



Projet fin d'année 2023/2024

# Conception et réalisation d'un moteur de recherche semantique MFI



Presenté par :

**Ait elfatmi Najah**

Encadré par :

**Mr.bencharef omar**



# Remerciements

Je tiens à exprimer ma profonde gratitude à chacun d'entre vous pour votre précieuse contribution à mon stage. Votre soutien, vos encouragements et votre expertise ont été inestimables tout au long de cette expérience enrichissante. À ma mère, Zouhri Nazha, et à mon père, Ait elfatmi Ahmed, je vous remercie pour votre amour, votre soutien indéfectible et vos conseils avisés qui m'ont permis de persévérer et d'atteindre mes objectifs. À l'équipe de ALiDanTeK, je suis reconnaissante pour l'opportunité que vous m'avez offerte de m'épanouir professionnellement et pour les enseignements précieux que j'ai pu tirer de cette expérience. Enfin, à mon encadrante, Mr bencharef Omar, je vous adresse un remerciement spécial. Votre guidance, vos conseils éclairés et votre patience ont été essentiels pour ma croissance professionnelle. Votre expertise m'a non seulement permis d'acquérir de nouvelles compétences, mais elle m'a aussi inspiré à atteindre l'excellence. Ce stage a été une étape cruciale dans mon parcours, et cela n'aurait pas été possible sans vous tous. Vos efforts ont contribué à mon développement personnel et professionnel, et je suis profondément reconnaissant de vous avoir à mes côtés.

# Abstract

This project, conducted as part of my final-year internship at ALiDanTeK, focuses on the design and implementation of a semantic search engine for Canadian tax documents. The system is designed to help tax professionals efficiently access relevant legal information by simplifying the search process.

We utilized web scraping techniques with Python to extract legal texts from the CANLII website, a major repository of Canadian legal documents. With the help of generative artificial intelligence and large language models (LLMs), we generated questions based on the scraped texts. These questions, along with their corresponding answers, were stored in an Excel file and later transferred to a Qdrant database. When users submit a query, an API searches the database for similar questions and retrieves the most relevant answers.

The project followed Scrum and Agile methodologies to enable efficient project management and iterative development. By integrating modern technologies such as web scraping, natural language processing, and a semantic search model, the system offers a user-friendly solution that enhances the productivity of tax professionals by providing quick access to accurate legal information.

# Résumé

Dans le cadre de mon projet de fin d'année chez ALiDanTeK, j'ai participé à la conception et à la réalisation d'un moteur de recherche sémantique destiné à simplifier la recherche d'informations fiscales canadiennes. Ce projet répond au besoin des fiscalistes d'accéder rapidement et efficacement à des réponses précises issues de la législation canadienne.

Nous avons utilisé des techniques de scraping avec Python pour extraire les textes juridiques du site CANLII, une ressource majeure pour les documents législatifs canadiens. Ensuite, à l'aide de l'intelligence artificielle générative et LLM (large language models), nous avons généré automatiquement des questions basées sur ces textes. Chaque texte et ses questions générées ont été organisés dans un fichier Excel, puis stockés dans une base de données vectorielle, Qdrant.

Lorsque l'utilisateur pose une question via l'interface, un API interroge la base de données pour trouver les réponses les plus pertinentes en se basant sur les questions similaires . Cette approche permet de trouver des réponses confidentielles et de manière pertinente, tout en rendant l'expérience utilisateur fluide et intuitive.

En suivant les méthodologies Scrum et Agile, nous avons pu itérer sur les différentes phases du projet, allant de la collecte des données à l'implémentation du back-end et front-end. Ce projet offre une solution innovante, pratique et efficace pour améliorer l'accès à l'information fiscale.

# Liste des acronymes

- API Application Programming Interface  
CANLI I Institut canadien d'information juridique  
CMMI Capability Maturity Model Integration  
CRM Customer Relationship Management  
EMSI Ecole Marocaine Des Sciences de L'ingenieurs  
ERP Enterprise Resource Planing  
IA Intelligence artificielle  
IIR Ingénieur en Informatique et Réseaux  
LLM Les grands modèles de langage  
MFI Mon Fiscaliste Intelligent  
MIAGE Méthodes Informatiques Appliquées à la Gestion d'Entreprise  
MTP Mo del-View-Presenter  
MVC Mo del-View-Controller  
MVVM Mo del-View-Viewmo del  
MVW Mo del-View-Whatever  
PFA Projet de Fin d'Année  
UML Unified Modeling Language

# Glossaire

- AI: Discipline informatique visant à créer des systèmes capables d'effectuer des tâches normalement réservées à l'intelligence humaine.
- API: Interface qui permet à des applications de communiquer entre elles en utilisant des requêtes et des réponses.
- Back-end : Partie de l'application qui gère la logique serveur, le stockage des données et les requêtes.
- Base de données vectorielle : Type de base de données qui stocke des vecteurs pour permettre des recherches de similarité, comme dans le cadre de la recherche sémantique.
- CANLII : Plateforme en ligne qui offre un accès gratuit aux textes de loi et décisions judiciaires canadiens.
- Front-end : Partie visible de l'application avec laquelle les utilisateurs interagissent, ici développée pour permettre aux utilisateurs de poser des questions.
- LLM: Modèles de traitement du langage naturel à grande échelle, utilisés pour générer ou comprendre du texte
- Requête (Query) : Demande d'information envoyée par l'utilisateur au moteur de recherche pour obtenir une réponse pertinente.
- SCRAPING : Technique utilisée pour extraire automatiquement des données à partir de sites web
- Web Scraping : Procédé d'extraction automatique de contenu à partir de sites web.

# Table des matières

Table des matières	18	
2.6 Cas d'utilisation globale .....	20	
Conclusion.....	21	
Chapitre 3 : Analyse et conception . . . . .	22	
Introduction.....	22	
3.1	Le formalisme UML.....	22
3.2	Le choix de UML.....	22
3.3 Modèle-Vue-Contrôleur.....	24	
3.4 Conception	28	
Conclusion.....	29	
Chapitre 4 : Technologies utilisées . . . . .	30	
Introduction.....	30	
4.1	Framework FLASK.....	30
4.2	4.1.1 Caractéristiques principales du Framework FLASK ..	31
4.3	4.1.2 Pourquoi utiliser Flask pour Mon projet MFI?....	31
4.4	PYTHON.....	32
4.5	Reactjs.....	33
4.6	XAMPP.....	34
	UML.....	34
	SCRAPING.....	35
4.7 Outil de conception.....	36	
4.8	Qdrant.....	37
4.9	PHPMYADMIN.....	37
	4.10 Visual Studio Code.....	38
	Conclusion.....	39
Chapitre 5 : Réalisation et mise en œuvre . . . . .	40	
5.1	Introduction.....	40
	5.2 Présentation de l'application MFI: .....	40
5.2.1	Paged'accueilMFI.....	41
5.2.2	PageAboutusMFI.....	41
5.2.3	PagedecontacterMFI.....	42
	5.2.4 Page de localisation MFI : .....	42
	5.2.5 Page d'authentification : .....	44
	5.2.6 interface du moteur de recherche : .....	45
Affichage de réponses:.....	46	
5.2.8 Affichage du texte juridique : .....	46	
5.2.9 Enregistrement de réponse: .....	47	
5.2.10 Enregistrement de la question dans la base donnée : ..	47	
5.2.11 Historique des questions posées : .....	48	
5.2.12 Affichage des réponses enregistrer de chaque question :	49	
5.2.13 Base de données vectorielles .....	50	
Conclusion et Perspectives . . . . .	50	
Bibliographie . . . . .	50	

# Liste des figures

Figure1.1:Alidanteklogo.....	3
Figure 1.2 : Développement ALiDanTeK .....	5
Figure 1.3 : Présence de l'entreprise ALiDanTeK .....	5
Figure 1.4 : Fiche Technique de l'entreprise ALiDanTeK ...	6
Figure 1.5 : logo du Site Web de la législation .....	8
Figure1.6:MéthodeSCRUM.....	9
Figure1.7:EquipeScrum.....	10
Figure1.8:Sprints.....	11
Figure 1.9 : Diagramme de Gantt.....	12
Figure 2.1 : Architecture du front-end .....	17
Figure 2.2 : Architecture backend flask .....	18
Figure3.1:Modèle-Vue-Contrôleur.....	23
Figure 3.2 : Diagramme de cas d'utilisations du système MFI.	24
Figure 3.3 : Diagramme de classe MFI .....	25
Figure 3.4 : Diagramme de séquence d'authentification ....	26
Figure 3.5 : Diagramme de séquence d'une question .....	27
Figure 4.1 : Logo du Framework Flask .....	30
Figure4.2:LogodePython.....	31
Figure4.3:Logodereactjs.....	32
Figure4.4:LogodeXAMPP.....	33
Figure4.5:Logod'UML.....	34
Figure4.6:Shemascraping.....	34
Figure4.7:LogostarUML.....	35
Figure4.8:LogodeQdrant.....	36
Figure 4.9 : Logo de PHP MY ADMIN.....	37
Figure 4.10 :Logo de Visual Studio Code.....	38
Figure5.1:Paged'accueilMFI.....	40
Figure5.2:PageAboutusMFI.....	41
Figure 5.3 : Page de contacter MFI.....	41
Figure 5.4 : Page de localisation MFI.....	42
Figure 5.5 : Page d'authentification.....	42
Figure 5.6 : Base de donnée utilisateur.....	43
Figure 5.7 : Page de localisation MFI.....	44
Figure 5.8 : moteur de recherche MFI.....	44
Figure5.9:Pagedereponses.....	45
Figure5.10:texte juridique.....	45
Figure5.11:Enregistrement reponse.....	46

Figure 5.12 :base de donnée question. . . . .	4
Figure 5.13 :Historique des questions. . . . .	6
Figure 5.14 :page de details. . . . .	4
Figure5.15:Qdrant DB. . . . .	7
	4
	7
	4
	8

# Introduction générale

Avec la croissance massive du secteur informatique, les exigences des clients, la criticité des flux de données traitées et la nécessité de limiter les défauts des logiciels sont devenues des préoccupations majeures. Ces enjeux peuvent conduire à la frustration des utilisateurs, voire à la perte de clients, si les solutions ne répondent pas à leurs attentes en termes de performance et de fiabilité. C'est dans ce contexte que ALiDanTeK a adopté un plan stratégique visant à s'aligner sur l'évolution rapide du domaine du génie logiciel, tout en mettant les besoins des clients au centre de ses priorités.

Une des composantes clés de cette stratégie est l'importance accrue accordée au test logiciel, un processus essentiel pour s'assurer que les produits développés respectent les spécifications fonctionnelles et offrent une expérience utilisateur optimale. Cette phase, bien que coûteuse en termes de ressources humaines et financières, est cruciale pour garantir la qualité des logiciels et éviter les anomalies susceptibles d'affecter les performances du produit final.

C'est dans ce cadre d'innovation et de recherche d'amélioration continue que s'inscrit mon projet de fin d'études, intitulé : « Conception et réalisation d'un moteur de recherche sémantique ». Ce projet a pour objectif de faciliter et d'accélérer la recherche dans les documents fiscaux canadiens, en permettant aux fiscalistes de trouver des informations pertinentes en posant des questions en langage naturel. Grâce à l'intégration des technologies modernes telles que le scraping, une ressource majeure pour les documents législatifs canadiens. Ensuite, à l'aide de l'intelligence artificielle générative et LLM , et une interface utilisateur intuitive, cette solution propose une approche novatrice pour améliorer l'efficacité et la productivité des utilisateurs.

Ce rapport comporte les Chapitres décrit ci-dessous :

Chapitre 1 : Contexte général du projet

Chapitre 2 : Etude Préliminaire et Fonctionnelle

Chapitre 3 : Analyse et Conception

Chapitre 4 : Technologies utilisées

Chapitre 5 : Réalisation et mise en œuvre

---

# Chapitre 1

## Contexte général

---

## Introduction

Ce premier chapitre est une présentation générale de l'entreprise ALiDanTek. Nous abordons également la dématérialisation fiscale et nous décrivons le déroulement du stage.

### 1.1 Présentation de L'entreprise ALIDANTEK



Figure 1.1 – Alidantek logo.

Notre entreprise dispose d'une équipe de professionnels expérimentés et passionnés, qui mettent leur savoir-faire et leur créativité au service de chaque projet. Nous proposons une approche personnalisée et adaptée à vos besoins, en prenant le temps de comprendre vos objectifs et votre audience cible...

Des solutions sur mesure : Nous vous proposons des solutions, qui vous permettront de mettre en avant votre entreprise et de vous démarquer de vos concurrents.

N'hésitez pas à nous contacter : pour plus d'informations sur nos services et pour discuter de vos projets. Nous serons heureux de vous aider à atteindre vos objectifs !

AliDanTek est une société canadienne de consulting et de développement de solutions informatiques, avec des bureaux à Montréal et à Marrakech. ALiDanTeK est une entreprise multinationale qui aide les sociétés à se démarquer de ses concurrents en utilisant la technologie pour automatiser et optimiser les opérations, et transforme les entreprises de la plupart des grands secteurs avec des solutions numériques puissantes et adaptables qui répondent aux besoins d'aujourd'hui et de demain

## 1.2 Nos meilleurs Services

(IA) Intelligence Artificielle

Integration de système ERP

Formation Professionnelle

Certification Professionnelle

Centre de recherche et d'innovation en technologies de l'information

→ Cellules de recherches pour des projets innovants :

Équipe de chercheurs (professeurs et PHD).

→ Incubateurs de projets innovants :

Intelligence Artificielle, Machine Learning, Blockchain, détection d'image, « Webscraping », etc... .

Prise en charge des dossiers de recherche scientifique et développement expérimental (RSDE).

→ Avantages :

Moyen de financement pour réduire les coûts de développement informatique

centre d'expertise en génie logiciel (développement d'applications)

→ Développement personnalisé :

Équipe de plus de 15 Ingénieurs logiciels.

→ Utilisation des meilleures pratiques de Génie Logiciel et d'Agilité.

→ Certification en génie Logiciel au niveau 3 du CMMI (en cours).

→ Avantages :

Approche de livraison graduelle des projets selon le maximum de valeur pour le client.

Avec un échéancier et budget fixe.

Tarif très concurrentiel.

centre de support des produits

→ Prise en charge des produits logiciels développés par nos clients :

Hébergement des produits.

Analyse de la capacité des équipements.

Disponibilité des produits à 99,9

Support à la production 24 heures / 7 jours sur 7.

→ Avantages :

Approche client en main.

Minimise les investissements pour nos clients.

Tarif fixe pas de surprise.

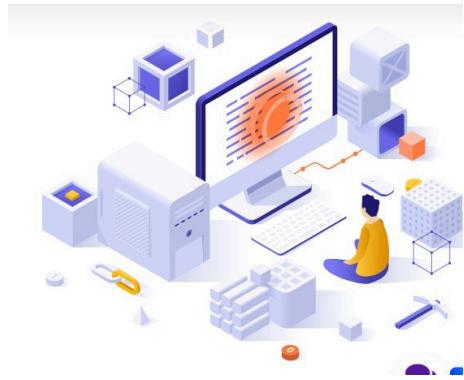


Figure 1.2 – Développement ALiDanTeK

### 1.3 Présence mondiale



Figure 1.3 – Présence de l'entreprise ALiDanTeK

### 1.4 POURQUOI CHOISIR ALiDanTeK

#### Pour la tranquillité d'esprit

Dans plus de la moitié des développements, les livrables ont de grands retards et les coûts explosent. Avec ALiDanTeK et grâce à leur approche unique vous êtes certains que les livrables se font dans un délai raisonnable et nos coûts sont fixes. Les membres d'ALiDanTeK sont stables et compétentes et ainsi vous êtes à l'abri des aléas du marché concernant la pénurie et la volatilité de la main d'œuvre. Nous sommes convaincus que notre entreprise est la solution idéale pour vous aider à développer votre présence en ligne. Nous utilisons différents outils logiciels pour nos développements. Voici des exemples de technologie de pointe utilisées : « Front end » : React, Angular, Flutter, etc... « Back end » : Laravel , SpringBoot, .net, Node.js avec webservices, etc...

AliDanTek - 1er choix pour les solutions informatiques :

Plus de 25 ans de travail dans les services informatiques développant des applications logicielles et des applications mobiles pour des clients du monde entier. Pour votre industrie très spécifique, nous avons des solutions informatiques sur mesure. Cadre du stage :

### 1.5 Cadre du stage :

Le stage a été effectué à l'entreprise ALiDanTeK au sein de la division de Marrakech, spécialiste en

Recherche et innovation en technologies de l'informatique

### 1.6 Fiche technique d'ALiDanTeK :

<u>Nom de l'entreprise</u>	ALiDanTeK
<u>Forme Juridique</u>	Sarl (Société commerciale où la responsabilité des associés est limitée au montant de leurs apports)
<u>Siège Social</u>	63 Carson, Dollard-Des-Ormeaux (Montréal) QC, Canada H9B 2J4
<u>Activité</u>	Centre de recherche, Génie logiciel (développement d'applications sur mesure en AI ou autre), Centre de support et d'évolution de produits, Conseil et formation (CRM, ERP, PMP, AI, Agilité, etc.)
<u>Produits</u>	Centre d'évolution des <b>produits</b> . → Prise en charge de l'évolution des <b>produits</b> : Gestion des problèmes et correctifs.
<u>Création</u>	2021
<u>Site web</u>	<a href="https://www.alidantek.ca/">https://www.alidantek.ca/</a>

Figure 1.4 – Fiche Technique de l'entreprise ALiDanTeK

### 1.7 Déroulement du stage :

Le déroulement du stage peut être divisé en deux phases, 2 semaines de formation et 2 mois dédiés à la réalisation du projet de stage.

. Formations :

le full stack : Une série de formations sur le front et le backend , comprenant également une explication détaillée du projet MFI (Mon Fiscaliste Intelligent).

## 1.8 Contexte du projet

Le domaine fiscal est une composante cruciale pour les entreprises et les individus, nécessitant une compréhension précise et à jour des lois et règlements en vigueur. Cependant, l'accès à cette information juridique peut s'avérer complexe, notamment en raison de la masse considérable de textes législatifs, des mises à jour régulières et de la terminologie souvent difficile à comprendre pour les non-spécialistes.

Dans ce contexte, les outils traditionnels de recherche juridique, comme les bases de données statiques ou les moteurs de recherche génériques, montrent rapidement leurs limites. La nécessité d'une solution capable de comprendre le langage naturel et d'offrir des réponses pertinentes en fonction des besoins spécifiques des utilisateurs devient essentielle.

C'est dans cet environnement que s'inscrit le projet Mon Fiscaliste Intelligent (MFI). Il vise à répondre à ces enjeux en développant un moteur de recherche sémantique spécialisé dans le domaine fiscal canadien, offrant des réponses juridiques pertinentes et contextualisées. En intégrant des technologies avancées telles que une ressource majeure pour les documents législatifs canadiens et à l'aide de l'intelligence artificielle générative et LLM et la base de données Qdrant pour le stockage sémantique, MFI ambitionne de révolutionner la manière dont les professionnels du droit accèdent aux informations fiscales.

## 1.9 Problématique

La problématique à laquelle ce projet tente de répondre est la suivante : Comment optimiser l'accès à l'information fiscale canadienne à travers un outil de recherche intelligent capable de fournir des réponses juridiques précises et contextualisées, tout en tenant compte de la complexité et de l'évolution constante des textes de loi?

Les utilisateurs – qu'il s'agisse d'avocats, de fiscalistes ou de citoyens cherchant à mieux comprendre les obligations fiscales – font face à plusieurs défis :

La difficulté de trouver des informations précises dans un volume immense de données légales. Le manque d'outils qui puissent interpréter les questions posées en langage naturel et fournir des réponses adaptées. L'absence d'une solution capable de s'adapter aux évolutions constantes des textes fiscaux.

## 1.10 Les objectifs du projet

L'objectif principal de ce projet est de concevoir et mettre en œuvre un moteur de recherche sémantique innovant, capable de :

Comprendre et interpréter les questions posées en langage naturel par les utilisateurs. Extraire et fournir des réponses juridiques pertinentes, basées sur les textes de loi canadiens. Améliorer l'efficacité des recherches fiscales en offrant une expérience utilisateur simple, rapide et précise grâce à une interface web intuitive. Automatiser la collecte et la structuration des informations fiscales à partir de bases de données comme CANLII, tout en assurant la mise à jour continue des données. Ce projet vise également à exploiter les avancées en matière d'intelligence artificielle et de traitement du langage naturel pour offrir un outil évolutif, capable de s'adapter aux futures évolutions des textes de loi, tout en restant accessible et utile pour une grande variété d'utilisateurs.



Figure 1.5 – logo du Site Web de la législation

## 1.11 Gestion de projet

### 1.11.1 La méthode AGILE SCRUM

La méthode Scrum est un cadre de gestion de projet qui entre dans la catégorie des méthodes agiles. Les méthodes agiles prennent à cœur la flexibilité dans le processus de développement, l'implication du client dans ce processus et la prise en compte de ses remarques tout au long du projet. Par rapport à une méthode classique de gestion de projet, on ne rencontre pas uniquement un chef de projet et une ou plusieurs équipes dans l'organisation. Trois rôles Scrum principaux sont définis dans le processus Scrum :

- Le Scrum Master a pour rôle de vérifier l'application des principes de Scrum, s'assure que la communication fonctionne correctement au sein de l'équipe et explore toutes les pistes d'améliorations en termes de productivité et de savoir-faire ;
- L'équipe Scrum regroupe généralement une dizaine de personnes (mais de très gros projets peuvent rassembler plusieurs dizaines de personnes) chargées de la réalisation des différentes tâches. Chaque membre est au même niveau, il n'y a pas de hiérarchie et chacun peut contribuer en fonction de ses compétences ;
- Enfin, le Product Owner est un expert du processus métier du client, définit les spécifications fonctionnelles et arbitre les priorités dans les réalisations

ma kaynch had ghant f figures/diagrammes choufi souri dialy uploaditih deja ? oui



Figure 1.6 – Méthode SCRUM.

Avant chaque sprint, il faut faire une planification du sprint pour définir les objectifs et les tâches à accomplir. Cette planification du sprint permet à l'équipe de développement de définir le travail à réaliser et de s'assurer que tout le monde est sur la même longueur d'onde.

Pendant le sprint, l'équipe de développement travaille de manière autonome pour terminer les tâches assignées.

On prépare des réunions quotidiennes, que l'on appelle « daily scrum », pour garder l'équipe à jour sur l'avancement du programme. Mais également, pour identifier les éventuels problèmes ou obstacles.

Après la réunion de revue de sprint, on organise une rétrospective pour évaluer le sprint écoulé.

Elle permet aussi d'identifier les améliorations possibles et apporter des ajustements pour les sprints futurs.

En adoptant une approche itérative, l'équipe de développement peut livrer des fonctionnalités à valeur ajoutée de manière régulière, tout en restant flexible pour s'accorder aux changements et aux besoins des futurs usagers. [4]

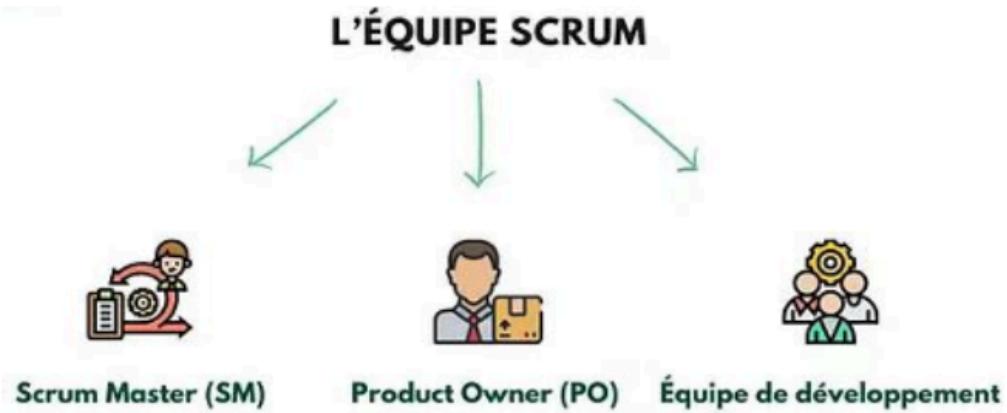


Figure 1.7 – Equipe Scrum

- Le Scrum Master : Facilite l'équipe SCRUM, assure le respect des principes SCRUM, améliore la communication et la productivité.
- Le Product Owner : Représente les clients, gère le carnet de produit, définit les objectifs du projet, et priorise les éléments.
- L'équipe de développement : Auto-organisée et interdisciplinaire, réalise le travail du projet, collabore pour atteindre les objectifs du sprint, et chaque membre est polyvalent.

### 1.11.2 Conduite du projet

Le projet a été conduit selon les méthodologies Scrum et Agile, permettant des cycles de développement courts avec des révisions régulières et des ajustements en fonction des retours des utilisateurs et des parties prenantes. Chaque sprint a permis de valider des fonctionnalités spécifiques, comme l'intégration de la base de données ou l'authentification des utilisateurs.

## Présentation des sprints :

Sprint	Durée	Tâches principales	Livrables
Sprint 1	Semaine 1 à 2	<ul style="list-style-type: none"> <li>- Analyse des besoins et spécifications fonctionnelles</li> <li>- Définition de l'architecture du projet</li> <li>- Répartition des tâches</li> <li>- Choix des technologies</li> </ul>	<ul style="list-style-type: none"> <li>- Document des spécifications</li> <li>- Diagramme d'architecture</li> </ul>
Sprint 2	Semaine 3 à 4	<ul style="list-style-type: none"> <li>- Développement du scraping avec Python pour extraire les documents fiscaux du site CANLII</li> <li>- Nettoyage et préparation des données extraites</li> </ul>	<ul style="list-style-type: none"> <li>- Script de scraping fonctionnel</li> <li>- Données extraites propres</li> </ul>
Sprint 3	Semaine 5 à 6	<ul style="list-style-type: none"> <li>- Génération automatique des questions à l'aide de ChatGPT-4</li> <li>- Organisation des questions et réponses dans un fichier Excel</li> <li>- Validation des résultats</li> </ul>	<ul style="list-style-type: none"> <li>- Questions générées et fichier Excel finalisé</li> </ul>
Sprint 4	Semaine 7 à 9	<ul style="list-style-type: none"> <li>- Conception de la base de données Qdrant</li> <li>- Importation des données (questions et réponses) dans la base</li> <li>- Tests de requêtes et d'indexation</li> <li>- Développement du back-end avec Python et Flask</li> </ul>	<ul style="list-style-type: none"> <li>- Base de données Qdrant configurée</li> <li>- Tests de requêtes</li> <li>- API fonctionnelle</li> </ul>
Sprint 5	Semaine 10 à 12	<ul style="list-style-type: none"> <li>- Création d'une API pour interroger la base de données</li> <li>- Connexion du back-end à la base de données</li> <li>- Développement du front-end en React.js</li> <li>- Création de l'interface utilisateur pour la recherche et l'affichage des résultats</li> <li>- Tests d'intégration</li> </ul>	<ul style="list-style-type: none"> <li>- Connexion entre backend et Qdrant</li> <li>- Interface utilisateur prête</li> <li>- Application testée et intégrée</li> </ul>

Figure 1.8 – Sprints

- Diagramme de GANTT :

Il s'agit d'un outil pratique pour la planification de projets, car il fournit une vue d'ensemble des tâches prévues et permet à toutes les personnes impliquées de savoir quelle activité doit être effectuée et à quelle date précise. Ce type de diagramme présente la date de début et de fin du projet. Dans notre cas, étant donné que nous avons suivi une méthodologie agile, les tâches de développement ont été regroupées en sprints (sprint backlog).[1]



Figure 1.9 – Diagramme de Gantt.

## Conclusion

En somme, le projet MFI s'inscrit dans un cadre plus large d'amélioration de l'accès à l'information juridique grâce aux avancées technologiques. Il représente une solution novatrice pour répondre aux besoins croissants en matière de traitement automatisé des données fiscales et juridiques. Le moteur de recherche sémantique développé lors de ce projet apporte une réelle valeur ajoutée pour les professionnels du droit, en leur permettant de gagner du temps dans leurs recherches tout en garantissant une plus grande précision des résultats obtenus. Ce projet montre comment l'alliance entre l'intelligence artificielle et les bases de données avancées peut transformer et moderniser la manière dont les informations juridiques sont traitées et consultées.

---

## Chapitre 2

### Étude Préliminaire et Fonctionnelle

---

## Introduction

L'étude préliminaire est une étape cruciale dans la réalisation de tout projet informatique. Elle permet de définir clairement les besoins, de comprendre les enjeux du projet, et d'établir les bases pour les prochaines phases de développement. Cette étude s'attache à analyser l'existant, à identifier les exigences fonctionnelles et non fonctionnelles, ainsi qu'à proposer un cadre conceptuel pour la solution à développer. Ce chapitre présente les différentes analyses menées en amont du développement du moteur de recherche sémantique Mon Fiscaliste Intelligent (MFI), en se concentrant sur les aspects fonctionnels, techniques, et les attentes des utilisateurs.

### 2.1 Étude préliminaire

L'objectif de cette étude préliminaire est de comprendre le contexte dans lequel s'inscrit le projet MFI, d'analyser les besoins des utilisateurs et de déterminer les contraintes techniques à prendre en compte. Les moteurs de recherche classiques ne sont pas adaptés à des domaines spécifiques tels que la fiscalité, où une compréhension fine du langage et du contexte est essentielle pour fournir des résultats pertinents.

En amont du projet, plusieurs points ont été étudiés :

Analyse de l'existant : une revue des outils de recherche juridique actuels a montré leurs limitations, notamment en termes de précision et de compréhension des questions posées en langage naturel. Étude des technologies : l'utilisation de modèles d'intelligence artificielle est une ressource majeure pour les documents législatifs canadiens. Ensuite, à l'aide de l'intelligence artificielle générative et LLM et une base de données vectorielles comme Qdrant a été jugée pertinente pour pallier ces manques et offrir une solution évolutive et intelligente. Analyse des utilisateurs : les principaux utilisateurs ciblés sont des fiscalistes, avocats, juristes et citoyens cherchant des réponses rapides et précises à des questions fiscales complexes.

### 2.2 Etude fonctionnelle

L'étude fonctionnelle vise à définir les fonctionnalités attendues du système. Le moteur de recherche sémantique doit permettre à l'utilisateur de poser des questions en langage naturel et de recevoir des réponses contextualisées sous forme de documents ou de recommandations pertinentes.

#### 2.2.1 Capture des besoins fonctionnels

Les besoins fonctionnels sont les fonctionnalités essentielles que le système doit offrir pour répondre aux attentes des utilisateurs. Dans le cadre du projet MFI, les besoins fonctionnels identifiés sont les suivants :

Recherche en langage naturel : l'utilisateur doit pouvoir formuler des questions en langage naturel (par exemple : "Quelles sont les obligations fiscales pour une entreprise en

2024 ?"). Génération automatique de réponses : le système doit fournir des réponses pertinentes, extraites de la base de données juridique, notamment à partir des textes de loi canadiens. Affichage des résultats : les résultats doivent être affichés sous forme de documents ou de résumés accompagnés de liens vers les textes de loi complets. Mise à jour automatique des données : les textes de loi et autres informations doivent être mis à jour automatiquement pour garantir la pertinence des réponses. Filtrage des résultats : les utilisateurs doivent pouvoir filtrer les réponses selon différents critères (par exemple : année, type d'entreprise, type d'impôt).

### 2.2.2 Capture des besoins non fonctionnels

Les besoins non fonctionnels concernent les caractéristiques techniques et qualitatives que le système doit respecter. Ils sont cruciaux pour garantir une expérience utilisateur fluide et fiable. Voici les besoins non fonctionnels du projet MFI :

Performances : le système doit fournir des réponses en moins de 5 secondes pour garantir une expérience utilisateur fluide. Scalabilité : la solution doit pouvoir s'adapter à une augmentation du nombre d'utilisateurs ou de données sans perte de performance. Sécurité : les données échangées, notamment les questions posées par les utilisateurs, doivent être protégées pour respecter la confidentialité des informations sensibles. Fiabilité : le système doit fonctionner de manière continue avec un temps d'indisponibilité minimum. Accessibilité : l'interface utilisateur doit être accessible aux personnes ayant des besoins spécifiques, telles que les utilisateurs souffrant de déficiences visuelles. Compatibilité : le système doit être compatible avec différents types d'appareils et navigateurs web.

## 2.3 Caractéristiques Principales:

Scraping des documents : Extraction automatisée des textes juridiques à partir de sources fiables comme CANLII, avec des mises à jour régulières pour garantir l'actualité des informations fiscales. Base de connaissances : Un référentiel structuré regroupant des articles de loi et des informations fiscales, optimisé pour des requêtes rapides et précises. Génération de questions avec l'intelligence artificielle : L'intelligence artificielle génère des questions à partir des textes juridiques extraits, et ces questions sont associées à des réponses pertinentes, toutes stockées dans une base de données vectorielle Qdrant. Interface conviviale : Développée en React.js, l'interface permet aux fiscalistes de poser des questions en langage naturel et d'obtenir des réponses claires, sans nécessiter de compétences techniques. Sécurité renforcée : Un système de gestion des accès sécurisé garantit que seules les personnes autorisées peuvent accéder aux fonctionnalités de la plateforme.

## 2.4 Architecture du Système :

L'architecture de l'application MFI repose sur plusieurs composants clés, assurant un fonctionnement fluide et efficace :

**Front-end (React.js)** : L'utilisateur pose une question via l'interface utilisateur développée en React.js. L'interface est conçue pour être intuitive, permettant aux utilisateurs de formuler leurs questions en langage naturel. **Authentification** : Avant de traiter la requête, l'utilisateur doit passer par un processus d'authentification sécurisé pour garantir l'accès aux fonctionnalités de la plateforme. **Back-end (Flask)** : Une fois l'authentification réussie, la requête de l'utilisateur est envoyée au serveur back-end développé en Flask. Le back-end traite la requête et interroge la couche de données pour obtenir des réponses pertinentes. **Data Layer (Qdrant)** : Le back-end utilise Qdrant, une base de données vectorielle, pour rechercher des questions similaires dans la base de données. Qdrant permet de récupérer rapidement les questions associées et les réponses pertinentes, optimisant ainsi le temps de réponse. **Retour à l'utilisateur** : Les résultats trouvés sont renvoyés au front-end, où ils sont présentés à l'utilisateur sous forme de réponses claires et compréhensibles.

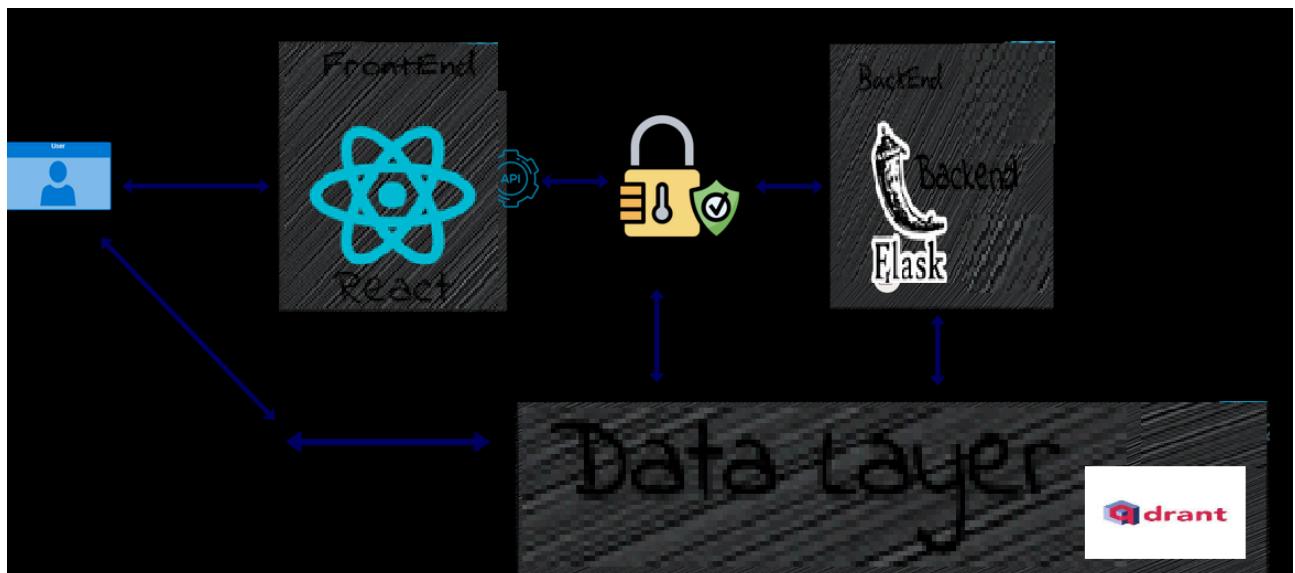


Figure 2.1 – Architecture du front-end

## 2.5 Architecture du Back-end avec Flask

L'architecture de l'application Flask repose sur une structure bien définie, facilitant la gestion des requêtes et l'interaction avec la base de données :

**Gestion des Requêtes** : Flask reçoit une requête lorsque l'utilisateur demande la recherche d'une question. La requête est d'abord traitée par un contrôleur qui s'assure que les paramètres nécessaires sont bien fournis. **Appel au Service** : Le contrôleur invoque un service spécifique à l'application, qui contient la logique métier nécessaire pour traiter la requête. Ce service exécute une méthode préalablement codée, responsable de la recherche des questions dans la base de données. **Communication avec la Base de Données** : La

méthode du service nécessite d'accéder aux données stockées dans la base de données. Pour ce faire, l'application ne se connecte pas directement à la base de données. Au lieu de cela, elle utilise une couche de repository. Repository : Le repository est une couche d'abstraction qui gère la communication entre l'application et la base de données. Cette couche permet d'effectuer des opérations de lecture et d'écriture sur la base de données, tout en encapsulant la logique d'accès aux données. Ainsi, lorsque le service en a besoin, il interroge le repository pour récupérer les données pertinentes. Retour des Données : Une fois que le repository a renvoyé les données au service, celui-ci les traite et retourne les résultats au contrôleur. Le contrôleur renvoie finalement les résultats au front-end, où ils sont affichés à l'utilisateur.

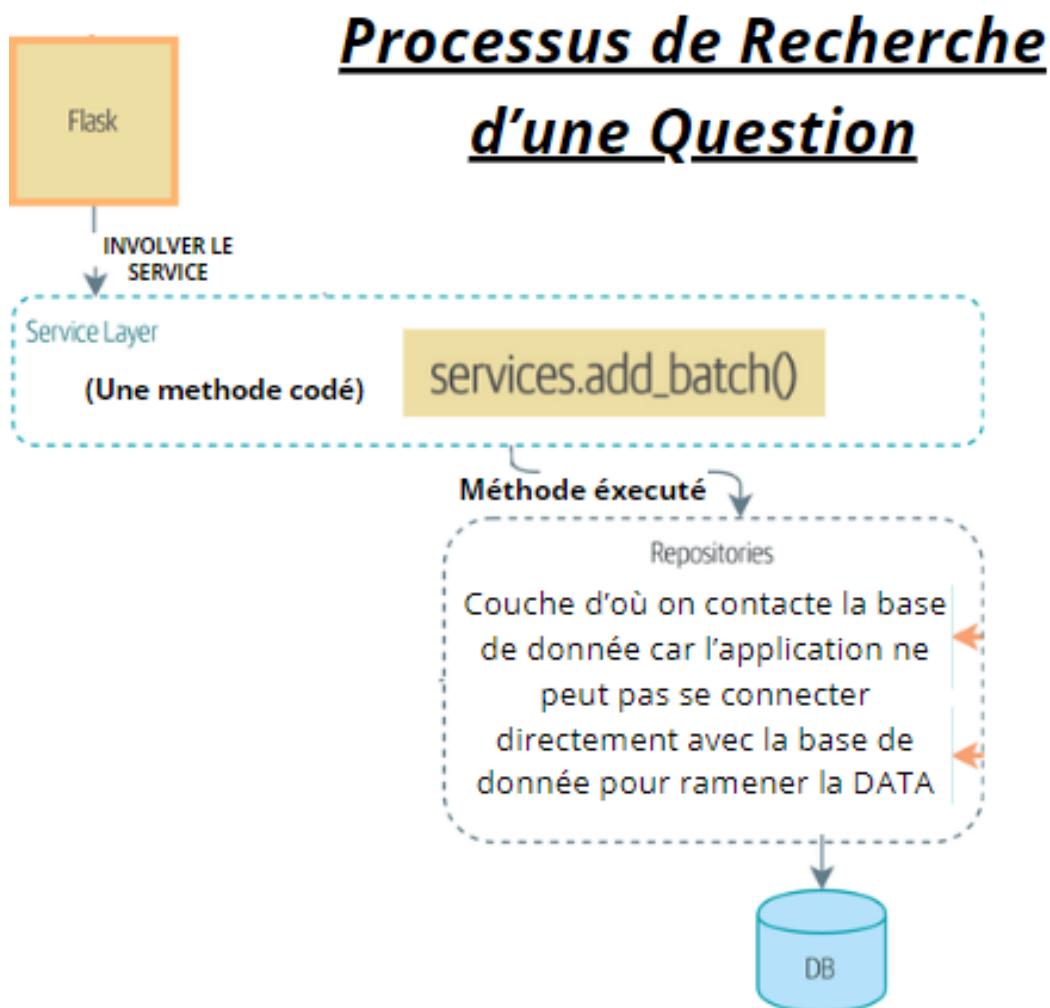


Figure 2.2 – Architecture backend flask

## 2.6 Cas d'utilisation globale

Un Dans le projet Mon Fiscaliste Intelligent (MFI), les utilisateurs doivent s'authentifier avant d'accéder aux fonctionnalités du moteur de recherche. Après connexion, ils peuvent poser des questions, consulter des réponses, enregistrer ces réponses, et accéder à leur historique de recherches. Le diagramme suivant décrit les interactions entre l'utilisateur

et le système, ainsi que les fonctionnalités accessibles après authentification.

Acteurs identifiés pour MFI :

- Utilisateur : Ce sont les fiscalistes, avocats, ou citoyens qui utilisent le moteur de recherche pour obtenir des réponses juridiques.
- Administrateur : Il gère le système, les utilisateurs et les mises à jour des données.  
Cas d'utilisation pour l'utilisateur :
  - Se connecter : L'utilisateur s'authentifie avant d'utiliser le moteur de recherche.
  - Poser une question : L'utilisateur soumet une question en langage naturel après authentification.
  - Consulter les réponses : L'utilisateur visualise les réponses pertinentes fournies par le système.
  - Enregistrer une réponse : L'utilisateur a la possibilité de sauvegarder une réponse qu'il juge importante.
  - Accéder à l'historique des recherches : L'utilisateur peut consulter les questions qu'il a posées auparavant, avec leurs réponses associées.
- Description : Ce cas d'utilisation décrit le processus permettant à un utilisateur de se connecter au système avant d'accéder au moteur de recherche sémantique.

Acteur principal :

- Utilisateur : Il s'agit du fiscaliste ou tout autre professionnel qui souhaite utiliser le moteur de recherche pour obtenir des réponses juridiques sur la fiscalité canadienne.
- Scénario principal :  
S'authentifier (inclus) :  
L'utilisateur doit s'authentifier pour accéder à la plateforme. Il entre son identifiant et son mot de passe dans l'interface de connexion. Le système vérifie les informations d'identification de l'utilisateur (validation). Si l'authentification réussit, l'utilisateur est redirigé vers l'interface principale du moteur de recherche.
- Accéder au moteur de recherche :  
Une fois connecté, l'utilisateur accède à la barre de recherche où il peut poser des questions en langage naturel. Le moteur de recherche sémantique traite la requête et retourne les réponses les plus pertinentes, basées sur les textes de loi contenus dans la base de données.

- Extensions (scénarios alternatifs) : L'utilisateur ne pourra pas accéder au moteur de recherche tant que l'authentification n'est pas réussie.
- Relations avec d'autres cas d'utilisation : L'authentification est incluse dans l'accès au moteur de recherche. Cela signifie que l'utilisateur ne pourra utiliser le moteur de recherche qu'une fois qu'il a validé son identité.

## Conclusion

L'étude préliminaire a permis de poser les bases du projet Mon Fiscaliste Intelligent en identifiant les fonctionnalités essentielles et les exigences techniques. Le moteur de recherche sémantique qui en résulte vise à offrir une expérience utilisateur fluide, performante et sécurisée, tout en répondant aux besoins spécifiques des professionnels du droit fiscal. Ces analyses permettent de cadrer les futures étapes de développement, en assurant que la solution proposée soit à la fois pertinente et évolutive.

---

# Chapitre 3

## Analyse et conception

---

## Introduction

Ce chapitre se concentre sur le périmètre d'intervention, c'est-à-dire l'étendue de notre domaine d'action. Dans cette section, on exposera la conception générale du projet, offrant ainsi une vision d'ensemble du travail requis.

### 3.1 Le formalisme UML

Le langage UML (Unified Modeling Language) est constitué de diagrammes intégrés utilisés par les développeurs informatiques pour la représentation visuelle des objets, des états et des processus dans un logiciel ou un système. Le langage de modélisation peut servir de modèle pour un projet et garantir une architecture d'information structurée; il peut également aider les développeurs à présenter leur description d'un système d'une manière compréhensible pour les spécialistes externes. UML est principalement utilisé dans le développement de logiciels orientés objet. Les améliorations apportées à la norme dans la version 2.0 la rendent également adaptée à la représentation des processus de gestion.[6] UML n'étant pas une méthode, l'utilisation des diagrammes est laissée à l'appréciation de chacun. Le diagramme de classes est généralement considéré comme l'élément central d'UML. Des méthodes, telles que le processus unifié proposé par les créateurs originels de UML, utilisent plus systématiquement l'ensemble des diagrammes et axent l'analyse sur les cas d'utilisation (« use case ») pour développer par itérations successives un modèle d'analyse, un modèle de conception, et d'autres modèles. D'autres approches se contentent de modéliser seulement partiellement un système, par exemple certaines parties critiques qui sont difficiles à déduire du code.

### 3.2 Le choix de UML

Aujourd'hui, les diagrammes UML figurent toujours comme outils incontournables et standards pour les développeurs, ainsi que pour les gestionnaires de projet, les chefs d'entreprise, les entrepreneurs technologiques et les professionnels de tous les secteurs. Parmi les raisons pour lesquelles on a choisi ce langage de modélisation :

- L'UML assure la bonne communication;
- L'UML simplifie la complexité et améliore la qualité du travail;
- UML n'est pas spécifique à un langage de programmation objet, elle peut donc être utilisée avec n'importe quel langage.

### 3.3 Modèle-Vue-Contrôleur

MVC (Model-View-Controller ou Modèle-Vue-Contrôleur) est un modèle dans la conception de logiciels. Il met l'accent sur la séparation entre la logique métier et l'affichage du logiciel. Cette «séparation des préoccupations» permet une meilleure répartition du travail et une maintenance améliorée. Certains autres modèles de conception sont basés

sur MVC, tels que MVVM (Model-View-Viewmodel), MTP (Model-View-Presenter) et MVW (Model-View-Whatever).

Les 3 parties du modèle de conception de logiciel MVC peuvent être décrites comme suit :

- Model (modèle) : gère les données et la logique métier.
- View (vue) : gère la disposition et l'affichage.
- Controller (contrôleur) : achemine les commandes des parties "model" et "view".

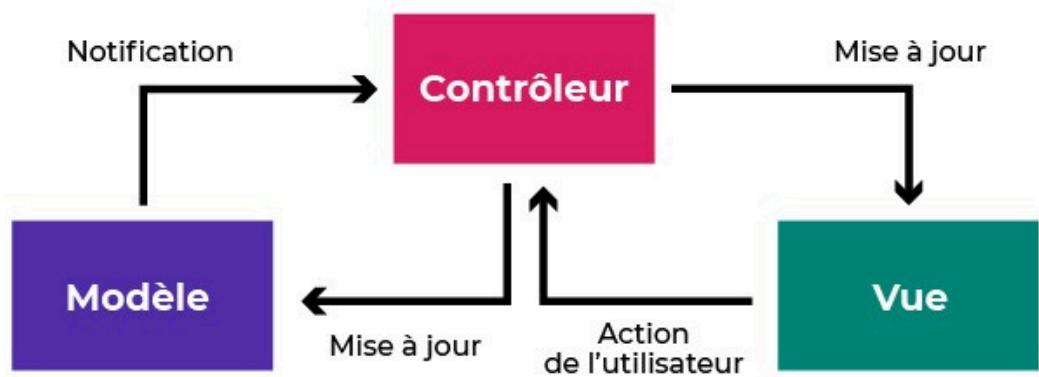


Figure 3.1 – Modèle-Vue-Contrôleur.

- Model Vue Contrôleur Le patron de conception MVC [3] (Modèle-Vue-Contrôleur) est un modèle architectural qui divise une application en trois composants principaux : le Modèle, la Vue et le Contrôleur;
- Le Modèle représente les données et la logique métier de l'application. Il est responsable de la manipulation, de la validation et de la gestion des données. Le Modèle peut interagir avec des sources de données telles que des bases de données ou des API externes pour récupérer ou stocker des données [2];
- La Vue est responsable de l'affichage des données au format approprié pour l'utilisateur ;
- Le Contrôleur assure les échanges entre la vue et le modèle notamment grâce à des composants métiers.

## 3.4 Conception

### Diagramme de cas d'utilisation

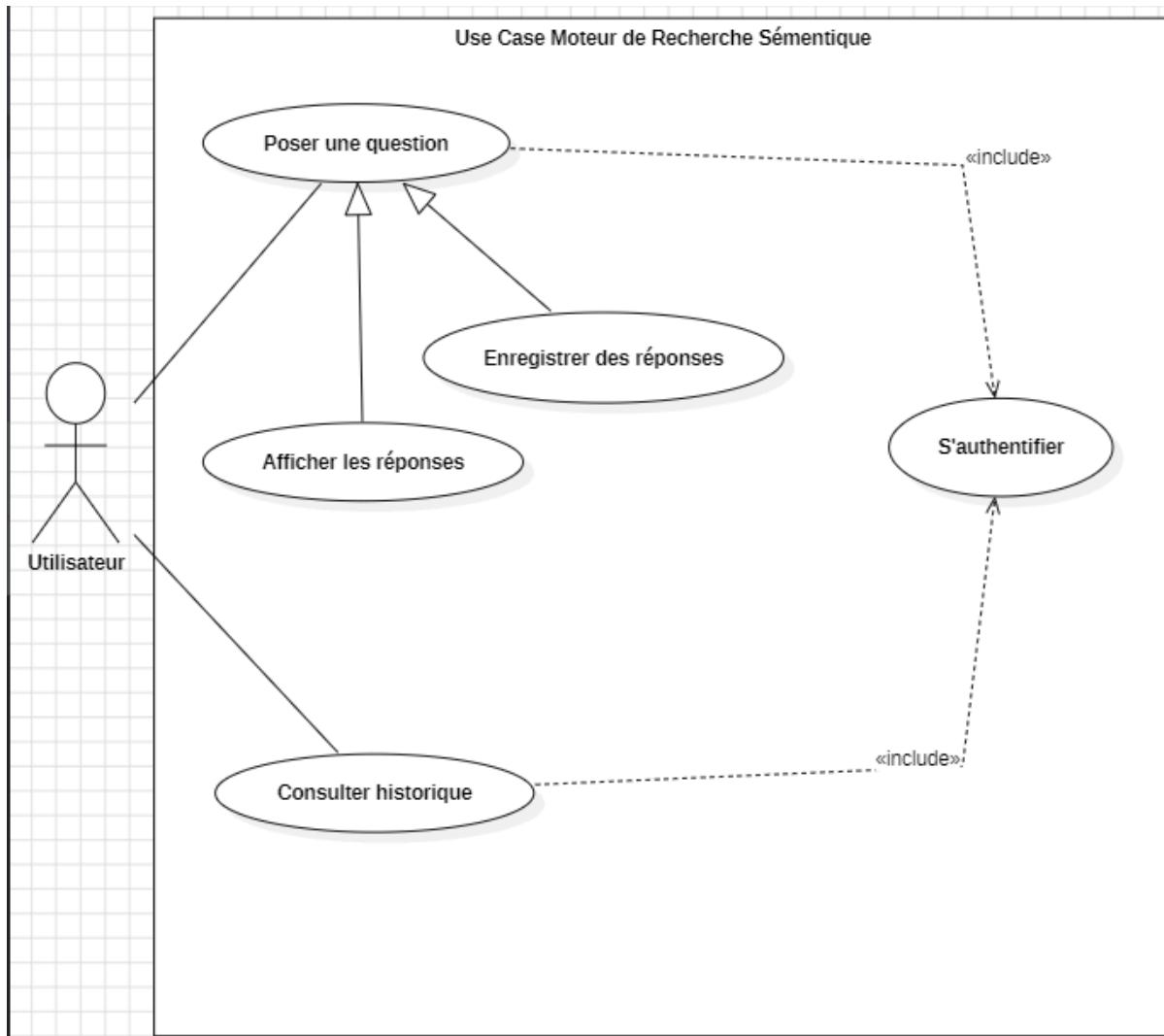


Figure 3.2 – Diagramme de cas d'utilisations du système MFI.

Dans le cadre de l'utilisation du moteur de recherche sémantique, l'authentification est une étape cruciale pour garantir que seuls les utilisateurs autorisés peuvent interagir avec le système. L'utilisateur doit d'abord se connecter en fournissant son identifiant et son mot de passe via l'interface de connexion. Une fois ces informations validées par le système, l'utilisateur accède à la plateforme, où il peut poser des questions en langage naturel et obtenir des réponses issues des textes de loi stockés dans la base de données. L'authentification est un processus intégré à l'accès au moteur de recherche, ce qui signifie que sans connexion préalable, l'utilisateur ne pourra pas interagir avec le moteur. Cette étape permet de protéger les données sensibles et d'assurer un suivi des interactions utilisateurs. En cas d'échec d'authentification, l'utilisateur sera invité à réessayer ou à réinitialiser ses informations d'identification.

- L'utilisateur doit s'authentifier pour accéder à la plateforme. Il entre son identifiant

et son mot de passe dans l'interface de connexion.

- Le système vérifie les informations d'identification de l'utilisateur (validation).
- Si l'authentification réussit, l'utilisateur est redirigé vers l'interface principale du moteur de recherche.
- Une fois connecté, l'utilisateur accède à la barre de recherche où il peut poser des questions en langage naturel.
- Le moteur de recherche sémantique traite la requête et retourne les réponses les plus pertinentes, basées sur les textes de loi contenus dans la base de données.

Diagramme de classe :

Diagramme de classe modélise le système de gestion de projet, mettant en avant les interactions entre trois rôles principaux : Employé, Chef de produit (Product Manager), et Contrôle Qualité (Quality Control). Chaque rôle a des responsabilités distinctes

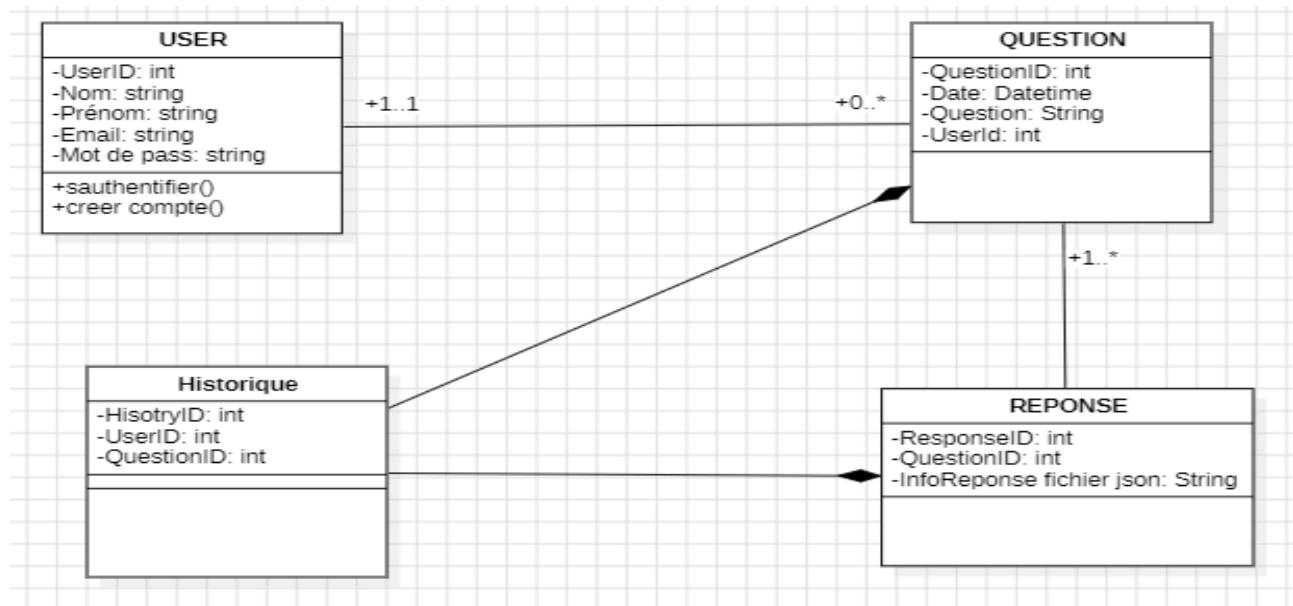


Figure 3.3 – Diagramme de classe MFI

Résumé explicatif du diagramme de classe UML du projet MFI :

Le système est basé sur plusieurs classes principales qui interagissent entre elles pour permettre à un utilisateur de poser des questions fiscales et d'obtenir des réponses pertinentes.

- Utilisateur : Un utilisateur du système doit s'authentifier pour accéder au moteur de recherche et poser des questions. Chaque utilisateur a un Role (par exemple, simple utilisateur ou autre, selon les droits du système).
- Role : Définit le rôle attribué à chaque utilisateur. Dans ce projet, bien qu'il n'y ait pas d'administrateur, le rôle est attribué pour gérer les autorisations d'accès.

- Question : Représente les questions posées par l'utilisateur, qui sont ensuite associées à des Réponses.

- Réponse : Contient les réponses aux questions posées, extraites des textes de lois canadiens, grâce au moteur de recherche sémantique.
- Historique : Enregistre les interactions de l'utilisateur, notamment les questions posées et les réponses obtenues, afin de permettre un suivi personnalisé.

### Diagramme de Séquence

- Diagramme de séquence d'authentification :

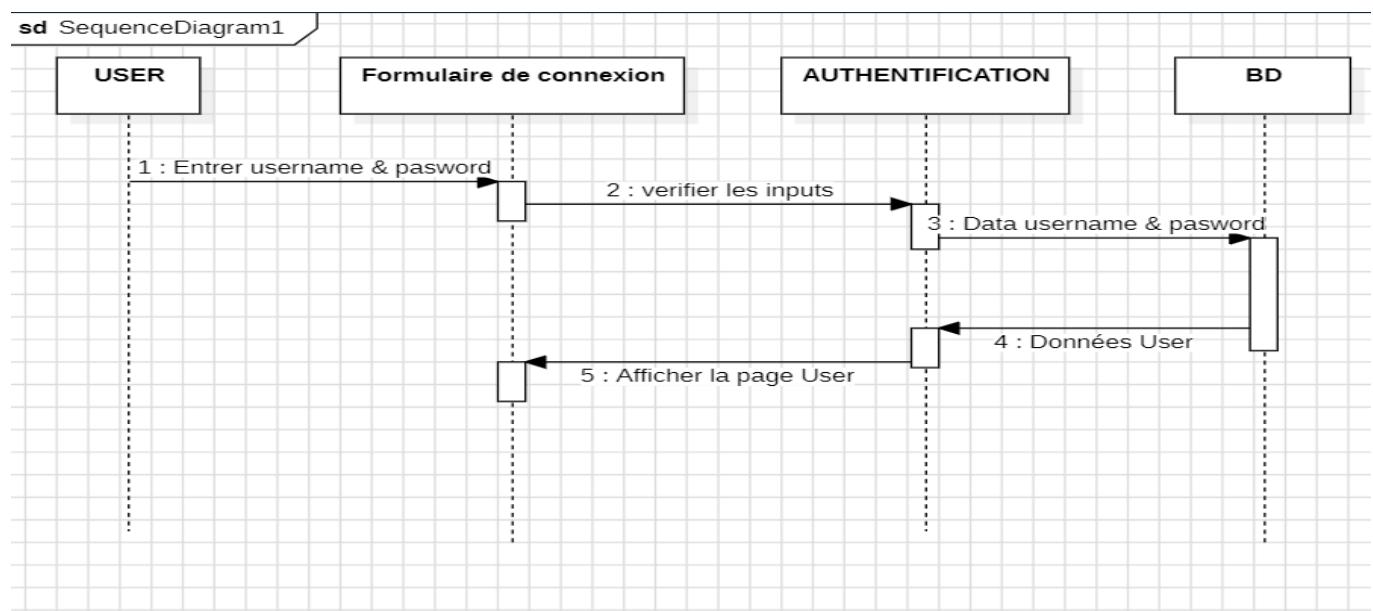


Figure 3.4 – Diagramme de séquence d'authentification

- Diagramme de séquence de poser des questions et générer rapport :

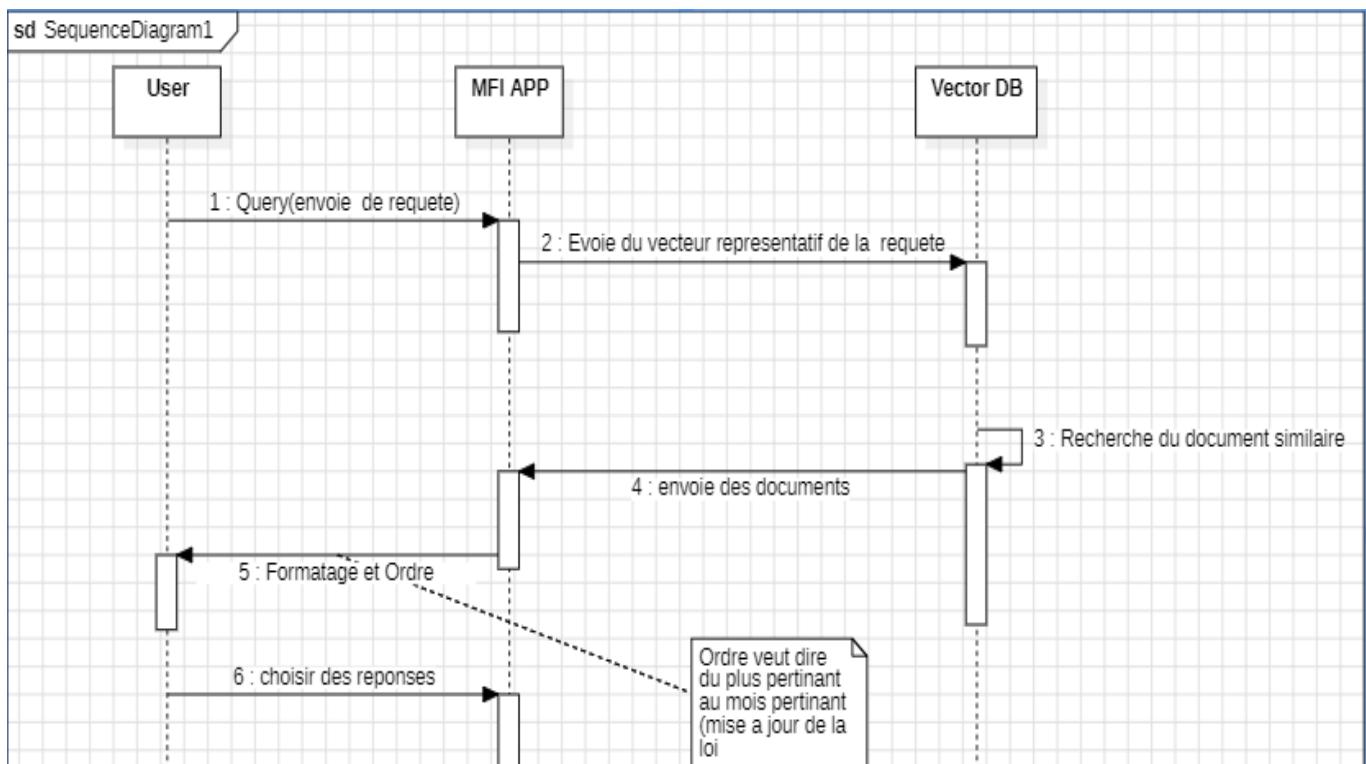


Figure 3.5 – Diagramme de séquence d'une question

Dans ce diagramme de séquence, nous modélisons l'interaction entre les principaux acteurs et systèmes pour traiter une requête utilisateur via le moteur de recherche sémantique MFI.

- L'utilisateur envoie une requête (Query) à l'application MFI. Cette requête correspond à une question posée en langage naturel concernant un sujet fiscal.
- L'application MFI prend la requête et génère un vecteur représentatif de cette requête (en utilisant des techniques de traitement sémantique) qu'elle envoie à la base de données vectorielle (Vector DB).
- La base de données vectorielle (Vector DB) compare ce vecteur à ceux déjà stockés dans sa base de données, représentant des documents ou des réponses similaires. Elle cherche les documents ou réponses les plus proches en termes de sémantique.
- La Vector DB renvoie à l'application MFI les documents ou réponses trouvés, triés par ordre de pertinence.
- MFI reçoit ces résultats et les formate en un ordre allant de la réponse la plus pertinente à la moins pertinente.
- MFI envoie ensuite ces résultats à l'utilisateur pour qu'il puisse les consulter.

L'utilisateur sélectionne la réponse qu'il juge la plus adéquate parmi les options proposées, puis l'enregistre dans l'historique pour une consultation ultérieure.

## Conclusion

Dans ce chapitre, on a mené l'analyse et la conception de notre application. On a présenté le pattern MVC et le modèle trois-tiers qui structurent notre architecture applicative en séparant les données, la logique et la présentation. On a expliqué le choix de ce langage, offrant une notation standardisée, expressive et évolutive. De plus, on a utilisé le formalisme UML pour représenter les différents aspects de notre système, tels que les cas d'utilisation, les classes et les séquences. Ce chapitre nous a ainsi permis de passer de la conception à la réalisation de notre projet.

Dans le prochain chapitre, on se concentrera sur la citation des différents outils et technologies utilisées pour la mise en œuvre de ce projet.

---

# Chapitre 4

## Technologies utilisées

---

## Introduction

Ce chapitre explorera en détail les technologies variées que l'on a employées tout au long de la réalisation de notre projet MFI. On décrira en profondeur les outils, les langages de programmation et les frameworks que l'on a choisis, en mettant en évidence leur rôle central dans la conception et le développement de notre solution. En outre, on discutera des défis et des avantages liés à l'utilisation de ces technologies, ainsi que de leur adéquation avec les objectifs spécifiques de notre projet. En somme, ce chapitre constituera un guide exhaustif des fondements technologiques de notre réalisation.

### 4.1 Framework FLASK

Flask est un framework léger et minimaliste en Python, idéal pour le développement d'applications web. Il est classé comme microframework car il est très léger. Flask a pour objectif de garder un noyau simple mais extensible.[flask]

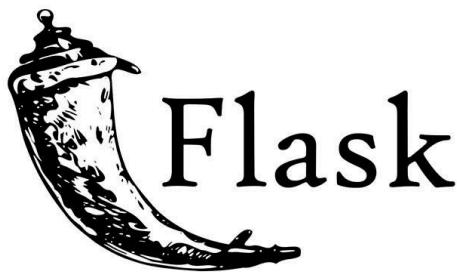


Figure 4.1 – Logo du Framework Flask

#### 4.1.1 Caractéristiques principales du Framework FLASK

Simplicité et légèreté : Flask ne force pas une structure rigide, ce qui te permet de concevoir l'architecture de ton application comme tu le souhaites. Extensibilité : Il est facile d'ajouter des extensions pour des fonctionnalités comme l'authentification, les bases de données, les API REST, etc. Flexibilité : Parfait pour des applications de petite à moyenne échelle, ce qui est adapté à ton projet de moteur de recherche sémantique.

- Simplicité et légèreté : Flask ne force pas une structure rigide, ce qui te permet de concevoir l'architecture de ton application comme tu le souhaites.
- Extensibilité : Il est facile d'ajouter des extensions pour des fonctionnalités comme l'authentification, les bases de données, les API REST, etc.
- Flexibilité : Parfait pour des applications de petite à moyenne échelle, ce qui est adapté à ton projet de moteur de recherche sémantique.

#### 4.1.2 Pourquoi utiliser Flask pour Mon projet MFI?

- API RESTful : Flask permet de créer des API facilement, ce qui est crucial pour communiquer avec la base de données vectorielle (Qdrant) et le backend qui gère les requêtes.
- Gestion des requêtes utilisateur : Flask peut gérer les requêtes des utilisateurs et envoyer des réponses formatées à partir des données traitées dans la base de données.  
Authentification : Il est facile d'intégrer des modules d'authentification avec Flask (par exemple Flask-Login) pour sécuriser l'accès au moteur de recherche.  
Rapidité et modularité : Adapté à des besoins simples tout en permettant une personnalisation, ce qui est important pour un projet comme le tien qui ne nécessite pas un framework lourd.

## 4.2 PYTHON

Python est un langage de programmation interprété, polyvalent et orienté objet, créé par Guido van Rossum et publié pour la première fois en 1991. Il est conçu pour être facile à lire et à écrire, avec une syntaxe simple et claire, ce qui le rend accessible aussi bien aux débutants qu'aux développeurs expérimentés[7]



Figure 4.2 – Logo de Python.

- Facilité d'apprentissage : Python est réputé pour sa syntaxe simple et lisible, ce qui facilite le développement rapide et la maintenance du code. Richesse des bibliothèques : Il existe un large éventail de bibliothèques et frameworks en Python, utiles pour des tâches comme le web scraping, le traitement des données, et la création d'API.
  - Traitement du langage naturel : Avec des bibliothèques comme spaCy, NLTK, ou transformers (de Hugging Face), Python est parfait pour les projets liés à l'analyse et au traitement sémantique
- Intégration avec GPT-4 : Le framework s'appuie sur un modèle d'injection de dépendances permettant une gestion efficace des services. Ces derniers facilitent l'intégration avec les API REST pour échanger des données avec un backend distant.

- Modularité et réutilisabilité : Python permet d'intégrer facilement des modèles comme l'intelligence artificielle générative et LLM via des API pour générer les questions à partir des textes de loi canadiens.

Python est utilisé dans le projet MFI (Mon Fiscaliste Intelligent) pour plusieurs aspects essentiels. Il permet d'extraire les textes de loi canadiens à partir du site CANLII en utilisant des techniques de scraping. Ensuite, Python est utilisé pour générer des questions à partir de ces textes de loi en s'appuyant sur l'intelligence artificielle générative et LLM. Ces questions, ainsi que leurs réponses correspondantes, sont stockées dans un fichier Excel, puis importées dans une base de données Qdrant. Lorsqu'un utilisateur pose une question, une API interroge la base de données pour trouver des questions similaires et en récupérer les réponses appropriées. Python joue donc un rôle clé dans le traitement des données, la génération de questions et l'interaction avec la base de données.

### 4.3 Reactjs

React.js est une bibliothèque JavaScript développée par Facebook pour construire des interfaces utilisateur interactives et dynamiques, notamment des applications web à page unique (SPA). Elle permet de créer des composants réutilisables qui gèrent leur propre état et se mettent à jour automatiquement lorsque les données changent. React utilise un concept appelé "Virtual DOM" (Document Object Model virtuel), qui améliore les performances en minimisant les manipulations du DOM réel. En rendant les mises à jour de l'interface utilisateur plus efficaces, React facilite le développement d'applications web rapides et réactives.|9|

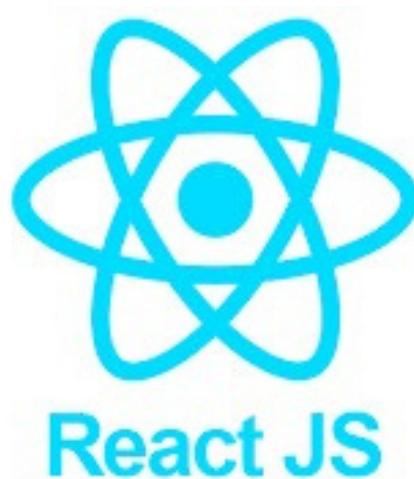


Figure 4.3 – Logo de reactjs

- Composants : React se base sur des composants, des morceaux modulaires de l'interface utilisateur qui peuvent être réutilisés et combinés pour construire des applications complexes. Chaque composant gère son propre état et ses propres props (propriétés). Virtual DOM : React utilise un Virtual DOM pour améliorer les performances. Lorsqu'un composant change, React met à jour le Virtual DOM et compare cette version à l'ancienne. Ensuite, il effectue les mises à jour nécessaires uniquement sur le DOM réel, réduisant ainsi le nombre de manipulations coûteuses.

- JSX : React utilise JSX (JavaScript XML), une syntaxe qui permet de combiner le code JavaScript avec le HTML. JSX rend le code plus lisible et plus facile à écrire en permettant l'intégration de balises HTML directement dans le JavaScript.

- Écosystème : React possède un riche écosystème de bibliothèques et d'outils pour la gestion d'état (comme Redux ou Context API), le routage (comme React Router), et la gestion des effets secondaires (comme React Query).

## 4.4 XAMPP

XAMPP permet aux développeurs de configurer un environnement de serveur local sur leur propre machine pour tester et développer des sites web ou des applications avant de les déployer sur un serveur en ligne. Il est disponible pour Windows, macOS et Linux. La configuration est généralement simple grâce à un panneau de contrôle qui permet de démarrer et d'arrêter les différents services inclus dans le paquet.



Figure 4.4 – Logo de XAMPP

Apache : Un serveur web qui permet de servir des pages web aux utilisateurs.

MySQL (ou MariaDB) : Un système de gestion de bases de données relationnelles qui permet de stocker et de gérer les données.

PHP : Un langage de script côté serveur utilisé pour développer des applications web dynamiques.

Perl : Un autre langage de script qui peut être utilisé pour divers objectifs, bien que son utilisation soit moins courante dans le développement web moderne.

## 4.5 UML

Acronyme pour « Unified Modeling Language ». On le traduit par « Langage de modélisation unifié ». La notation UML est un langage visuel constitué d'un ensemble de schémas, appelés diagrammes, qui donnent chacun une vision différente du projet à traiter.[6]



Figure 4.5 – Logo d'UML.

UML nous fournit donc des diagrammes pour représenter le logiciel à développer : son fonctionnement, sa mise en route, les actions susceptibles d'être effectuées par le logiciel.

## 4.6 SCRAPING

Les bots de scraping, ou scrapers, cherchent, pour leur part, à obtenir des données sur les sites web, indépendamment des différentes tentatives de restriction d'accès. une technique utilisée pour extraire des données à partir de sites web. Cela implique généralement d'automatiser la récupération d'informations depuis des pages web, souvent en utilisant des scripts ou des logiciels spécialisés.[8]

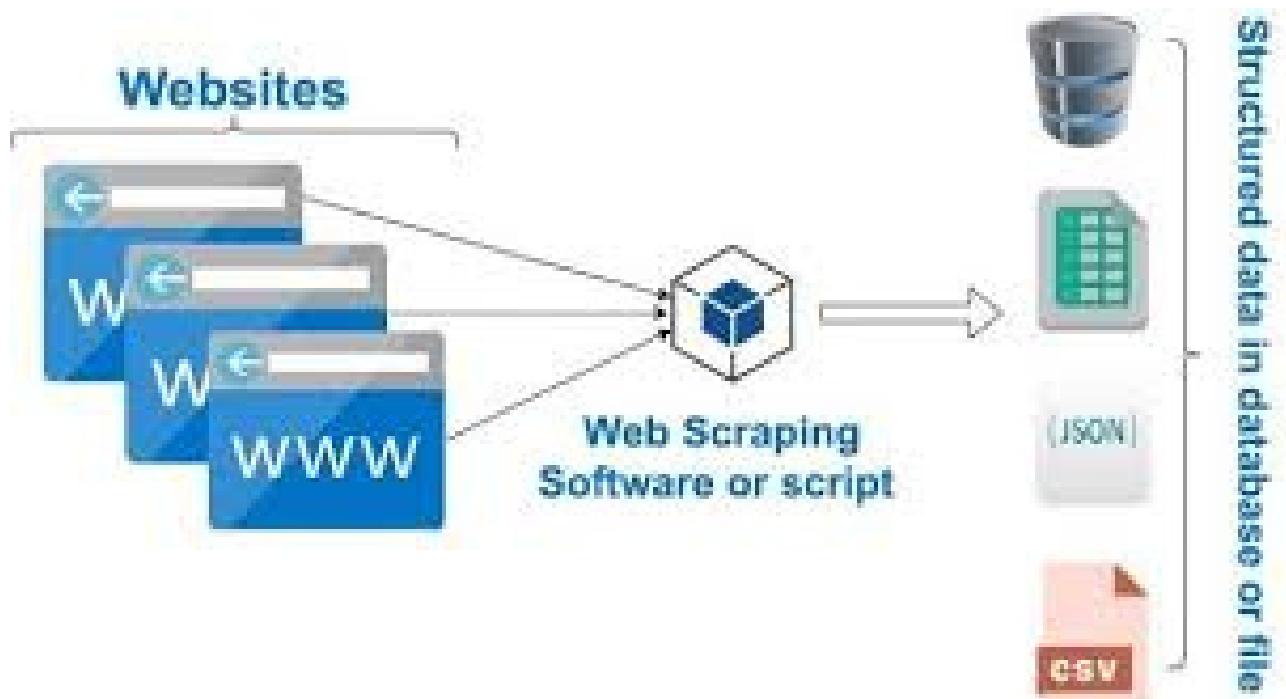


Figure 4.6 – Schéma scraping

- Extraction de Données :  
d'une ou plusieurs pages web, comme du texte, des images, ou des informations structurées. Outils et Langages : Les langages de programmation comme Python sont souvent utilisés pour le scraping, grâce à des bibliothèques comme BeautifulSoup, Scrapy, ou Selenium. Ces outils permettent de parcourir les pages web, d'extraire les données et de les structurer pour une utilisation ultérieure.
- HTML Parsing    Le scraping implique généralement l'analyse du code HTML des pages web pour identifier et extraire les données désirées. Les données extraites peuvent ensuite être stockées dans des formats comme CSV, JSON, ou des bases de données.
- Respect des Règles : Il est important de respecter les conditions d'utilisation des sites web et les règles concernant le scraping. Certains sites interdisent le scraping dans leurs termes de service, et il est essentiel de respecter les politiques de robots.txt qui indiquent quelles parties du site peuvent être explorées par des robots ou des scripts.

## 4.7 Outil de conception

StarUML :



Figure 4.7 – Logo starUML

StarUML est un logiciel de modélisation UML, qui a été « cédé comme open source » par son éditeur, à la fin de son exploitation commerciale (qui visiblement continue ...), sous une licence modifiée de GNU GPL. StarUML gère la plupart des diagrammes spécifiés dans la norme UML 2.0.

## 4.8 Qdrant

Qdrant est une base de données vectorielle conçue pour la gestion et la recherche de données non structurées, particulièrement adaptées pour les applications d'apprentissage automatique et d'intelligence artificielle.[qdrant]



Figure 4.8 – Logo de Qdrant

En tant que système de contrôle de révision distribué, il est destiné à la vitesse, intégrité de données et le support pour les développements non linéaires (milliers de branches parallèles). [qdrant]

- Recherche Vectorielle :Qdrant est optimisé pour la recherche basée sur des vecteurs, ce qui est utile pour des tâches telles que la recherche de similarité, la recherche par proximité, et la gestion des embeddings générés par des modèles de machine learning.
- Indexation Efficace : Il utilise des structures de données et des algorithmes avancés pour indexer et rechercher rapidement des vecteurs dans de grands ensembles de données. Cela permet d'effectuer des recherches rapides même avec des millions de vecteurs.
- API RESTful Qdrant offre une API RESTful pour interagir avec la base de données, ce qui facilite l'intégration avec des applications et des services web.
  - Scalabilité : Conçu pour être scalable, Qdrant peut gérer de grandes quantités de données et est souvent utilisé dans des applications nécessitant une recherche et une récupération rapide d'informations à partir de données complexes et volumineuses.

## 4.9 PHPMYADMIN

est une application Web de gestion pour les systèmes de gestion de base de données MySQL et MariaDB, réalisée principalement en PHP et distribuée sous licence GNU GPL.



Figure 4.9 – Logo de PHP MY ADMIN.

phpMyAdmin est souvent utilisé dans des environnements de développement web et est généralement intégré dans des solutions comme XAMPP, WAMP, ou MAMP pour offrir une gestion facile des bases de données locales.

## 4.10 Visual Studio Code

Grâce à Visual Studio Code, vous disposerez d'un environnement de développement complet et configurable selon vos besoins. L'éditeur prend en charge une large gamme de langages de programmation, notamment JavaScript, TypeScript et React, qui seront essentiels pour le développement de votre application. Vous pourrez ainsi bénéficier d'une assistance intelligente, comme la complétion de code, la vérification en temps réel et la navigation rapide dans votre projet.



Figure 4.10 – Logo de Visual Studio Code.

De plus, Visual Studio Code offre une intégration transparente avec les principaux outils et services utilisés dans le développement web et mobile, tels que les gestionnaires de versions (Git), les plates-formes de construction (comme webpack) et les environnements d'exécution (Node.js). Cela vous permettra de coordonner efficacement votre workflow et de bénéficier d'une expérience de développement fluide et cohérente. [5]

## Conclusion

Dans ce chapitre, on a pris le temps de vous présenter les outils et les technologies que l'on a soigneusement sélectionnés pour créer notre projet. Chacun de ces éléments a été choisi en fonction de son utilité pour atteindre notre objectif final.

Ces ressources sont diverses, allant des langages de programmation aux outils informatiques spécialisés. Chacun d'eux a un rôle spécifique dans la construction, l'amélioration et la mise en place de notre solution.

Dans le prochain chapitre, on vous montrera comment on a utilisé ces éléments pour concrétiser notre projet.

---

# Chapitre 5

## Réalisation et mise en œuvre

---

## 5.1 Introduction

Ce chapitre détaille le processus de réalisation et de mise en œuvre de l'application MFI (Mon Fiscaliste Intelligent). Après avoir défini les bases théoriques et les spécifications fonctionnelles dans les chapitres précédents, nous nous concentrerons ici sur la transformation de ces concepts en un produit opérationnel.

## 5.2 Presentation de l'application MFI :

Mon Fiscaliste Intelligent (MFI) est une application innovante conçue pour offrir une assistance juridique spécialisée dans le domaine fiscal. Elle vise à fournir des réponses précises et pertinentes aux questions fiscales des utilisateurs, en utilisant une technologie avancée pour simplifier l'accès à des informations complexes.

### 5.2.1 Page d'accueil MFI

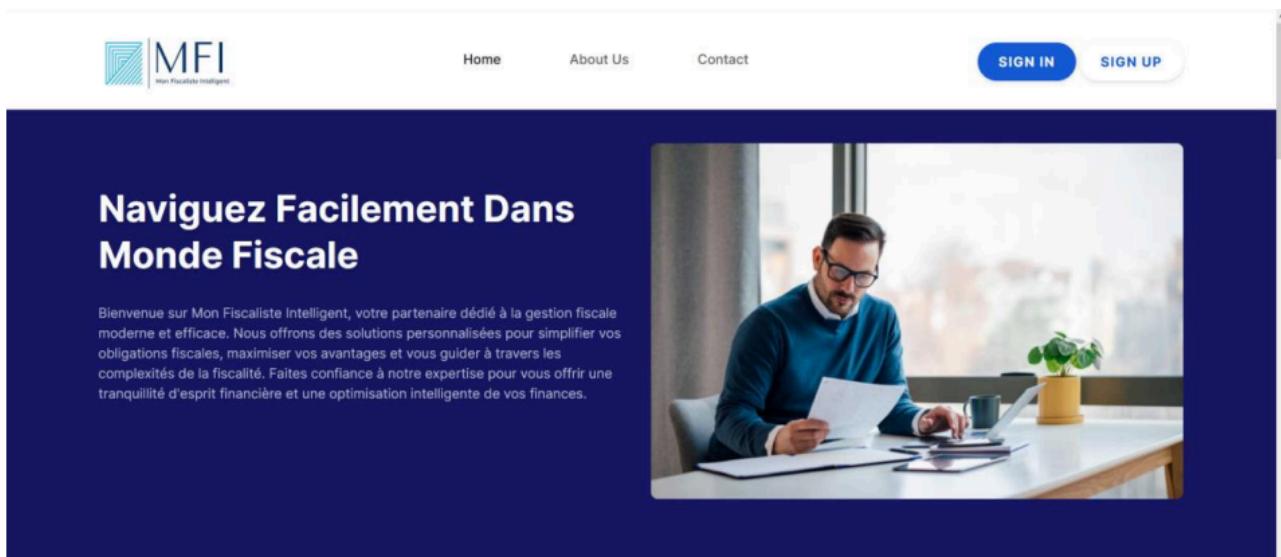


Figure 5.1 – Page d'accueil MFI.

Cette interface représente la page d'accueil de l'application MFI

## 5.2.2 Page About us MFI

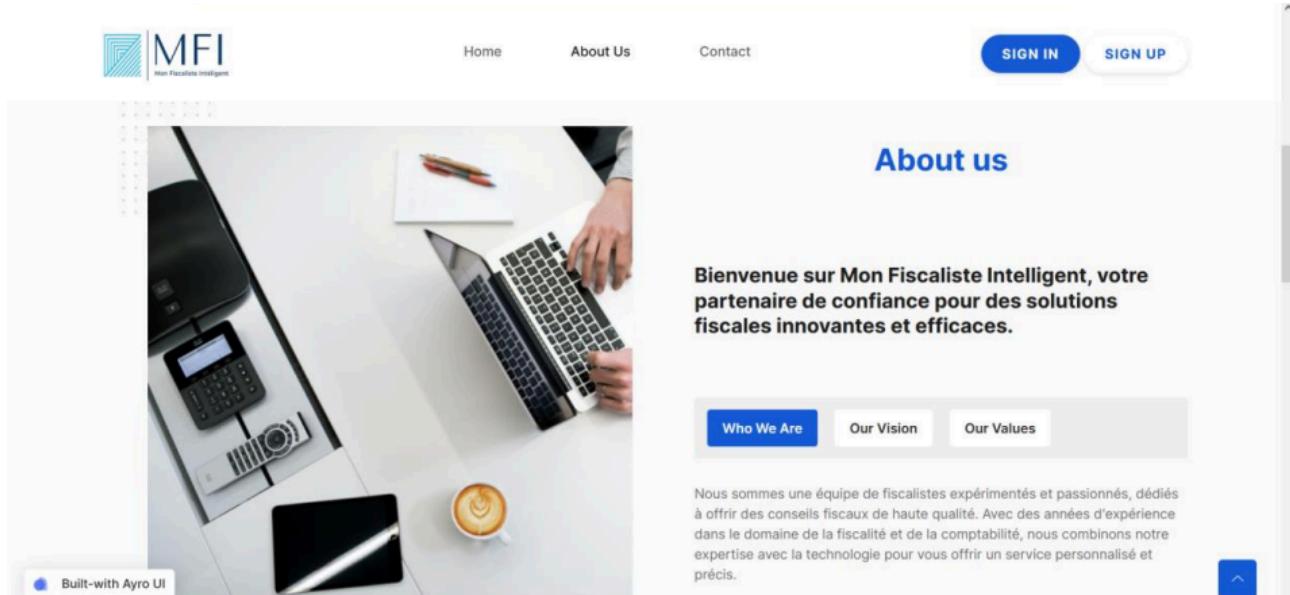


Figure 5.2 – Page About us MFI.

Cette interface représente la page ABOUT US de l'application MFI

## 5.2.3 Page de contacter MFI

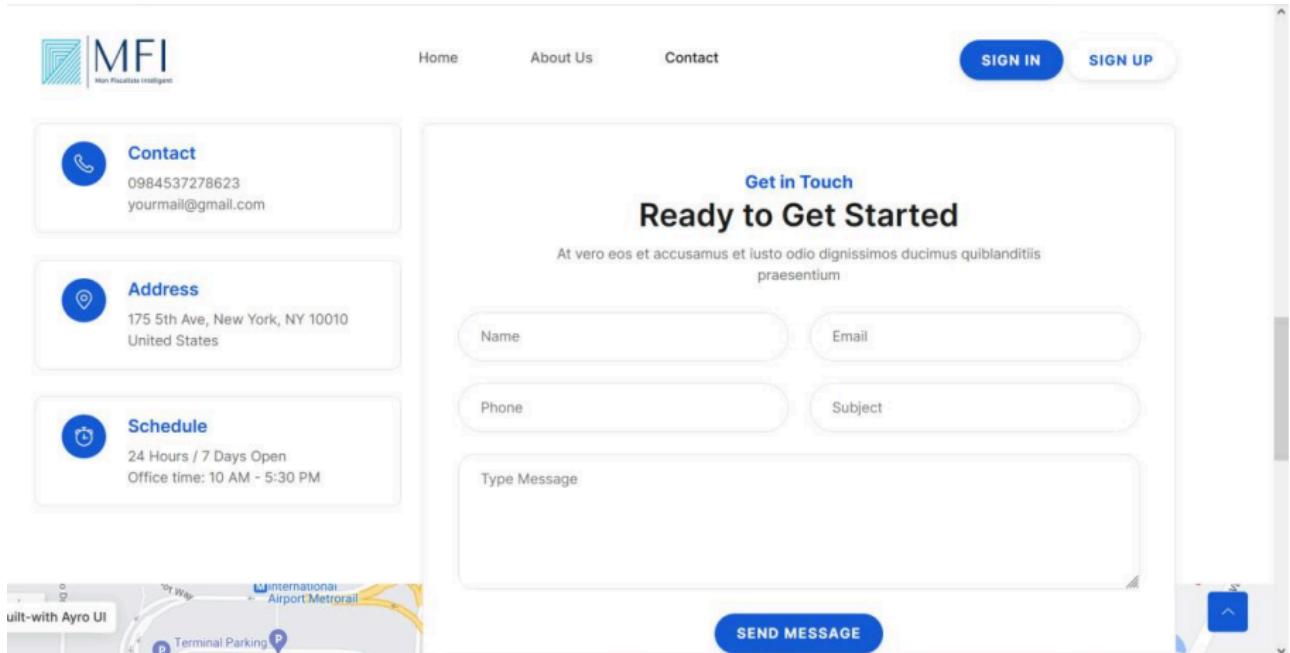


Figure 5.3 – Page de contacter MFI.

Cette interface représente la page de contact de l'application MFI

## 5.2.4 Page de localisation MFI :

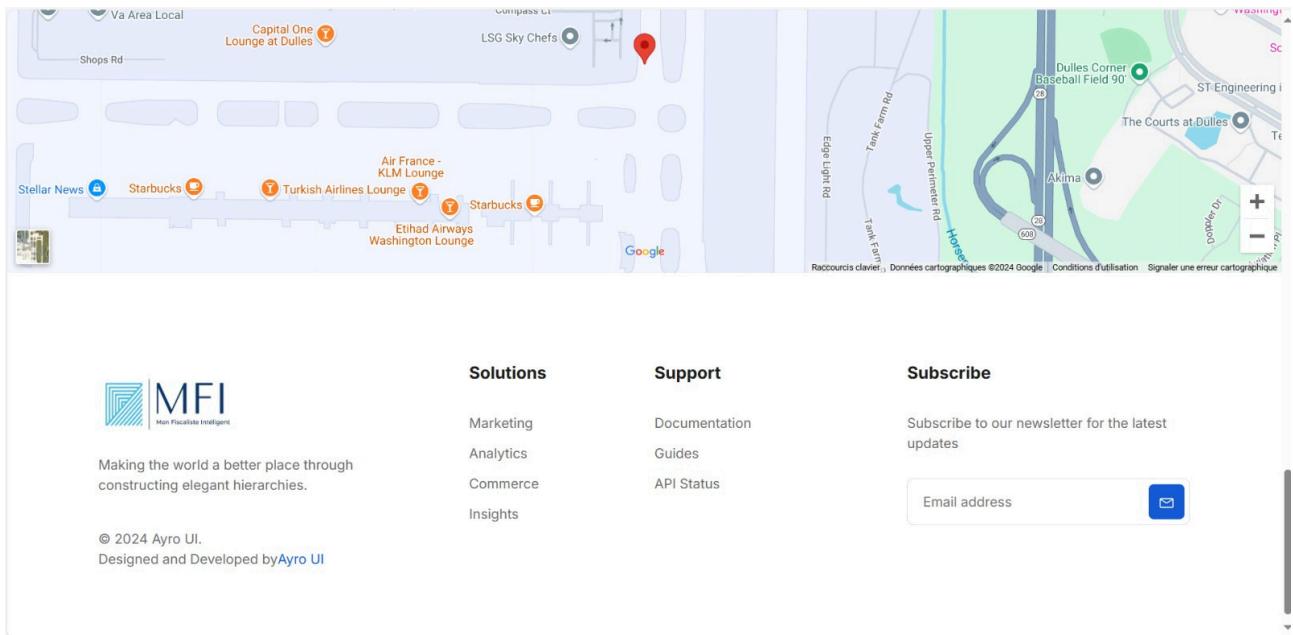


Figure 5.4 – Page de localisation MFI.

Cette interface représente la localisation

## 5.2.5 Page d'authentification :

- Interface création de compte

Figure 5.5 – Page d'authentification.

Cette interface est une interface de creation de compte dont le client rentre ses information comme le nom ,le telephone, l'adresse mail et le mot de pass.

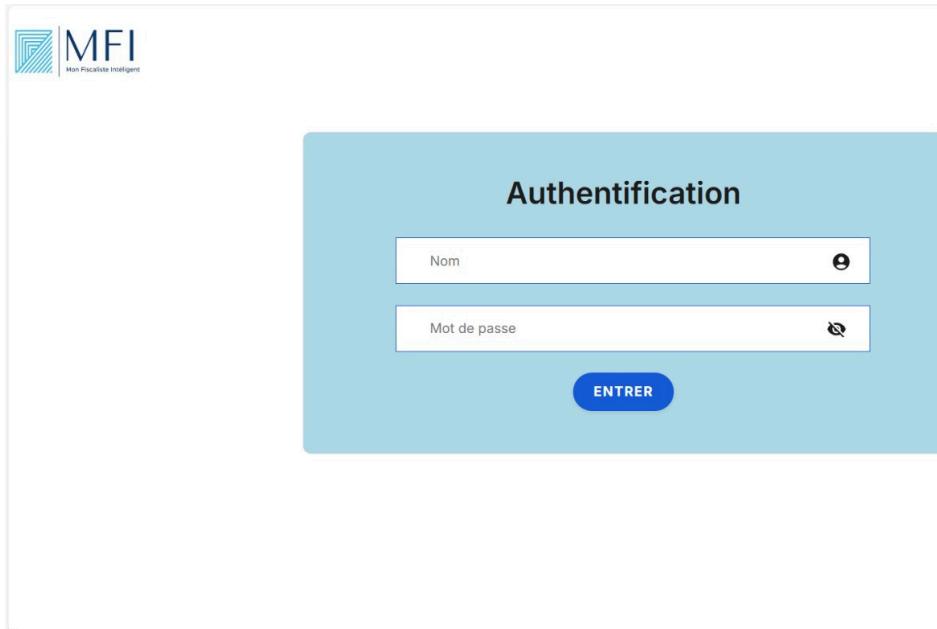
- Enregistrement de l'utilisateur

<input type="checkbox"/> Tout afficher	Nombre de lignes :	25	<input type="button" value="Filtrer les lignes: Chercher dans cette table"/>	Trier par clé :	Aucun(e)
Options supplémentaires					
	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	<b>id</b>	<b>name</b>
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	1	najah
123456	najah@gmail.com	\$2b\$12\$52Q10fVnGekeUdqDEZVaAe.RyAsPiWqy0lUzvP0W/Dp...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	2	Malak
1234556	najah@gmail.com	\$2b\$12\$/Xj_6qAfIj4zdMHL5MC.pX3FkTufu7hSGLZHUGNRF...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	3	Najah
98765443	najah@gmail.com	\$2b\$12\$P8YrTRxDnh8Sp3KUCt8G20C80xoH3c2TVVlbUfhWqEi...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	4	hajar
123456	aaaa@gmail.com	\$2b\$12\$45mD5LW5mKZPylznwfaquxmowKyigMutWP2sgNUjGO...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	5	AZERT
123456	gffr@gmail.com	\$2b\$12\$XxLdr90s0Spjl7ibr9GzMFOyvZjbJK0iay/68oLous...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	6	sdfg
123456	najah@gmail.com	\$2b\$12\$yTC9Z4y6lWrl/OCZ/e0bDOnfldS19ESbP1l6pc/2ZP7...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	7	ARTYHGF
123456	ZSDF@gmail.com	\$2b\$12\$sjRlKj4VcNGXTDlvghXhgO96uBAq2loNb5w6lv4TwhM...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	8	Najah
123456	najah@gmail.com	\$2b\$12\$mcY14AlFaGRGRHSiijZredJki4bubUnFePcz8UD4cK...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	9	fati
456789	najhd@gmail.com	\$2b\$12\$75B1t9JfsgcOmwuT1/v0Au00uj/B1BHZeYole5/...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	10	fatimazehra
676543	fati@gmail.com	\$2b\$12\$SE6N6VG33H8QXNozEw7kukTY4hbDm4hG2hBaIDbWR...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	11	Hajar
692371234	Hajar@gmail.com	\$2b\$12\$NNW.ArHR0azp1YRavzu.tT.PT1FCh37gaAp1KFMS...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	12	ZE
1234567890	A@gmail.com	\$2b\$12\$G3OVOsu2FuojhJcsJfTup8ZL2kLB86CYrWXYYRRgg...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	13	Malak
696371234	naitelefatimi@gmail.com	\$2b\$12\$1X818nR83nk0Br0NjeJxu8l5jlrUhLBVtyEWBCN5...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	14	er
660306585	nait@gmail.com	\$2b\$12\$raK3TxkB0/rHKEW18hBO3vQXKsW6TLbgBXPXLw3r...			
<input type="checkbox"/>	<input type="button" value="Éditer"/>	<input type="button" value="Copier"/>	<input type="button" value="Supprimer"/>	15	user
696371234	naitelefatimi@gmail.com	\$2b\$12\$ztm.spMaN2Je5UoomdChEuhoudVG1iC5G9GEJ8bUp...			

Figure 5.6 – Base de donnée utilisateur.

Cette image représente la Base de donnée des users et comme on peut voir l'utilisateur est affiché dans la base de donnée

- Interface de connexion



The image shows the 'Authentification' (Authentication) page of the MFI (Mon Fiscaliste Intelligent) application. At the top left is the MFI logo. The main title 'Authentification' is centered above two input fields: 'Nom' (Name) and 'Mot de passe' (Password). To the right of each input field is a small circular icon with a character symbol. Below the inputs is a blue 'ENTRER' (Enter) button.

Figure 5.7 – Page de localisation MFI.

Cette interface représente la connexion de l'utilisateur; il doit rentrer ses informations personnelles comme le nom et le mot de passe

#### 5.2.6 interface du moteur de recherche :



The image shows the search interface of the MFI application. At the top right is a user profile icon with the text 'Bonjour fati'. The MFI logo is at the top center. Below it is a search bar with the placeholder 'Rechercher...' and a blue 'Rechercher' (Search) button.

Figure 5.8 – moteur de recherche MFI.

Après le succès de l'authentification on bascule directement vers cette interface du moteur de recherche MFI où l'utilisateur peut poser des questions

### 5.2.7 Affichage de reponses :

<b>Response 1</b> 13(4) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 2</b> 69.73 <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 3</b> 54 <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>
<b>Response 4</b> 18(1) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 5</b> 147 <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 6</b> 67.6 <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>
<b>Response 7</b> 220(3.1) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 8</b> 60(i),146 <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 9</b> 110.6(6) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>
<b>Response 10</b> 15(2.4)e),15(2.4)f) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 11</b> 256(2.1) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>	<b>Response 12</b> 248(1) <b>ENREGISTRE</b> <b>SHOW RESPONSE</b>

Figure 5.9 – Page de reponses.

Cette interface represente les réponses affichées après le lancement de la recherche de la question

### 5.2.8 Affichage du texte juridique :

**Response 10**  
15(2.4)e),15(2.4)f)  
**ENREGISTRE** **SHOW RESPONSE**

**Response 4**  
18(1)  
**ENREGISTRE** **SHOW RESPONSE**

**Response 7**  
220(3.1)  
**ENREGISTRE** **SHOW RESPONSE**

**Response 10**  
15(2.4)e),15(2.4)f)  
**ENREGISTRE** **SHOW RESPONSE**

**Response 11**  
256(2.1)  
**ENREGISTRE** **SHOW RESPONSE**

**Response 12**  
248(1)  
**ENREGISTRE** **SHOW RESPONSE**

Figure 5.10 – texte juridique.

Cette image represente la reponse sous forme de texte juridique de CANLII

### 5.2.9 Enregistrement de réponse :

Response 1	Response 2	Response 3
13(4)  ENREGISTRE SHOW RESPONSE	69.73  ENREGISTRE SHOW RESPONSE	54  ENREGISTRE SHOW RESPONSE
Response 4	Response 5	Response 6
18(1)  ENREGISTRE SHOW RESPONSE	147  ENREGISTRE SHOW RESPONSE	67.6  ENREGISTRE SHOW RESPONSE
Response 7	Response 8	Response 9
220(3.1)  ENREGISTRE SHOW RESPONSE	60i),146  ENREGISTRE SHOW RESPONSE	110.6(6)  ENREGISTRE SHOW RESPONSE
Response 10	Response 11	Response 12
15(2.4)e),15(2.4)f)  ENREGISTRE SHOW RESPONSE	256(2.1)  ENREGISTRE SHOW RESPONSE	248(1)  ENREGISTRE SHOW RESPONSE

Figure 5.11 – Enregistrement reponse.

Dans cette image l'utilisateur a essayé d'enregistrer une reponse et un message d'enregistrement s'est affiché

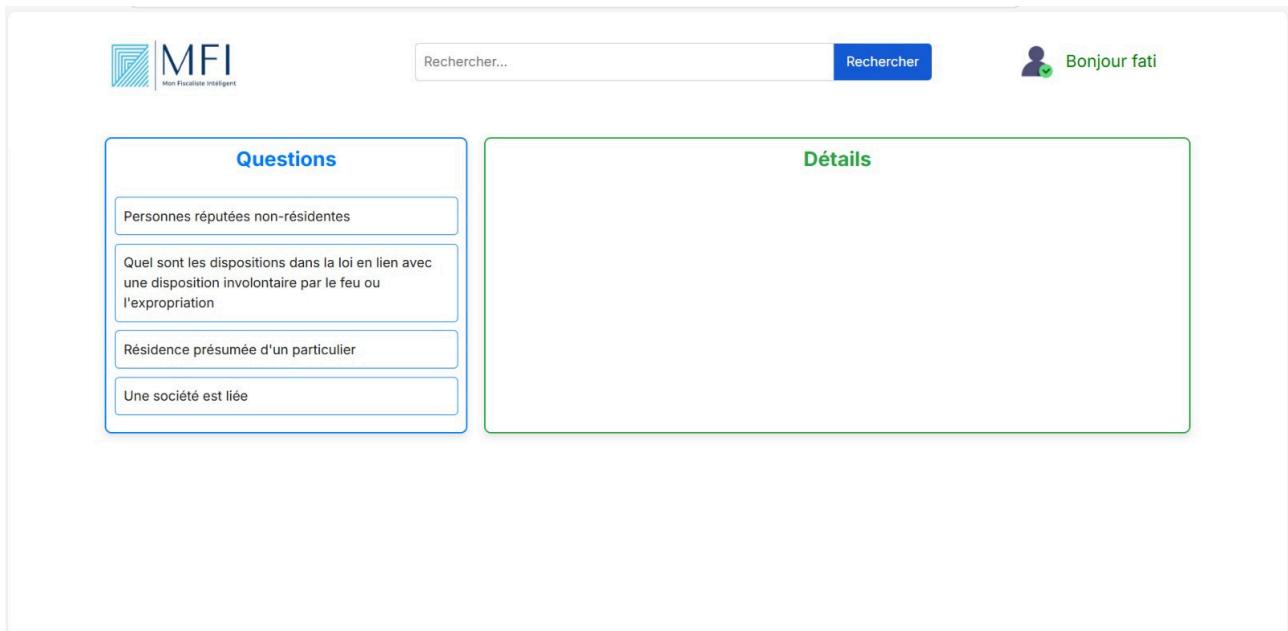
### 5.2.10 Enregistrement de la question dans la base donnée :

	id	user_id	question	response	created_at
<input type="checkbox"/>	1	9	Résidence présumée d'un particulier	45(3)	2024-09-11 14:54:27
<input type="checkbox"/>	2	9	Personnes réputées non-résidentes	250(1)	2024-09-11 16:10:06
<input type="checkbox"/>	3	9	Personnes réputées non-résidentes	250(1)	2024-09-11 16:37:49
<input type="checkbox"/>	4	9	Résidence présumée d'un particulier	250(1)	2024-09-11 21:10:19
<input type="checkbox"/>	5	9	Une société est liée	251(2)b)(ii)	2024-09-13 16:16:05
<input type="checkbox"/>	6	9	Une société est liée	256(2)	2024-09-13 16:16:12
<input type="checkbox"/>	7	9	Une société est liée	50(1)a)	2024-09-13 16:16:18
<input type="checkbox"/>	8	14	Deux sociétés sont liées	251(2)c)	2024-09-13 16:47:15
<input type="checkbox"/>	9	14	Deux sociétés sont liées	186(2),186(4)	2024-09-13 16:47:23
<input type="checkbox"/>	10	15	Période normale de nouvelle cotisation	152(4)	2024-09-15 16:54:22
<input type="checkbox"/>	12	9	Résidence présumée d'un particulier	18(1)h)	2024-09-16 03:28:34
<input type="checkbox"/>	13	9	Quel sont les dispositions dans la loi en lien ave...	13(4)	2024-09-17 17:26:57

Figure 5.12 – base de donnée question.

Cette image représente la base de donnée où les questions posées sont stockées

### 5.2.11 Historique des question posées :

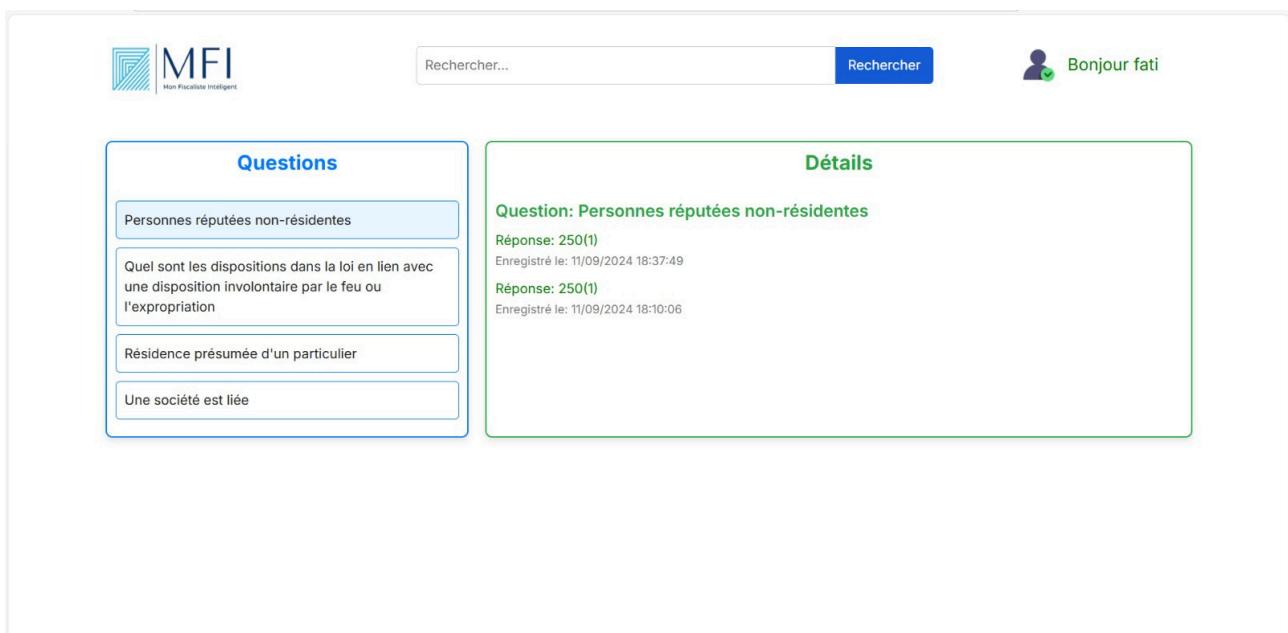


The screenshot shows the MFI (Mon Fiscaliste Intelligent) platform interface. At the top, there is a logo for IRISI and a search bar with a 'Rechercher...' placeholder and a blue 'Rechercher' button. To the right of the search bar is a user profile icon with the name 'Bonjour fati'. On the left side, there is a sidebar titled 'Questions' containing four items: 'Personnes réputées non-résidentes', 'Quel sont les dispositions dans la loi en lien avec une disposition involontaire par le feu ou l'expropriation', 'Résidence présumée d'un particulier', and 'Une société est liée'. The main content area is titled 'Détails' and is currently empty.

Figure 5.13 – Historique des questions.

Cette interface représente l'historique des questions recherchées.

### 5.2.12 Affichage des reponses enregistrer de chaque question :



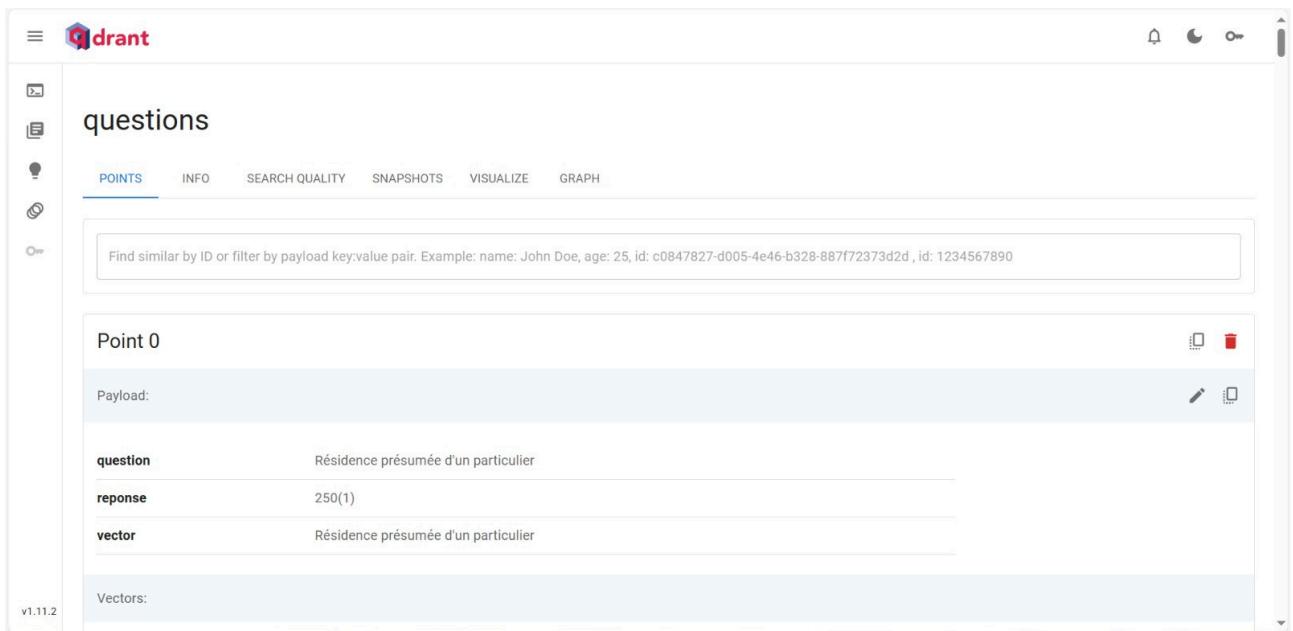
The screenshot shows the MFI platform interface. At the top, there is a logo for IRISI and a search bar with a 'Rechercher...' placeholder and a blue 'Rechercher' button. To the right of the search bar is a user profile icon with the name 'Bonjour fati'. On the left side, there is a sidebar titled 'Questions' containing four items: 'Personnes réputées non-résidentes', 'Quel sont les dispositions dans la loi en lien avec une disposition involontaire par le feu ou l'expropriation', 'Résidence présumée d'un particulier', and 'Une société est liée'. The main content area is titled 'Détails' and shows a section for 'Question: Personnes réputées non-résidentes'. It displays the response 'Réponse: 250(1)' and the date and time of registration 'Enregistré le: 11/09/2024 18:37:49'. Below this, it shows another entry for the same question with the response 'Réponse: 250(1)' and the date and time of registration 'Enregistré le: 11/09/2024 18:10:06'.

Figure 5.14 – page de details.

La zone de details représente les différents détails sur les questions comme les réponses que l'utilisateur a enregistré sur cette question et la date d'enregistrement et l'heure de

l'enregistrement

### 5.2.13 Base de donnée vectorielle



The screenshot shows the Qdrant DB interface. The top navigation bar includes icons for file, edit, and settings, followed by the Qdrant logo and the title "questions". Below the title, there are tabs: POINTS (which is selected), INFO, SEARCH QUALITY, SNAPSHOTS, VISUALIZE, and GRAPH. A search bar below the tabs contains the placeholder text: "Find similar by ID or filter by payload key:value pair. Example: name: John Doe, age: 25, id: c0847827-d005-4e46-b328-887f72373d2d , id: 1234567890". The main content area displays a single point named "Point 0". The "Payload:" section shows the following fields:

question	Résidence présumée d'un particulier
reponse	250(1)
vector	Résidence présumée d'un particulier

Below the payload section, there is a "Vectors:" section which is currently empty.

Figure 5.15 – Qdrant DB.

Cette image représente la base de donnée vectorielle où nous avons stocké les questions générées avec leurs réponses.

# Conclusion et Perspectives

Le développement du moteur de recherche sémantique pour les documents fiscaux canadiens représente une avancée significative dans la gestion et la recherche d'informations juridiques complexes. Grâce à l'utilisation de technologies modernes telles que le web scraping et la base de données vectorielle, ainsi qu'à l'aide de l'intelligence artificielle générative et des LLM (large language models), nous avons pu générer automatiquement des questions basées sur les textes fiscaux collectés. Cela a permis de créer une solution efficace qui répond directement aux besoins des fiscalistes en leur offrant un gain de temps considérable et une expérience utilisateur optimisée. La combinaison de Python et Flask pour le backend et de React.js pour le front-end a assuré une application robuste, performante et évolutive.

Les différentes étapes du projet, de la collecte des données à l'implémentation des fonctionnalités de recherche et d'authentification, ont montré l'importance d'une gestion de projet structurée selon les méthodologies Scrum et Agile. Le système d'authentification sécurisé mis en place garantit également une protection optimale des données utilisateur, un aspect essentiel dans tout système de gestion d'information.

Enfin, l'intégration future de l'apprentissage automatique permettrait d'envisager un système capable d'améliorer automatiquement la précision des réponses en fonction des interactions des utilisateurs. En développant ces perspectives, le moteur de recherche sémantique pourrait devenir un outil incontournable pour les fiscalistes et d'autres professionnels, en leur offrant un accès rapide et précis à des informations fiables, facilitant ainsi leur travail quotidien.

# Bibliographie

- [1] Gantt.<https://www.gantt.com/fr/>. [En ligne; consulté le 20.07.2023].
- [2] LeMVC-ModelViewController.<https://adventy.org/fr/mvc>. [En ligne; consulté le 19.03.2023].
- [3] ArchitectureModèleVueContrôleur.<https://www.irif.fr/~carton/Enseignement/InterfacesGraphiques/MasterInfo/Cours/Swing/mvc.html>. [En ligne; consulté le 21.06.2023].
- [4] Méthode Scrum : Le guide complet pour tout comprendre. <https://chef-de-projet.fr/methodologie-scrum/>. [En ligne; consulté le 20.05.2023].
- [5] Matúš Sulíretal. "Visual augmentation of source code editors: A systematic mapping study". In : Journal of Visual Languages Computing (2018). url : <https://www.sciencedirect.com/science/article/pii/S1045926X18301861>.
- [6] UML : un langage de modélisation de type graphique. <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/uml-un-langage-de-modelisation-pour-la-programmation-orientee-objet/>. [En ligne; consulté le 20.05.2023].
- [7] Python : un langage de programmation polyvalent. [En ligne]. Disponible : <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/python-un-langage-de-programmation-polyvalent/>.
- [8] Web Scraping : comment extraire des données du web." [En ligne]. Disponible : <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/web-scraping-comment-extraire-des-donnees-du-web/>. [Consulté le : 20 mai 2023].
- [9] React : un framework JavaScript pour construire des interfaces utilisateur." [En ligne]. Disponible : <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/react-un-framework-javascript-pour-construire-des-interfaces-utilisateur/>.