

Mémoire de Projet de Fin d'Etudes

Présenté en vue de l'obtention du

Diplôme National de Licence en Science
d'Informatique

Author Name

Date

ABSTRACT

A brief summary of the project.

DÉDICATION

To my parents.

TABLE DES MATIÈRES

1	Étude Préliminaire et Contextualisation du Projet	1
1	Contexte général et cadre académique du projet	2
1.1	Cadre académique du projet	2
1.2	Présentation de l'entité d'accueil	2
1.3	Identification de la problématique	4
1.4	Idée générale du projet	4
1.5	Concepts de base liés à notre projet	5
2	Analyse de l'existant et perspectives critiques	6
2.1	Étude des solutions existantes	6
2	Fondements de l'Intelligence Artificielle et de l'Infrastructure Cloud dans notre Plateforme	9
1	L'intelligence artificielle dans notre plateforme	10
1.1	Définitions	10
1.2	Explication des concepts clés	11
1.3	Algorithme d'Apprentissage Adopté	13
2	Le Cloud dans notre plateforme	15
2.1	Présentation du cloud computing	16
2.2	Type de cloud	16
2.3	Avantages de l'adoption du cloud dans notre projet	17

TABLE DES MATIÈRES

2.4	Service cloud adoptés	17
-----	---------------------------------	----

TABLE DES FIGURES

1.1	Logo du « Globe Services Informatique (GSI) »	3
1.2	Logo de « Google Classroom »	6
1.3	Logo de « Piazza »	7
1.4	Logo de « ChatGPT »	7
1.5	Logo de « poe »	8
2.1	Le déroulement et les étapes de Dynamic RAG	13
2.2	Le déroulement de l'algorithme Sparse Mixture of Experts	14

LISTE DES TABLEAUX

2.1	Descriptions des services de cloud computing	18
-----	--	----

Introduction

A long introduction.

CHAPITRE 1

ÉTUDE PRÉLIMINAIRE ET CONTEXTUALISATION DU PROJET

Contents

1	Contexte général et cadre académique du projet	2
1.1	Cadre académique du projet	2
1.2	Présentation de l'entité d'accueil	2
1.3	Identification de la problématique	4
1.4	Idée générale du projet	4
1.5	Concepts de base liés à notre projet	5
2	Analyse de l'existant et perspectives critiques	6
2.1	Étude des solutions existantes	6

Introduction

Ce chapitre vise à exposer l'étude préliminaire de notre projet. Tout d'abord, nous présenterons le contexte général et l'entité qui nous a accueilli pour faire un stage de fin d'études. Puis, nous aborderons la problématique et l'idée générale de notre projet. Ensuite, nous entreprendrons une analyse approfondie de l'existant, mettant en évidence les avantages et les limites des solutions similaires présentées sur le marché, afin de nous en inspirer et de retenir une solution plus raffinée. Enfin, nous concluons ce chapitre en exposant la méthodologie de développement la plus adaptée à la réalisation de notre projet, et le diagramme de Gantt illustrant le planning général de celui-ci.

1 Contexte général et cadre académique du projet

Dans cette partie, nous représenterons le contexte de notre projet, qui inclut le cadre académique, la présentation de l'organisme d'accueil, la problématique et l'idée générale de notre projet.

1.1 Cadre académique du projet

Notre projet, intitulé « *Conception et développement d'une plateforme E-learning intégrant un système intelligent d'élucidation intelligente des ressources pédagogiques* », s'inscrit dans le cadre de la préparation d'un projet de fin d'études en vue de l'obtention du diplôme national de Licence en Sciences d'Informatique à « **l'Institut Supérieur d'Informatique et de Mathématiques de Monastir (ISIMM)** » pour l'année universitaire 2023/2024. Le stage a été effectué au sein de la société **Globe Services Informatique GSI** durant une période de 4 mois.

1.2 Présentation de l'entité d'accueil

L'idée initiale de notre projet est proposée par nous-même et elle est adoptée par **STE Globe Services Informatique (GSI)**, une entreprise établie à Houmet

Souk, Monastir, en Tunisie. Fondée le 20 mai 2000 en tant que société à responsabilité limitée (SARL), **GSI** se concentre principalement sur le commerce et la maintenance de matériel informatique, ainsi que sur le développement web.



FIGURE 1.1 – Logo du « Globe Services Informatique (GSI) »

Les secteurs d'activités de GSI sont :

- La vente et la maintenance de matériel informatique.
- Le développement web et mobile.
- La formation en informatique.
- La vente de logiciels.

Identité de l'organisation :

- **Nom** : Globe Services Informatique (GSI).
- **Adresse** : 5 Rue de la République, Houmet Souk, Monastir, Tunisie.
- **Téléphone** : +(216) 73 447 447.
- **Fax** : (+216) 73 447 447.
- **E-mail** : commercial@gsi.com.tn.
- **Site web** : www.gsi.com.tn.

1.3 Identification de la problématique

Dans le quotidien étudiantin et dans le déroulement ordinaire d'un cours, où l'enseignant anime la classe et les étudiants s'investissent pleinement dans l'acquisition et la compréhension des connaissances, des perturbations diverses sont fréquemment rencontrées. En effet, les étudiants se heurtent souvent à des obstacles dans la compréhension de leurs cours ce qui entrave leur progression académique.

De plus, les supports pédagogiques fournis par les enseignants tels que les cours, exercices, corrections et examens peuvent parfois se révéler insuffisants pour une assimilation complète. Par conséquent, de nombreux étudiants sont amenés à chercher d'autres ressources par eux-mêmes. Néanmoins, l'accès à ces ressources présente un autre défi incontournable. Cette situation limite considérablement leur capacité à acquérir efficacement des connaissances.

Exemple : Le coût élevé de l'accès à certains documents et ressources pédagogiques d'intérêt constitue un obstacle financier majeur qui prive les étudiants d'outils essentiels fondamentaux à leur épanouissement éducatif.

1.4 Idée générale du projet

À l'origine de notre projet, nous avons puisé notre inspiration dans notre expérience de la vie étudiantine. Ainsi, nous cherchons à relever les défis auxquels nos pairs sont confrontés, tels que la compréhension de leurs cours, l'accès aux ressources pédagogiques adaptées, et le besoin de soutien académique personnalisé.

Notre idée consiste à concevoir et mettre en œuvre une plateforme interactive qui révolutionne l'expérience d'apprentissage des étudiants et des enseignants en fournissant des contenus multimédias et des outils d'intelligence artificielle. De plus, cette plateforme favorise l'apprentissage personnalisé et collaboratif, tout en demeurant une solution riche en services pertinents, intelligents et open-source.

Nous visons à exploiter des technologies avancées de l'intelligence artificiel (IA) et

du cloud pour mettre en place une solution robuste.

Avant de nous plonger dans l'analyse et la critique des solutions existantes sur le marché, il convient de préciser que notre projet est basé sur une idée préliminaire. Toutefois, cette idée peut être affinée et améliorée en s'inspirant des meilleures pratiques et des fonctionnalités offertes par les plateformes similaires. Dans un premier temps, nous allons donc expliquer les concepts de base liés à notre domaine de travail, afin de poser les fondations nécessaires à la compréhension de notre projet.

1.5 Concepts de base liés à notre projet

Notre projet s'appuie sur deux piliers fondamentaux qui sont le cloud computing et l'intelligence artificielle (IA). Ces concepts jouent un rôle central de notre ambition de révolutionner l'expérience d'apprentissage des étudiants et des enseignants, en offrant des solutions innovantes et performantes.

- **Cloud computing** : Le cloud computing révolutionne la manière dont les services informatiques sont fournis et consommés. En exploitant le cloud, notre projet peut tirer parti d'une infrastructure évolutive et flexible, permettant un accès rapide et sécurisé aux ressources informatiques. Cette approche offre également une réduction des coûts opérationnels et une amélioration de l'efficacité grâce à la mise à l'échelle automatique des ressources en fonction des besoins.
- **Intelligence artificielle (IA)** : L'intelligence artificielle constitue le cœur de notre projet, offrant des fonctionnalités avancées telles que l'analyse de documents, la recommandation de contenu personnalisé et l'assistance virtuelle. Grâce à l'IA, notre plateforme peut comprendre, interpréter et répondre aux besoins des utilisateurs de manière intelligente, offrant ainsi une expérience d'apprentissage plus personnalisée et efficace.

En combinant les capacités du cloud computing et de l'IA, notre projet vise à fournir une solution innovante et intelligente pour répondre aux défis de l'éducation contemporaine.

2 Analyse de l'existant et perspectives critiques

L'étude de l'existant est une étape primordiale qui permet de définir les points forts et les points faibles des systèmes similaires actuellement en place. Alors, cette section sera dédiée à faire une étude approfondie et critique des solutions existantes.

2.1 Étude des solutions existantes

Dans le domaine des technologies éducatives, plusieurs plateformes se démarquent par leur contribution à l'apprentissage. Nous étudierons ces solutions pour comprendre leurs forces. Nous identifierons également les opportunités d'amélioration que notre solution pourrait exploiter.

Étant donné qu'il n'existe pas de solutions tunisiennes similaires à notre projet, notre revue se concentre sur le marché international.

Ainsi, nous avons porté notre attention sur les solutions étrangères les plus connues qui s'alignent avec le contexte de notre projet. Notre étude se focalise, particulièrement sur *Google Classroom*, *Poe*, *ChatGPT* et *Piazza*, les quatre solutions les plus populaires et adaptées dans le monde entier.

- **Google Classroom** est une plateforme éducative qui permet aux enseignants de créer des salles de classe virtuelles pour leurs étudiants, partager des documents, des devoirs et communiquer avec les apprenants. Le logo de Google Classroom est illustré à la figure 1.2, permet de :



FIGURE 1.2 – Logo de « Google Classroom »

- Gérer les cours ainsi que les ressources pédagogiques pour les étudiants.

- Faire des discussions dans les classes.
- **Piazza** est une plateforme de collaboration en ligne conçue pour faciliter la communication entre les étudiants et les professeurs. Le logo de cette plateforme est affiché dans la figure 1.3, permet de :



FIGURE 1.3 – Logo de « Piazza »

- Stocker les documents.
- Organiser des discussions entre les étudiants et les enseignants.
- **ChatGPT** ou Chat Generative Pretrained Transformer, est un chatbot doté d'intelligence artificielle qui fournit des réponses textuelles instantanées aux questions des utilisateurs. Il se positionne comme un outil polyvalent répondant aux besoins diverse. Le logo de ChatGPT est illustré à la figure 1.4, permet de :



FIGURE 1.4 – Logo de « ChatGPT »

- Répondre aux questions et aux requêtes en se basant sur le contexte de la conversation.
- Fournir des informations sur divers sujets.
- **Poe** est un chatbot qui fonctionne à partir de textes, offrant des interactions conversationnelles pour répondre aux questions des utilisateurs et fournir une assistance. Le logo de Poe est illustré à la figure 1.5, permet de :
 - Résoudre des problèmes en décrivant votre contexte.
 - Obtenir des réponses personnalisées à vos besoins.

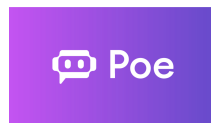


FIGURE 1.5 – Logo de « poe »

CHAPITRE 2

FONDEMENTS DE L'INTELLIGENCE ARTIFICIELLE ET DE L'INFRASTRUCTURE CLOUD DANS NOTRE PLATEFORME

Introduction

Ce chapitre explore les fondements de l'Intelligence Artificielle (IA) et de l'infrastructure cloud dans notre plateforme. Nous aborderons des définitions et explications des concepts clés de l'AI et de l'apprentissage profond, notre choix de l'algorithme MoE et l'intégration du cloud. Cette compréhension nous permettra de saisir les capacités de notre plateforme.

1 L'intelligence artificielle dans notre plateforme

Cette section sera consacrée à la présentation des concepts de base sur l'intelligence artificielle.

1.1 Définitions

Dans cette partie, nous allons définir l'intelligence artificielle et l'apprentissage profond.

1.1.1 Intelligence artificielle

L'intelligence artificielle concerne la création de machines capables de penser et d'agir comme des êtres humains [X]. En d'autres termes, c'est la science qui vise à créer des programmes informatiques et des machines qui peuvent imiter le raisonnement humain, apprendre de l'expérience et accomplir des tâches variées de manière autonome.

1.1.2 Apprentissage profond

L'apprentissage profond, connu sous le nom de deep learning, est une branche de l'IA qui se concentre sur l'entraînement de réseaux de neurones artificiels [x]. Ces réseaux de neurones sont organisés en couches. Chaque couche transforme l'entrée qu'elle reçoit pour produire une sortie. Ces transformations sont ajustées

automatiquement pendant l'apprentissage pour améliorer les performances du système.

1.2 Explication des concepts clés

Dans cette section, nous allons clarifier quelques concepts fondamentaux.

1.2.1 Natural Language Model (NLP)

Les modèles de langage naturel (NLP) sont des composants de l'intelligence artificielle conçus pour comprendre et générer un langage humain naturel. Ils sont largement utilisés dans des applications telles que la traduction automatique, la génération de texte et l'analyse du sentiment. Ces modèles sont entraînés sur de grandes quantités de données textuelles afin d'apprendre les structures linguistiques et de capturer les nuances du langage.

1.2.2 Large Model Language (LLM)

Les modèles de langage de grande taille (LLM) sont des modèles d'intelligence artificielle qui ont été entraînés sur de vastes ensembles de données textuelles pour acquérir une compréhension approfondie du langage naturel. Ces modèles sont capables de générer du texte cohérent et de qualité et sont souvent utilisés pour une variété de tâches en NLP.

1.2.3 Tokenizers

Les tokenizers sont des outils essentiels en traitement automatique du NLP qui découpent le texte en unités plus petites appelées "tokens". Ces tokens peuvent être des mots, des sous-mots ou même des caractères individuels. Ils servent de points de départ pour l'analyse et le traitement du texte.

1.2.4 Retrieval Augmented Generation (RAG)

RAG est un concept utilisé dans le domaine de l'intelligence artificielle et le traitement du langage naturel. Il utilise des techniques de récupération d'informa-

tions pour améliorer la génération de texte par des modèles de langage.

En d'autres termes, au lieu de simplement générer du texte en fonction de ce que le modèle a appris pendant son entraînement, un système RAG va d'abord chercher dans une grande base de données de textes pour trouver des informations pertinentes. Il utilisera ensuite ces informations pour générer une réponse plus informée et précise.

Notons que cette base de données de textes peut être vectorielle. C'est-à-dire que les textes sont représentés sous forme de vecteurs dans un espace vectoriel. Ceci permet de les comparer et de les rechercher de manière efficace.

1.2.5 Dynamic Retrieval Augmented Generation (Dynamic RAG)

Dynamic RAG (Retrieval Augmented Generation) est une approche avancée de RAG qui permet d'adapter dynamiquement la récupération d'informations en fonction du contexte de la requête de l'utilisateur.

Dans un système RAG standard, la stratégie de récupération d'informations est généralement fixe et déterminée à l'avance. Par exemple, le système doit toujours chercher dans la même base de données de textes.

Alors que dans un système Dynamic RAG, le système peut adapter sa stratégie de récupération d'informations en fonction de la tâche à accomplir. Par exemple, pour une tâche de réponse à des questions, le système peut chercher dans une base de données de textes spécifique.

Cette approche permet au système d'être plus flexible et adaptable, ce qui lui permet d'améliorer ses performances sur une variété de tâches.

La Figure 2.1. synthétise les étapes et le déroulement de Dynamic RAG.

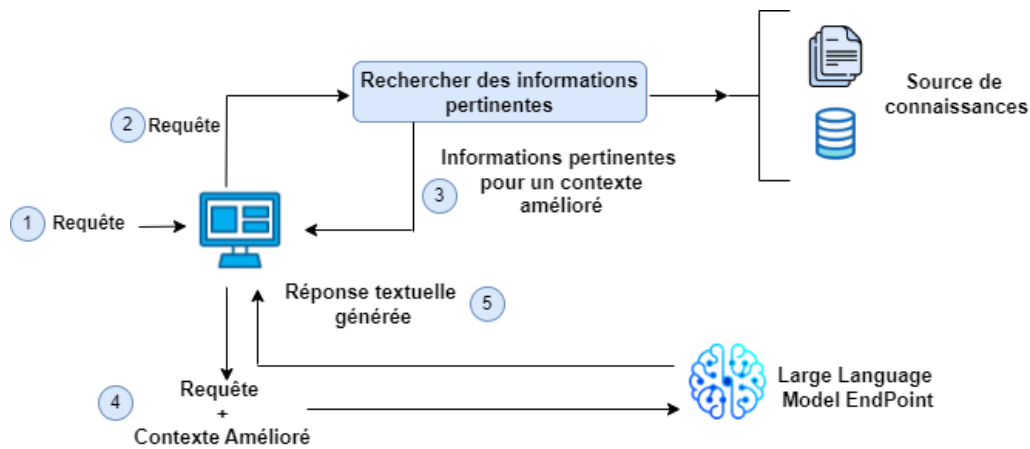


FIGURE 2.1 – Le déroulement et les étapes de Dynamic RAG

1.3 Algorithme d'Apprentissage Adopté

Dans cette section, nous expliquerons le Sparse Mixture of Experts (MoE), l'algorithme d'apprentissage adopté. Nous aborderons l'introduction à cet algorithme, la définition de la couche Sparse MoE et son architecture ainsi que le processus de formation et d'inférence associé.

Notons que les informations de cette section sont issues de la référence suivante.[x]

1.3.1 Couche Sparse MoE et son architecture

La couche Sparse MoE est une composante clé des modèles de transformers remplaçant les couches traditionnelles de réseaux de neurones denses. Chaque couche Sparse MoE est constituée de multiples "experts" qui sont essentiellement des réseaux de neurones. Ces experts sont souvent des réseaux feed-forward (FFN) et peuvent aussi être plus complexes ou former un autre MoE permettant ainsi des MoEs hiérarchiques.

En plus des experts, une couche Sparse MoE inclut un réseau de portes ou "routeur". Ce routeur détermine comment les tokens sont distribués aux experts. Il est important de noter qu'un même token peut être attribué à plusieurs experts.

Le routeur est constitué de paramètres appris et entraîné avec le reste du réseau.

La Figure 2.2. synthétise l'architecture de l'algorithme Sparse Mixture of Experts.

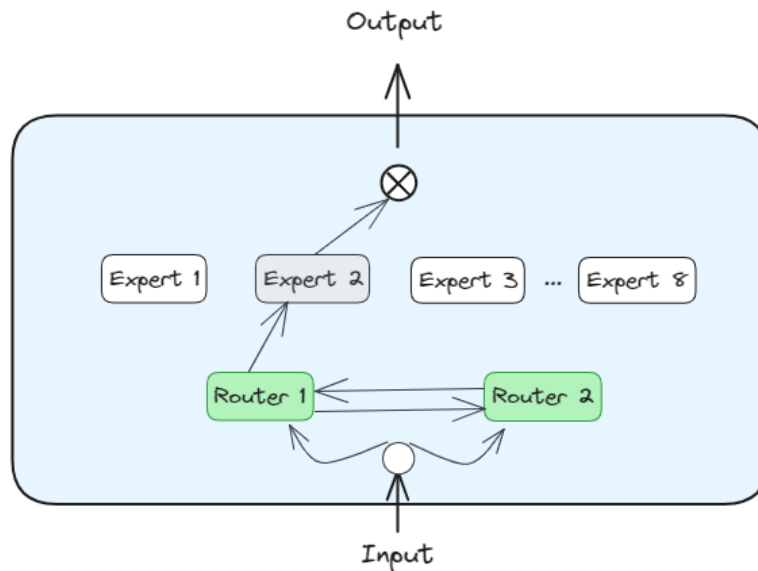


FIGURE 2.2 – Le déroulement de l’algorithme Sparse Mixture of Experts

1.3.2 Fonctionnement de l’algorithme Sparse Mixture of Experts

L’algorithme sparse Mixture of Experts est une méthode avancée qui combine les avantages des modèles experts spécialisés avec la flexibilité des modèles généralistes.

Contrairement aux réseaux neuronaux traditionnels qui utilisent une seule architecture de modèle pour traiter toutes les entrées, le MoE divise le processus d’apprentissage en plusieurs experts spécialisés. Ces experts sont ensuite combinés de manière pondérée pour générer une prédiction finale (voir Figure 2.2). Chaque expert est responsable de traiter une portion spécifique de l’entrée en apportant ainsi sa propre expertise à la résolution du problème. Ces experts peuvent être considérés comme des sous-modèles qui se spécialisent dans différentes parties de l’espace d’entrée.

L’approche sparse du MoE signifie que seuls quelques experts sont activés pour chaque exemple d’entrée. Elle permet une utilisation efficace des ressources computationnelles et une adaptation dynamique aux différentes conditions d’entrée.

Cette activation sélective des experts est déterminée par des routeurs qui dirigent l'entrée vers les experts les plus pertinents en fonction de son contenu.

1.3.3 Processus de formation et d'inférence

Le processus d'entraînement de Mixture of Experts (MoE) implique simultanément l'apprentissage des paramètres des experts et du routeur.

Pendant la formation, chaque jeton d'entrée est dirigé vers un ou plusieurs experts en fonction des poids appris par le routeur. Les experts traitent ces jetons et retournent leurs sorties respectives. Ces sorties sont combinées pour former la sortie finale du MoE utilisée pour calculer la perte et mettre à jour les paramètres du modèle. En phase d'inférence, seuls les experts les plus pertinents sont sélectionnés selon les poids du routeur en réduisant ainsi la complexité de calcul et améliorant l'efficacité du MoE. Bien que tous les paramètres doivent être chargés en mémoire, seuls ceux nécessaires sont utilisés pendant l'inférence.

1.3.4 Raisons du Choix de l'Algorithme Sparse Mixture of Experts

Le choix d'adopter l'algorithme Sparse Mixture of Experts (MoE) repose sur sa capacité à optimiser l'utilisation des ressources computationnelles en ne faisant fonctionner que les parties pertinentes du modèle. Cela permet d'augmenter l'efficacité de l'apprentissage et réduire la charge de calcul qui est essentiel pour des applications à grande échelle nécessitant des modèles complexes.

En intégrant le Sparse MoE dans notre plateforme, nous sommes en mesure de fournir des réponses plus précises et adaptées aux besoins spécifiques de chaque utilisateur en améliorant ainsi l'expérience globale d'apprentissage.

2 Le Cloud dans notre plateforme

Dans cette partie, nous allons présenter le cloud computing, les types de cloud ainsi que le type de cloud adopté dans notre plateforme.

2.1 Présentation du cloud computing

Le cloud computing est une technologie qui permet d'accéder à des ressources informatiques en utilisant l'internet.

L'infrastructure cloud joue un rôle crucial dans le bon fonctionnement et l'évolutivité de notre application. Elle nous permet de bénéficier de services performants et sécurisés en optimisant nos coûts d'exploitation.

2.2 Type de cloud

Dans cette section, nous présenterons les types de cloud, suivi par le type adopté.

2.2.1 Types d'infrastructures de cloud computing

Le cloud computing est divisé en trois types principaux : public, privé et hybride.

- **Cloud privé**

Un cloud privé est un environnement de cloud computing dédié à une seule organisation. Dans un cloud privé, toutes les ressources sont isolées et sous le contrôle d'une seule organisation. Ainsi, le cloud privé est également appelé cloud interne ou d'entreprise. [x]

- **Cloud publique**

Un cloud public est une infrastructure informatique dans laquelle un fournisseur de services met des ressources à la disposition du public via internet. Les ressources varient selon le fournisseur mais peuvent inclure des capacités de stockage, des applications ou des machines virtuelles. [x]

- **Cloud hybride**

Un cloud hybride est un environnement informatique mixte dans lequel des applications s'exécutent à l'aide d'une combinaison de ressources de calcul, de stockage et de services dans différents environnements (clouds publics et

clouds privés, y compris des centres de données sur site ou en périphérie).
[x]

2.2.2 Type de cloud adopté

Notre application repose sur une architecture Cloud hybride qui combine à la fois des services de Cloud public et de Cloud privé. Cette approche nous permet de tirer parti des avantages de chaque modèle en fonction des besoins spécifiques de notre application.

2.3 Avantages de l'adoption du cloud dans notre projet

L'infrastructure Cloud offre de nombreux avantages pour notre projet. Tout d'abord, elle nous permet de bénéficier d'une grande scalabilité en adaptant automatiquement les ressources allouées en fonction des besoins de notre application. De plus, elle garantit une haute disponibilité et une redondance des données grâce à la répartition géographique des serveurs. Enfin, l'infrastructure Cloud nous permet de réduire nos coûts d'exploitation.

2.4 Service cloud adoptés

Parmi les services Cloud intégrés dans notre plateforme, nous retrouvons le Runtime Edge, le Serverless et le CDN (Content Delivery Network). Les services cloud utilisés dans notre plateforme sont présentés dans le tableau 2.1.

TABLE 2.1 – Descriptions des services de cloud computing

Service	Description
Runtime Edge	Edge Runtime est idéal pour la diffusion de contenu dynamique et personnalisé avec une faible latence en utilisant de petites fonctions simples. Sa rapidité provient de l'utilisation minimale des ressources, mais cela peut être limité dans de nombreux scénarios. Dans notre application, Edge Runtime est utilisé pour optimiser la diffusion de contenu personnalisé pour garantir ainsi une expérience utilisateur rapide et réactive.
Serverless	Les architectures sans serveur, ou Serverless, permettent de déléguer la gestion des serveurs à un fournisseur de services Cloud. Cette approche offre plusieurs avantages tels qu'une réduction des coûts d'exploitation, une scalabilité automatique et une simplification de la maintenance. Dans notre application, le Serverless est utilisé pour gérer les fonctions backend telles que le traitement des données et la gestion des Application Programming Interface (API).
Content Delivery Network	Un Content Delivery Network (CDN) est un réseau de serveurs distribués géographiquement qui permet de distribuer du contenu à grande échelle. L'utilisation d'un CDN permet d'améliorer les performances de notre application en réduisant la latence et en optimisant la bande passante. Dans notre cas, le CDN est utilisé pour distribuer les ressources statiques de notre application telles que les images, les feuilles de style et les scripts.

Conclusion

Notre application exploite des techniques d'apprentissage profond et une infrastructure cloud avancée pour offrir une expérience utilisateur optimale et répondre aux exigences de performance et de fiabilité. Le chapitre suivant approfondira les exigences fonctionnelles et non fonctionnelles pour concevoir notre solution adaptée.