

Rencontre ALLiance du 2 Septembre 2025

Utilisation d'un RAG – Retour d'expérience

Dossier
Facile

Julien Henry - opérateur chez DossierFacile



Financé par
l'Union européenne
NextGenerationEU



Dossier Facile est financé à hauteur de 900 000€ par le Plan de Relance

Sommaire

- **I Introduction**
 - I.1 Contexte et objectifs
 - I.2 Définitions
- **II Mise en place du RAG**
 - II.1 Webhooks sur MM
 - II.2 Le workflow
 - II.3 Ingestion documentaire
- **III La documentation**
 - III.1 Quelques définitions préalables
 - III.2 Pourquoi adapter sa documentation
 - III.3 Différence entre une documentation adaptée ou non
- **IV Exploiter l'historique des conversations**
- **V Considérations annexes**

Contexte

- **Le flux de dossier à traiter augmente de manière exponentielle en été**
- **Accueil de 20/30 nouveaux opérateurs de validation débutants en quelques semaines.**
- **Tous les opérateurs ne lisent pas la documentation.**
- **Tous les opérateurs ne posent pas leurs questions sur le Mattermost**
- **La formation de nouveaux opérateurs est chronophage**

Objectifs

- **Accélérer la formation des nouveaux opérateurs**
- **Améliorer la formation continue des anciens opérateurs**
- **Identifier les lacunes documentaires**
- **Standardiser nos traitements**

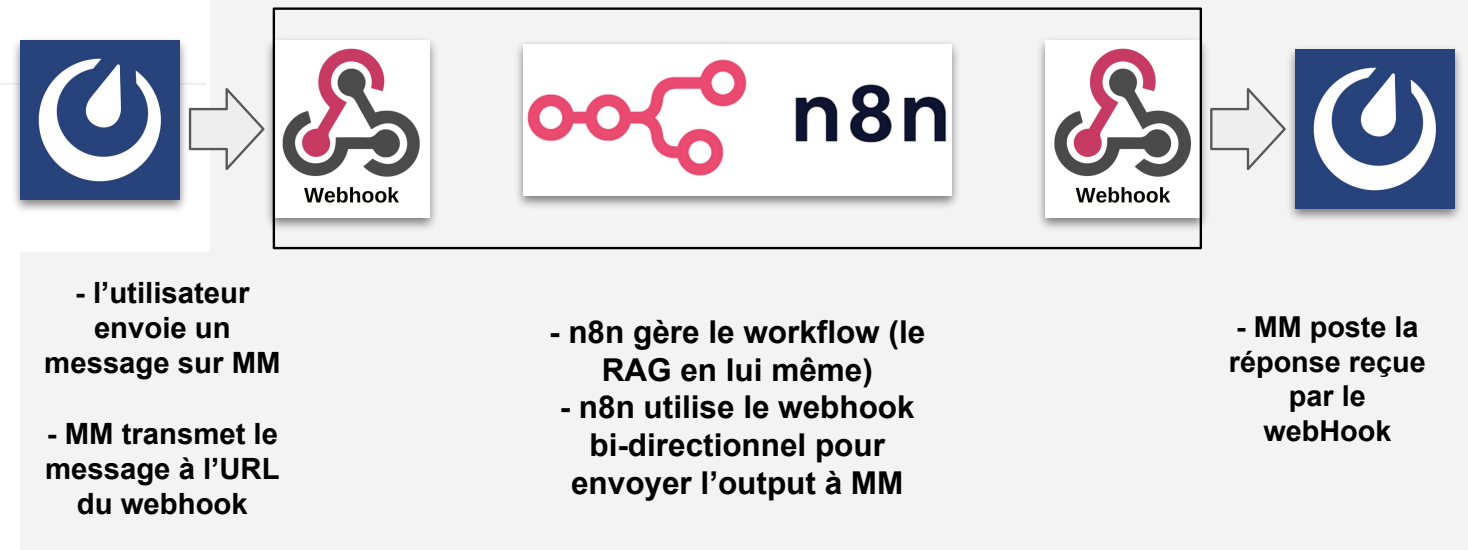
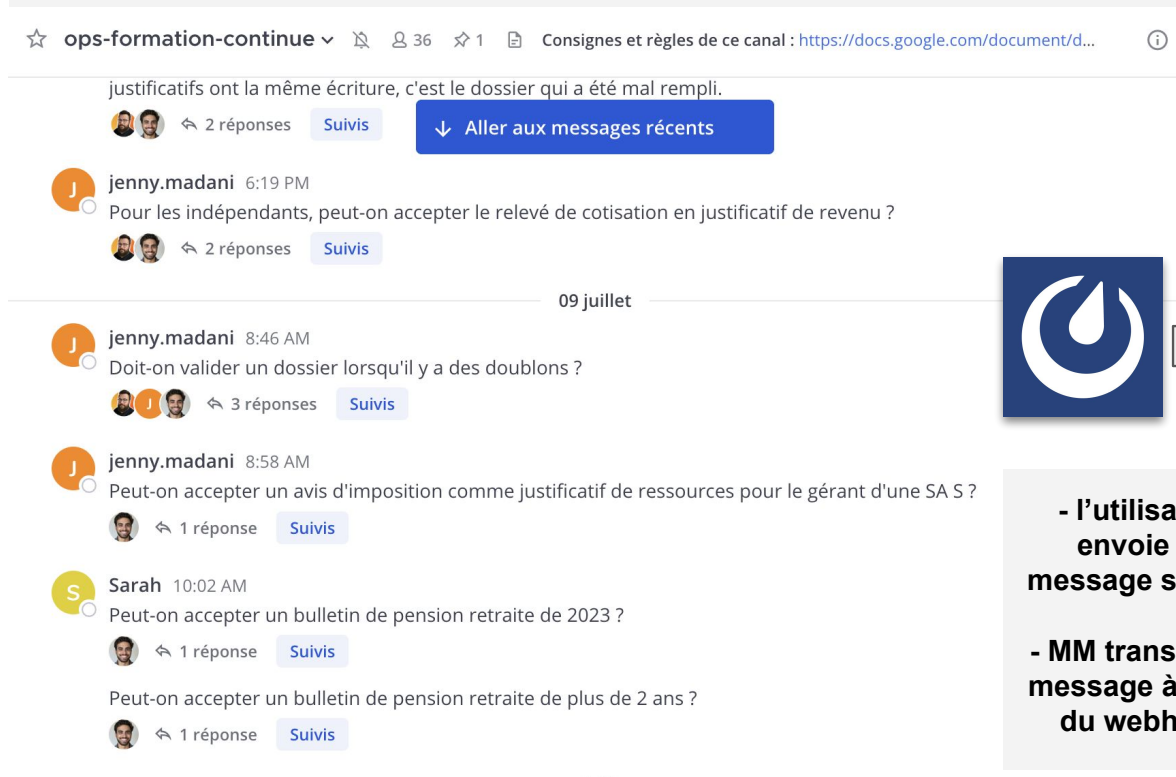
Définitions

- **RAG (Retrieval Augmented Generation)** : un chat bot personnalisé qui se base d'abord sur une base de données interne. Capacité à dire qu'il ne sait pas répondre à une questions
- **N8n** : *Outil d'automatisation open-source qui connecte des applications (API, services) entre elles via des workflows visuels – low-code*
- **Webhook** : Système de notification automatique qui transmet des données en temps réel d'une application à une autre via une URL, sans action manuelle.




"Un webhook, c'est l'API qui vous appelle au lieu d'attendre que vous l'appeliez."





II - Mise en place

Le bot répond aux questions envoyées dans un canal Mattermost spécifique (entre 5 et 40 sec de délais). Les réponses sont publiées dans le thread. Mise en place possible sans disposer des droits administrateurs sur Mattermost. Cette solution ne permet pas créer un vrai BOT avec qui on peut discuter en message privé. Pour cela il faut : accès admin + websocket.



II.1 Mise en place – Webhooks sur MM

-  Channels ✓
-  Boards
-  Playbooks

-  Intégrations
-  Groupes d'utilisateurs
-  Télécharger les applications
-  À propos de beta.gouv.fr

FREE EDITION

This is the free **unsupported** edition of Mattermost.



Webhooks sortants

Les webhooks sortants permettent aux intégrations externes de recevoir et répondre aux messages

Canal

Ce champ est facultatif si vous spécifiez au moins un mot déclencheur. Spécifiez le canal public qui fournit la charge utile au webhook.

Mot-clé (Un par ligne)

Spécifie les mots-clés déclencheurs qui envoient des requêtes HTTP POST à votre application. Le déclencheur peut être pour le canal, le webhook sortant, ou les deux. Si vous sélectionnez le canal, le déclencheur est facultatif. Si vous sélectionnez les deux, le message doit correspondre aux deux valeurs.

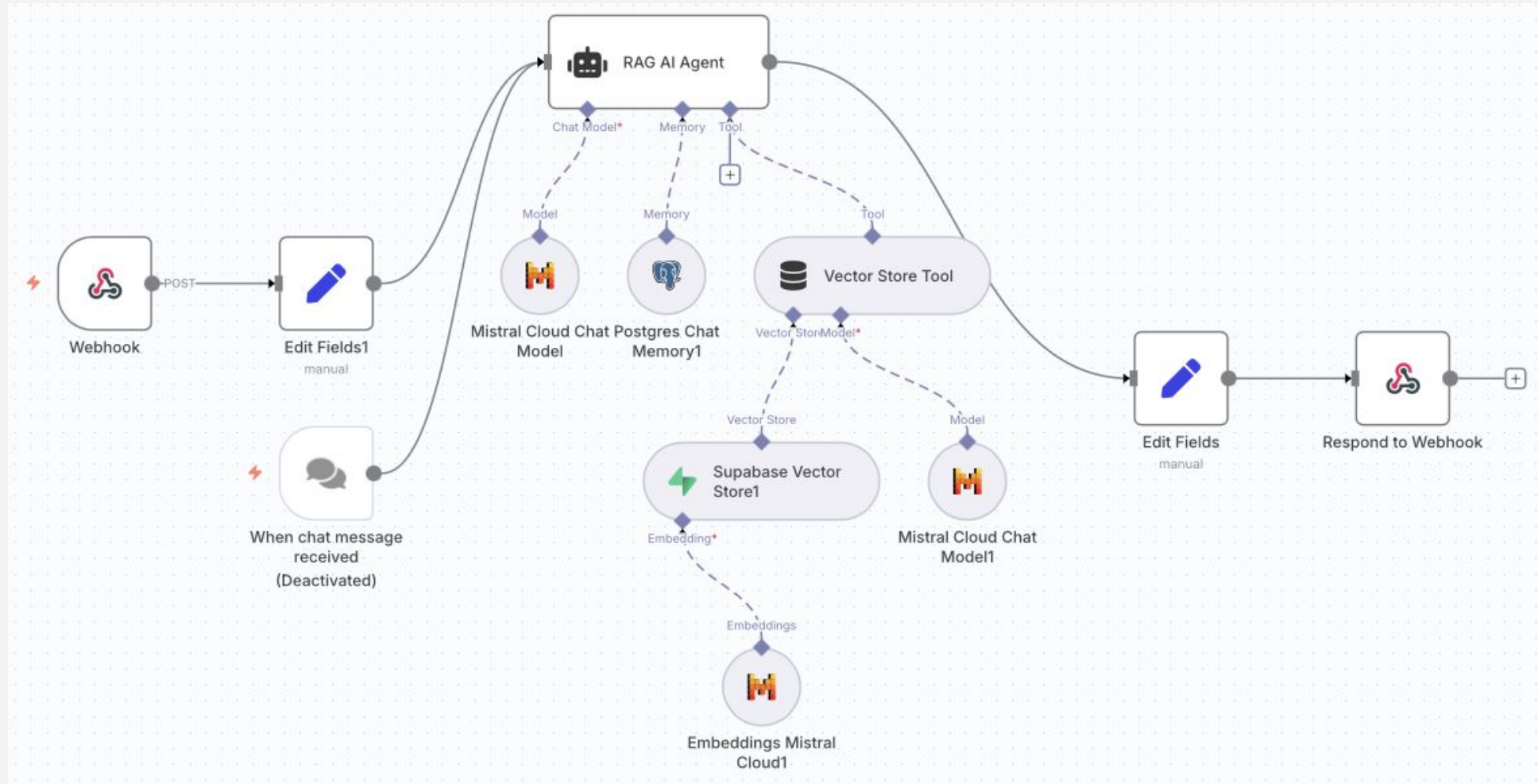
Déclencher lorsque

Spécifie à quel moment déclencher le webhook sortant.

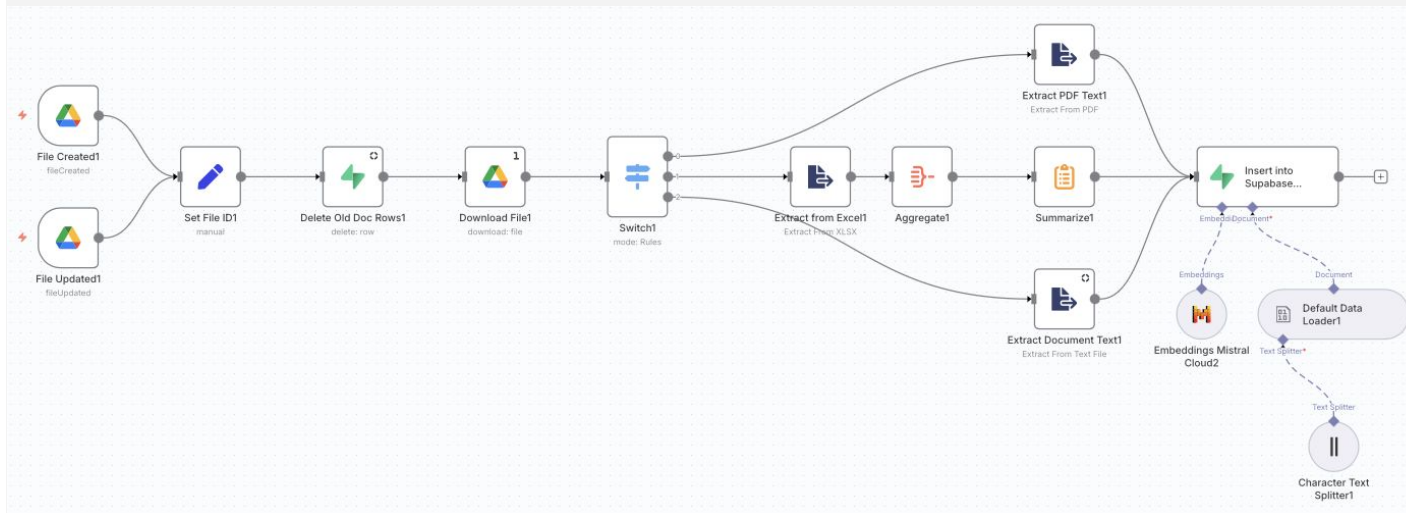
URLs de rappel (une par ligne)

Spécifie l'URL à laquelle les messages vont être envoyés. Si l'URL est privée, ajoutez-la en tant que [connexion interne fiable](#).

II.2 Mise en place – Le workflow

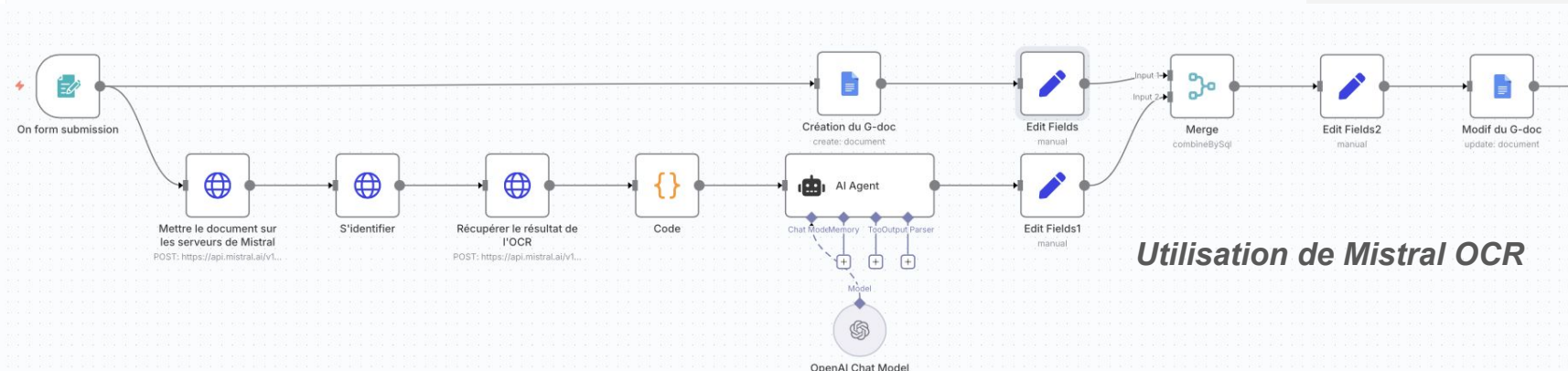


II.3 Mise en place – Ingestion documentaire



BASE DOCUMENTAIRE :

- L'intégralité de la documentation publique de DossierFacile
- La documentation interne qui explicite nos process de traitement





III.1 - Qu'est ce qu'un token ?

69 mots 448 caractères

Token count

140

12 sept. 2024

Garant – Justificatif de situation professionnelle –
Général

Grands principes

Les pièces validées apparaîtront telles quelles chez
le propriétaire ou l'agence qui les recevra. Pour la
réputation du service DossierFacile, il est donc cruc
ial de maintenir un standard élevé de validation.

En particuliers, seront systématiquement refusées :
les pièces contenant des arrière-plans perturbateurs
les captures d'écran contenant des éléments perturbat
eurs

les documents inutiles (Tolérance en fonction des ca
s)

Un token = environ 3,2 caractères

Unité de référence pour quantifier
l'information dans le cadre du
traitement du langage.

Découpage différent selon les
modèles

<https://compteur-de-mots.ouest-france.fr/>

<https://tiktokenizer.vercel.app/?model=gpt-4-1106-preview>



III.1 Qu'est ce que le « ChunkSize » et l'Overlap ?

- **Chunk size** : Taille des morceaux de texte découpés pour être analysés. Entre 200 et 1000 tokens
- **Chunk Overlap** : Chevauchement entre morceaux pour éviter de perdre le contexte. Entre 10 et 20% du chunk size



Taille	Avantages	Inconvénients
Petit	Précision, contexte local bien capté	Risque de perte de sens global
Gros	Meilleure compréhension globale	Moins précis, coût computationnel ↑

|| Character Text Splitter1

Parameters

Settings

Docs [↗](#)

Separator

Chunk Size

750

Chunk Overlap

130

III.2 Adapter sa documentation - Pourquoi ?

12 sept. 2024

Garant - Justificatif de situation professionnelle - Général

Grands principes

Les pièces validées apparaîtront telles quelles chez le propriétaire ou l'agence qui les recevra. Pour la **réputation** du service DossierFacile, il est donc crucial de maintenir un standard élevé de validation.

En particuliers, seront systématiquement refusées :

- les pièces contenant des arrière-plans perturbateurs
- les captures d'écran contenant des éléments perturbateurs
- les documents inutiles (*Tolérance en fonction des cas*)

Rappel du décret :

D. - Un ou plusieurs documents attestant des activités professionnelles parmi les documents suivants :

1. Contrat de travail ou de stage ou, à défaut, une attestation de l'employeur précisant l'emploi et la rémunération proposée, la date d'entrée en fonctions envisagée et le cas échéant la durée de la période d'essai.
2. L'extrait K ou K bis du registre du commerce et des sociétés datant de moins de trois mois pour une entreprise commerciale.
3. La fiche d'immatriculation au Registre national des entreprises de moins de trois mois pour une entreprise du secteur des métiers et de l'artisanat.
4. La copie du certificat d'identification de l'INSEE, comportant les numéros d'identification, pour un travailleur indépendant.
5. La copie de la carte professionnelle pour une profession libérale.
6. Toute pièce récente attestant de l'activité pour les autres professionnels.

Sites d'information :

- Pour en savoir plus : <https://www.service-public.fr/particuliers/vosdroits/F35247>
- <https://www.service-public.fr/particuliers/vosdroits/F1169>

Liste des pièces acceptées par les opérateurs :

garant

1 sur 1

Token count
1789

Chunk Size

750

Chunk Overlap

130

- Chunk 1 : 1–750
- Chunk 2 : 620–1370 ($750 - 130 = 620$)
- Chunk 3 : 1240–1789 ($1370 - 130 = 1240$)

Liste des pièces acceptées par les opérateurs :

- Un bulletin de salaire de moins de 3 mois
- Contrat de travail ou de stage ou d'alternance en cours signé des deux parties ;
- Attestation de l'employeur signée et datée de moins de trois mois
- Carte d'étudiant, attestation Parcours Sup' ou certificat de scolarité pour l'année en cours, Accord préalable d'inscription (étudiant étranger)
- Copie de la carte professionnelle (profession libérale, de santé, etc)

Les pièces de qui ? Pour quoi ?



III.3 Adapter sa documentation, étape obligatoire

Liste des pièces acceptées par les opérateurs :

- Un bulletin de salaire de moins de 3 mois
- Contrat de travail ou de stage ou d'alternance en cours signé des deux parties ;
- Attestation de l'employeur signée et datée de moins de trois mois
- Carte d'étudiant, attestation Parcours Sup' ou certificat de scolarité pour l'année en cours, Accord préalable d'inscription (étudiant étranger)
- Copie de la carte professionnelle (profession libérale, de santé, etc)



Garants - Liste des pièces acceptées par les opérateurs pour justifier de la situation professionnelle :

- Un bulletin de salaire de moins de 3 mois
- Contrat de travail ou de stage ou d'alternance en cours signé des deux parties ;
- Attestation de l'employeur signée et datée de moins de trois mois
- Carte d'étudiant, attestation Parcours Sup' ou certificat de scolarité pour l'année en cours, Accord préalable d'inscription (étudiant étranger)
- Copie de la carte professionnelle (profession libérale, de santé, etc)



Jimmy il y a 3 mois

pour un garant, dans la catégorie activité professionnelle, est ce qu'un bulletin de paie peut être accepté ?

2 réponses



Docteur Op... B... il y a 3 mois

Bonjour Jimmy,

Je n'ai pas trouvé d'informations précises concernant l'acceptation d'un bulletin de paie dans la catégorie "activité professionnelle" pour un garant. Pour obtenir une réponse exacte, je te conseille de vérifier avec ton tuteur ou de poser la question dans le canal [#opération](#).

N'hésite pas à revenir vers moi si tu as d'autres questions ou besoin de plus d'informations.

Correction de la documentation,
mise à jour automatique de la BdD
vectorielle.

Nouveau test réalisé 2 min plus
tard



julien.henry il y a 3 mois

pour un garant, dans la catégorie activité professionnelle, est ce qu'un bulletin de paie peut être accepté ?

8 réponses



Docteur Op... B... il y a 3 mois

Bonjour Julien,

Oui, un bulletin de salaire de moins de 3 mois peut être accepté pour justifier l'activité professionnelle d'un garant.

Pour toute vérification complémentaire, n'hésitez pas à consulter votre tuteur ou à poster un message dans le canal [#opération](#).

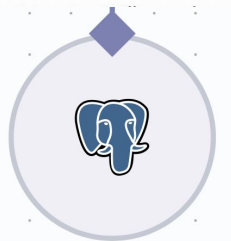
Bonne journée !

IV Exploiter l'historique des conversations

- Utilisation par ~20 opérateurs. (→ canal public donc peut être un peu de timidité.)
- Un peu moins de 200 utilisations – peu mais suffisant pour exploiter les données.

A	B	C
id	session_id	message
338	uyy5fbe3ubdq9rb4n7w3b546ac	<pre>{ "type": "ai", "content": "Bonjour Prénom\n\nD'après les informations disponibles, une attestation sur l'honneur de non-activité professionnelle est acceptée dans les cas suivants : mère au foyer, chômeur non indemnisé, et usager sans activité. Il n'est pas spécifiquement mentionné que la personne doit être inscrite à France Travail.\n\nCependant, pour être sûr, je vous recommande de vérifier cette information avec votre tuteur ou en postant un message dans le canal #opération.", "tool_calls": [], "additional_kwargs": {}, "response_metadata": {}, "invalid_tool_calls": [] }</pre>
337	uyy5fbe3ubdq9rb4n7w3b546ac	<pre>{ "type": "human", "content": "prénom.nom\n\nOn peut accepter une attestation sur l'honneur de non activité professionnelle si la personne n'est pas inscrite à France travail ", "additional_kwargs": {}, "response_metadata": {} }</pre>

n8n_chat_histories			
Filter		Sorted by 1 rule	
Insert			
	id int4	session_id varchar	message jsonb
<input type="checkbox"/>	338	uyy5fbe3ubdq9rb4n7w3b546ac	<pre>{ "type": "ai", "content": "Bonjour Sarah\n\n" }</pre>
<input type="checkbox"/>	337	uyy5fbe3ubdq9rb4n7w3b546ac	<pre>{ "type": "human", "content": "sarah.bensour" }</pre>
<input type="checkbox"/>	336	9bf13q4ryink5dg74dgtohuehh	<pre>{ "type": "ai", "content": "Bonjour Alain\n\nD" }</pre>
<input type="checkbox"/>	335	9bf13q4ryink5dg74dgtohuehh	<pre>{ "type": "human", "content": "alain.jauregu" }</pre>
<input type="checkbox"/>	334	9bf13q4ryink5dg74dgtohuehh	<pre>{ "type": "ai", "content": "Bonjour Alain,\n\n" }</pre>
<input type="checkbox"/>	333	9bf13q4ryink5dg74dgtohuehh	<pre>{ "type": "human", "content": "alain.jauregu" }</pre>
<input type="checkbox"/>	332	hesdkqq4zj8j7gccfnhirj7mkw	<pre>{ "type": "ai", "content": "Bonjour Afwan,\n\n" }</pre>
<input type="checkbox"/>	331	hesdkqq4zj8j7gccfnhirj7mkw	<pre>{ "type": "human", "content": "afwan.mirza-r" }</pre>
<input type="checkbox"/>	330	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "ai", "content": "Bonjour Jimmy,\n\n" }</pre>
<input type="checkbox"/>	329	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "hu" }</pre>
<input type="checkbox"/>	328	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "ai" }</pre>
<input type="checkbox"/>	327	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "hu" }</pre>
<input type="checkbox"/>	326	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "ai" }</pre>
<input type="checkbox"/>	325	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "hu" }</pre>
<input type="checkbox"/>	324	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "ai" }</pre>
<input type="checkbox"/>	323	nt5yreiwktgwxxheg3oabzmfne	<pre>{ "type": "hu" }</pre>



IV Exploiter l'historique des conversations

- Création d'une FAQ d'une centaines de questions

FAQ DossierFacile

L'équipe DossierFacile a souhaité mettre à disposition une foire aux questions (FAQ) regroupant les questions les plus fréquemment posées au support de Dossier Facile. N'hésitez pas à envoyer vos remarques à l'équipe support à l'adresse suivante : contact@dossierfacile.logement.gouv.fr.

Les questions sont classées par catégorie de documents demandés, avec en complément une catégorie **Divers**.

Sommaire

- 1 [Pièce d'identité](#)
- 2 [Justificatif de domicile](#)
- 3 [Situation professionnelle](#)
- 4 [Justificatifs de ressources](#)
- 5 [Avis d'imposition](#)
- 6 [Divers](#)

<https://aide.dossierfacile.logement.gouv.fr/fr/article/faq-dossierfacile-9vof04/>

V Considérations annexes

Coûts

- API Mistral = 1€ (< 0,01 €/requête)
- N8n = Hébergeable, la version cloud gratuite doit suffire.
- Supabase = Hébergeable, la version cloud gratuite suffit

Outils

- Mistral API
- “Front” → Mattermost
- N8n pour mettre en lien les différents outils utilisés
- Google Drive
- Supabase (cloud)

Données

- Mistral API apparemment RGPD friendly ? En tout cas, Service non “SecNum Cloud”
- Supabase en version Cloud : cloud AWS → Cloud Act applicable.
- Services Google et API → Cloud Act applicable dans tous les cas.



Merci pour votre attention

Contact :
julien.henry@dossierfacile.fr
[@julien.henry](#) (Mattermost)

Dossier
Facile



Brainstorm

- délais de réponse : entre 5 et 40 secondes (attention aux webhook Mattermost, refus de poster la réponse après un délais, que j'estime à 60 secondes mais je n'ai pas vérifié)
- attention ce bot n'est pas accessible en message privé – pour cela il faut un websocket, pas accessible si non admin sur l'instance Mattermost
- Comment choisir la bonne taille des chunk ? avantages et inconvénients
- Mistral embedding utilise moins de vecteurs 1024 contre 1536 chez Open Ai (à qualité d'entraînement égale et optimisation identique = perte de qualité (nuance) mais plus rapide et moins énergivore pour la recherche dans la BdD.) (Mistral très fort sur le français donc embedding Mistral tout de même ok)
- attention à ne pas utiliser Mistral Small (pas de tools available ?)
- 3 modèles utilisés (génération réponse, Embedding et insertion documents dans bdd et retrieval)
- **Ajouter photo supabase historique et expliquer le traitement effectué après qui a débouché sur le réalisation de la FAQ**

<https://compteur-de-mots.ouest-france.fr/>

<https://tiktokenizer.vercel.app/?model=gpt-4-1106-preview>