



Rapport de Projet Intelligence Artificielle et NoSQL

CertiProfile – Chatbot Intelligent de Recommandation de Certifications

**Filière : Génie Informatique – Spécialité Big Data et Intelligence
Artificielle**

Par :

Ait Sassi Maroua

Benmakhlouf Chaymae

Bouaroua Rania

Fateh Asmae

*Professeur du CM :
Hafidi Hakim*

*Professeur du TP :
Rabeh Ayoub*

2025 -2026

Table des matières

<i>Introduction</i>	5
Chapitre 1	6
.....	6
Présentation Générale du Projet	6
1. Contexte	7
2. Problématique	7
3. Objectifs du projet.....	7
4. Idée et concept	8
5. Importance du projet.....	8
Chapitre 2	9
Technologies et Outils Utilisés	9
.....	9
1. Frontend – React	10
2. Backend – FastAPI	11
3. Base de données relationnelle – MySQL	13
4. Base de données NoSQL – Neo4j	14
5. Recherche sémantique – Sentence-Transformers.....	15
6. Intelligence Artificielle – Modèle de langage (LLM)	15
7. Outils de développement	17
Chapitre 3	19
Base de Données NoSQL (Neo4j)	19
1. Présentation de Neo4j	20
1.1. Justification du choix de Neo4j	20
1.2. Modélisation des données	20
1.3. Exploitation de Neo4j dans le chatbot.....	21
1.4. Avantages de l'utilisation de Neo4j.....	22
Chapitre 4	23
.....	23
Artificielle Intelligence du Chatbot	23
1. Rôle de l'Intelligence Artificielle	24
2. Modèle de langage (LLM).....	24

3.	Approche RAG (Retrieval-Augmented Generation)	25
4.	Recherche sémantique avec Sentence-Transformer.....	25
5.	Reranking des résultats	26
6.	Génération de recommandations personnalisées	26
7.	Modèle de recommandation basé sur RAG et graphe Neo4j.....	27
	Chapitre 5 Fonctionnalités de l'application	29
1.	Authentification des utilisateurs	30
2.	Interface du chatbot.....	30
3.	Gestion des sessions de discussion.....	31
4.	Recommandation de certifications	32
5.	Filtrage selon le profil de l'utilisateur	33
6.	Historique des conversations	34
7.	Suppression des conversations	34
8.	Analyse du CV (PDF)	35
9.	Affichage des scores de correspondance	35
	Chapitre 6	37
	Architecture globale du système.....	37
1.	Vue d'ensemble de l'architecture	38
2.	Composants principaux	38
3.	Flux de données	38
4.	Avantages de l'architecture	38
	Chapitre 7.....	40
	Difficultés rencontrées et perspectives d'amélioration	40
1.	Difficultés rencontrées.....	41
2.	Perspectives d'amélioration	41
	Conclusion Générale.....	43

Table des figures

FIGURE 1: REACT – FRONTEND WEB FRAMEWORK	10
FIGURE 2: INTERFACE UTILISATEUR POUR LA RECOMMANDATION DE CERTIFICATIONS DANS CERTIPROFILE	10
FIGURE 3: FASTAPI – BACKEND REST API	11
FIGURE 4: ARCHITECTURE DU BACKEND FASTAPI ET DU MOTEUR DE RECOMMANDATION CERTIPROFILE	12
FIGURE 5: MYSQL – RELATIONAL DATABASE FOR USER MANAGEMENT	13
FIGURE 6: TABLE DES UTILISATEURS DANS LA BASE DE DONNÉES MYSQL (PHPMYADMIN)	13
FIGURE 7: NEO4J – BASE DE DONNÉES NoSQL ORIENTÉE GRAPHE	14
FIGURE 8: MODÈLE DE DONNÉES DES CERTIFICATIONS DANS NEO4J.....	14
FIGURE 9: RECHERCHE SÉMANTIQUE AVEC SENTENCE-TRANSFORMERS.....	15
FIGURE 10: GROQ LLM (LLAMA-3.1-8B-INSTANT).....	15
FIGURE 11: EXEMPLE DE DIALOGUE ENTRE L’UTILISATEUR ET LE CHATBOT	16
FIGURE 12: RECOMMANDATION DE CERTIFICATIONS À PARTIR D’UN CV (PDF)	17
FIGURE 13: ENVIRONNEMENT DE DÉVELOPPEMENT DU PROJET	18
FIGURE 14: PIPELINE DE RECOMMANDATION DU CHATBOT CERTIPROFILE	28
FIGURE 15: PAGE D’AUTHENTIFICATION DES UTILISATEURS DU CHATBOT CERTIPROFILE	30
FIGURE 16: INTERFACE DE DISCUSSION DU CHATBOT.....	31
FIGURE 17: GESTION DES CONVERSATIONS DE L’UTILISATEUR	32
FIGURE 18: EXEMPLE DE RECOMMANDATION DE CERTIFICATIONS.....	33
FIGURE 19: FILTRAGE DES CERTIFICATIONS SELON LE PROFIL DE L’UTILISATEUR.....	33
FIGURE 20: HISTORIQUE DES DISCUSSIONS	34
FIGURE 21: SUPPRESSION D’UNE SESSION DE DISCUSSION	34
FIGURE 22: ANALYSE DU CV ET RECOMMANDATIONS BASÉES SUR LE PDF	35
FIGURE 23: SCORES DE CORRESPONDANCE DES CERTIFICATIONS.....	36

Introduction

Dans un contexte de transformation numérique accélérée, les compétences technologiques sont devenues un élément essentiel pour l'insertion professionnelle et l'évolution de carrière. Les domaines du Cloud Computing, de la Data Science et de l'Intelligence Artificielle occupent aujourd'hui une place centrale dans les entreprises, ce qui a entraîné une forte demande de profils qualifiés dans ces secteurs. Pour répondre à cette demande, de nombreuses organisations proposent des certifications professionnelles permettant de valider les compétences techniques et d'améliorer l'employabilité.

Cependant, face au grand nombre de certifications disponibles, les étudiants et les professionnels rencontrent souvent des difficultés à identifier celles qui correspondent réellement à leur niveau, à leur domaine de spécialisation et à leurs objectifs de carrière. Les informations sont généralement dispersées sur plusieurs plateformes, peu structurées et parfois difficiles à comparer, ce qui rend le processus de décision long et complexe.

C'est dans ce contexte que s'inscrit le projet **CertiProfile**, un chatbot intelligent de recommandation de certifications. Ce projet a pour objectif d'accompagner les utilisateurs dans le choix de certifications adaptées à leur profil, dans les domaines du Cloud Computing, de la Data Science et de l'Intelligence Artificielle. Le système repose sur la combinaison de deux technologies majeures :

d'une part, l'**Intelligence Artificielle**, qui permet une interaction naturelle avec l'utilisateur, et d'autre part, une **base de données NoSQL orientée graphe (Neo4j)**, qui permet de modéliser les relations entre les certifications, les compétences, les niveaux et les domaines.

Grâce à une approche basée sur la **recherche sémantique et le raisonnement sur graphe**, CertiProfile est capable d'analyser le profil de l'utilisateur, d'identifier ses besoins et de proposer des recommandations personnalisées et pertinentes. Cette combinaison entre données structurées et intelligence générative permet d'offrir une expérience proche de celle d'un conseiller en orientation, tout en automatisant le processus de recherche et de sélection des certifications.

Le projet CertiProfile vise ainsi à fournir une solution innovante et intelligente pour faciliter l'orientation professionnelle et aider les utilisateurs à construire un parcours de formation cohérent, adapté à leurs objectifs et aux exigences du marché du travail.

Chapitre 1

Présentation Générale du Projet

Ce chapitre présente le contexte général du projet CertiProfile, en mettant en évidence la problématique liée au choix des certifications professionnelles. Il décrit également les objectifs du projet, son idée principale ainsi que le concept global du chatbot. Cette première partie permet de comprendre le besoin auquel répond l'application et la solution proposée, avant d'aborder les aspects techniques dans les chapitres suivants.

1. Contexte

Le développement rapide des technologies numériques a profondément transformé le marché du travail. Les entreprises recherchent aujourd’hui des profils disposant de compétences solides dans des domaines tels que le **Cloud Computing**, la **Data Science** et l’**Intelligence Artificielle**. Dans ce contexte, les certifications professionnelles sont devenues un moyen important de valider les compétences techniques et d’améliorer l’employabilité des candidats.

Cependant, face à la diversité des certifications proposées par de nombreux organismes (AWS, Microsoft, Google, IBM, etc.), les étudiants et les professionnels rencontrent souvent des difficultés à identifier celles qui sont réellement adaptées à leur profil, à leur niveau et à leurs objectifs de carrière.

2. Problématique

La recherche d’informations sur les certifications est généralement longue, dispersée et peu structurée. Les utilisateurs doivent consulter plusieurs plateformes, comparer des contenus techniques et interpréter des descriptions parfois complexes. Ce processus rend le choix difficile et peut conduire à des décisions inadaptées, telles que le choix d’une certification trop avancée, trop coûteuse ou non alignée avec les objectifs professionnels.

Il existe donc un besoin réel pour un **outil intelligent** capable de centraliser les informations, de comprendre le profil de l’utilisateur et de proposer des recommandations personnalisées de manière simple et efficace.

3. Objectifs du projet

Le projet **CertiProfile** a pour objectif principal de développer un chatbot intelligent capable de :

- Analyser le profil de l'utilisateur (niveau, domaine, objectifs),
- Comprendre ses besoins et ses préférences,
- Exploiter une base de données structurée de certifications,
- Recommander des certifications adaptées dans les domaines du Cloud, de la Data Science et de l'Intelligence Artificielle.

L'objectif global est de faciliter l'orientation professionnelle et d'aider les utilisateurs à construire un parcours de formation cohérent, efficace et adapté au marché du travail.

4. Idée et concept

L'idée du projet est de concevoir un **assistant virtuel intelligent** qui joue le rôle d'un conseiller en orientation. Le chatbot interagit avec l'utilisateur en langage naturel et s'appuie sur une base de données **NoSQL orientée graphe (Neo4j)** pour exploiter les relations entre certifications, compétences, niveaux et domaines.

Grâce à cette combinaison entre **Intelligence Artificielle** et **base de données graphe**, le système est capable de fournir des recommandations plus pertinentes et mieux adaptées au profil de chaque utilisateur.

5. Importance du projet

Ce projet présente une forte valeur ajoutée car il :

- Facilite l'accès à l'information sur les certifications,
- Aide les utilisateurs à mieux planifier leur parcours professionnel,
- Permet d'identifier des certifications adaptées à leur budget, à leur carrière et à leurs objectifs personnels,
- Exploite des technologies modernes telles que l'Intelligence Artificielle et les bases de données NoSQL,
- Répond à un besoin réel dans le domaine de la formation et de l'orientation professionnelle.

Grâce à ces fonctionnalités, **CertiProfile** constitue un outil intelligent d'aide à la décision pour toute personne souhaitant investir efficacement dans sa formation.

Chapitre 2

Technologies et Outils Utilisés

*Ce chapitre présente les différentes technologies et outils utilisés pour la conception et le développement du chatbot **CertiProfile**. Ces technologies ont été sélectionnées afin d'assurer un système performant, sécurisé et capable de fournir des recommandations intelligentes. Elles permettent également de garantir une bonne interaction entre l'utilisateur, l'Intelligence Artificielle et les bases de données.*

1. Frontend – React

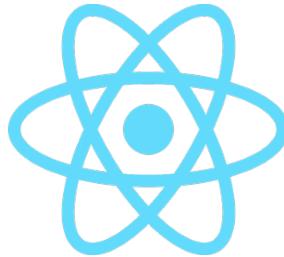


Figure 1:React – Frontend
Web Framework

Le frontend de l'application est développé avec **React**, une bibliothèque JavaScript moderne utilisée pour la création d'interfaces utilisateur interactives. React permet de concevoir une interface fluide, rapide et dynamique pour le chatbot. Grâce à cette technologie, l'utilisateur peut facilement interagir avec le système, poser des questions, consulter les recommandations et gérer ses sessions de discussion.

L'interface comprend plusieurs composants, notamment la page de connexion, la zone de discussion, la barre latérale des conversations et les boutons de gestion des sessions. React permet également de mettre à jour l'interface en temps réel sans recharger la page, ce qui améliore considérablement l'expérience utilisateur.

A screenshot of the CertiProfile AI web application. The interface has a dark theme with a pink header. On the left, there's a sidebar with 'Sessions' (Sessions), '+ Nouveau chat' (New Chat), 'Déconnexion' (Logout), and 'Nouvelle conversation' (New Conversation). The main area features a large AI icon and the text 'Bienvenue sur CertiProfile AI'. Below it, there are three buttons: 'Cloud (AWS, Azure, GCP)', 'Data Engineering & Analytics', and 'IA & Machine Learning'. To the right, there's a sidebar titled 'Filtres de recherche' (Search filters) with dropdowns for 'DOMAINÉ' (All domains), 'NIVEAU' (All levels), and 'BUDGET MAXIMUM' (Ex: 300). Below that are sections for 'Analyse des compétences' (Skill analysis) and 'Recommandations' (Recommendations). At the bottom, there's a text input field 'Écris ta question ici (ex: "Quelle certif Azure pour débuter ?")', and buttons for 'Importer CV' (Import CV), 'Cache' (Cache), and 'Envoyer' (Send).

Figure 2:Interface utilisateur pour la recommandation de certifications dans CertiProfile

2. Backend – FastAPI



Figure 3: FastAPI – Backend REST API

Le backend du projet repose sur **FastAPI**, un framework Python rapide et moderne destiné au développement d'API REST. FastAPI est utilisé pour gérer la logique métier du projet, la communication entre le frontend et les bases de données, ainsi que l'intégration de l'Intelligence Artificielle.

Il permet de recevoir les requêtes envoyées par l'utilisateur, de les analyser, d'interroger les bases de données et de retourner les réponses générées par le chatbot. FastAPI joue ainsi un rôle central dans l'orchestration du système et dans le fonctionnement du moteur de recommandation.

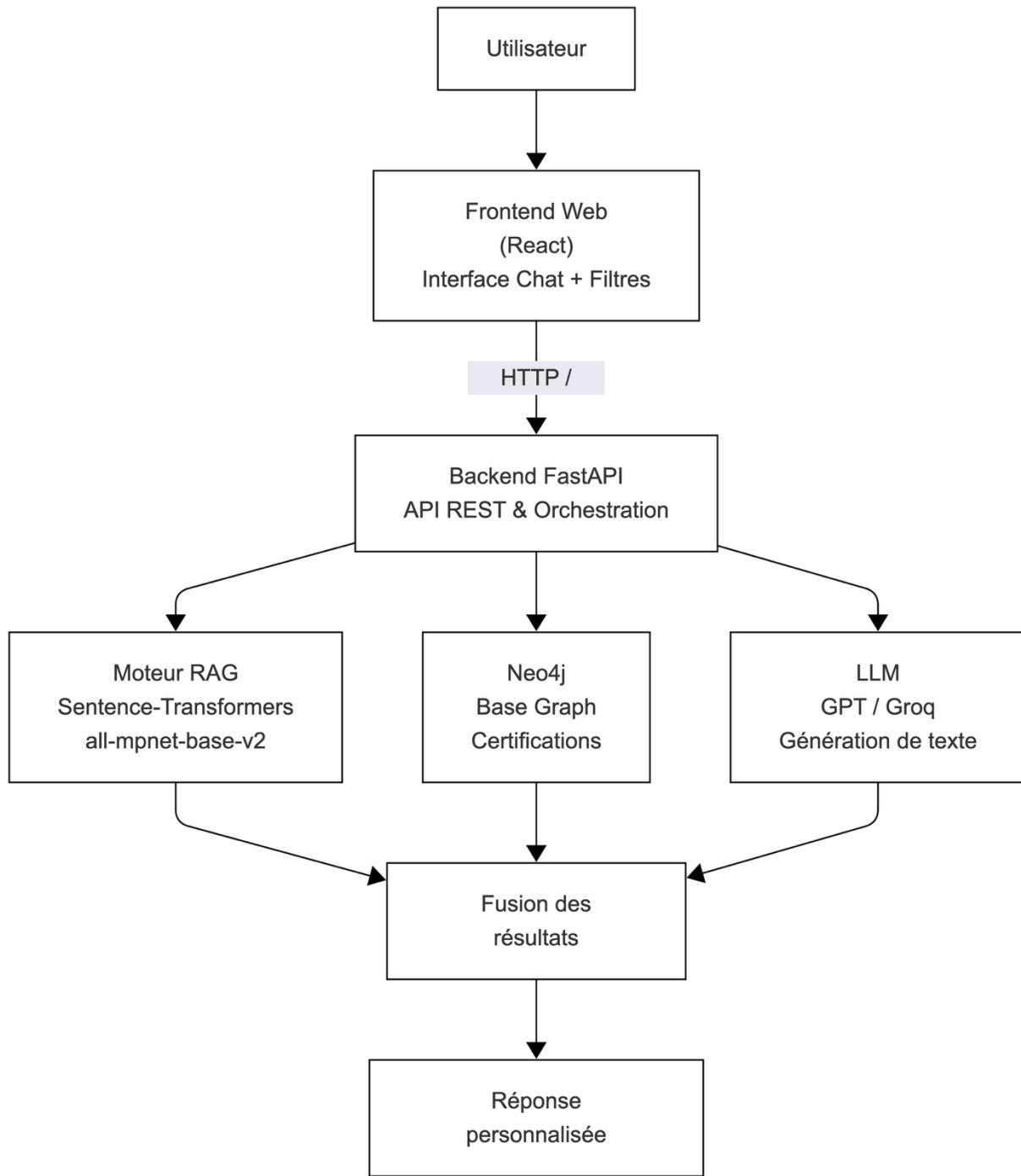


Figure 4: Architecture du backend FastAPI et du moteur de recommandation CertiProfile

3. Base de données relationnelle – MySQL



Figure 5: MySQL – Relational Database for User Management

Une base de données **MySQL** est utilisée pour stocker les informations relatives aux utilisateurs, notamment l'email, le mot de passe chiffré et les dates de création de compte. Cette base permet d'assurer une gestion sécurisée des comptes, l'authentification des utilisateurs et l'isolation des données entre les différents profils.

L'utilisation de MySQL garantit également la fiabilité, la cohérence et l'intégrité des données utilisateurs.

	id	email	created_at	password_hash
<input type="checkbox"/>	Éditer	Copier	Supprimer	3 chaymae@test.com 2020-01-10 15:34:08 \$2b\$12\$KZZEjgQV99AtJINXc3Uy.VxHxOJ4wBVVaihXjnFD...
<input type="checkbox"/>	Éditer	Copier	Supprimer	4 maroua@test.com 2020-01-10 16:14:20 \$2b\$12\$Ob0i2VgjEscePghfbf22ulpk2ywANRSSZLZh90SE...
<input type="checkbox"/>	Éditer	Copier	Supprimer	5 testuser@example.com 2020-01-11 00:24:59 \$2b\$12\$NvqulIZQraXQVUrz4Yhh.EIXWoEyrWMr6bFOW8IiOR...
<input type="checkbox"/>	Éditer	Copier	Supprimer	6 rania@test.com 2020-01-11 13:03:00 \$2b\$12\$Gw3Q4TSBlq2MHcj.cNNUYzz3dOs1ZRbhWoB8oVn...

Figure 6: Table des utilisateurs dans la base de données MySQL (phpMyAdmin)

4. Base de données NoSQL – Neo4j



Figure 7: Neo4j – Base de données NoSQL orientée graphe

Le système de recommandation repose principalement sur **Neo4j**, une base de données NoSQL orientée graphe. Neo4j permet de représenter les certifications sous forme de nœuds reliés entre eux par des relations telles que le domaine, le niveau, les compétences requises et le fournisseur.

Ce modèle en graphe est particulièrement adapté pour ce type d'application, car il facilite l'exploration des relations entre les données et permet de réaliser des requêtes complexes afin de proposer des recommandations personnalisées basées sur le profil de l'utilisateur.

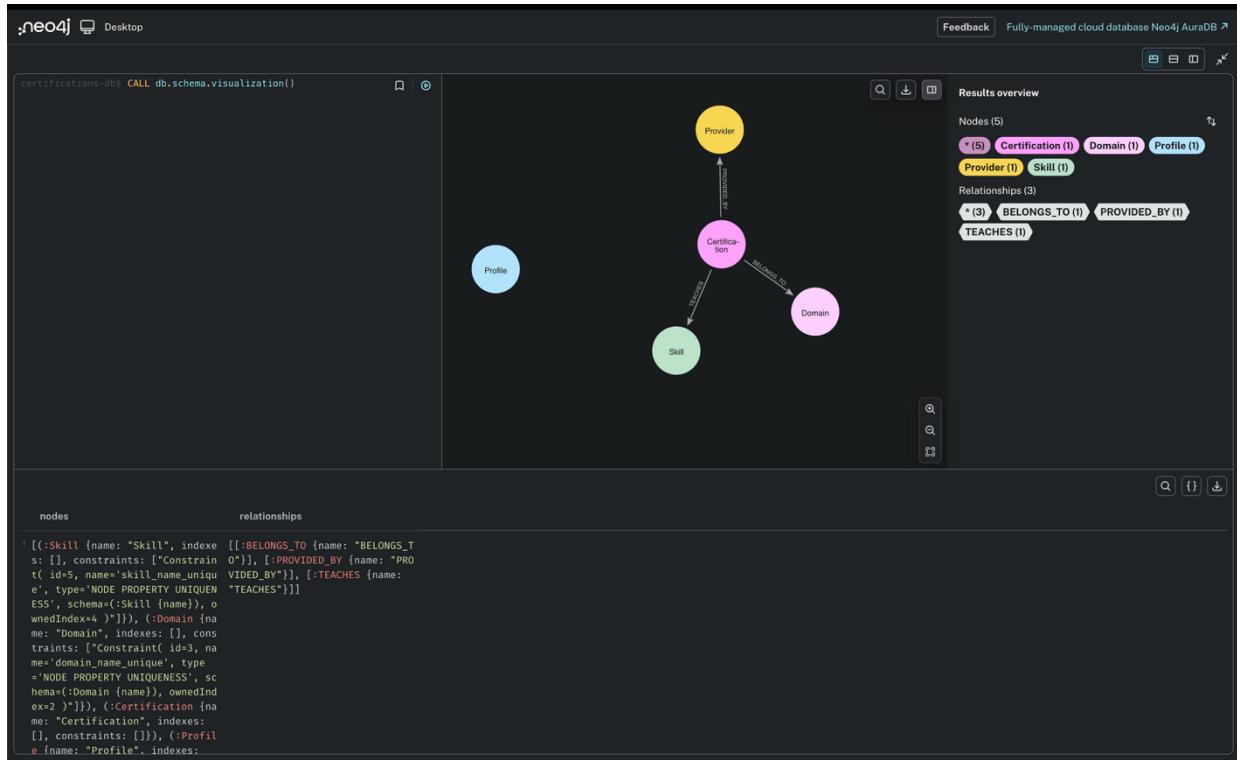


Figure 8: Modèle de données des certifications dans Neo4j

5. Recherche sémantique – Sentence-Transformers

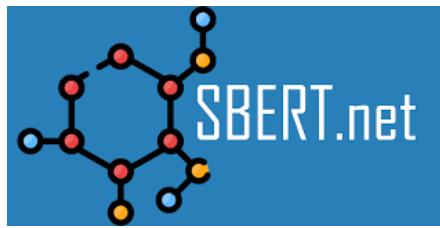


Figure 9: Recherche sémantique avec Sentence-Transformers

Le système de recommandation de CertiProfile repose sur un moteur de **recherche sémantique** basé sur la bibliothèque **Sentence-Transformers**, en particulier le modèle *all-mpnet-base-v2*. Ce modèle permet de transformer les textes, qu'il s'agisse des questions de l'utilisateur ou des descriptions des certifications, en vecteurs numériques appelés *embeddings*.

Ces représentations vectorielles permettent de mesurer la similarité sémantique entre la requête de l'utilisateur et les certifications stockées dans Neo4j. Grâce à cette approche, le système peut identifier les certifications les plus pertinentes même lorsque les mots utilisés sont différents. Sentence-Transformers constitue ainsi un composant essentiel du moteur **RAG (Retrieval-Augmented Generation)** utilisé dans CertiProfile.

6. Intelligence Artificielle – Modèle de langage (LLM)



Figure 10: Groq LLM (llama-3.1-8b-instant)

Le chatbot CertiProfile utilise un modèle de langage de type **LLM (Large Language Model)** fourni via l'API **Groq**, basé sur le modèle **llama-3.1-8b-instant**.

Ce modèle est utilisé pour comprendre les questions posées par l'utilisateur, analyser ses intentions et générer des réponses en langage naturel.

Le LLM est chargé de :

- interpréter la demande de l'utilisateur (domaine, niveau, budget, objectifs),
- reformuler les résultats fournis par le moteur de recommandation,
- présenter les certifications de manière claire, structurée et compréhensible.

Le modèle de langage ne fonctionne pas de manière isolée : il est utilisé conjointement avec les données extraites depuis la base de données **Neo4j** et les résultats du moteur **RAG**.

Cette combinaison entre **données structurées (Neo4j)** et **génération de texte (LLM Groq)** permet de produire des réponses précises, cohérentes et basées sur des informations réelles, offrant ainsi une interaction intelligente et personnalisée similaire à celle d'un conseiller en orientation professionnelle.

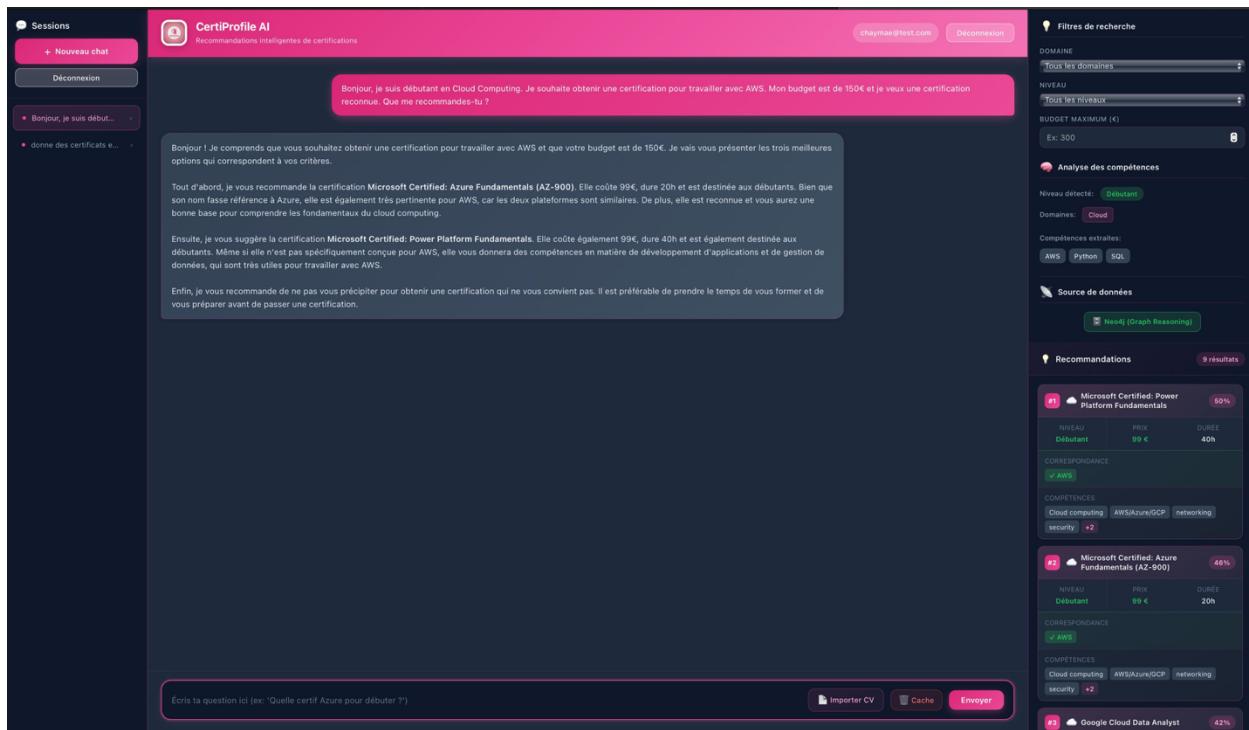


Figure 11: Exemple de dialogue entre l'utilisateur et le chatbot

The screenshot shows the CertiProfile AI platform's user interface. On the left, there's a sidebar with 'Sessions' (Sessions), '+ Nouveau chat' (New chat), and 'Déconnexion' (Logout). The main area displays a CV titled 'CV_Data_Engineer_Marie_Laurent.pdf' (366 mots). Below the CV, a text input field says 'Écris ta question ici (ex: "Quelle certif Azure pour débuter ?")'. To the right, there are three recommended certifications:

- 1. Certified Kubernetes Administrator (CKA)**: Prix : 395€, Durée : 4 mois, Niveau : Intermédiaire. Correspondance: ✓ Kubernetes, ✓ Docker, ✓ Kubernetes.
- 2. AWS Certified Solutions Architect – Associate**: Prix : 150€, Durée : 80h, Niveau : Intermédiaire. Correspondance: ✓ Apache Spark, ✓ AWS, ✓ GCP, ✓ Cosmos DB.
- 3. Microsoft Certified: Azure Data Fundamentals**: Prix : 99€, Durée : 20h, Niveau : Débutant. Correspondance: ✓ AWS, ✓ GCP.

On the far right, there's a sidebar for 'Filtres de recherche' (Search filters) with dropdowns for 'Domaine' (Tous les domaines), 'Niveau' (Tous les niveaux), and 'Budget Maximum' (Ex: 300). It also includes sections for 'Analyse des compétences' (Skill analysis), 'Source de données' (Data source), and 'Recommandations' (Recommendations) with 10 résultats.

Figure 12: Recommandation de certifications à partir d'un CV (PDF)

7. Outils de développement

Plusieurs outils ont été utilisés pour le développement du projet :

- **Python** pour le développement du backend et la gestion de l'IA
- **JavaScript** pour le développement du frontend
- **GitHub** pour la gestion et la collaboration sur le code source
- **Postman** pour le test des API
- **Visual Studio Code** comme environnement de développement

Ces outils ont permis d'assurer un développement structuré, collaboratif et efficace.

The screenshot shows the Visual Studio Code interface with the following details:

- EXPLORER**: Shows the project structure under "CHATBOT-CERTIFICATIONS". Key files include:
 - backend/app/routers/chat_rag.py
 - backend/app/services/recommend.py
 - backend/app/utils.py
 - backend/auth.py
 - backend/config.py
 - backend/database.py
 - backend/main.py
 - backend/models.py
 - backend/schemas.py
 - backend/test_reco.py
 - docs/README.md
 - frontend/dist
 - frontend/node_modules
 - frontend/public/vite.svg
 - frontend/src/assets/react.svg
 - frontend/src/components/ChatBox.css
 - frontend/src/components/ChatBox.jsx
- EDITOR**: Displays a Python file named "chat_rag.py" with code related to user preferences and conversation memory.
- TERMINAL**: Shows the command line output for running the backend and frontend development environments.
- STATUS BAR**: Shows the current file path as "chatbot-certifications", the file name "chat_rag.py", and the line number "18".

Figure 13: Environnement de développement du projet.

Cette figure illustre l'environnement utilisé pour le développement de CertiProfile, comprenant Visual Studio Code pour l'édition du code source, le terminal pour l'exécution du backend FastAPI et du frontend React (Vite), ainsi que la structure complète du projet.

Chapitre 3

Base de Données NoSQL (Neo4j)

*Ce chapitre présente la base de données **NoSQL Neo4j** utilisée dans le projet **CertiProfile**. Il décrit les raisons du choix de cette technologie, la manière dont les données ont été modélisées, ainsi que son rôle central dans le fonctionnement du système de recommandation. Neo4j permet d'organiser les certifications et leurs relations de façon efficace afin de fournir des recommandations personnalisées et pertinentes aux utilisateurs.*

1. Présentation de Neo4j

Neo4j est une base de données **NoSQL orientée graphe**. Contrairement aux bases de données relationnelles classiques qui utilisent des tables, Neo4j stocke les données sous forme de **nœuds**, de **relations** et de **propriétés**. Cette structure permet de représenter naturellement des données fortement connectées, ce qui est particulièrement adapté aux systèmes de recommandation.

Dans le projet CertiProfile, Neo4j est utilisée pour stocker et organiser les certifications, les domaines, les niveaux et les relations entre ces éléments.

1.1. Justification du choix de Neo4j

Le choix de Neo4j repose sur plusieurs raisons :

- Les certifications sont fortement liées entre elles par des relations (domaine, niveau, prérequis).
- La structure en graphe permet d'effectuer des recherches rapides et complexes.
- Neo4j facilite la recommandation de parcours de formation cohérents.

Grâce à Neo4j, il est possible de parcourir le graphe pour identifier les certifications les plus adaptées au profil de l'utilisateur.

1.2. Modélisation des données

Dans la base Neo4j, les principales entités sont :

- **Certification**
- **Domaine** (Cloud, Data Science, Intelligence Artificielle)
- **Niveau** (Débutant, Intermédiaire, Avancé)
- **Fournisseur** (AWS, Azure, Google, etc.)
- **Skill** (Compétence)
- **Provider** (AWS, Google, IBM, etc).

Le niveau (Débutant, Intermédiaire, Avancé) est stocké comme une **propriété** du nœud `Certification` et non comme une entité séparée.

Ces entités sont reliées par des relations explicites dans le graphe Neo4j :

- `BELONGS_TO` : relie une certification à son domaine (Cloud, Data, IA).
- `PROVIDED_BY` : relie une certification à son fournisseur (AWS, Google, IBM, etc.)
- `TEACHES` : relie une certification aux compétences qu'elle permet d'acquérir

Cette modélisation permet de représenter de manière claire et efficace les liens entre les certifications, leurs domaines, leurs fournisseurs et les compétences enseignées.

1.3. Exploitation de Neo4j dans le chatbot

Lorsque l'utilisateur pose une question, le backend FastAPI interroge Neo4j afin de récupérer les certifications correspondant à son profil. La base de données est structurée de façon à garantir une classification unique des certifications par **domaine** (Cloud, Data Science, Intelligence Artificielle) et par **niveau** (Débutant, Intermédiaire, Avancé), stocké directement dans chaque certification.

Cette structuration permet au moteur RAG de filtrer avec précision les certifications en fonction du profil de l'utilisateur, notamment :

- son niveau
- son domaine d'intérêt
- son budget
- et ses compétences.

Le chatbot s'appuie ainsi sur un moteur de recommandation basé sur le graphe Neo4j et un pipeline **RAG (Retrieval-Augmented Generation)** pour proposer des recommandations personnalisées, pertinentes et adaptées aux besoins de chaque utilisateur. Les résultats récupérés depuis Neo4j sont ensuite utilisés par le modèle d'Intelligence Artificielle afin de générer une réponse claire, structurée et compréhensible.

1.4. Avantages de l'utilisation de Neo4j

L'utilisation de Neo4j permet :

- une meilleure organisation des données,
- des requêtes rapides,
- des recommandations plus précises,
- et une meilleure évolutivité du système.

Grâce à cette base NoSQL, CertiProfile peut facilement être enrichi avec de nouvelles certifications et de nouveaux domaines.

Chapitre 4

Artificielle Intelligence du Chatbot

Ce chapitre présente le rôle de l’Intelligence Artificielle dans le fonctionnement du chatbot CertiProfile. Il décrit comment les modèles de langage, la recherche sémantique et l’approche **RAG** sont combinés afin d’analyser les demandes des utilisateurs et de générer des recommandations de certifications pertinentes, personnalisées et basées sur des données réelles.

1. Rôle de l'Intelligence Artificielle

L'Intelligence Artificielle constitue le cœur du chatbot **CertiProfile**. Elle permet d'analyser les messages de l'utilisateur, d'identifier ses besoins, d'interpréter ses intentions et de produire des réponses adaptées en langage naturel.

Grâce à l'IA, le chatbot peut simuler le comportement d'un conseiller en orientation professionnelle en proposant des recommandations personnalisées, cohérentes et pertinentes selon le profil de l'utilisateur.

2. Modèle de langage (LLM)

CertiProfile utilise un modèle de langage de type **LLM (Large Language Model)** fourni via l'API **Groq**, basé sur le modèle :

llama-3.1-8b-instant

Ce modèle est chargé de :

- comprendre les questions de l'utilisateur,
- interpréter son profil (niveau, domaine, objectifs),
- expliquer les certifications,
- formuler les recommandations sous forme de réponses claires et naturelles.

Le LLM ne décide pas seul des certifications : il s'appuie sur les données récupérées depuis Neo4j et le moteur RAG pour générer des réponses fiables et basées sur des informations réelles.

3. Approche RAG (Retrieval-Augmented Generation)

Le projet CertiProfile repose sur l'approche **RAG (Retrieval-Augmented Generation)**, qui combine :

- une base de données (Neo4j),
- un moteur de recherche sémantique,
- et un modèle de langage (LLM).

Le fonctionnement global est le suivant :

1. L'utilisateur pose une question.
2. Le backend FastAPI analyse la demande.
3. Les certifications sont récupérées depuis Neo4j selon le domaine, le niveau et le budget.
4. Les textes sont transformés en vecteurs sémantiques.
5. Les certifications les plus pertinentes sont sélectionnées.
6. Le LLM génère une réponse personnalisée à partir de ces données.

Cette approche garantit des réponses à la fois intelligentes, fiables et basées sur des données structurées.

4. Recherche sémantique avec Sentence-Transformer

Pour la partie RAG, le système utilise le modèle :

sentence-transformers/all-mnpt-base-v2

pour convertir :

- les questions de l'utilisateur,
- et les descriptions des certifications en **vecteurs numériques** appelés **embeddings**.

Ces embeddings permettent de mesurer la similarité sémantique entre la demande de l'utilisateur et les certifications stockées dans Neo4j.

Grâce à cette approche, le chatbot peut retrouver des certifications pertinentes même lorsque les mots utilisés sont différents.

5. Reranking des résultats

Après la recherche sémantique, CertiProfile applique un modèle de **reranking** basé sur:

ms-marco-MiniLM-L-6-v2

Ce modèle permet de réordonner les certifications selon leur pertinence réelle par rapport à la question de l'utilisateur.

Le score final d'une certification est calculé à partir :

- de la similarité sémantique,
- de la correspondance des compétences,
- et du score de reranking.

Cela permet d'obtenir un classement précis des meilleures certifications.

6. Génération de recommandations personnalisées

La génération finale des recommandations dans CertiProfile repose sur la combinaison de plusieurs composants intelligents.

Tout d'abord, **Neo4j** fournit les données structurées contenant les certifications, leurs domaines, leurs niveaux, leurs prix et leurs compétences associées. Cette base de données permet d'extraire uniquement les certifications correspondant aux critères de l'utilisateur.

Ensuite, **Sentence-Transformers** est utilisé pour effectuer une recherche sémantique. Il transforme la demande de l'utilisateur et les descriptions des certifications en vecteurs numériques, ce qui permet de mesurer leur similarité et d'identifier les certifications les plus proches du besoin réel de l'utilisateur.

Un module de **reranking** est ensuite appliqué afin de réordonner les certifications obtenues en fonction de leur pertinence réelle. Ce mécanisme permet d'améliorer la précision du classement final en privilégiant les certifications les plus adaptées.

Enfin, le **LLM Groq** intervient pour générer la réponse finale. Il prend les certifications sélectionnées et les transforme en une réponse en langage naturel, claire, explicative et adaptée au profil de l'utilisateur.

Le système filtre les certifications selon plusieurs critères : le **domaine**, le **niveau**, le **budget** et les **compétences** de l'utilisateur. Cette combinaison garantit que les recommandations fournies sont à la fois pertinentes, cohérentes et personnalisées.

7. Modèle de recommandation basé sur RAG et graphe Neo4j

Le chatbot CertiProfile repose sur une architecture avancée combinant une base de données orientée graphe **Neo4j** et un moteur de génération de texte basé sur l'approche **RAG (Retrieval-Augmented Generation)**.

Chaque certification est décrite dans Neo4j à l'aide de plusieurs attributs tels que le domaine, le niveau, le prix, la durée, les compétences, le fournisseur et l'objectif professionnel.

Cette structuration permet au système de représenter efficacement les relations entre les certifications et de filtrer les résultats en fonction du profil de l'utilisateur, notamment son domaine d'intérêt, son niveau de compétences, son budget et ses objectifs.

Le pipeline de recommandation fonctionne de la manière suivante : l'utilisateur saisit une demande via l'interface web, qui est transmise au backend FastAPI.

Celui-ci interroge la base Neo4j afin de récupérer les certifications correspondant aux critères du profil. Les descriptions des certifications et la question de l'utilisateur sont ensuite transformées en embeddings par le moteur de recherche sémantique.

Ces représentations vectorielles permettent d'identifier et de classer les certifications les plus pertinentes.

Enfin, les résultats sélectionnés sont transmis au modèle de langage (LLM), qui génère une réponse claire, structurée et personnalisée destinée à l'utilisateur.

Pipeline de recommandation du chatbot CertiProfile

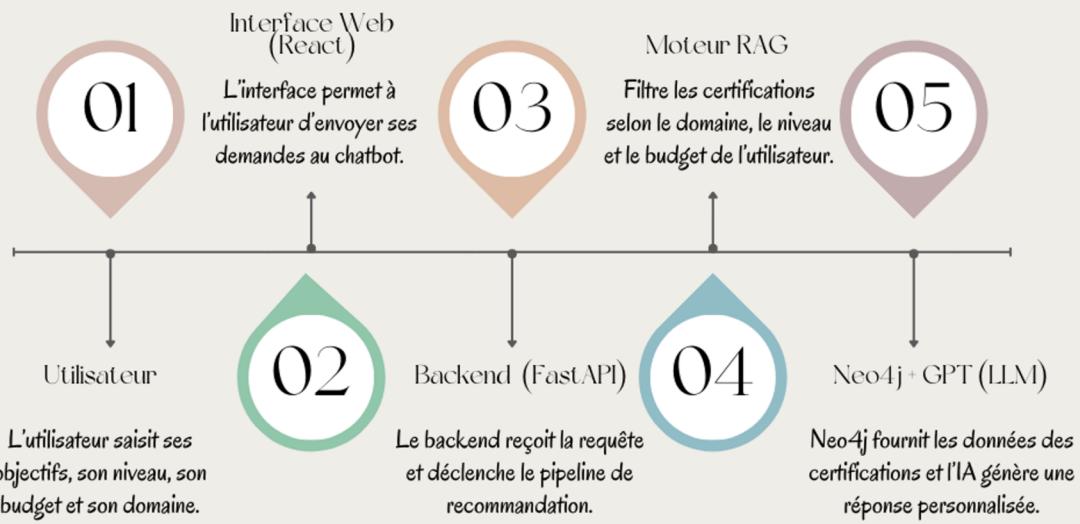


Figure 14: Pipeline de recommandation du chatbot CertiProfile

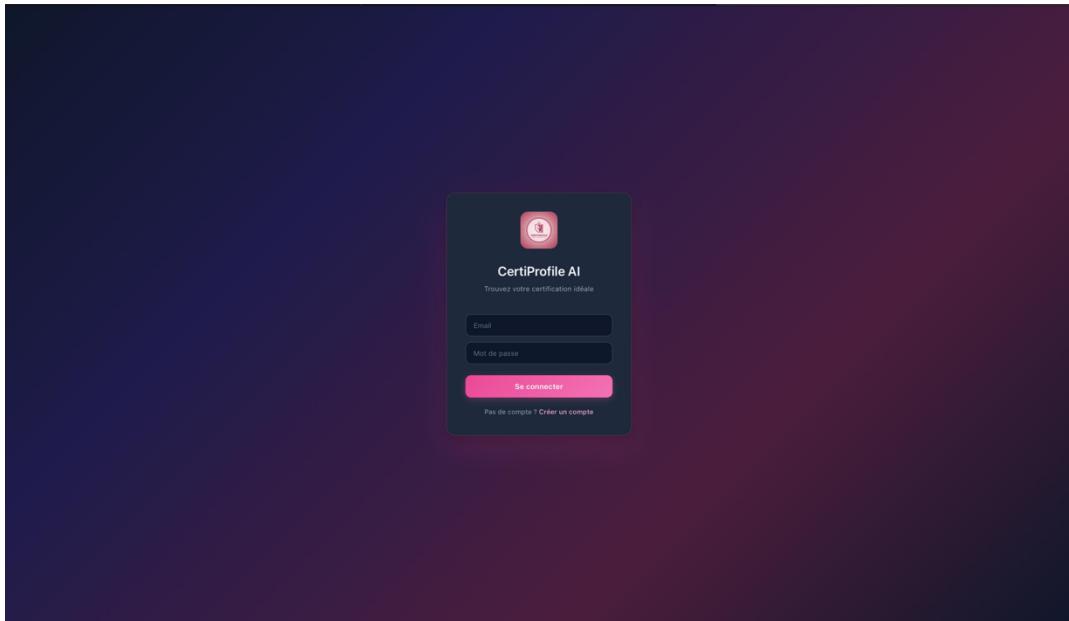
Chapitre 5

Fonctionnalités de l'application

Ce chapitre présente les principales fonctionnalités offertes par l'application **CertiProfile**. Il décrit les différentes interactions possibles entre l'utilisateur et le chatbot, depuis l'authentification jusqu'à la consultation des recommandations. Il met également en évidence les services mis à disposition pour faciliter la recherche, le filtrage et la recommandation de certifications adaptées au profil et aux objectifs de chaque utilisateur.

1. Authentification des utilisateurs

L'application propose un système d'authentification permettant aux utilisateurs de créer un compte et de se connecter de manière sécurisée. Chaque utilisateur dispose d'un identifiant unique, ce qui permet de garantir la confidentialité des données et d'assurer une gestion personnalisée des sessions de discussion.



*Figure 15: Page d'authentification des utilisateurs du chatbot
CertiProfile*

2. Interface du chatbot

CertiProfile offre une interface conviviale permettant à l'utilisateur d'interagir facilement avec le chatbot. L'utilisateur peut saisir ses questions, consulter les réponses générées par l'IA et recevoir des recommandations de certifications en temps réel.

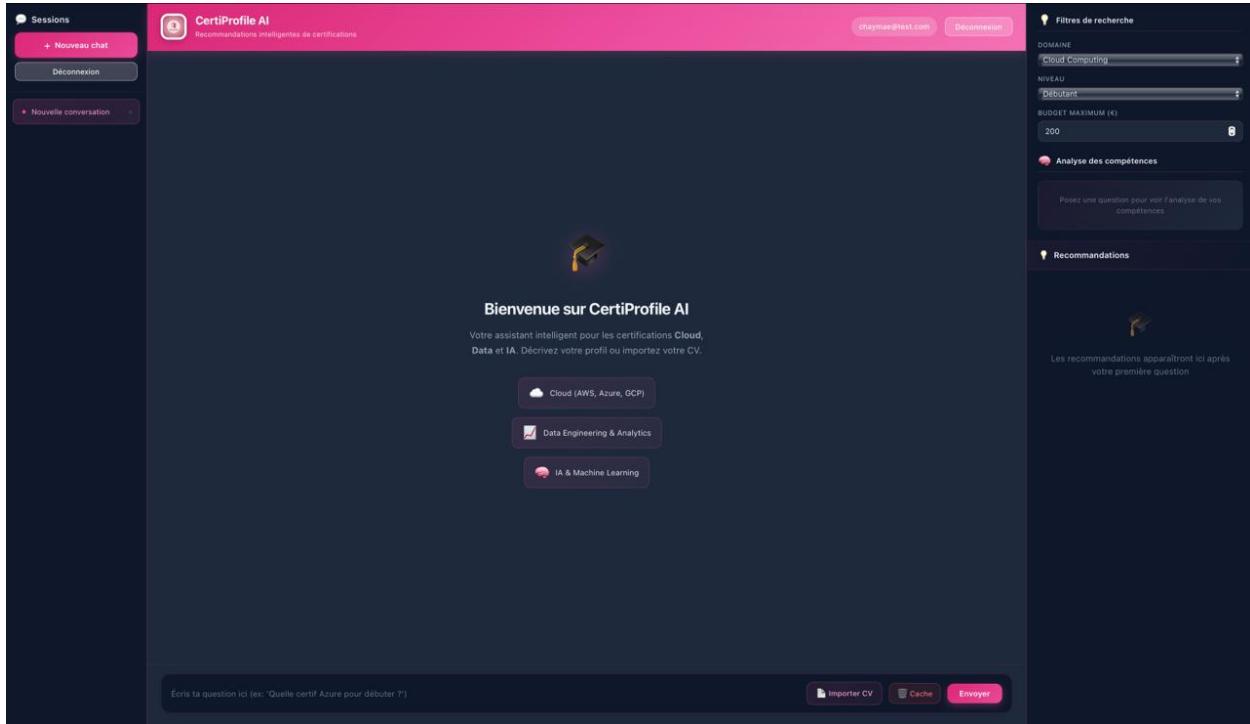


Figure 16: Interface de discussion du chatbot

3. Gestion des sessions de discussion

L’application permet la création de plusieurs sessions de discussion. Chaque session correspond à une conversation distincte, ce qui permet à l’utilisateur de reprendre une recherche antérieure ou de lancer une nouvelle demande sans perdre ses précédentes interactions.

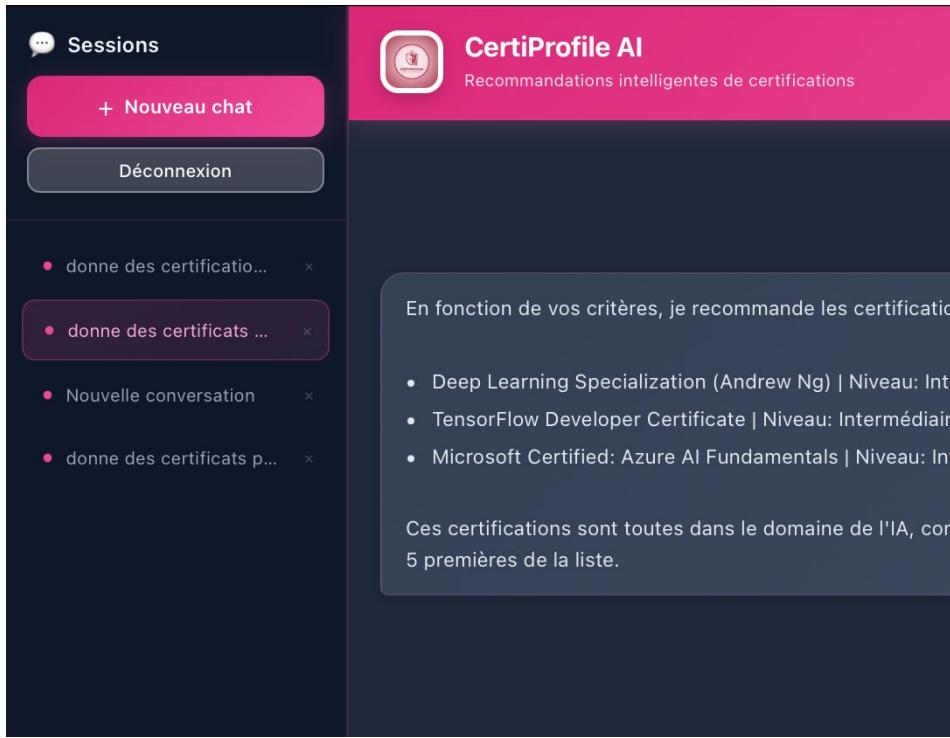


Figure 17: Gestion des conversations de l'utilisateur

L'utilisateur peut gérer plusieurs sessions de discussion afin de conserver l'historique de ses recherches et organiser ses demandes de certifications.

4. Recommandation de certifications

La fonctionnalité principale de CertiProfile est la recommandation intelligente de certifications. En fonction du profil de l'utilisateur (domaine, niveau, budget), le chatbot propose une liste de certifications pertinentes accompagnées d'explications claires.

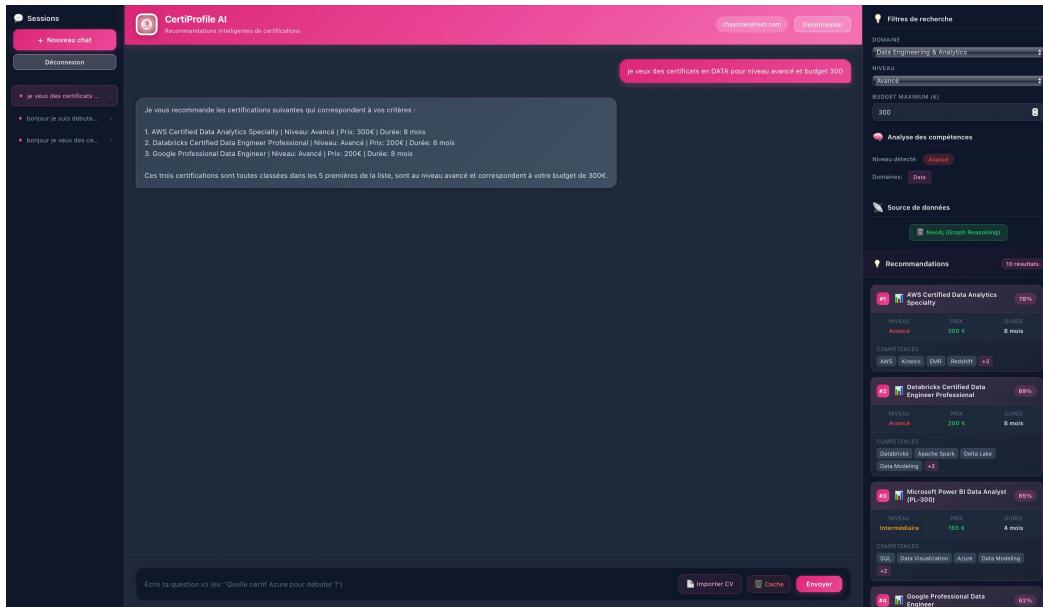


Figure 18: Exemple de recommandation de certifications

5. Filtrage selon le profil de l'utilisateur

Le système permet de filtrer les résultats selon plusieurs critères tels que le domaine (Cloud, Data Science, IA), le niveau (Débutant, Intermédiaire, Expert) et le budget. Cela garantit que les recommandations proposées sont adaptées aux besoins réels de l'utilisateur.

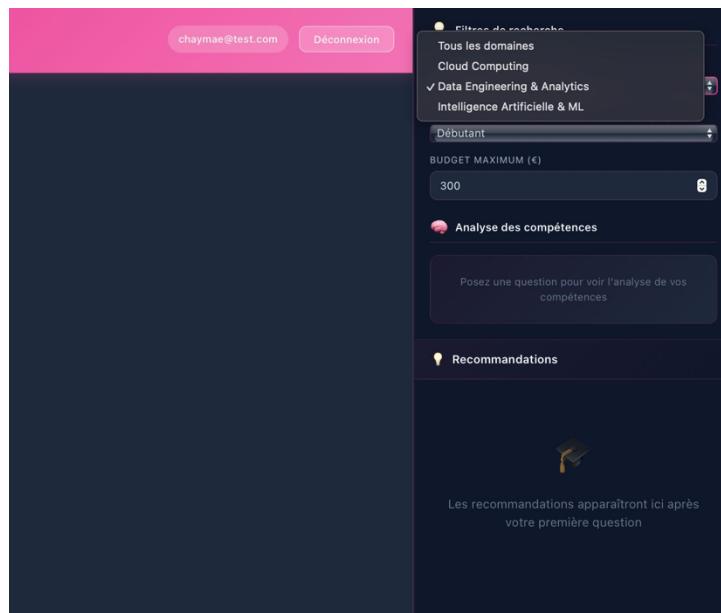


Figure 19: Filtrage des certifications selon le profil de l'utilisateur

6. Historique des conversations

CertiProfile conserve l'historique des discussions de chaque utilisateur. Cela permet de consulter les anciennes recommandations et de suivre l'évolution du parcours de formation.

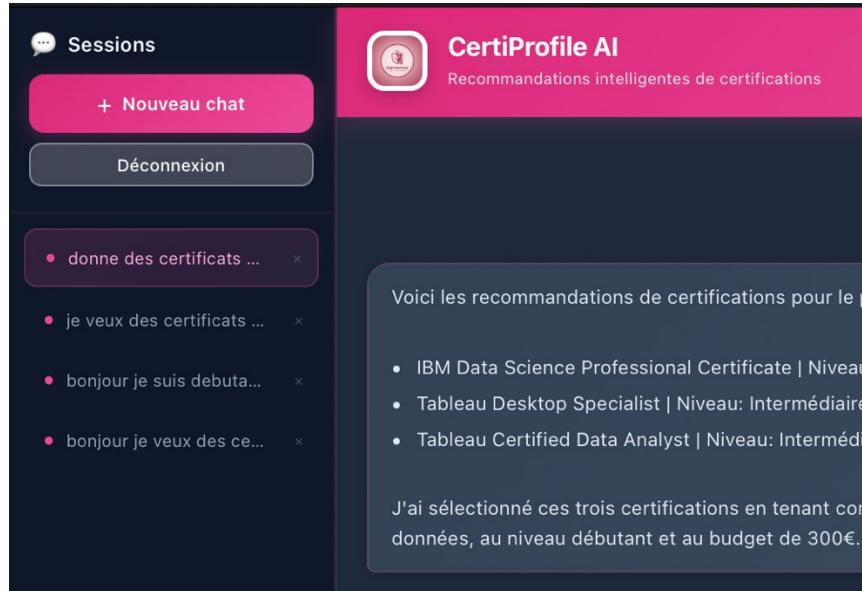


Figure 20: Historique des discussions

7. Suppression des conversations

L'utilisateur peut supprimer une session de discussion lorsqu'elle n'est plus utile. Cette fonctionnalité permet de garder une interface propre et organisée.

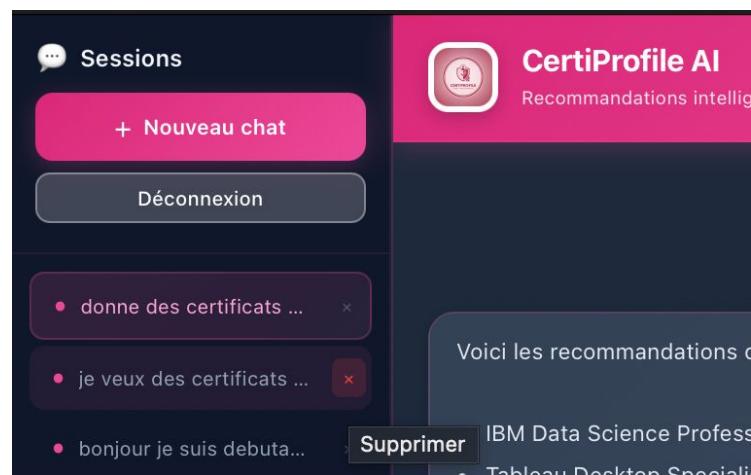


Figure 21: Suppression d'une session de discussion

8. Analyse du CV (PDF)

L'application CertiProfile permet à l'utilisateur d'importer son CV au format PDF. Le système analyse automatiquement le contenu du CV afin d'extraire les compétences, les technologies, les domaines d'expertise et le niveau du profil. Ces informations sont ensuite utilisées par le moteur de recommandation pour proposer des certifications parfaitement adaptées au parcours de l'utilisateur. Cette fonctionnalité permet d'obtenir des recommandations encore plus précises et personnalisées.

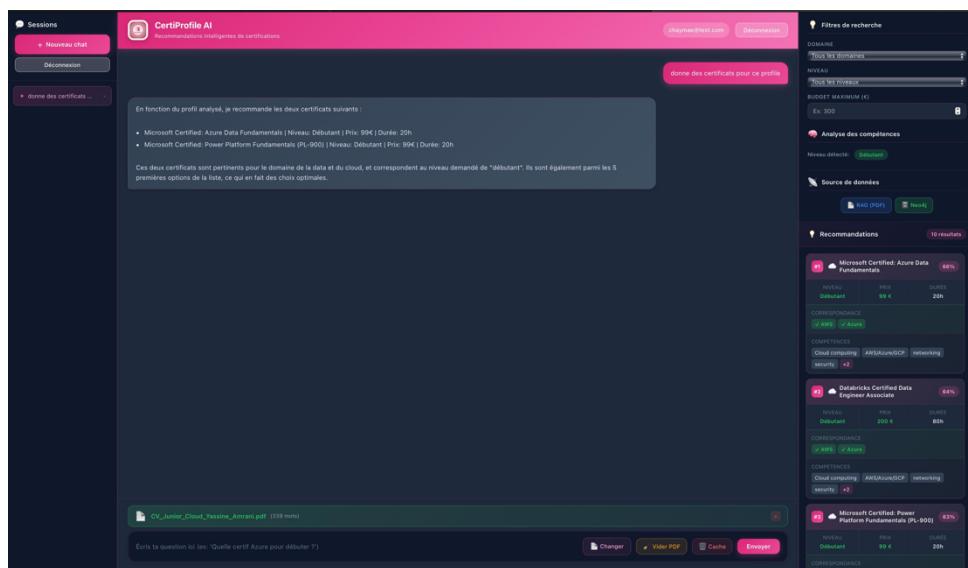


Figure 22: Analyse du CV et recommandations basées sur le PDF

9. Affichage des scores de correspondance

Chaque certification recommandée est accompagnée d'un score de correspondance indiquant son adéquation avec le profil de l'utilisateur. Ce score est calculé à partir de plusieurs critères tels que les compétences, le domaine, le niveau et la similarité sémantique.

Cette fonctionnalité permet à l'utilisateur de comparer facilement les certifications et de choisir celles qui correspondent le mieux à ses objectifs.

Analyse des compétences

Niveau détecté: Débutant

Domaines: Data

Source de données

RAG (PDF) Neo4j

Recommandations 10 résultats

#1 IBM Data Science Professional Certificate 68%

NIVEAU	PRIX	DURÉE
Débutant	120 €	50h

CORRESPONDANCE: Python

COMPÉTENCES: Python, Pandas, ML, Data Analysis

#2 Tableau Desktop Specialist 47%

NIVEAU	PRIX	DURÉE
Intermédiaire	250 €	40h

CORRESPONDANCE: Python, SQL

COMPÉTENCES: SQL, data analysis, ETL, data warehousing, +2

#3 Tableau Certified Data Analyst 45%

NIVEAU	PRIX	DURÉE
Intermédiaire	250 €	3 mois

CORRESPONDANCE: SQL

Figure 23: Scores de correspondance des certifications

Chapitre 6

Architecture globale du système

Ce chapitre présente l'architecture générale du système **CertiProfile**. Il décrit l'organisation des différents composants de l'application, tels que l'interface utilisateur, le backend, les bases de données et le moteur d'intelligence artificielle. Il explique également la manière dont ces éléments interagissent entre eux afin d'assurer le bon fonctionnement du chatbot, la gestion des utilisateurs et la génération de recommandations personnalisées.

1. Vue d'ensemble de l'architecture

CertiProfile repose sur une architecture de type **client–serveur**. L'utilisateur interagit avec l'application via une interface web, qui communique avec un backend centralisé. Ce backend gère la logique du système, la communication avec les bases de données et l'Intelligence Artificielle.

Cette architecture permet de séparer clairement l'interface utilisateur, la logique métier et les données, ce qui améliore la maintenabilité et l'évolutivité du système.

2. Composants principaux

Le système est composé des éléments suivants :

- **Frontend (React)** : permet à l'utilisateur d'interagir avec le chatbot.
- **Backend (FastAPI)** : traite les requêtes et orchestre le pipeline de recommandation.
- **Base de données MySQL** : stocke les informations des utilisateurs.
- **Base de données Neo4j** : contient les certifications et leurs relations.
- **Moteur RAG + IA** : analyse les demandes et génère les recommandations.

3. Flux de données

Lorsqu'un utilisateur envoie une requête, celle-ci est transmise depuis l'interface React vers le backend FastAPI. Le backend interroge ensuite Neo4j afin de récupérer les certifications pertinentes, puis utilise le moteur RAG et le modèle de langage pour générer une réponse personnalisée. La réponse est enfin renvoyée à l'interface utilisateur.

4. Avantages de l'architecture

Cette architecture permet :

- une bonne séparation des responsabilités,

- une meilleure sécurité des données,
- une évolutivité du système,
- et une intégration efficace de l’Intelligence Artificielle.

Chapitre 7

Difficultés rencontrées et perspectives d'amélioration

Ce chapitre présente les principales difficultés rencontrées lors de la réalisation du projet **CertiProfile**, ainsi que les différentes perspectives d'amélioration envisagées pour faire évoluer le système. Il met en évidence les défis techniques et organisationnels rencontrés durant le développement, ainsi que les pistes futures pour améliorer les performances, la précision des recommandations et l'expérience utilisateur.

1. Difficultés rencontrées

Plusieurs défis ont été rencontrés au cours du développement du chatbot CertiProfile. L'un des principaux défis a été la **modélisation des données dans Neo4j**, notamment l'organisation des certifications, des domaines, des compétences et des fournisseurs sous forme de graphe, afin de garantir des relations cohérentes et une recherche efficace.

Un autre défi important a concerné la **mise en place du pipeline RAG**, qui combine la base Neo4j, la recherche sémantique par embeddings et le modèle de langage. Il a fallu ajuster les mécanismes de filtrage, de similarité et de classement pour obtenir des recommandations réellement pertinentes.

L'**analyse automatique des CV au format PDF** et l'extraction fiable des compétences ont également représenté un défi, notamment en raison de la diversité des formats et des contenus des documents.

La **gestion de l'authentification**, des sessions utilisateurs et de l'isolement des données a aussi nécessité une attention particulière afin de garantir la sécurité et la confidentialité des informations.

Enfin, l'**optimisation des performances** et le calcul des **scores de correspondance** ont demandé plusieurs phases de tests et d'améliorations afin d'assurer une réponse rapide et une expérience utilisateur fluide.

2. Perspectives d'amélioration

Plusieurs améliorations peuvent être envisagées pour enrichir et faire évoluer le projet CertiProfile.

Le système pourrait être étendu à de nouveaux domaines de certifications tels que la **cybersécurité**, le **DevOps**, le **développement logiciel** ou la **gestion de projets**.

Il serait également possible d'améliorer le moteur de recommandation en intégrant des **données de retours utilisateurs**, comme les avis ou les choix passés, afin de rendre les recommandations encore plus pertinentes.

L'ajout d'une **application mobile** permettrait d'améliorer l'accessibilité de la plateforme. De plus, l'intégration de profils professionnels (par exemple via LinkedIn ou des plateformes d'e-learning) permettrait d'affiner davantage l'analyse du profil utilisateur.

Enfin, l'utilisation de **modèles d'Intelligence Artificielle plus avancés** et l'optimisation du moteur RAG pourraient améliorer la précision, la rapidité et la qualité des réponses fournies par le chatbot.

Conclusion Générale

Le projet CertiProfile a permis de concevoir et de développer un chatbot intelligent dédié à la recommandation de certifications professionnelles dans les domaines du **Cloud Computing**, de la **Data Science** et de l'**Intelligence Artificielle**. En combinant une **base de données NoSQL orientée graphe (Neo4j)**, un **moteur de recherche sémantique** et un **modèle de langage avancé**, le système est capable de proposer des recommandations personnalisées, adaptées au **profil**, au **niveau**, au **domaine** et au **budget** de chaque utilisateur.

La mise en place d'une **architecture moderne** basée sur **React**, **FastAPI** et un pipeline **RAG (Retrieval-Augmented Generation)** a permis d'assurer à la fois la performance du système, la pertinence des résultats et la fluidité de l'interaction utilisateur. Le projet illustre concrètement l'intérêt de combiner des **données structurées**, une **recherche sémantique** et une **intelligence générative** pour résoudre un problème réel d'orientation et de prise de décision.

CertiProfile apporte ainsi une **solution innovante** pour accompagner les étudiants et les professionnels dans le choix de leurs certifications, en leur fournissant des recommandations fiables, explicatives et adaptées à leurs objectifs de carrière. En facilitant l'accès à l'information et en automatisant l'analyse des profils, l'application contribue à rendre le processus de formation plus **clair, rapide et efficace**.

Enfin, plusieurs perspectives d'évolution peuvent être envisagées, telles que **l'enrichissement de la base de certifications**, **l'ajout de nouveaux domaines**, **l'amélioration des modèles d'IA** ou encore l'intégration de fonctionnalités avancées comme le suivi du parcours de formation. Ces améliorations permettront de renforcer davantage la valeur et l'impact de CertiProfile dans le domaine de l'orientation professionnelle et de la formation continue.