

# T-ESP-700

---

# Big Brain

Lucas FIXARI  
Pierre ROCHETTE  
William WOZIWODA

# Plan de Gestion et Qualité - IA d'Analyse Documentaire

<b>1. Objectifs du Plan de Gestion du Projet</b>	<b>3</b>
<b>2. Déroulement et Phases du Projet</b>	<b>3</b>
Phase 1 : Préparation et Recherche (Semaines 1-4)	3
Phase 2 : Développement des Modules de Base (Semaines 5-12)	3
Phase 3 : Interaction en Langage Naturel et Recherche (Semaines 13-18)	3
Phase 4 : Synthèse, Extraction Sémantique et Interface Utilisateur (Semaines 19-24)	3
Phase 5 : Test et Optimisation (Semaines 25-28)	4
<b>3. Plan de Qualité et Critères d'Acceptation</b>	<b>4</b>
Processus de Contrôle Qualité	4
Critères d'Acceptation des Livrables	4
<b>4. Gestion des Risques</b>	<b>4</b>
Risques Identifiés	5
Plans d'Action	5
<b>5. Suivi et Reporting</b>	<b>5</b>
Fréquence des Rapports d'Avancement	5
Indicateurs de Performance (KPIs)	5
<b>6. Plan de Communication et Gestion des Rôles</b>	<b>6</b>
Rôles et Responsabilités	6
Plan de Communication	6

# 1. Objectifs du Plan de Gestion du Projet

Ce plan de gestion du projet a pour objectif de :

- Structurer le déroulement du projet pour maximiser l'efficacité des équipes et des ressources.
- Définir les étapes clés, milestone, et livrables du projet.
- Garantir la qualité des livrables en respectant les critères de précision, de pertinence et de fiabilité.
- Assurer la compatibilité Python et la performance des modules développés.
- Minimiser les risques par une gestion proactive et des plans de contournement.

## 2. Déroulement et Phases du Projet

### Phase 1 : Préparation et Recherche (Semaines 1-4)

- **Objectifs** : Identification des outils et bibliothèques en Python, élaboration du PBS, planification du projet.
- **Livrables** : Liste des technologies, PBS, cahier des charges détaillé.
- **KPI** : Document de PBS validé, sélection des outils.
- **Risques** : Incompatibilité des bibliothèques.
- **Contournement** : Prévoir des alternatives dès le début.

### Phase 2 : Développement des Modules de Base (Semaines 5-12)

- **Objectifs** : Développer les modules de **Prétraitement des Données** et de la **Base de Connaissances**.
- **Livrables** : Code pour l'extraction de texte, indexation et stockage structuré dans la Base de Connaissances avec SQLite et Whoosh.
- **KPI** : Temps d'extraction de texte inférieur à 5 secondes pour les PDF.
- **Risques** : Performance de l'indexation.
- **Contournement** : Optimisation par une configuration allégée.

### Phase 3 : Interaction en Langage Naturel et Recherche (Semaines 13-18)

- **Objectifs** : Intégrer le **Traitement du Langage Naturel** pour interpréter les requêtes et comprendre les entités.
- **Livrables** : Module NLP avec spaCy ou Transformers pour la compréhension des questions, la reconnaissance d'entités et l'analyse syntaxique.
- **KPI** : 80% de précision dans la compréhension des questions courantes.
- **Risques** : Difficulté d'interprétation des requêtes complexes.
- **Contournement** : Simplifier initialement les requêtes.

### Phase 4 : Synthèse, Extraction Sémantique et Interface Utilisateur (Semaines 19-24)

- **Objectifs** : Développer l'**Interface Utilisateur** et intégrer le **Module de Synthèse** pour la génération de résumés, ainsi que l'**Extraction Sémantique** pour contextualiser les réponses.
- **Livrables** : Interface en Flask, FastAPI ou tkinter ; module de génération de résumés ; module d'Extraction et Intégration Sémantique.
- **KPI** : Temps de réponse inférieur à 5 secondes pour des requêtes simples.
- **Risques** : Saturation de l'interface.
- **Contournement** : Créer une version CLI pour tests initiaux.

## Phase 5 : Test et Optimisation (Semaines 25-28)

- **Objectifs** : Tester chaque module, optimiser la performance et la précision des résultats.
- **Livrables** : Rapport de tests, document de recherche, prototype fonctionnel.
- **KPI** : Au moins 90% de précision dans les tests d'interaction et de restitution.
- **Risques** : Limitation de la performance en cas de gros volumes
- **Contournement** : Ajouter des limites de taille de fichiers pour la première version.

## 3. Plan de Qualité et Critères d'Acceptation

### Processus de Contrôle Qualité

- **Contrôle à chaque étape** : Exécution de tests unitaires après le développement de chaque module.
- **Tests de Prétraitement** : Vérifier l'extraction correcte de texte et images pour chaque format de document.
- **Tests de Performance NLP** : Assurer la précision du NLP pour interpréter les questions et identifier les intentions.
- **Tests de Synthèse et Restitution** : Vérifier que les résumés sont clairs, pertinents, et accompagnés des sources.
- **Tests d'Indexation et Recherche** : Vérifier le temps de réponse des recherches et la capacité à traiter de grands volumes de documents.

### Critères d'Acceptation des Livrables

- **Analyse de Documents** : Extraire avec succès le texte et les images des fichiers, avec un taux d'erreur inférieur à 2%, et conversion en JSON.
- **Base de Connaissances** : Structuration et indexation des données, avec un temps de réponse maximal de 5 secondes.
- **Interaction NLP** : Précision de 80% dans la compréhension des questions.
- **Synthèse et Restitution** : Présentation des résumés, avec mention des sources, pour les documents pertinents.

## 4. Gestion des Risques

## Risques Identifiés

**Compatibilité des Bibliothèques Python** : Risque d'incompatibilité entre les bibliothèques d'extraction (PyMuPDF, python-docx, pytesseract) et d'indexation (SQLite, Whoosh).

- **Contournement** : Prévoir des alternatives dès la phase de recherche.

**Précision de l'OCR et NLP** : La précision de l'extraction (OCR) et du traitement NLP peut varier selon la qualité des documents.

- **Contournement** : Limiter l'analyse à des documents de bonne qualité pour la première version.

**Limitations des Performances Locales** : La gestion des ressources pour de gros volumes de fichiers peut réduire la performance.

- **Contournement** : Optimiser l'indexation et mettre en place des contraintes de taille et de type de fichiers.

## Plans d'Action

- **Suivi continu des performances** : Effectuer des tests de performance après chaque ajout de module.
- **Réévaluation des outils** : Si les résultats des bibliothèques choisies ne sont pas satisfaisants, tester des alternatives.

## 5. Suivi et Reporting

### Fréquence des Rapports d'Avancement

- **Rapports mensuels** : Inclure une synthèse des tâches réalisées, difficultés rencontrées, et évolution des KPIs.
- **Rapports de Milestones** : Détail des étapes atteintes et conformité aux critères d'acceptation.

### Indicateurs de Performance (KPIs)

- **Précision de l'OCR** : Au moins 98% de précision sur des documents de qualité standard.
- **Temps de réponse NLP** : Temps de réponse NLP inférieur à 3 secondes pour des questions simples.
- **Temps d'indexation** : Temps maximal d'indexation inférieur à 5 secondes pour des documents de 10 pages.
- **Temps de réponse de l'indexation dans la Base de Connaissances** : Temps de réponse maximal de 5 secondes pour une recherche.

- **Satisfaction Utilisateur** : Retours utilisateurs positifs (>80%) sur la précision et l'interface.

## 6. Plan de Communication et Gestion des Rôles

### Rôles et Responsabilités

- **Promoteur (WPL\*)** : Responsable principal du suivi et des décisions clés, organisation des réunions de suivi.
- **Développeur de l'Analyse et du Prétraitement des Données** : Responsable des modules de **Prétraitement des Données**, incluant l'extraction de texte, OCR, et la normalisation.
- **Développeur de la Base de Connaissances** : Responsable de l'indexation et du stockage des données avec SQLite et Whoosh.
- **Développeur NLP** : Responsable de la compréhension des requêtes et de l'analyse NLP, en incluant les entités, étiquetage des parties du discours, et analyse syntaxique.
- **Développeur de l'Extraction Sémantique** : Responsable de l'**Extraction et Intégration Sémantique** pour extraire les relations entre entités et structurer les connaissances.
- **Responsable de l'Interface Utilisateur** : Responsable du développement de l'interface en Flask, FastAPI ou tkinter et de la restitution des réponses en résumés.

### Plan de Communication

- **Réunions hebdomadaires** : Point rapide pour aligner les tâches et identifier les obstacles.
- **Compte-rendu mensuel** : Envoyé à tous les membres et sauvegardé pour un suivi de l'évolution.
- **Feedback Utilisateurs** : Sessions de tests régulières pour adapter l'interface et la pertinence des résultats.

\*WPL = Work Package Leader