

T-ESP-700

Big Brain

Lucas FIXARI
Pierre ROCHETTE
William WOZIWODA

Analyse de la Pertinence et des Bénéfices du Projet - IA d'Analyse Documentaire

1. Introduction	3
2. Analyse du Besoin	3
2.1 Contexte et Besoin	3
2.2 Les Défis Actuels	3
3. Objectifs et Bénéfices Anticipés	3
3.1 Objectifs du Projet	3
3.2 Bénéfices Clés	3
4. Publics Cibles et Cas d'Utilisation	4
4.1 Particuliers	4
4.2 Municipalités et Administrations Publiques	5
4.3 Entreprises et Secteur Privé	5
4.4 Secteurs Médicaux et Pharmaceutiques	5
6. Comparaison avec les Solutions Existantes	6
6.1 Solutions de Recherche Documentaire dans le Cloud	6
6.2 Logiciels de Gestion Documentaire Locaux	6
6.3 Outils de Recherche et d'Indexation Open Source avec IA (Ex. Apache Solr et Elasticsearch)	7
6.4 Valeur Ajoutée de l'IA d'Analyse Documentaire Locale	7

1. Introduction

Ce document présente l'analyse de la pertinence et des bénéfices du projet d'IA d'Analyse Documentaire Locale. Il vise à démontrer comment ce projet répond à des besoins spécifiques en recherche documentaire contextuelle, sécurisée et performante pour des environnements locaux, en ciblant différents types d'utilisateurs et leurs cas d'usage.

2. Analyse du Besoin

2.1 Contexte et Besoin

Que vous soyez un particulier, une administration, une entreprise ou toute autre organisation vous accumulez au fil des ans une quantité croissante de documents sous divers formats (PDF, DOCX, Excel, images, etc.). Accéder rapidement à une information spécifique dans cette masse de données devient un défi, surtout lorsque les informations sont réparties entre de nombreux fichiers et formats et encore plus quand le nommage du fichier n'indique en rien le contenu du document.

2.2 Les Défis Actuels

1. **Confidentialité et Sécurité** : Les solutions de recherche documentaire dans le cloud posent des problèmes de confidentialité pour les utilisateurs qui ne peuvent pas externaliser leurs documents sensibles.
2. **Accès et Recherche Contextuelle** : Il est souvent difficile de retrouver une information précise sans connaître l'organisation ou le contenu exact des documents.
3. **Perte de Temps** : Les recherches manuelles dans les documents sont longues, laborieuses, et ne garantissent pas une précision ou une exhaustivité des résultats.

L'IA d'Analyse Documentaire Locale répond à ces défis en permettant une recherche locale, rapide et sécurisée, avec une compréhension contextuelle et la capacité de répondre aux questions en langage naturel.

3. Objectifs et Bénéfices Anticipés

3.1 Objectifs du Projet

L'objectif est de fournir une IA locale, flexible et sécurisée, capable d'analyser et d'indexer automatiquement des documents pour permettre aux utilisateurs de poser des questions en langage naturel et obtenir des réponses précises et contextualisées.

3.2 Bénéfices Clés

1. **Confidentialité et Contrôle des Données**

- **Bénéfice** : En stockant et traitant les documents localement, le projet offre une sécurité accrue et réduit les risques de fuite d'informations.
 - **Impact Anticipé** : Cette confidentialité est cruciale pour les secteurs administratifs, juridiques, et financiers, où la protection des données est primordiale.
2. **Gain de Temps et Productivité**
- **Bénéfice** : En quelques secondes, l'IA fournit des réponses précises à partir d'une question en langage naturel, réduisant le temps de recherche et permettant de se concentrer sur des tâches à plus forte valeur ajoutée.
 - **Impact Anticipé** : Bénéfique pour les professionnels, tels que les analystes financiers ou les juristes, qui doivent rapidement accéder à des informations complexes ou quantitatives.
3. **Recherche Intuitive et Personnalisée**
- **Bénéfice** : L'IA permet une recherche en langage naturel, rendant l'outil accessible et intuitif même pour des utilisateurs sans compétences techniques.
 - **Impact Anticipé** : Utile pour divers profils d'utilisateurs, des professionnels aux particuliers cherchant des informations dans leurs archives personnelles.
4. **Optimisation de la Gestion Documentaire**
- **Bénéfice** : Grâce à l'indexation automatique et à la reconnaissance contextuelle, l'IA optimise la gestion documentaire, permet des requêtes complexes, et facilite la recherche thématique.
 - **Impact Anticipé** : Les entreprises peuvent mieux structurer et exploiter leur base documentaire, facilitant l'accès aux informations stratégiques sans une gestion manuelle lourde.

4. Publics Cibles et Cas d'Utilisation

4.1 Particuliers

Les particuliers conservent souvent une multitude de documents personnels (factures, contrats, relevés, rapports médicaux, etc.) et peuvent tirer parti de cette IA pour gérer et retrouver des informations spécifiques dans leurs documents personnels sans perdre de temps.

● Exemples de Cas d'Utilisation :

1. **Recherche de Documents Financiers** : Un utilisateur peut demander des informations sur ses relevés bancaires ou factures pour retrouver des informations de dépense précises, comme "montant total des dépenses en électricité sur un an".
2. **Accès aux Dossiers Médicaux** : Lorsqu'un utilisateur cherche un traitement spécifique ou une date de consultation dans ses rapports médicaux, l'IA peut identifier et extraire les informations pertinentes.

3. **Rappel des Contrats et Garanties** : Un particulier peut poser des questions comme “quand expire ma garantie pour tel appareil ?” et l’IA localise le contrat correspondant.

4.2 Municipalités et Administrations Publiques

Les administrations gèrent une grande quantité de documents relatifs aux dossiers citoyens, aux permis de construire, aux budgets et autres informations publiques. L’IA permettrait aux fonctionnaires de retrouver des informations spécifiques sans connaître le document exact.

- **Exemples de Cas d’Utilisation :**

1. **Gestion des Dossiers Citoyens** : Un agent municipal peut poser une question sur l’historique de demandes d’un citoyen (ex. “quelles sont les autorisations de permis délivrées pour ce citoyen ?”).
2. **Recherche Budgétaire** : L’IA aide à retrouver des montants ou détails sur des dépenses dans des rapports budgétaires, pour des questions comme “montant total alloué aux infrastructures l’année dernière”.
3. **Dossiers de Travaux Publics** : Lors de la recherche d’informations sur un projet ou une demande, l’IA peut fournir un résumé ou des détails pertinents sans que l’agent doive parcourir manuellement les archives.

4.3 Entreprises et Secteur Privé

Les entreprises produisent et conservent des milliers de documents (contrats, rapports financiers, fiches produits, documents RH) et peuvent tirer parti de cette IA pour un accès rapide aux informations internes, améliorant la productivité et la gestion documentaire.

- **Exemples de Cas d’Utilisation :**

1. **Recherche de Données Client** : Dans une entreprise de service, les employés peuvent interroger l’IA pour obtenir un historique des transactions d’un client (ex. “quels sont les montants facturés à ce client en 2023 ?”).
2. **Contrats et Juridique** : Pour des questions juridiques, les responsables peuvent retrouver rapidement des clauses spécifiques d’un contrat ou vérifier les dates de renouvellement.
3. **Rapports RH et Salaires** : Dans les services RH, l’IA permet de rechercher des informations sur les évaluations ou des changements dans les politiques internes, ou bien sur des éléments précis d’un dossier salarié.

4.4 Secteurs Médicaux et Pharmaceutiques

Les hôpitaux, cliniques, et laboratoires gèrent des dossiers patients, des rapports de recherche, et des réglementations. Le projet d’IA d’Analyse Documentaire Locale est particulièrement adapté à ce domaine où la sécurité des données est cruciale.

- **Exemples de Cas d’Utilisation :**

1. **Recherches dans les Dossiers Patients** : Permet aux professionnels de santé de retrouver des informations médicales spécifiques (ex. "quels traitements ont été administrés à tel patient en 2022 ?").
2. **Documents de Recherche** : Les chercheurs peuvent poser des questions dans des publications ou rapports pour retrouver des informations spécifiques sur des études ou résultats.
3. **Normes et Réglementations** : Dans les laboratoires, l'IA permet de rechercher des informations sur les normes et protocoles, facilitant la conformité réglementaire sans devoir parcourir manuellement des documents.

6. Comparaison avec les Solutions Existantes

6.1 Solutions de Recherche Documentaire dans le Cloud

De nombreuses solutions cloud, comme **Google Drive**, **Microsoft OneDrive**, et **Dropbox**, offrent des capacités de recherche documentaire avec indexation des contenus. Ces plateformes permettent la recherche de mots-clés dans des documents stockés en ligne et bénéficient parfois de fonctionnalités avancées, telles que la reconnaissance de texte (OCR) intégrée.

- **Limites pour l'utilisateur :**

1. **Problèmes de Confidentialité** : Les utilisateurs doivent transférer leurs données sur des serveurs externes, ce qui pose des risques en termes de sécurité et de confidentialité. Cela est particulièrement problématique pour des documents sensibles, comme les informations financières, juridiques, ou de santé.
2. **Dépendance au Réseau** : Ces solutions nécessitent une connexion Internet et sont donc dépendantes de l'accessibilité et de la qualité du réseau.
3. **Coûts d'Abonnement** : Certaines fonctionnalités avancées (indexation étendue, OCR) sont payantes, ce qui limite l'accès à un usage gratuit ou abordable.

6.2 Logiciels de Gestion Documentaire Locaux

Des logiciels comme **DocFetcher**, **Copernic Desktop Search**, et **X1 Search** offrent des fonctionnalités de recherche de fichiers en local, souvent avec une indexation basée sur des mots-clés. Ces outils permettent de parcourir des documents stockés sur un ordinateur sans nécessiter de connexion Internet.

- **Limites pour l'utilisateur :**

1. **Pas de Recherche Contextuelle** : Ces logiciels sont majoritairement limités à une recherche par mots-clés et ne peuvent pas interpréter des questions posées en langage naturel ni fournir de réponses contextuelles.

2. **Peu d'Intelligence Artificielle** : Les fonctionnalités de traitement du langage naturel (NLP) sont souvent absentes, rendant impossible l'analyse approfondie du contexte ou la reconnaissance d'entités.
3. **Organisation Manuelle** : Ces outils demandent souvent une organisation manuelle des fichiers ou une connaissance préalable de l'emplacement des documents recherchés, ce qui rend la recherche inefficace pour des utilisateurs ayant de vastes archives.

6.3 Outils de Recherche et d'Indexation Open Source avec IA (Ex. Apache Solr et Elasticsearch)

Des solutions open source comme **Apache Solr** et **Elasticsearch** permettent une indexation avancée et des recherches performantes en local ou sur serveur. Ces outils sont couramment utilisés dans des environnements professionnels nécessitant une recherche rapide et évolutive dans de grands volumes de données.

- **Limites pour l'utilisateur :**

1. **Complexité de Mise en Œuvre** : Ces solutions nécessitent une configuration technique avancée, ce qui les rend difficiles à déployer pour un usage individuel ou dans des petites entreprises sans équipe technique dédiée.
2. **Absence de Langage Naturel** : Bien qu'ils offrent des options d'indexation puissante, ces outils ne prennent pas en charge les questions posées en langage naturel sans personnalisation poussée.
3. **Pas d'Interface Utilisateur Intuitive** : Ces outils manquent d'interfaces conviviales pour les utilisateurs finaux et sont souvent utilisés en back-end d'applications plus complexes, ce qui limite leur utilisation directe par des non-techniciens.

6.4 Valeur Ajoutée de l'IA d'Analyse Documentaire Locale

Le projet d'IA d'Analyse Documentaire Locale combine les avantages de solutions de recherche locales avec des fonctionnalités avancées de traitement du langage naturel, ce qui le distingue des options existantes :

1. **Sécurité et Confidentialité** : Contrairement aux solutions cloud, ce projet garantit que toutes les données restent stockées localement, minimisant les risques de violation de données et respectant les contraintes de confidentialité.
2. **Recherche en Langage Naturel** : Là où les solutions existantes se limitent souvent à des recherches par mots-clés, cette IA interprète des questions en langage naturel, permettant des recherches contextuelles avancées sans connaissances techniques.
3. **Accessibilité et Simplicité d'Utilisation** : Conçue pour être simple à installer et intuitive à utiliser, elle ne requiert ni compétences techniques avancées ni configuration complexe, ce qui la rend accessible aux particuliers et aux organisations sans support technique.