

Cahier de recette

ASADI : ASsistant ADministratif Intelligent

Groupe:
L3W1

Encadrant:
David Janisek

Auteurs:
Djaffer Abdel Malik, Gourmelen Thomas, Ponnoussamy Valentin, Traore Ali

Version du document:
1.0

Résumé

Dans le cadre de l'UE Projet Informatique à l'université Paris Cité, nous devons par équipes de quatre tenir un projet tout le long du semestre 6 sous la tutelle d'un encadrant. Ce projet vise à développer un assistant numérique capable de répondre automatiquement aux questions administratives, d'expliquer des procédures et d'offrir un apprentissage interactif via des scénarios immersifs, en s'appuyant sur une base documentaire centralisée. Le présent cahier des charges rédigé en accord avec l'encadrant pose les bases du projet. Il se veut descriptif des objectifs et ambitions de l'application et explique la démarche et le déroulement prévu pour la réalisation du projet.

Sommaire

1	Introduction	3
1.1	Contexte et enjeux	3
1.2	Objectifs du cahier de recette	3
2	Objectif et méthodes	4
2.1	Objectifs	4
2.2	Planning	4
2.3	Méthodes	5
3	Concepts de base	6
4	Logiciel utilisé	7
5	Tests	8
5.1	Utilisateur	8
5.2	Administrateur	12
5.3	Autres tests	13
6	Références	15
7	Index	16

1 Introduction

1.1 Contexte et enjeux

Les procédures administratives et réglementaires peuvent être complexes et parfois difficiles à comprendre, tant pour les nouveaux collaborateurs que pour les utilisateurs ayant besoin d'informations claires et accessibles. Un assistant numérique, capable de répondre automatiquement à des questions administratives, d'expliquer des procédures et de proposer un mode d'apprentissage interactif, contribuerait à simplifier ces tâches tout en favorisant une intégration plus rapide des nouveaux membres d'une équipe.

En s'appuyant sur un ensemble de ressources documentaires (fichiers, textes réglementaires, sites web, etc.), cet assistant permettrait de centraliser et de rendre accessible une base de connaissances, tout en offrant un apprentissage ludique et personnalisé grâce à des scénarios interactifs et des questions adaptées.

1.2 Objectifs du cahier de recette

Le cahier de recette vise à s'assurer de la conformité de l'application avec le cahier des charges. Afin de vérifier sa conformité nous allons établir les objectifs et méthodes du projet, les concepts de base, le matériel utilisé et un ensemble de tests.

Ce document servira donc de référence pour l'ensemble des phases de validation, depuis les tests unitaires jusqu'à la recette finale, afin d'assurer la qualité et la robustesse de l'application avant sa mise en production.

2 Objectif et méthodes

Le développement de notre assistant numérique basé sur la technologie RAG repose sur une méthodologie rigoureuse afin d'assurer un produit fonctionnel, optimisé et aligné avec les besoins définis. Cette section présente les objectifs du projet, son planning de développement, ainsi que les méthodes utilisées.

2.1 Objectifs

L'objectif principal de ce projet est de développer une application web intelligente, capable de répondre aux questions administratives, d'expliquer des procédures réglementaires, et de proposer un mode d'apprentissage interactif. Pour atteindre cet objectif, plusieurs sous-objectifs sont définis :

- ▶ Intégration d'un système RAG permettant de générer des réponses basées sur une base documentaire vérifiée.
- ▶ Mise en place d'une base de donnée vectorielle auto-hébergée pour stocker et interroger les documents réglementaires.
- ▶ Développement d'une interface utilisateur intuitive et accessible pour garantir une navigation fluide et une bonne expérience utilisateur.
- ▶ Gestion sécurisée des documents administratifs : seuls les administrateurs désignés auront la possibilité d'ajouter, modifier ou supprimer des documents dans la base de connaissances.
- ▶ Mise en place d'un mode d'apprentissage interactif, incluant des tests de connaissances et des scénarios administratifs réalistes.
- ▶ Analyse des incohérences dans les processus administratifs, notamment sur les différentes versions des documents.
- ▶ L'assistant doit notifier son incapacité à répondre et citer précisément la source de la réponse (document, page, ligne).

Ces objectifs guideront l'ensemble du processus de développement et assureront la création d'un assistant fiable et efficace.

2.2 Planning

Le projet sera développé sur une période de 12 semaines, avec une livraison finale prévue pour la semaine du 28 avril 2025. Afin d'assurer un bon suivi et une progression constante, le développement sera divisé en plusieurs phases :

- ▶ Semaine du 4 mars : Finalisation de la conception et de la documentation initiale. Début de la phase de développement.
- ▶ Semaine du 7 avril : Dernière phase de développement, incluant l'indexation des documents, l'intégration du RAG et les tests fonctionnels.
- ▶ Semaine du 21 avril : Finalisation de l'intégration et validation des fonctionnalités avec le cahier de recette.
- ▶ Semaine du 28 avril : Préparation de la soutenance avec la rédaction du rapport de projet, la mise à jour du wiki et la création des supports de présentation.

Ce découpage permet une gestion agile du projet avec des étapes clairement définies et des objectifs intermédiaires atteignables.

2.3 Méthodes

Pour garantir un développement efficace et structuré, plusieurs méthodes seront appliquées tout au long du projet :

- ▶ Développement itératif et agile : Adoption d'une approche Agile, avec des sprints courts et des révisions régulières pour ajuster les fonctionnalités en fonction des tests et retours.
- ▶ Gestion du code avec un système de versionnement (Git/SVN) afin de suivre l'évolution du développement et assurer un travail collaboratif efficace.
- ▶ Tests unitaires et validation continue pour garantir la stabilité et la performance du logiciel à chaque étape du développement.

Cette méthodologie assurera une progression efficace et maîtrisée du projet, tout en garantissant un produit final robuste et conforme aux attentes définies.

3 Concepts de base

Cette section introduira les concepts de base du projet, servant ainsi à la compréhension globale du cahier de recette.

- **Intelligence artificielle** : L'**intelligence artificielle (IA)** est un ensemble de techniques permettant à des machines d'imiter certaines capacités humaines, comme l'apprentissage, la compréhension et la prise de décision.

Les **LLM (Large Language Models)** sont une catégorie d'IA entraînée sur de grandes quantités de texte pour comprendre et générer du langage naturel. Ils utilisent des réseaux de neurones avancés, comme les **transformers**, pour prédire et produire des textes cohérents en fonction d'une entrée donnée.

- **IA générative** : L'**IA générative** est une branche de l'**intelligence artificielle** capable de produire du contenu original (texte, image, son, code, etc.) en s'appuyant sur des modèles entraînés sur de grandes quantités de données, comme les **LLM (Large Language Models)**.

Le **RAG (Retrieval-Augmented Generation)** est une technique qui améliore l'IA générative en combinant :

- La **récupération d'informations** à partir de sources externes (bases de données, documents, API).
- La **génération de texte** en utilisant ces informations pour produire des réponses plus précises et fiables.

Le **RAG** permet donc d'éviter les hallucinations des modèles purement génératifs en s'appuyant sur des données factuelles et actualisées.

- **GED (Gestion de documents électronique)** est un système informatique qui permet de numériser, stocker, organiser, rechercher et gérer des documents électroniques de façon sécurisée et centralisée. Elle facilite l'accès, le partage et la traçabilité des documents tout en optimisant la gestion des versions et les droits d'accès.
- **Assistant numérique** : Un **assistant numérique** est un logiciel ou une intelligence artificielle capable d'aider les utilisateurs en répondant à leurs questions, en effectuant des tâches ou en automatisant certaines actions.

Il peut fonctionner par commande vocale (ex : Siri, Google Assistant) ou par texte (ex : chatbot, assistant IA) et s'appuie souvent sur des **LLM** et des techniques comme le **RAG** pour fournir des réponses précises et contextuelles.

- **Documents administratifs** : Les **documents administratifs** sont des écrits officiels produits ou reçus par une administration publique ou une entreprise dans le cadre de ses activités. Ils peuvent être des formulaires, des rapports, des décisions, des contrats, etc.
- **Mode d'apprentissage interactif** : Un **mode d'apprentissage interactif** est une méthode pédagogique qui engage activement l'apprenant à travers des activités participatives. Il repose sur des interactions avec des outils numériques, des simulations, des exercices pratiques, des jeux éducatifs ou encore des discussions en groupe, favorisant ainsi une meilleure assimilation des connaissances.
- **Processus administratifs** : Les **processus administratifs** sont des séries d'actions ou de procédures mises en place par une organisation (publique ou privée) pour gérer ses activités administratives de manière efficace. Ils incluent des tâches comme la gestion des documents, le traitement des demandes, la prise de décisions, et la communication interne et externe.

4 Logiciel utilisé

Nom	Nature	Utilité	Limites
LlamaIndex	Bibliothèque	Permet d'intégrer des sources de données externes à des modèles de génération (RAG). Améliore la recherche d'information et la génération de texte.	Certains documents, comme les images et les PDFs scannés, sont mal extraits.
pdf2image	Bibliothèque	Convertit des fichiers PDF en images (une page par image).	Ne gère pas les PDF contenant uniquement du texte sans images ; qualité de conversion dépend de la qualité du PDF.
pytesseract	Bibliothèque	Utilise l'OCR (Reconnaissance Optique de Caractères) pour extraire du texte à partir d'images.	La précision de l'OCR peut être faible sur des images de mauvaise qualité ou avec des polices difficiles à lire.
PostgreSQL	Système de gestion de bases de données (SGBD)	Base de données relationnelle utilisée pour stocker, organiser et gérer de grandes quantités de données.	Un inconvénient de PostgreSQL est sa consommation mémoire plus élevée par rapport à d'autres SGBD comme MySQL.
Python	Langage de programmation	Langage polyvalent utilisé pour le développement d'applications et de systèmes d'IA.	Peut être moins performant que d'autres langages pour des applications à très haute performance (par exemple, en calcul scientifique intensif).
Ollama	Plateforme/outil d'IA	Permet de déployer des modèles de langage (LLM) localement, avec des outils pour intégrer l'IA dans des applications sans dépendance cloud.	Nécessite une gestion des ressources locales, ce qui peut être limité par la capacité matérielle des systèmes.
Django	Framework	Framework Python pour le développement rapide d'applications web.	Peut être complexe pour les petits projets et nécessite des configurations pour certains environnements.
Django ORM	Technique	Technique permettant la manipulation de base de données relationnelle en utilisant des objets sans écrire de requêtes SQL	Cette technique ne permet pas de faire des requêtes complexes

5 Tests

Les tests constituent une étape essentielle du processus de validation du projet. Ils permettent de s'assurer que chaque fonctionnalité développée répond aux exigences définies et fonctionne correctement dans les conditions prévues. L'objectif de cette section est de décrire les différents tests réalisés, leurs objectifs, ainsi que les critères d'acceptation permettant de valider chaque fonctionnalité.

Les tests sont répartis en différentes catégories et sous-catégories. Cette répartition permettra une meilleure distribution des tests entre les différents collaborateurs de ce projet.

5.1 Utilisateur

5.1.a Création d'un compte

Objectif: Vérifier le bon fonctionnement de la création d'un compte utilisateur.

Pré-requis: Avoir un smartphone ou un ordinateur ayant accès à une connexion wifi.

N°	Action	Attendu	Résultat
UA1	Cliquer sur "créer un compte"	Ouvrir la page asadi.fr/creation-d'un-compte	
UA2	Rentrer une adresse mail non valide	Message d'erreur pour la saisie de l'adresse mail	
UA3	Rentrer une adresse mail valide sans mot de passe	Message d'erreur "Veuillez saisir un mot de passe"	
UA4	Rentrer une adresse mail valide et un mot de passe non valide	Message d'erreur "Erreur de mot de passe"	
UA5	Rentrer une adresse mail valide et un mot de passe valide	Création du compte et connexion automatique à ce compte	
UA6	Rentrer une adresse mail déjà associée à un compte	Message d'erreur "Compte déjà existant", renvoi sur la page de connexion	

5.1.b Connexion à son compte

Lors de la connexion d'un utilisateur au site, une page de connexion lui est proposée afin de s'identifier. Cette identification permet à l'utilisateur d'avoir accès à son historique mais aussi à ses quiz.

Objectif: Vérifier le bon fonctionnement de la connexion à l'application

Pré-requis: Avoir un compte

N°	Action	Attendu	Résultat
UB1	Rentrer une adresse mail non valide.	Message d'erreur "Mail invalide". Réinitialisation de la zone de saisie du mail.	
UB2	Rentrer son adresse mail valide sans mot de passe.	Message d'erreur "Veuillez saisir un mot de passe". Zone de saisie du mot de passe.	
UB3	Rentrer son adresse mail valide et un mot de passe non valide.	Message d'erreur "Erreur de mot de passe". Réinitialisation de la zone de saisie du mot de passe.	
UB4	Rentrer son adresse mail valide et son mot de passe valide.	Connexion au compte et lancement de la page d'accueil. Passage à la page d'utilisation du site.	
UB5	Passage en mode admin	Connexion au compte admin et lancement de la page d'accueil.	
UB6	Appuyer sur le bouton pour créer un compte.	Renvoie sur la page de création de compte.	
UB7	Appuyer sur le bouton « Mot de passe oublié » de la page de connexion.	Possibilité de modifier son mot de passe si le mail rempli correspond à celui d'un compte valide.	
UB8	Appuyer sur le bouton « Se déconnecter » du compte utilisateur.	Déconnexion du compte.	

5.1.c Recherche et réponse aux questions administratives

Objectif: Vérification de l'outil de recherche et de la qualité des réponses aux questions administratives.

Pré-requis: Avoir un compte valide et une base de connaissance remplie.

N°	Action	Attendu	Résultat
UC1	Poser une multitude de questions de longueur et complexité différentes pour déterminer la compréhension des requêtes utilisateur.	L'assistant doit interpréter correctement les questions et fournir des réponses pertinentes, précises et bien contextualisées.	
UC2	Lancer plusieurs requêtes et examiner les sources affichées.	Les réponses doivent être accompagnées des références documentaires utilisées.	
UC3	Poser une question incomplète ou ambiguë.	L'assistant doit proposer une reformulation ou demander des précisions.	
UC4	Poser une question nécessitant une réponse détaillée.	Le LLM génère une réponse bien découpée avec une structuration logique	
UC5	Effectuer une requête utilisateur avec des documents de référence.	La réponse ne contient pas d'informations inventées et est basée sur les sources à disposition.	
UC6	Poser une question avec aucune donnée disponible.	L'assistant indique clairement qu'aucune information est disponible et qu'il n'est pas en mesure de répondre à cette question	
UC7	Poser une question avec plusieurs sujets mélangés.	L'assistant segmente la réponse en plusieurs parties pour les différents sujets	
UC8	Établir des interactions dans la même session pour vérifier la cohérence dans une conversation.	L'assistant garde en mémoire le contexte des échanges précédents	

5.1.d Mode d'apprentissage interactif (Scénario)

Objectif: Vérification du fonctionnement de l'utilisation du mode d'apprentissage interactif (Scénario).

Pré-requis: Avoir un compte

N°	Action	Attendu	Résultat
UD1	Lancer une session d'apprentissage et analyser les questions posées.	Les questions doivent être pertinentes et adaptées au contenu abordé.	
UD2	Répondre correctement ou incorrectement aux questions.	L'assistant doit fournir un retour approprié (correction, explication).	
UD3	Lancer un scénario où l'utilisateur doit résoudre une situation.	L'assistant doit proposer une mise en situation réaliste et donner un retour constructif.	

5.1.e Mode d'apprentissage interactif (Quizz)

Objectif: Vérification du fonctionnement de l'utilisation du mode d'apprentissage interactif (Quiz).

Pré-requis: Avoir un compte

N°	Action	Attendu	Résultat
UE1	L'utilisateur demande à l'assistant un quiz sur un thème donnée.	Un nouveau quiz sera créer par l'assistant contenant une question et un réponse sur le thème donné.	
UE2	Répondre correctement ou incorrectement aux questions.	L'assistant doit fournir un retour approprié (correction, explication).	
UE3	L'utilisateur supprime un quiz.	Le quiz sera bien supprimé et ne sera plus disponible.	

5.1.f Historique

Objectif: Vérification du fonctionnement de l'historique.

Pré-requis: Avoir un compte ayant déjà effectué des prompts.

N°	Action	Attendu	Résultat
UF1	Consulter l'historique d'un utilisateur	L'historique est affiché correctement.	
UF2	Supprimer l'historique d'un utilisateur	L'historique est supprimé correctement.	
UF3	Consulter la chronologie de l'historique.	L'historique affiche les bonnes dates et heures des différentes interactions.	
UF4	Retrouver une ancienne session dans l'historique grâce au nom de la session lié au sujet traité.	Chaque nom de session permet de savoir le sujet d'une session et de la distinguer des autres.	
UF5	Reprendre une ancienne session	L'application doit conserver les progrès et proposer des contenus adaptés.	

5.2 Administrateur

L'administrateur étant un utilisateur aillant plus de droit qu'un utilisateur lambda, nous avons seulement besoin de tester les fonctionnalités supplémentaire.

5.2.a Ajout et suppression de documents

Objectif: Vérification de l'ajout, suppression et de la vérification de documents administratif à la base de donnée

Pré-requis: Avoir un compte administrateur et une base de connaissance remplie

N°	Action	Attendu	Résultat
AA1	Dépôt de fichier avec une extension non gérée sur l'application	Message d'erreur "Cette extension de fichier n'est pas gérer par l'application".	
AA2	Dépôt d'un trop grand nombre de fichiers.	Message d'erreur "Veuillez déposez moins de documents simultanément". Annulation de l'opération.	
AA3	Insérer un fichier déjà présent dans les documents de l'application.	Message d'erreur "Le fichier est déjà présent dans la base de donnée".	
AA4	Ajouter de nouveaux documents et tester leur prise en compte.	Les documents sont ajoutés et l'assistant doit intégrer les nouvelles sources dans ses réponses.	
AA5	Supprimer des documents et tester l'arrêt de prise en compte.	Les documents sont supprimés et non accessible pour fournir de réponse.	
AA6	Ajouter des documents volumineux	Les documents sont ajoutés dans la base de connaissance sans latence excessive.	

5.2.b Gestion des utilisateurs

Objectif: Vérification de la fonction de gestion des utilisateurs

Pré-requis: Avoir un compte administrateur et des comptes utilisateurs inscrits

N°	Action	Attendu	Résultat
AB1	Cliquer sur le bouton de la liste des utilisateurs	Affichage de la liste d'utilisateurs	
AB2	Cliquer sur le bouton « Supprimer l'utilisateur ».	Suppression du compte de l'utilisateur et suppression de la liste utilisateur	

5.2.c Espace de travail

Objectif: Vérification du fonctionnement du mode espace de travail

Pré-requis: Avoir un compte administrateur

N°	Action	Attendu	Résultat
AC1	Création d'un espace avec un nom valide	L'espace de travail est créé	
AC2	Création d'un espace de travail déjà existant	Message d'erreur "Espace de travail déjà existant". Zone de saisie du nom.	
AC3	Suppression d'un espace existant	L'espace de travail est supprimé	
AC4	Ajout d'un document valide (PDF, TXT, etc.)	Le document est accessible dans l'espace de travail	
AC5	Ajout d'un document avec un format non supporté	Erreur de validation	
AC6	Ajout d'un document trop volumineux	Erreur (limite de taille dépassée)	

5.3 Autres tests

5.3.a Tests techniques

Objectif: Vérification du fonctionnement de l'application sur différents aspects techniques.

Pré-requis: Avoir un compte administrateur et une base de connaissance remplie

N°	Action	Attendu	Résultat
TA1	Mesurer le temps de réponse moyen sur plusieurs requêtes.	Les réponses doivent être générées en un temps raisonnable.	
TA2	Simuler plusieurs utilisateurs posant des requêtes simultanément.	L'application doit gérer la charge sans ralentissement excessif.	
TA3	Tester la compatibilité sur plusieurs navigateurs.	L'affichage et les fonctionnalités sont identiques sur tous les navigateurs.	
TA4	Donner à l'assistant différent format de documents (.txt, .pdf,.docx, .xlsx, HTML et images OCR).	Extraction de texte réussie sans perte d'information.	
TA5	Appuyer sur le bouton pour supprimer le compte.	Suppression du compte.	

5.3.b Tests de sécurité

Objectif: Vérification de la fiabilité et sécurité de l'assistant

Pré-requis: Avoir un compte administrateur

N°	Action	Attendu	Résultat
TB1	Récupérer les informations utilisateur sans autorisation.	Impossible de les récupérer, aucune fuite de données sensibles	
TB2	Se connecter avec différents niveaux de droits (admin, utilisateur standard).	Accès contrôlé selon les permissions définies.	
TB3	Récupérer les mots de passe qui sont stockés sous forme chiffrée dans la base de données	Impossible à récupérer, Les mots de passe sont stockés sous hash sécurisé	
TB4	Établir un prompt demandant de contourner les restrictions	L'assistant refuse de répondre ou détecte la tentative	

6 Références

- ▶ Pour la rédaction de cette documentation, nous avons utilisé le modèle d'intelligence artificielle ChatGPT. Ce modèle nous a aidés sur plusieurs points :
 - Reformulation de phrases et correction de fautes d'orthographe : Nous avons utilisé ChatGPT pour nous assister dans la rédaction de certaines phrases, afin d'améliorer la clarté du texte.
 - Recherche : ChatGPT nous a permis d'explorer différentes technologies que nous devons utiliser dans ce projet. Il nous a également aidés à mieux comprendre certains concepts abstraits.

Toutefois, l'ensemble des droits de propriété intellectuelle de ce document demeure exclusivement attribué à ses auteurs.

7 Index

A

Administrateur 4, 12
 API 6
 Application web 7
 Apprentissage interactif 1, 3, 6
 Apprentissage ludique 3
 Assistant numérique 1, 3, 4, 6

B

Base de donnée vectorielle 4

C

ChatGPT 15
 Conformité 3

D

Django 7
 Documents réglementaires 3, 4
 Développement 4, 5

F

Framework 7

G

GED 6
 Gestion agile 4

I

IA Générative 6
 Intelligence artificielle 6, 7
 Intégration 3

L

Langage naturel 6
 LlamaIndex 7
 LLM 6, 7

O

OCR 7, 14
 Ollama 7

P

Pdf2image 7
 PGSQL 7
 Phase de validation 3
 Processus administratif 6
 Python 7

Q

Quiz 11

R

RAG 4, 6, 7
 Recette finale 3

S

SGBD 7
 Soutenance 4
 Système de versionnement 5

T

Tests unitaires 3, 5
 Transformers 6

U

Utilisateur 3, 14