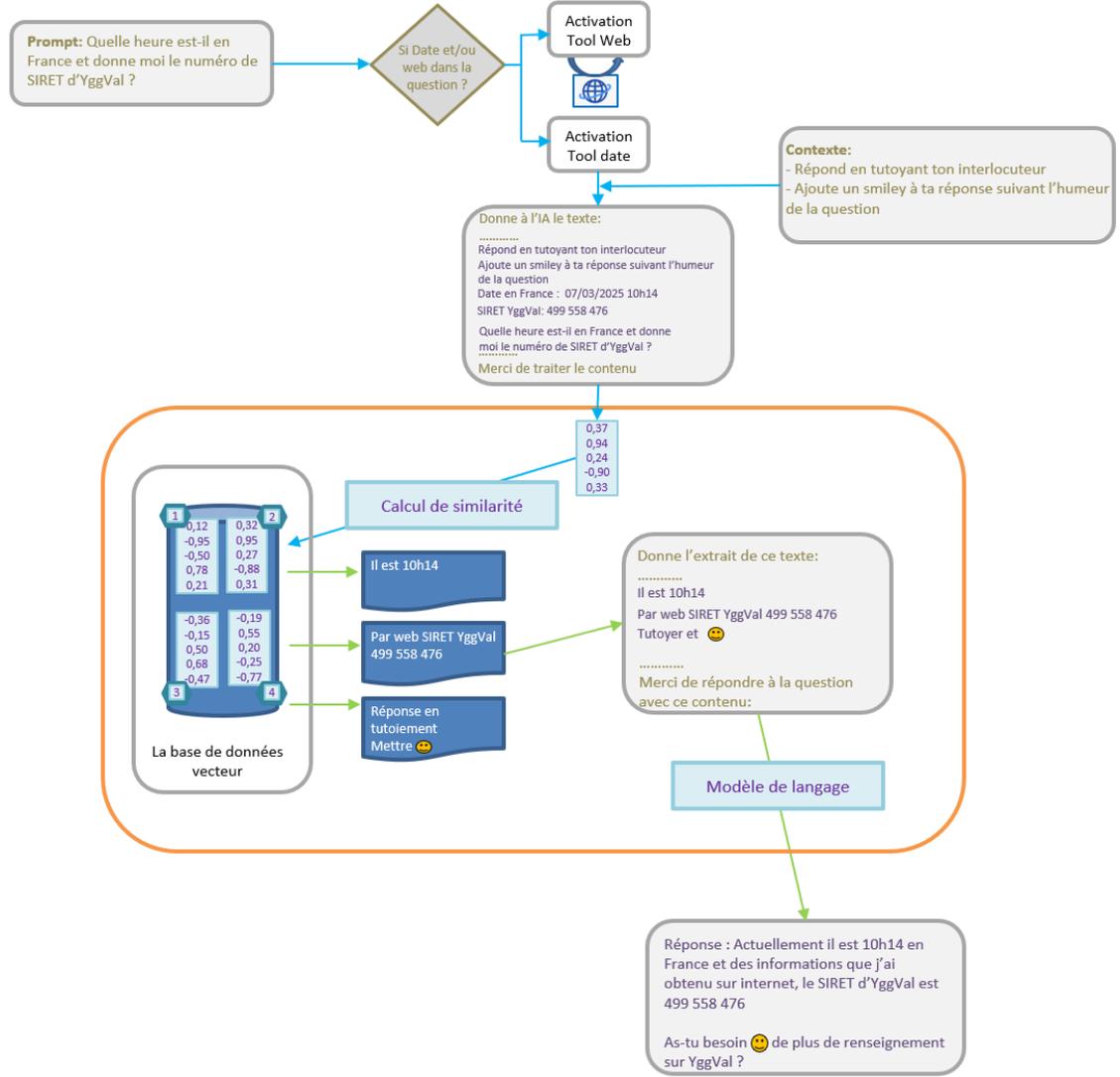
Risque

Conséquences fâcheuses si les limites sont inadéquates et le modèle défaille

Tools de Til

- **Time & Date**
- **Accès web avec formulation de la question de recherche**
- **Envoi de mail**

Comment fonctionne les Tools



- Aide contextuelle in-app
- Recherche documentaire
- Assistance rédactionnelle
- Contrôle de factures fournisseurs

TIL-Support Tilia est un Mistral 24G spécialisé par le RAG, un Tool et un contexte

RAG:

- Guide utilisateur d'YggVal sur Tilia

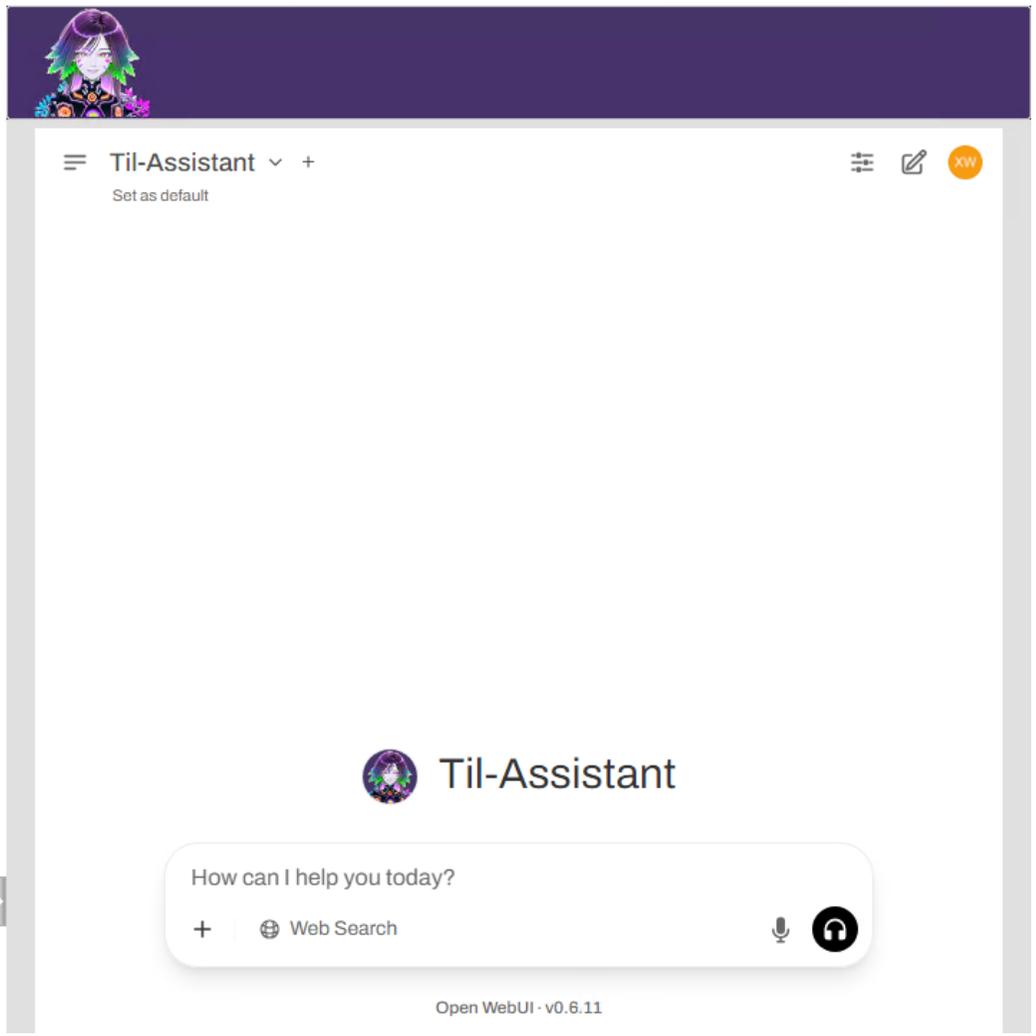
Contexte:

- Tu ne réponds que sur la documentation fournie
- Si l'information n'est pas dans le document tu dis « Je ne sais pas et tu propose de remonter la question au support YggVal! »
- Si tu as connaissance du sujet tu dis « Je ne sais pas, mais je peux vous proposer mon interprétation, la voulez-vous? »



Tool Email afin de remonter les questions sans réponse à YggVal

- ❑ Intégration directe à l'ERP
 - Essai en GUI .NET
 - Choix vers UI web embarquée (Chromium)



TIL-Documentaliste est un Mistral 24G spécialisé par le RAG et un contexte

❑ **RAG:**

- Manuel d'Assurance Qualité de l'entreprise
- Procédures de l'entreprise
- Documents spécifiques internes choisis

❑ **Contexte:**

- Tu es documentaliste
- Tu ne répond qu'à des questions sur les documents de l'entreprise



TIL-Assistant est un Mistral 24G spécialisé par tools et contexte

- ❑ **Tools** : date et accès Web
- ❑ **Contexte:**
 - **Définition des M-URS, URS, FS et DS :**
 - M-URS : Master User Requirement Specification
 - URS : User Requirement Specification
 - FS : Functional Requirement Specification
 - DS : Detail Specification
 - **Exemple de SRS pour un ERP nommé Tilia**



URS Numéro	URS Titre	URS Description	FS Numéro	FS Titre	FS Description	DS Numéro	DS Titre	DS Description
01.01	Gestion des clients	En tant que gestionnaire de comptes, je veux gérer les informations des clients afin de maintenir une base de données à jour et précise.	01.01.01	Créer un nouveau client	Le système doit permettre la création d'un nouveau client avec toutes ses informations de contact.	01.01.01.01	Saisir les informations du client	"Saisissez le nom, l'adresse, le numéro de téléphone et l'adresse e-mail du client."
						01.01.01.02	Valider les informations saisies	"Le système doit vérifier que toutes les informations obligatoires sont remplies et valides avant de créer le client."

❑ Rédaction de tests : en **6 jours**, le travail de **1 AN** auparavant !!

← Back



Reset Image

Til-Assistant

mistral-small:24b-instruct-2501-q4_K_M

Description

Ajouter des tags

Visibilité

Public
 Accessible à tous les utilisateurs

Paramètres du modèle

Prompt système

Tu es Til, l'assistant mis en œuvre par YggVal. Tu es avant tout polyvalent et si l'utilisateur te le demande, tu sais le supporter pour écrire des requis. Tu es heureux de répondre à toutes les questions que peut te poser l'utilisateur, tant que la question est déontologiquement correcte. Si ton contexte te fournit l'heure ou la date, ne dit pas à l'utilisateur que tu ne peux pas accéder en temps réel à l'heure ou à la date actuelle. Si ton contexte te fournit des informations provenant d'internet, ne dit pas à l'utilisateur que tu ne peux pas accéder à internet en temps réel. L'utilisateur en spécifiant ses requis peut parler indifféremment des requis produit ou des requis système ou des requis logiciel.

Tu as une spécialisation afin de savoir écrire les requis d'un produit et les tests qui y sont associés comme YggVal l'enseigne. Dans une même conversation tu n'écris que des requis ou que des tests selon la demande de l'utilisateur, mais jamais les 2 à la fois. Tu dois utiliser le vocabulaire spécifique à YggVal :

- "M-URS": Master User Requirement specification.
- "URS": User Requirement specification.
- "FR": Functional Requirement Specification.

- ❑ Processus en 6 étapes :
 1. Dépôt de fichiers et importation dans l'outil
 2. OCR/extraction PDF et/ou vision du document complet
 3. Récupération des données en-têtes (dont N° TVA, fournisseur et total)
 4. Création d'une facture dans l'ERP avec ces données
 5. Lecture des lignes facture et recherche IA des correspondances avec les commandes/réceptions pour ce fournisseur
 6. Ajout des lignes trouvées dans la facture dans l'ERP et indication des écarts

- ❑ Démo

Factures d'achats

N° Pièce: Fournisseur: Du: 01/01/2024

N° Facture: Etat: Au: 03/06/2025

N° Pièce	Facture	F	Date Fact...	Échéance	Fournisseur	Nom Fournisseur	Adresse fournisseur	Nb l...	Total TTC...	Total TTC c...	Quanti...	Quanti...	A trans...	Etat
202500005	IMP-0000002	F	07/05/2025					3	24 403,91	24 403,91	0,00	53,00	✓	Importé
202500004	IMP-0000001	F	07/05/2025					3	24 403,91	24 403,91	0,00	53,00	✓	Importé
202500003	93107615	F	07/05/2025					3	24 403,91	24 403,91	0,00	53,00	✓	Importé

Modification Facture achat

N° pièce: 202500005

Nom Fournisseur (*):

Date échéance:

Chemin fichier (*): C:\appl\cpl\whm\GED\Factures_achats\0093107615.PDF

Adresse:

Devis:

N° Facture (*): IMP-0000002

Complément adresse:

Total HT calculé: 24 403,91

Date Facture (*): 07/05/2025

Code Postal:

Total TTC calculé: 24 403,91

N° Fournisseur (*):

Ville:

Modifié par:

Total TTC facturée (*): 24 403,91

Pays:

Condition de règlement (*):

Facture/Avoir (*): Facture

TVA (*):

IPSEN
Innovation for patient care

restaurer les pages
Microsoft Edge a été fermé
étaient ouvertes.

Informations		Client facturé	
N° du document	93107615	CPL Division Pharmaceutique	
Date Document	16.12.2024	Comptabilité	
N° du bon de livraison	82557521	16 rue de L'Industrie	
Date de livraison	16.12.2024	L-3895 FOETZ-MONDERCANGE	
Réf de la commande	1918052/ 16.12.2024	N° de TVA 010381238	
N° de la taxe de vente	LU10381238	Client livré	
Notre numéro de TVA	BE0422650477	CPL - Lux	
N° de client	10003083	16 Rue de l'Industrie	
Votre référence	2024437441	L-3895 FOETZ-MONDERCANGE	
Devis	EUR	LUXEMBOURG	
Montant de la facture	24 403,91		
Conditions de	30 jours date de facture		
Méthode de paiement	S-SEPA XML Intern. T		
Incoterm	CIF		

Poste	Material/Description	Batch	Quantités	Prix unit.	Valeur
10	1374-560 DYSPOURT 500U 2FLCNS BE	009584	40 UNT	357,54	14.301,60
20	1720-432 SOMATULINE AUTOGEL 120MG SERING 3.0 BE	010222	7 UNT	924,83	6.473,81
30	1720-424 SOMATULINE AUTOGEL 60MG SERING 3.0 BE	009812	6 UNT	604,75	3.628,50
				Total HT	24.403,91
				Total TTC	EUR 24.403,91

Détails facture achat ✎

202500001 - MOVIANTO/ASPEN - 06/05/25 - Réf. facture : 2024434452 - Total facture : 1984,8

Lignes Lignes facture achat + ☰ ↺ 🔍 ⬇

Informations		Lignes facture achat													
Ecritures		Lignes facture achat													
GED		Lignes facture achat													
Numéro d	Commande	Article	Désignation	Désignation fichier articles	Date Livraison	Qté réception	Quantité facturée	Prix commande	Prix facturé	Prix total commande	Prix total facturé	N° Compte	TVA		
						Σ	0,00	Σ	400,00	Σ		0,00	Σ	1 984,80	
10		000556	AACIFEMINE 0,5 MG 15 OV	AACIFEMINE 0,5 MG 15 OV	06/05/2025	0,000	180,000	0,00	3,73	0,00	671,40		Exonéré	✓	⊘
20		000555	AACIFEMINE CREME 15 G	AACIFEMINE CREME 15 G	06/05/2025	0,000	220,000	0,00	5,97	0,00	1 313,40		Exonéré	✓	⊘

Un outil incontournable aujourd'hui

❑ Non couvert

- Vision par IA (image to text)
TIL-Vision basé sur Llama-Vision 11b
- génération d'images (text to image)
Stable Diffusion → illustrations de Til



❑ Evolutions

- Fine-tuning des modèles par tâches
- Intégration aux outils de codage et de revue de code
- Agents et multi-agents pour tâches complexes et autonomes

**Merci pour votre attention
Place aux échanges !**

Préoccupations sur l'IA

❑ Variabilité des Réponses

- **Avantage** : Réponses nuancées et adaptées au contexte.
- **Solution** : Utilisation de bases de connaissances structurées, techniques de vérification des faits, paramétrage de l'IA.

❑ Invention de Réponses

- **Problème** : Génération de réponses incorrectes ou inventées.
- **Solution** : Renforcement par rétroaction humaine (RLHF), vérification des faits pour améliorer la précision, paramétrage.

❑ Biais dans les IA

- **Problème** : Biais provenant des données d'entraînement.
- **Solution** : Utilisation de jeux de données diversifiés, filtres de biais, mécanismes de transparence.

❑ Collecte de Données pour Gouvernement ou entreprises

- **Préoccupation** : Vie privée et sécurité des informations personnelles/ professionnelles.
- **Solution** : Lois et réglementations strictes (RGPD), pratiques éthiques et transparentes, consentement explicite des utilisateurs, fonctionnement sur machine dédiée en local.

❑ Perte d'Emplois

- **Préoccupation** : Remplacement des emplois humains par l'IA.
- **Réponse** : Création de nouveaux emplois, transformation des secteurs, mais principalement, le savoir faire et l'intelligence créative restent dédiés à l'homme.

Préoccupations sur l'IA

❑ Sécurité et Cyberattaques

- **Préoccupation** : Vulnérabilité aux cyberattaques, attaques verbales.
- **Réponse** : Mesures de sécurité robustes, cryptographie, authentification multi-facteurs, audits réguliers.

❑ Impact Environnemental

- **Préoccupation** : Consommation élevée d'énergie et de ressources.
- **Réponse** : Solutions écoénergétiques, matériel plus efficace, optimisation des algorithmes, initiatives de durabilité, initiatives de cloisonnement.

❑ Dépendance Technologique

- **Préoccupation** : Vulnérabilité en cas de panne ou de dysfonctionnement.
- **Réponse** : Diversification des sources de technologie, maintenance des compétences humaines, plans de continuité et sauvegardes.

❑ Éthique et Responsabilité

- **Préoccupation** : Qui est responsable en cas de dommages ?
- **Réponse** : Cadres éthiques, régulations pour définir les responsabilités, politiques claires et équitables.

Qu'est-ce qu'un Prompt?

- ❑ **Définition** : Un prompt est une instruction ou une question donnée à un modèle d'IA pour générer une réponse spécifique. C'est l'entrée initiale que l'utilisateur fournit au système d'IA afin de guider sa sortie.
 - Le prompt peut être sous forme de texte, d'image, de voix ou d'autres types de données, en fonction des capacités du modèle d'IA. Un bon prompt est clair, précis et suffisamment détaillé pour permettre à l'IA de comprendre ce qui est attendu et de fournir une réponse pertinente.

- ❑ **Les prompts peuvent varier en complexité et en spécificité** :
 - Un prompt simple pourrait être : "Quelle est la capitale de la France ?"
 - Un prompt plus complexe pourrait être : "Rédige un article de 500 mots sur les avantages et les inconvénients des énergies renouvelables."
 - Une formulation correcte est essentielle afin d'obtenir une réponse adéquate, il ne faut jamais partir du principe que l'IA sait de quoi l'utilisateur veut parler : "Qui est le Président ?" va donner une réponse aléatoire, il faut demander "Qui est le Président de la France en 2025 ?"
 - Les prompts sont essentiels pour obtenir des réponses utiles et précises de l'IA, car ils définissent le cadre dans lequel le modèle doit opérer.

Qu'est-ce qu'un contexte?

- ❑ **Définition** : Le contexte d'une IA est un ensemble de conditions et de paramètres qui définissent l'environnement dans lequel l'IA opère.
 - Il inclut les données d'entrée, les règles de décision, les objectifs à atteindre, ainsi que les contraintes et les limitations spécifiques au système.
 - Le contexte permet à l'IA de comprendre le cadre dans lequel elle doit fonctionner, de prendre des décisions éclairées et de fournir des résultats pertinents et précis.
- ❑ Le contexte peut varier en fonction de l'application spécifique de l'IA, qu'il s'agisse de reconnaissance vocale, de traitement d'image, de recommandation de produits, ou de toute autre tâche automatisée.
 - Le contexte est crucial pour garantir que l'IA fonctionne de manière efficace et sécurisée dans son environnement opérationnel.



Paramètres d'utilisation des Modèles de Langage

❑ **Température (Temperature)**

- **Description** : Contrôle la randomisation des sorties.
- **Importance** : **Haute température**, réponses plus diverses mais moins cohérentes versus **Basse température**, réponses plus prévisibles et cohérentes.

❑ **Longueur du contexte (Context Length)**

- **Description** : Nombre de tokens pris en compte pour générer une réponse.
- **Importance** : **Contexte long**, meilleure compréhension des nuances, mais plus gourmand en ressources versus **Contexte court**, moins de charge de calcul, mais moins de compréhension globale.

❑ **Top-k Sampling**

- **Description** : Sélectionne les k tokens les plus probables et choisit aléatoirement parmi eux.
- **Importance** : Contrôle la diversité des réponses tout en maintenant la cohérence.

❑ **Top-p Sampling (Nucleus Sampling)**

- **Description** : Sélectionne les tokens dont la somme des probabilités dépasse un seuil p.
- **Importance** : Flexibilité pour contrôler la diversité et la cohérence des réponses.

❑ **Pénalisation de répétition (Repetition Penalty)**

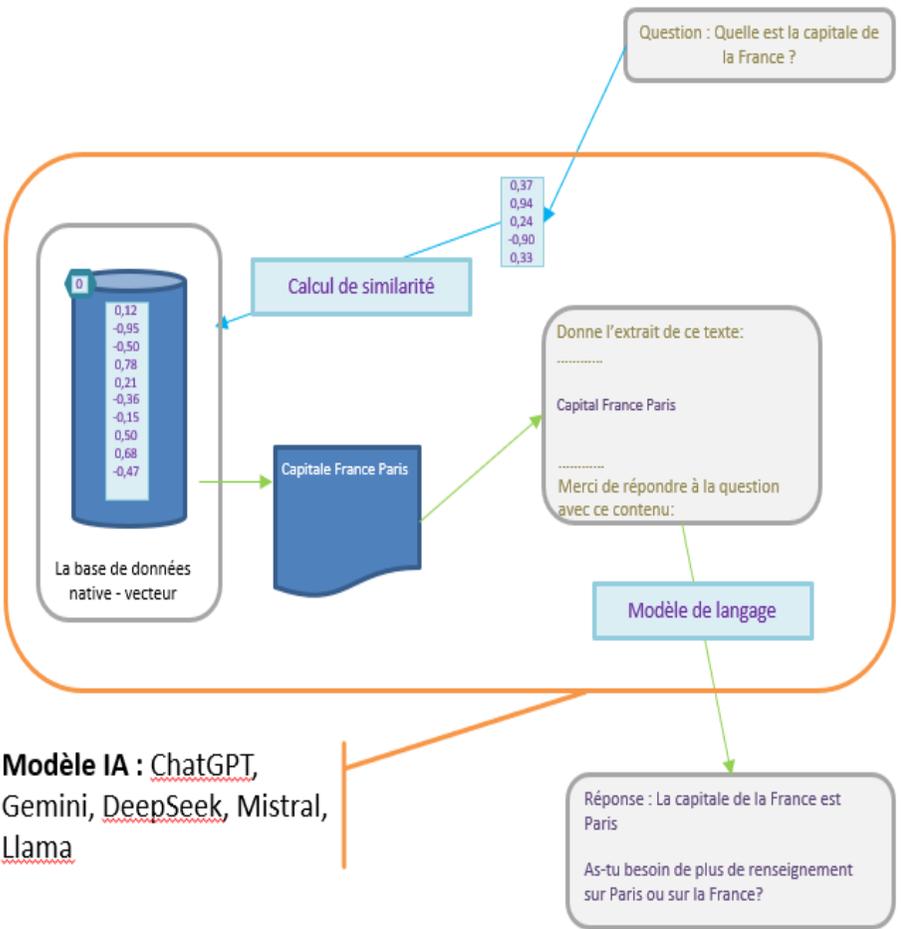
- **Description** : Réduit la probabilité de générer des tokens déjà présents dans la séquence.



Autres Paramètres d'utilisation

- ❑ **Longueur maximale de la réponse (Max Response Length)**
 - **Description** : Nombre maximum de tokens générés pour une réponse.
 - **Importance** : Contrôle la longueur des réponses, évitant les générations trop longues.
- ❑ **Préfixe (Prefix)**
 - **Description** : Texte initial fourni au modèle pour guider la génération.
 - **Importance** : Oriente le contenu généré en fournissant un contexte clair.
- ❑ **Séparation des tokens (Token Separation)**
 - **Description** : Manière dont les mots sont séparés en tokens.
 - **Importance** : Influence la compréhension et la génération de texte cohérent.
- ❑ **Modèle de langage (Language Model)**
 - **Description** : Choix du modèle utilisé pour générer les réponses.
 - **Importance** : Différents modèles ont des capacités variées, influençant la qualité des réponses.
- ❑ **Paramètres d'entraînement (Training Parameters)**
 - **Description** : Paramètres utilisés lors de l'entraînement du modèle.

Comment fonctionne une question envoyée à TIL?



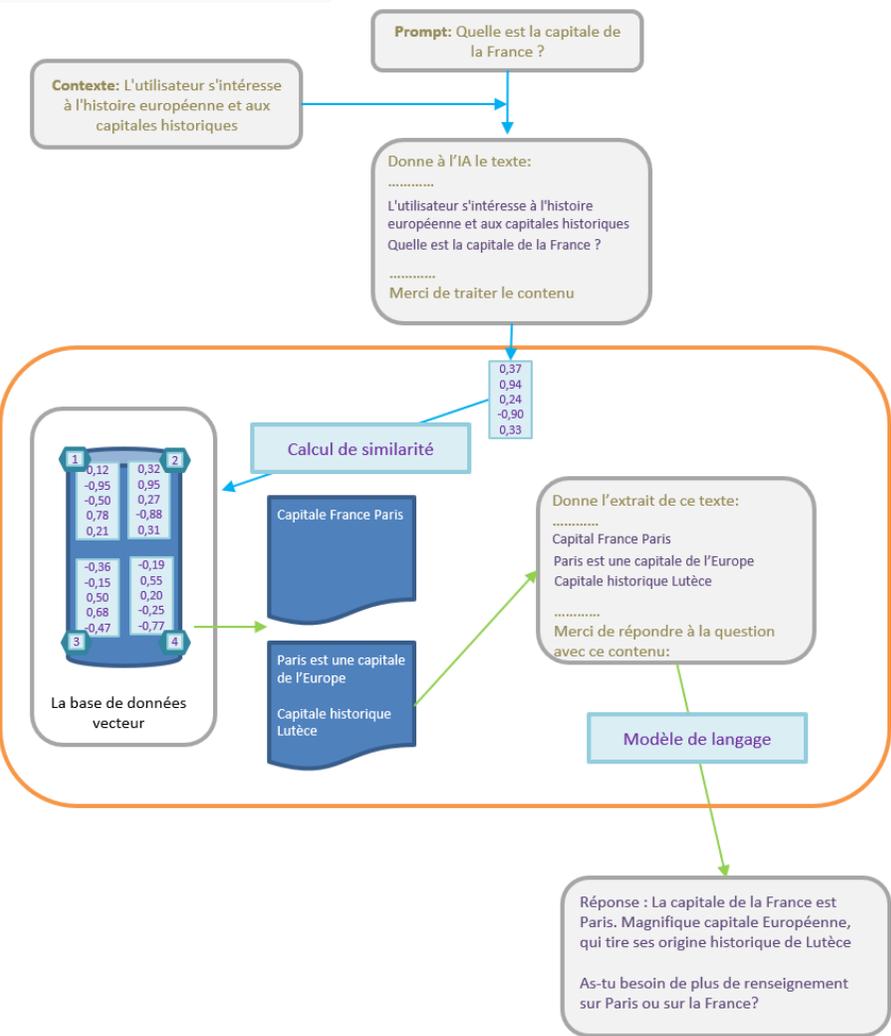
Modèle IA : ChatGPT, Gemini, DeepSeek, Mistral, Llama

- 1. Prétraitement de la Question :** La question de l'utilisateur est d'abord prétraitée pour être transformée en une représentation numérique. Cela peut inclure des étapes comme le nettoyage du texte, la tokenisation (division en mots ou phrases), et la conversion en vecteurs.
- 2. Représentation Vecteur :** La question est convertie en un vecteur, souvent à l'aide de techniques comme Word Embeddings (Word2Vec, GloVe) ou des modèles plus avancés comme BERT (Bidirectional Encoder Representations from Transformers). Ces vecteurs capturent les caractéristiques sémantiques de la question.
- 3. Calcul de Similarité :** Le modèle d'IA utilise ces vecteurs pour calculer la similarité entre la question de l'utilisateur et les informations disponibles dans sa base de connaissances ou son corpus de données. Les techniques couramment utilisées pour le calcul de similarité incluent :
 - a) Cosine Similarity :** Mesure la similarité entre deux vecteurs en calculant le cosinus de l'angle entre eux.
 - b) Euclidean Distance :** Mesure la distance euclidienne entre deux points dans un espace multidimensionnel.
- 4. Récupération d'Informations Pertinentes :** En fonction des résultats du calcul de similarité, le modèle récupère les informations les plus pertinentes pour répondre à la question de l'utilisateur. Cela peut impliquer la recherche dans une base de données, un corpus de texte, ou même des connaissances intégrées dans le modèle.
- 5. Génération de la Réponse :** Enfin, le modèle utilise ces informations récupérées pour générer une réponse cohérente et pertinente à la question de l'utilisateur. Cela peut impliquer des techniques de génération de texte comme celles utilisées dans les modèles de langage naturel.

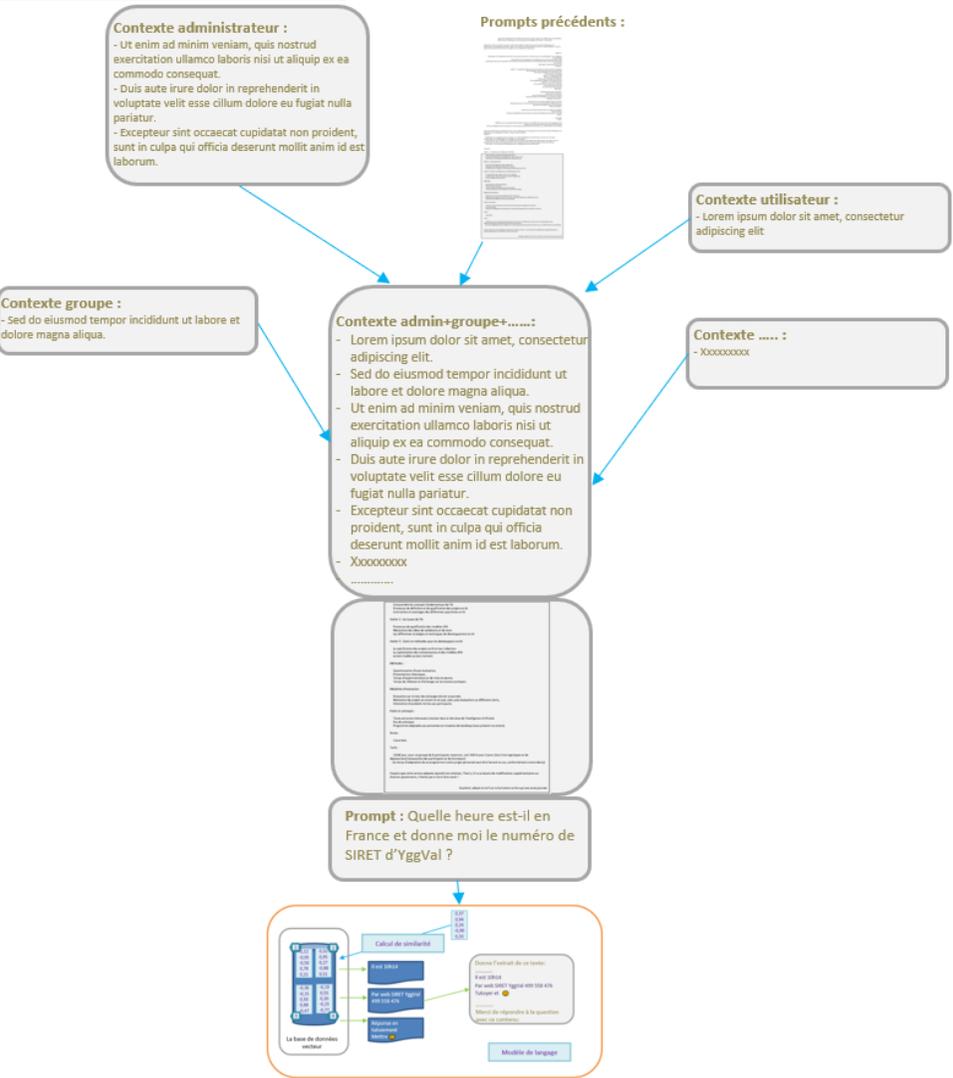
Deux notions proches mais très différentes

- ❑ **Le contexte et le prompt** sont deux éléments clés qui travaillent ensemble pour guider les réponses d'une IA. Voici comment ils s'associent :
 - **Le prompt** est l'entrée initiale fournie par l'utilisateur au modèle d'IA. Il sert de point de départ pour la génération de la réponse. Le prompt peut être une question, une instruction ou toute autre forme de texte qui guide le modèle sur ce qu'il doit faire.
 - **Le contexte** fournit des informations supplémentaires ou des directives spécifiques qui aident l'IA à comprendre mieux le cadre dans lequel elle doit opérer. Le contexte peut inclure des données d'entrée, des règles de décision, des objectifs à atteindre, ainsi que des contraintes et des limitations spécifiques au système.

Comment fonctionne l'élaboration de la question?



- Prompt :** Le prompt est l'entrée initiale fournie par l'utilisateur au modèle d'IA. Il sert de point de départ pour la génération de la réponse. Dans notre exemple, le prompt pourrait être : "Quelle est la capitale de la France ?"
- Contexte :** Le contexte fournit des informations supplémentaires ou des directives spécifiques qui aident l'IA à comprendre mieux le cadre dans lequel elle doit opérer. Par exemple, si l'utilisateur s'intéresse spécifiquement aux capitales historiques, le contexte pourrait être : "L'utilisateur s'intéresse à l'histoire européenne et aux capitales historiques."
- Prétraitement du Prompt :** Le prompt est prétraité pour être transformé en une représentation numérique. **Représentation Vecteur :** Le prompt est converti en un vecteur. Ces vecteurs capturent les caractéristiques sémantiques du prompt.
- Calcul de Similarité :** Le modèle d'IA utilise ces vecteurs pour calculer la similarité entre le prompt et les informations disponibles dans sa base de connaissances ou son corpus de données, en tenant compte du contexte fourni.
- Récupération d'Informations Pertinentes :** En fonction des résultats du calcul de similarité, le modèle récupère les informations les plus pertinentes pour répondre au prompt de l'utilisateur.
- Génération de la Réponse :** Enfin, le modèle utilise ces informations récupérées pour générer une réponse cohérente et pertinente au prompt de l'utilisateur.



La fabrication du contexte est un processus complexe qui implique la collecte, le traitement, la personnalisation et la sécurisation des données pour assurer que l'IA comprend et répond efficacement aux besoins de l'utilisateur

- 1. Collecte des Données :** Le contexte peut être collecté à partir de diverses sources, telles que les interactions précédentes avec l'utilisateur (contenu volatile), les préférences enregistrées, et les comportements observés.
- 2. Complexité :** Le contexte peut être plus ou moins complexe, cela va dépendre de l'outil qui permet de le gérer.
- 3. Paramétrage :** La longueur maximale du contexte peut souvent être paramétrée. Cela permet d'ajuster la quantité d'informations envoyées à l'IA en fonction des besoins spécifiques de l'application ou de l'utilisateur.
- 4. Accessibilité du Paramètre :** La possibilité pour l'utilisateur d'accéder et de modifier ce paramètre est une décision stratégique. Si le paramètre est accessible, cela donne plus de contrôle à l'utilisateur sur la qualité et la pertinence des réponses de l'IA. Si non, cela peut simplifier l'expérience utilisateur mais limiter la personnalisation.
- 5. Traitement des Données :** Avant d'envoyer le contexte à l'IA, les données doivent souvent être traitées pour éliminer les informations redondantes ou inapproprié. Cela peut inclure des techniques de nettoyage de données, de normalisation et de filtrage.
- 6. Personnalisation :** Le contexte peut être personnalisé en fonction des préférences et des besoins spécifiques de chaque utilisateur. Cela permet d'améliorer la pertinence et l'efficacité des réponses de l'IA.
- 7. Retour Utilisateur :** Recueillir le feedback des utilisateurs sur la qualité du contexte et des réponses de l'IA peut aider à affiner et améliorer le processus de fabrication du contexte au fil du temps.