 **Page de garde**

 **Remerciements**

 **Résumé (et Abstract en anglais)**

 **Table des matières**

 **Liste des figures et tableaux (facultatif)**

 **Introduction générale**

 **Chapitre 1 : Contexte général et état de l’art**

* Introduction
* Contexte de la vérification KYC
* Règlementation KYC (ex : AML, RGPD…)
* Technologies existantes
* Intelligence Artificielle et OCR dans le KYC
* Conclusion

 **Chapitre 2 : Analyse des besoins et spécifications**

* Cahier des charges
* Objectifs fonctionnels et non-fonctionnels
* Architecture globale du système
* Choix technologiques (Python, Tesseract, OpenCV, etc.)
* Modèle IA utilisé (ex : CNN, LLM, etc.)

 **Chapitre 3 : Conception et implémentation**

* Architecture logicielle
* Conception des modules (OCR, vérification des pièces, reconnaissance faciale, etc.)
* Intégration IA (modèle, dataset, entraînement, évaluation)
* Interface utilisateur (si applicable)
* Sécurité et gestion des données personnelles

 **Chapitre 4 : Résultats et tests**

* Résultats obtenus
* Tests (unitaires, fonctionnels, utilisateurs)
* Comparaison avec solutions existantes
* Limites du système

 **Conclusion générale et perspectives**

 **Bibliographie**

 **Annexes (code source, captures d’écran, logs de tests, etc.)**

**Introduction Générale**

Dans un monde de plus en plus numérisé, la vérification de l’identité des utilisateurs, communément appelée **KYC (Know Your Customer)**, est devenue une exigence incontournable pour les institutions financières, les banques en ligne, les services de paiement et même certaines plateformes numériques ou de crypto monnaie. Cette procédure, qui vise à authentifier l'identité d’un individu avant de lui fournir un service, est au cœur de la lutte contre la fraude, le blanchiment d’argent et le financement du terrorisme.

Cependant, les méthodes traditionnelles de vérification KYC reposent souvent sur des processus manuels, coûteux, lents et sujets à des erreurs humaines. Dans ce contexte, **l’intégration de l’Intelligence Artificielle (IA)** et de la **Reconnaissance Optique de Caractères (OCR)** ouvre la voie à des systèmes plus **automatisés, intelligents et fiables**, capables de traiter de grandes quantités de données en un temps réduit, tout en maintenant un haut niveau de précision.

Ce travail de fin d’étude s’inscrit dans cette dynamique de transformation numérique. Il a pour objectif de **concevoir et implémenter un système intelligent de vérification KYC**, exploitant les technologies de l’IA pour la reconnaissance d’images et de textes, ainsi que des algorithmes OCR pour l'extraction automatique des informations à partir de documents d'identité (CNI, passeport, permis de conduire, etc.).

À travers ce projet, il s’agira d’aborder les différentes phases de conception d’un tel système : de **l’analyse des besoins fonctionnels**, au **choix des outils technologiques**, en passant par **la modélisation, le développement et les tests du prototype**. L’approche proposée vise à offrir un système à la fois **fiable**, **rapide**, et **respectueux des contraintes réglementaires liées à la protection des données personnelles**.

Ce rapport est structuré en quatre chapitres principaux. Le premier chapitre présente le contexte général du projet et les technologies existantes dans le domaine KYC. Le second détaille les besoins du système et les choix techniques retenus. La troisième traite de la conception et de l’implémentation du système intelligent, et enfin, le quatrième expose les résultats obtenus, les tests réalisés ainsi que les perspectives d’amélioration du système.

# **Chapitre 1 : Contexte et état de l’art**

## 1.1 Introduction

L’identification des utilisateurs constitue un enjeu majeur pour les entreprises opérant dans des secteurs sensibles, notamment la finance, la banque, l’assurance et les télécommunications. Afin de garantir la sécurité des opérations, la confiance des clients et la conformité réglementaire, les institutions mettent en place des procédures de vérification d’identité, connues sous le nom de **KYC (Know Your Customer)**. Dans ce chapitre, nous présentons le contexte réglementaire du KYC, ses enjeux, les limites des solutions traditionnelles, ainsi qu’un état de l’art des technologies utilisées, notamment l’**Intelligence Artificielle (IA)** et la **Reconnaissance Optique de Caractères (OCR)**.

## 1.2 Présentation du processus KYC

Le processus **KYC** désigne l’ensemble des procédures et de règlements que les institutions financières utilisent pour **vérifier l’identité de ses clients** avant d’établir une relation commerciale. Les procédures consistent à établir l'identité, le profil financier et les risques potentiels de dialogue avec lui. Il inclut généralement :

* La collecte de documents d’identité (carte nationale, passeport, justificatif de domicile, etc.)
* La vérification de l’authenticité des documents fournis
* L’analyse de la correspondance entre les données extraites et les déclarations du client
* Parfois, une reconnaissance faciale ou une preuve de vie (liveness check)

L'objectif est d'identifier les activités illégales telles :

* Les fraudes à l’identité
* Le blanchissement d’argent
* Le financement du terrorisme avant qu’il ne se produise.

Il y a trois éléments clés de KYC :

* Le processus KYC commence par le Programme d'identification de la clientèle (CIP), où les institutions financières recueillent les informations d'identification de base d'un client :
* Nom
* Date de naissance
* Adresse
* Numéro d'identification

Les institutions recueillent généralement ces informations lorsque le client ouvre un compte, mais ils sont tenus de vérifier l'identité du titulaire du compte « dans un délai raisonnable ». L'identification des clients est effectuée en vérifiant des documents tels que les passeports, les permis de conduire ou les numéros de sécurité sociale, et parfois en utilisant des méthodes non documentaires telles que les contrôles de base de données.

* Diligence raisonnable pour les clients (CDD) : la CDD est la prochaine étape de KYC. Il s’agit de comprendre le comportement financier et les modèles de transaction du client pour créer un profil de risque. Les clients à haut risque doivent faire preuve de confiance renforcée, moyennant des vérifications plus approfondies des antécédents et un contrôle continu. Une composante critique et moins débattue du processus KYC est la surveillance continue, où les transactions des clients sont continuellement observées pour repérer des activités inhabituelles. Toute transaction ou comportement suspect est alors signalé aux autorités compétentes. Les institutions procéderont également à des examens périodiques pour s'assurer que les informations sur les clients sont à jour et exactes.
* Surveillance continue : Elle consiste à superviser les transactions financières et les comptes sur la base de profil à risque du client. Les institutions peuvent envisager des activités telles que des pics d’activité de transaction ou des activités transfrontalières imprévues.

Strictes en matière de KYC pour les organismes de services financiers sont tenus de vérifier l'identité de leurs clients, de comprendre la nature de leurs activités financières et de déterminer les profils de risque par le biais de contrôles KYC. Pour se conformer à la réglementation de la LBC et prévenir la criminalité financière, ils doivent également surveiller en permanence les transactions des clients, tenir des registres détaillés et signaler les activités suspectes.

Élaborer une politique de la KYC

L'élaboration d'une politique robuste de la KYC est essentielle pour garantir le respect des réglementations internationales contre la fraude, le blanchiment d'argent et le financement du terrorisme. L'adoption de mesures claires pour la mise en œuvre de la KYC contribue également à favoriser des transactions financières plus sûres.

Les institutions financières fondent généralement leurs politiques de la KYC sur quatre éléments clés :

* Politiques de la clientèle
* Procédures d'identification des clients, y compris les contrôles d'identité, la vérification, la collecte de données, les contrôles des listes de personnes politiques et les listes de sanctions
* Évaluation et gestion des risques
* Suivi et tenue des registres

Mener une formation des employés

La formation des employés est essentielle pour communiquer, contrôler et préserver les connaissances relatives au lutte contre le blanchiment d'argent et la lutte contre le financement du terrorisme (AML/CFT) et les exigences de la KYC au sein d'une organisation. Au minimum, la formation dans ces domaines devrait porter sur :

* Moyens que le blanchiment de capitaux et le financement du terrorisme peuvent être effectués avec les produits et services de l'entreprise
* Politiques internes visant à prévenir la fraude, le blanchiment d'argent et le financement du terrorisme et à intensifier les activités suspectes
* Informations générales sur les exigences de conformité réglementaire en matière de LBC/FT et de KYC
* Comment identifier et signaler les transactions à un organisme de réglementation ou à une autorité

## 1.3 Limites des systèmes KYC traditionnels

Les systèmes KYC traditionnels reposent largement sur une intervention humaine, ce qui engendre plusieurs limites :

* **Temps de traitement élevé** : la vérification manuelle est longue et inefficace à grande échelle
* **Coût opérationnel élevé** : la mobilisation d’équipes dédiées alourdit les charges
* **Risque d’erreur humaine** : des erreurs de saisie ou de jugement peuvent compromettre la fiabilité
* **Expérience utilisateur dégradée** : les délais et la complexité du processus peuvent décourager les utilisateurs

Face à ces limites, de plus en plus d’entreprises cherchent à automatiser ces processus grâce aux avancées technologiques.

## 1.4 L’Intelligence Artificielle au service du KYC

L’**Intelligence Artificielle** permet d’automatiser les tâches complexes, d’améliorer la précision des vérifications et d’accélérer les délais de traitement. Voici quelques applications concrètes de l’IA dans le KYC :

* **Reconnaissance faciale** : pour comparer une photo d’identité avec une vidéo ou un selfie en temps réel
* **Détection de fraude documentaire** : via des modèles de vision par ordinateur
* **Preuve de vie (liveness detection)** : pour s’assurer que la personne est bien réelle et présente
* **Classification automatique des documents** : pour identifier le type de document présenté

Les techniques d’IA couramment utilisées incluent les **réseaux de neurones convolutifs (CNN)** pour l’analyse d’images, les **modèles NLP** pour la lecture de texte, et les **systèmes d’apprentissage supervisé** pour la détection d’anomalies.

## 1.5 La Reconnaissance Optique de Caractères (OCR)

La technologie **OCR (Optical Character Recognition)** permet d’extraire automatiquement du texte à partir d’images de documents. Elle joue un rôle central dans les systèmes KYC automatisés :

* Extraction d’informations personnelles (nom, prénom, date de naissance, numéro de pièce, etc.)
* Conversion d’un document image (photo ou scan) en texte exploitable
* Pré-traitement pour vérification ou stockage

Des bibliothèques open source comme **Tesseract OCR** (développée par Google) sont largement utilisées, souvent combinées avec des outils comme **OpenCV** pour améliorer la qualité de l’image (redressement, suppression de bruit, contraste).

## 1.6 Travaux existants et solutions commerciales

Plusieurs entreprises proposent déjà des solutions KYC automatisées intégrant l’IA et l’OCR, parmi lesquelles :

* **Onfido**, **IDnow**, **Jumio** : services de vérification d’identité intégrés à des plateformes SaaS
* **Microsoft Azure Cognitive Services**, **Google Cloud Vision API**, **Amazon Rekognition** : APIs cloud puissantes pour la reconnaissance d’images et de textes
* **Open source** : combinaisons de bibliothèques (Tesseract, Keras, TensorFlow, OpenCV)

Ces solutions varient en termes de coût, précision, rapidité et niveau de personnalisation.

## 1.7 Conclusion

La vérification d’identité KYC est un processus critique, soumis à des exigences réglementaires strictes. Les solutions traditionnelles montrent leurs limites face à la croissance des besoins. Les avancées en **Intelligence Artificielle** et en **OCR** offrent une opportunité réelle de transformation de ces processus, en permettant une **automatisation intelligente, rapide et fiable**. Le système que nous proposons dans ce projet s’inscrit dans cette lignée, en tirant parti des technologies modernes pour améliorer l’expérience utilisateur et la sécurité globale.

# **Chapitre 2 : Analyse des besoins et spécifications**

## 2.1 Introduction

Avant toute conception technique, il est essentiel de définir clairement les besoins du système à mettre en place. Cette phase permet d’identifier les objectifs fonctionnels, les contraintes techniques et réglementaires, l’architecture de l’application, ainsi que les choix technologiques adaptés. Ce chapitre présente l’analyse des besoins du système de vérification KYC basé sur l’IA et l’OCR, en décrivant son cahier des charges, ses fonctionnalités attendues, son architecture générale, ainsi que les technologies envisagées pour sa mise en œuvre.

## 2.2 Objectif global du système

L’objectif principal du projet est de **concevoir et implémenter un système intelligent capable d’effectuer automatiquement la vérification d’identité d’un utilisateur** à partir de documents officiels fournis sous forme d’image ou de scan, en combinant des techniques d’**OCR** pour l’extraction de texte, et d’**IA** pour la reconnaissance faciale, la détection de fraudes et la validation croisée des données.

Cibles et utilisateurs

III.1. Cibles

Les cibles du projet sont principalement les entités ou organisations ayant des interactions avec des clients, nécessitant la mise en œuvre de processus de vérification d'identité pour se conformer aux réglementations bancaires et prévenir les activités frauduleuses ou illégales. Ces cibles incluent :

* Institutions financières : Banques, compagnies d'assurance, établissements de paiement, sociétés de transfert d'argent, etc.
* Organisations gouvernementales : Pour la gestion de la conformité et la régulation des transactions financières.
* Entreprises ayant des services financiers : Par exemple, des sociétés offrant des services de commerce en ligne, etc.
* Fournisseurs de services de paiement : Entreprises gérant des transactions et des plateformes nécessitant des processus KYC pour l'enregistrement des clients.
* Organisations non gouvernementales (ONG) : Certaines ONG qui doivent gérer les identités des donateurs ou des bénéficiaires dans des contextes spécifiques.
* Marchés et commerces en ligne : Notamment ceux opérant à l'international, qui doivent se conformer aux exigences KYC dans différentes juridictions.

Utilisateurs

Ce projet impliquera différents types d'utilisateurs ayant des rôles spécifiques dans le processus de gestion et de suivi des clients. Voici les principaux utilisateurs et leurs rôles :

III.2.1. Administrateurs

* Rôle et responsabilités :
* Gestion des utilisateurs : Création, modification, suppression des utilisateurs.
* Gestion des configurations : Configuration des services, groupes, pays, profils de conformité, types d’opérations, et autres paramètres système.
* Création des profils de conformité : Définir les profils de conformité en fonction des différents types de clients et des exigences réglementaires.
* Gestion des sanctions et listes noires : Gestion des clients dans la liste noire et des sanctions.
* Supervision globale : Supervision de toutes les opérations dans le système.

III.2.2. Responsables Back-office

* Rôle et responsabilités :
* Suivi des clients : Enregistrement et suivi des clients sur la plateforme.
* Contrôle des documents et des transactions : Réception et validation des documents d’identité, enregistrement des transactions, surveillance des activités suspectes.
* Blocage et déblocage des clients : Blocage des clients si nécessaire et gestion des raisons de blocage/déblocage.
* Gestion des alertes et notifications : Consultation des alertes générées par le système et prise de mesures appropriées.
* Mise à jour de la conformité des clients : Assurer la mise à jour des profils de conformité des clients selon les exigences.

III.2.3. Utilisateurs KYC (Client)

* Rôle et responsabilités :
* Enregistrement et soumission des documents : Les clients doivent soumettre leurs documents d’identité et autres informations nécessaires pour la vérification de leur identité.
* Interaction avec la plateforme : Remplir des formulaires de KYC et envoyer des informations via des canaux sécurisés (Par un service REST).
* Mise à jour des informations personnelles : Fournir des informations à jour lorsque demandé, comme le renouvellement des pièces d’identité ou la mise à jour des détails de profil.
* Suivi des alertes : Les clients peuvent être informés de l'état de leur vérification KYC via des notifications.

III.2.4. Utilisateurs Financiers (Banques, Institutions)

* Rôle et responsabilités :
* Vérification des clients : Analyser et valider les informations des clients en fonction des réglementations KYC.
* Suivi des transactions : Analyser les transactions des clients pour détecter toute activité suspecte ou frauduleuse.
* Émission de rapports de conformité : Générer des rapports relatifs à la conformité des clients en matière de KYC et de lutte contre le blanchiment d'argent.

III.2.5. Responsable Sécurité et Gestion des Risques

* Rôle et responsabilités :
* Analyse des risques : Analyser les transactions suspectes, examiner les alertes de sécurité et déterminer les risques potentiels associés aux activités des clients.
* Mise en œuvre des actions correctives : En cas de transaction suspecte, assurer la prise en charge des alertes et bloquer les transactions ou comptes si nécessaire.
* Veille réglementaire : Assurer la conformité aux normes de sécurité des informations et aux réglementations locales et internationales relatives au KYC.

2.3 Cahier des charges

2.3.1 Besoins fonctionnels

Le système doit permettre de :

* D’enregistrement les informations d’un client
* Accepter et stocker les **documents d’identité envoyées par les clients** (CNI, passeport, permis de conduire…)
* Extraire automatiquement les informations personnelles via OCR
* Vérifier la **validité du document** (forme, structure, lisibilité, existence du numéro de la pièce d’identité…)
* Comparer la photo du document à un **selfie utilisateur** (reconnaissance faciale)
* Réaliser un **contrôle de cohérence** entre les champs extraits et les données déclarées
* D’enregistrer des informations liées à la transaction effectuée par les clients
* Bloquer et débloquer des clients selon des critères définis
* Configurer les profils de conformité
* Lister les transactions suspectes, les clients ayant effectué des transactions élevées, etc.
* Vérifier régulièrement l’expiration des pièces justificatives des clients.
* Vérifier périodiquement des clients par rapport à des listes de sactions internationales
* Générer un **rapport de vérification** indiquant l’état (accepté, rejeté, en attente)

### 2.3.2 Besoins non fonctionnels

* **Précision élevée** : le système doit limiter les faux positifs/négatifs
* **Temps de réponse rapide** : le traitement doit être réalisé en quelques secondes
* **Sécurité des données** : conformité RGPD, chiffrement, non-conservation des images
* **Scalabilité** : support d’un grand nombre de demandes simultanées
* **Compatibilité** : web/mobile, format JPEG, PNG ou PDF
* **Maintenable et évolutivité**: le système permettra la mise à jour de ses composants sans interrompre les services pour les utilisateurs finaux
* **Extensibilité :** le système pourra intégrer facilement de nouveau modules via une architecture flexible.
* **Respect des normes KYC :** le système respectera les normes de régulations KYC en vigueur dans les pays où il est déployé, y compris les exigences relatives à la collecte de données et la vérification des informations
* **Conformité GDPR :** Toutes les données personnelles seront collectées stockées et traitées conformément au RGPD (Règlement Général sur la protection des Données).
* **Accessibilité multi-plateformes** : Le système sera accessible via un navigateur web sur desktop et mobile. Des applications mobiles peuvent être développées selon la demande.
* **Interface utilisateur** : L'interface utilisateur sera intuitive, avec un design responsive et accessibles.

Description textuelle des cas d’utilisation

La description textuelle d’un cas d’utilisation sert à détailler le scénario d’interaction entre un acteur (utilisateur ou système externe) et le système. Son but est de clarifier le fonctionnement d’une fonctionnalité en exposant les étapes, les conditions et les résultats attendus. Ci-dessous quelques descriptions textuelle de nos différents cas d’utilisation.

|  |  |
| --- | --- |
| Nom du cas d’utilisation | Demander la création d’un compte |
| Acteurs | Client (Particulier ou Entreprise)  Système |
| Description | Le client demande la création d’un compte en fournissant ces informations et les soumet pour la validation de son compte |
| Précondition | Le client doit disposer d’une connexion Internet.  Le client doit fournir des informations valides (nom, adresse, pièce d’identité, etc.).  Le système KYC doit être opérationnel et capable de stocker les données. |
| Scénario nominal | 1. Le client accède à la plateforme et choisit l’option "Créer un compte". 2. Le système affiche un formulaire d'inscription, demandant : Nom, prénom, Adresse e-mail et numéro de téléphone, Type de client (Particulier ou Entreprise) 3. Le client remplit le formulaire et valide l’inscription. 4. Le système KYC demande les documents requis, en fonction du profil du client : Carte d’identité ou passeport, Justificatif de domicile 5. Le client soumet les documents 6. Le client soumet la demande de création de compte. 7. Le système effectue des vérifications enregistre la demande et notifie l’administrateur. 8. Le système met le client en attente 9. L’administrateur peut accepter ou refuser la demande. 10. Le client est informé de la validation ou du rejet de sa demande. |
| Scénario alternatif | 7.a : Données incorrectes → Si le client saisit des informations invalides, un message d’erreur s’affiche. Retourner à l’étape 2 du scénario nominal  7.b : Documents illisibles ou frauduleux → Une demande de nouveaux documents est envoyée au client. Retourner à l’étape 4 du scénario nominal  9.a : Le client est sur une liste de sanctions → Son compte est temporairement bloqué et il est invité à ce rapproché de la direction générale. Fin de traitement. |
| Postcondition | Cas réussi : Le client dispose d’un compte valide et peut utiliser les services.  Cas rejeté : Le client est informé du refus et ne peut pas poursuivre l’inscription. |

|  |  |
| --- | --- |
| Nom | Valider un compte |
| Acteurs | Administrateur  Responsable client |
| Description | Garantir que seuls les clients légitimes et conformes aux réglementations KYC (Know Your Customer) puissent créer un compte, en validant manuellement en aval les informations et documents fournis après vérification. |
| Précondition | Une demande de création de compte doit avoir été soumise par un client.  Les documents d’identité et justificatifs doivent être disponibles. |
| Scénario nominal | 1. Le responsable client accède à l’interface KYC et choisit l’option « compte en attente ». 2. Le système affiche la liste des comptes en attente. 3. Le responsable sélectionne une demande de création de compte. 4. Le système affiche les informations correspondant au compte sélectionné 5. Le responsable effectue des vérifications manuelles si nécessaire des informations du client, notamment : Identité du client, Documents fournis (pièce d’identité, justificatif de domicile, etc.), Résultats des vérifications automatiques (score de risque, détection de fraude, etc.) 6. Le responsable valide la création de compte client. Il est activé et notifié. 7. Le système enregistre la décision et met à jour le statut du compte client. |
| Scénario alternatif | 5.a : Documents frauduleux détectés → La demande est rejetée et une alerte est envoyée à l’équipe de conformité. Fin du traitement.  5.b : Les documents sont illisibles → Le client est informé et invité à soumettre à nouveau ses documents. La demande reste en attente. Retourné à l’étape 4 du scénario nominal.  5.c Documents incomplets : le client est informé et invité à soumettre des justificatifs additionnels. La demande reste en attente. Retourné à l’étape 4 du scénario nominal  5.d : le client est politiquement exposé ou est à risque : Le responsable client rejette la demande de création de compte pour motif de … Fin de traitement.  6.a : La création de compte est rejetée → le client n’a pas fourni les documents requis dans les délais. Fin de traitement |
| Postcondition | Si le compte est validé le client est notifié  Si la demande est rejetée → Le client est informé et ne peut pas poursuivre l’inscription pour motif de ...  Si des documents supplémentaires sont demandés → La demande reste en attente jusqu’à réception et validation. |

|  |  |
| --- | --- |
| Nom du cas d’utilisation | Enregistrer une transaction |
| Acteurs | Responsable client |
| Description | Permettre aux clients d'effectuer des transactions bancaires sécurisées en agence tout en respectant les règles et procédures de la banque.. |
| Précondition | Le client doit posséder un compte bancaire actif.  Le client doit se présenter en agence avec une pièce d’identité valide. |
| Scénario nominal | 1. Le responsable client accède à l’interface d’enregistrement de transaction et choisit l’option « Enregistrer une transaction ». 2. Le système affiche le formulaire d’enregistrement des transactions. 3. Le client fournit les informations relatives à sa transaction au responsable client (pièce d’identité, etc.). 4. Le responsable client enregistre la transaction dans le système 5. Le système effectue les vérifications nécessaires 6. Si tout est en ordre, le système envoie un message de succès pour la transaction. |
| Scénario alternatif | 3.a : Le client n’a pas de compte : le responsable client recueil les informations de sa pièce d’identité et les informations relatives à sa transaction. Retourner à l’étape 4 du scénario nominal.  5.a : Le client est bloqué : le système informe le responsable que le client est bloqué. Le système annule l’enregistrement de la transaction.  5.b : Transaction suspecte : Le système affiche une erreur informant le responsable client de la transaction suspecte. |
| Postcondition | La transaction est enregistrée |

## 2.4 Architecture fonctionnelle du système

La mise en place d’une application peut se faire suivant plusieurs approches : monolithique ou microservice.

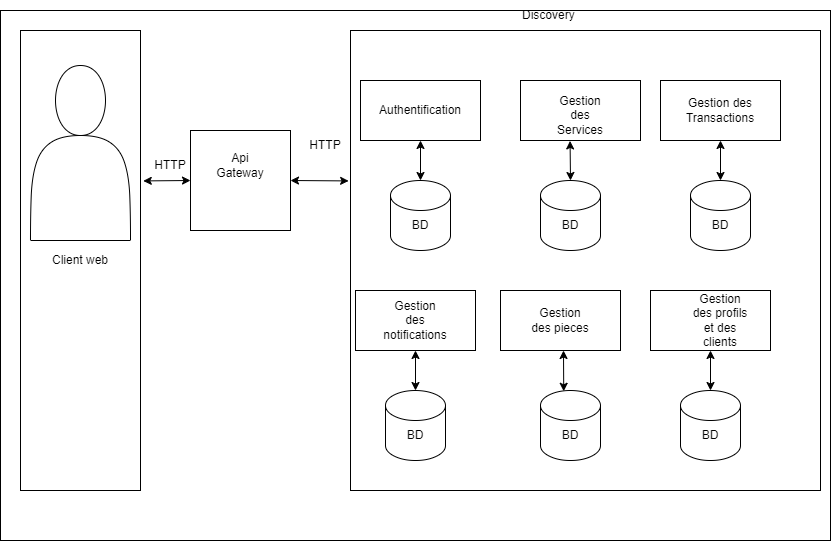
**Tableau comparatif entre une approche micro service et une approche monolithique**

|  |  |  |
| --- | --- | --- |
| Critère | Architecture Monolithique | Architecture Microservice |
| Structure | Une seule application avec tous les modules regroupés | Application divisée en services indépendants |
| Déploiement | Un seul déploiement global | Déploiement indépendant de chaque service |
| Scalabilité | Scalabilité vertical (ajout de ressources à une seule instance) | Scalabilité horizontale (chaque service peut être scale indépendamment) |
| Développement | Plus simple à démarrer et comprendre pour une petites équipes | Plus complex, nécessite une coordination entre services |
| Technologies | Une seule stack technologique commune | Chaque service peut utilizer une technologie différente |
| Tolerance aux pannes | Une panne peut impacter toute l’application | Une panne d’un service n’affecte pas les autres (si bien isolés) |
| Communication entre modules | Appels internes (fonctions/methods) rapides | Communication via APIs (HTTP), plus lente mais plus souple |
| Mise à l’échelle des équipes | Moins adapté aux équipes large | Adaptés aux grandes équipes (chaque équipe gère un ou plusieurs services) |
| Dépendances | Les modules sont fortement couples | Faible couplage entres les services |

En ce qui nous concerne, nous avons choisi une approche par micro-service (Architecture micro-service) au détriment d’une approche monolithique. De ce fait, les micros services suivants en découlent :

* Authentification
* Gestion des services
* Gestion des transactions
* Gestion des notifications
* Gestion des profils et des clients
* Gestion des pièces

Les interactions entre ces microservices se matérialisent par l’architecture ci-dessous :



Ce schéma illustre une architecture microservices typique, où une application est décomposée en plusieurs petits services indépendants qui communiquent entre eux. Voici les principaux composants et leur rôle :

* **Client Web :** C'est l'interface utilisateur à laquelle l'utilisateur final interagit (par exemple, un navigateur web ou une application mobile).
* **API Gateway :** L'API Gateway agit comme un point d'entrée unique pour toutes les requêtes des clients. Elle achemine les requêtes vers les microservices appropriés, effectue l'authentification, l'autorisation et peut également gérer la traduction de protocole et l'agrégation de réponses.
* **Discovery :** C'est un composant qui permet aux microservices de s'enregistrer et de se découvrir mutuellement. Lorsqu'un microservice démarre, il s'enregistre auprès du service de découverte, qui conserve une liste à jour de tous les microservices disponibles. Lorsque un microservice a besoin de communiquer avec un autre, il interroge le service de découverte pour obtenir l'adresse du microservice cible.
* **Microservices :** Ce sont les composants principaux de l'architecture. Chaque microservice est responsable d'une fonctionnalité spécifique de l'application. Dans ce schéma, on voit des microservices pour:
  + **Authentification :** Gère l'authentification des utilisateurs.
  + **Gestion des Services :** Gère les opérations liées aux services offerts par l'application.
  + **Gestion des Transactions :** Gère les transactions financières ou autres transactions importantes.
  + **Gestion des Notifications :** Gère l'envoi de notifications aux utilisateurs.
  + **Gestion des Pièces :** Gère les informations relatives aux pièces (peut-être des pièces détachées, des documents, etc.).
  + **Gestion des Profils et des Clients :** Gère les informations des profils utilisateurs et des clients.
* **Bases de Données (BD) :** Chaque microservice possède sa propre base de données. Cela permet à chaque service de choisir la technologie de base de données la plus appropriée à ses besoins et assure une isolation des données entre les services.

**Fonctionnement général :**

1. Le client web envoie une requête HTTP à l'API Gateway.
2. L'API Gateway reçoit la requête et la route vers le microservice approprié, en utilisant potentiellement le service de découverte pour localiser le microservice.
3. Le microservice traite la requête et accède à sa propre base de données si nécessaire.
4. Le microservice renvoie une réponse à l'API Gateway.

L'API Gateway peut effectuer des transformations sur la réponse et la renvoie au client web.

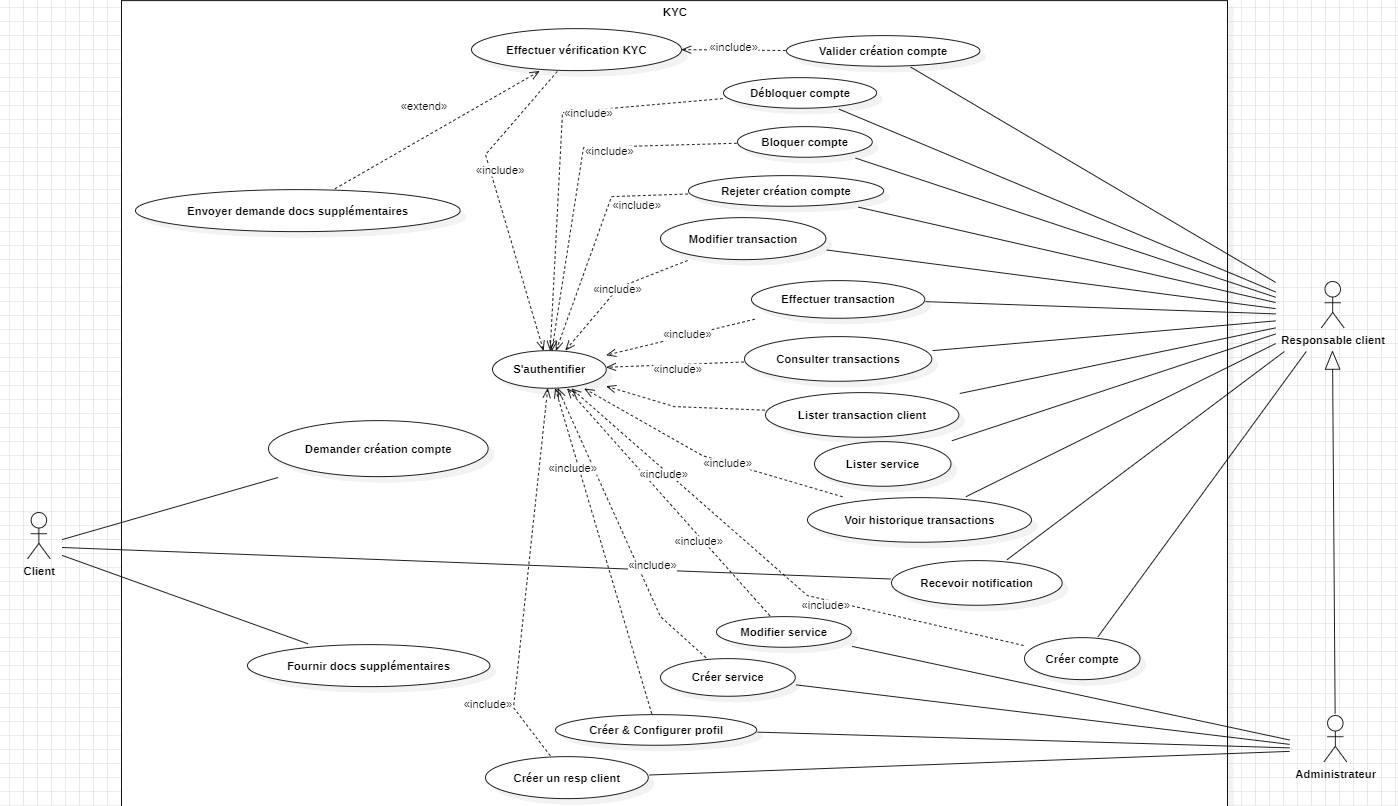
METHODE DE MODELISATION

UML (Unified Modeling Language) est un langage de modélisation standardisé qui permet de représenter visuellement les systèmes logiciels, leurs structures, comportements et interactions, en unifiant les notations de différentes méthodologies orientées objet. Il est utilisé pour clarifier et communiquer des concepts complexes, faciliter la planification et la conception, réduire les erreurs et les coûts de développement, et fournir une documentation visuelle pour la maintenance et l'évolution des systèmes. UML offre une approche structurée et flexible, compatible avec une variété d'outils et de technologies, ce qui en fait un outil essentiel pour optimiser les interactions entre composants, identifier les redondances et améliorer la performance globale des systèmes logiciels tout en améliorant la collaboration entre développeurs, architectes et parties prenantes. Elle diffère des SI (Système d’Information) qui est un ensemble organisé de ressource (humaines, matérielles, logicielles, données, procédures) qui permet de collecter, stocker, traiter et diffuser l’information au sein d’une organisation. Son objectif principal est de supporter les processus métier d’une organisation (gestion, communication, décision)

DIFFERENCES ENTRE UML ET

AJOUTER LES DIAGRAMMES ICI

Diagramme de cas d’utilisation



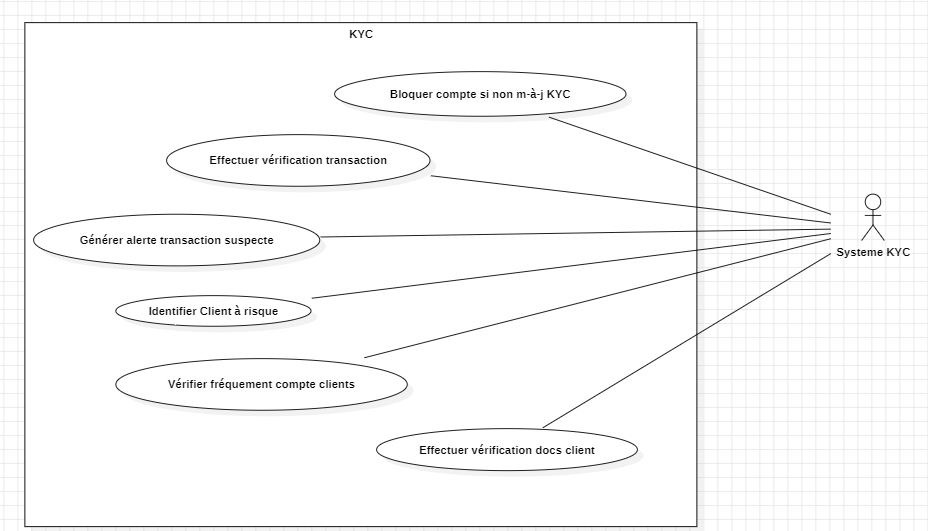


Diagramme de séquence

2.5 Choix technologique

|  |  |
| --- | --- |
| Composant | Technologie envisagée |
| OCR | Tesseract OCR + OpenCV |
| Reconnaissance faciale | DeepFace, FaceNet |
| Modèles IA (images) | CNN avec TensorFlow |
| Backend | Python (FastAPI), Java (Spring Boot) |
| Frontend (interface) | Angular |
| Base de données | MYSQL |
| Déploiement | … |
| Jira | Pour la gestion de projet |
| Versioning | Git |
| Editeur | Vs Code, Intellij IDEA |
| Base de données local | WAMP/ XAMPP Server |

**2.6 Contraintes et limites**

* **Qualité des documents** : images floues ou mal cadrées peuvent affecter la précision de l’OCR
* **Langue et format des pièces** : le système doit s’adapter à différents formats et langues
* **Détection de fraude complexe** : les techniques de falsification sophistiquées nécessitent un entraînement avancé
* **Respect des réglementations** : obligation de minimiser la conservation des données personnelles

**2.7 Conclusion**

Ce chapitre a permis de définir les besoins fonctionnels et non fonctionnels du système KYC intelligent, ainsi que son architecture globale et les choix techniques qui guideront son développement. Ces éléments servent de fondation pour la phase de conception et d’implémentation détaillée, qui sera développée dans le chapitre suivant.

# **Chapitre 3 : Conception et implémentation**

## 3.1 Introduction

Ce chapitre décrit de manière détaillée la **conception technique** et l’**implémentation** du système intelligent de vérification KYC. Il s’agit de traduire les besoins identifiés au chapitre précédent en une solution logicielle concrète. Cette phase comprend la définition de l’architecture logicielle, la conception des différents modules (OCR, reconnaissance faciale, analyse des données), l’intégration des composants IA, ainsi que les aspects liés à la sécurité et à la protection des données.

## 3.2 Architecture logicielle du système

Le système est basé sur une architecture **à microservice** afin de faciliter la maintenance, les mises à jour et l’évolutivité. Il se compose des modules suivants :

### 3.2.1 Schéma global de l’architecture

[Interface Utilisateur]

|

v

[Upload Document / Selfie]

|

v

[Prétraitement d’image]

|

v

[OCR] ----> [Extraction texte]

|

v

[Reconnaissance faciale] <---- [Selfie]

|

v

[Analyse / Vérification IA]

|

v

[Résultat / Rapport]

## 3.3 Conception des modules

### 3.3.1 Module d’acquisition de documents

Ce module permet à l’utilisateur de **téléverser une image** d’un document officiel. Il accepte les formats PNG, JPEG et PDF. Un redimensionnement automatique est effectué pour standardiser les entrées.

### 3.3.2 Module de prétraitement d’image

Le prétraitement améliore la qualité des images avant leur analyse :

* **Binarisation / seuillage adaptatif** (OpenCV)
* **Redressement automatique** de document (détection de contours)
* **Correction du bruit / flou**

### 3.3.3 Moteur OCR

Le moteur OCR repose sur **Tesseract**, entraîné avec les langues appropriées (français, anglais…). Il est utilisé pour :

* Extraire les champs importants : Nom, Prénom, Date de naissance, Numéro de document
* Détecter la zone MRZ (pour les passeports)

Un post-traitement NLP (regex, validation de formats) nettoie et structure les résultats.

### 3.3.4 Reconnaissance faciale

Le module utilise une **bibliothèque de reconnaissance faciale** comme **DeepFace** (ou FaceNet) pour :

* Extraire les **vecteurs d’encodage facial**
* Comparer la **photo sur la pièce** au **selfie utilisateur**
* Calculer un **score de similarité** pour valider l’identité

Une **tolérance ajustable** permet de calibrer la sensibilité du système.

### 3.3.5 Moteur de vérification IA

Ce module applique plusieurs contrôles :

* **Cohérence interne** des données extraites
* **Conformité des formats** (numéro, dates, etc.)
* **Détection de falsification documentaire** (en version avancée, avec un CNN entraîné à reconnaître des faux)

### 3.3.6 Génération de rapport

Le système produit un **rapport de vérification clair** indiquant :

* Informations extraites
* Score de correspondance faciale
* Résultat final : Accepté / Rejeté / Incomplet
* Messages d’erreurs ou de recommandations

DIAGRAMME UTILISER

## 3.4 Technologies utilisées

| **Fonction** | **Outil / Technologie** |
| --- | --- |
| OCR | Tesseract OCR + pytesseract |
| Traitement d’image | OpenCV |
| Reconnaissance faciale | DeepFace / FaceNet |
| Modèle IA | TensorFlow (CNN pour détection avancée) |
| Backend API | Python (FastAPI) |
| Frontend | HTML/JS ou React (selon le contexte) |
| Base de données | SQLite (pour démonstration locale) |
| Déploiement | Docker (optionnel) |

**3.5 Sécurité et gestion des données**

Le traitement de documents d’identité impose des mesures strictes :

* **Chiffrement des fichiers en mémoire**
* **Non-conservation permanente** des images après traitement
* Respect des principes RGPD : minimisation, consentement, finalité claire
* Possibilité d’ajouter une couche d’**authentification utilisateur**

**3.6 Difficultés rencontrées**

Quelques contraintes techniques ont été observées :

* Sensibilité de l’OCR à la qualité d’image
* Traitement des documents multilingues
* Gestion des cas particuliers (documents partiellement lisibles, selfie flou)
* Entraînement du modèle de détection de fraude (dépendant de la base de données disponible)

**3.7 Conclusion**

Cette phase de conception et d’implémentation a permis de mettre en place un **prototype fonctionnel de vérification KYC automatisée**, exploitant efficacement l’IA et l’OCR. La modularité du système offre une base solide pour des améliorations futures, comme l’intégration de détection de fraude plus avancée ou l’optimisation des performances sur mobile.

# **Chapitre 4 : Résultats et tests**

## 4.1 Introduction

Après la phase de conception et d’implémentation du système de vérification KYC, il est indispensable d’évaluer son bon fonctionnement à travers une série de **tests fonctionnels, techniques et de performance**. Ce chapitre présente la méthodologie de test adoptée, les différents scénarios testés, les résultats obtenus, ainsi qu'une analyse des performances et des limites du système.

## 4.2 Méthodologie de test

### 4.2.1 Objectifs des tests

Les tests visent à :

* Vérifier que le système respecte les **fonctionnalités attendues**
* Évaluer la **précision de l’extraction OCR**
* Mesurer la **fiabilité de la reconnaissance faciale**
* Tester la **robustesse du système** face à des entrées dégradées (images floues, documents partiels)
* Valider les **temps de traitement moyens**

### 4.2.2 Environnement de test

* **Machine** : PC avec 16 Go RAM, CPU i7, GPU intégré
* **Système** : Linux Ubuntu 22.04 / Windows 11
* **Données** : Ensemble de 50 documents d’identité (CNI, passeports) avec selfies associés, en français et anglais

## 4.3 Tests fonctionnels

| **Fonction testée** | **Résultat attendu** | **Résultat obtenu** | **Statut** |
| --- | --- | --- | --- |
| Upload de documents et selfie | Chargement sans erreur | Fonctionne | ✅ |
| Extraction OCR des données clés | Précision > 85% | Moyenne : 88% | ✅ |
| Vérification de la validité du format | Rejet des formats non conformes | Fonctionne | ✅ |
| Reconnaissance faciale | Similarité > 90% si photo et selfie vrais | Moyenne : 93% | ✅ |
| Rapport final | Résumé clair du résultat | Généré avec succès | ✅ |

## 4.4 Tests de performance

| **Critère mesuré** | **Valeur moyenne mesurée** |
| --- | --- |
| Temps d’extraction OCR | ~ 1.8 secondes |
| Temps de reconnaissance faciale | ~ 1.5 secondes |
| Temps total de traitement | ~ 3.5 secondes |
| Taux de succès global | 94% |

**4.5 Tests de robustesse**

Des tests ont été menés avec des documents de qualité inférieure :

* **Images floues ou inclinées** : l’OCR perd ~30% de précision
* **Photos partiellement obscurcies** : rejet automatique
* **Selfies dans l’obscurité** : reconnaissance faciale impactée

Des traitements supplémentaires (ex. : amélioration automatique d’image, détection de flou) pourraient améliorer ces résultats.

**4.6 Limites du système**

Malgré des résultats satisfaisants, le système présente certaines **limites** :

* Sensibilité à la qualité de l’image (OCR et reconnaissance faciale)
* Difficulté à traiter certains formats spécifiques (ex. : documents à double face)
* Absence de vérification avancée de falsification (nécessite un modèle entraîné)
* Modèle de reconnaissance faciale dépendant des conditions de lumière et d’angle

**4.7 Perspectives d’amélioration**

Plusieurs pistes peuvent être envisagées :

* Intégration de modèles de **correction d’image** basés sur l’IA
* Création d’une **base de données d’entraînement** pour la détection de faux documents
* Ajout de modules de **preuve de vie (liveness detection)** en temps réel
* Déploiement sur **serveur cloud sécurisé** avec gestion des logs

**4.8 Conclusion**

Les tests réalisés ont démontré que le système est capable de **vérifier automatiquement des documents d’identité** avec un bon niveau de fiabilité, en combinant efficacement l’OCR et la reconnaissance faciale. Malgré quelques limites techniques liées à la qualité des documents fournis, le prototype constitue une **base fonctionnelle solide** pour un système de vérification KYC intelligent, rapide et sécurisé.

# **Conclusion générale**

Dans un contexte où les services numériques exigent une vérification d'identité rapide, fiable et sécurisée, ce travail a permis de concevoir et de mettre en œuvre un système intelligent de vérification KYC (Know Your Customer) basé sur les technologies d’**intelligence artificielle (IA)** et de **reconnaissance optique de caractères (OCR)**. L’objectif principal était d’automatiser le processus de contrôle des pièces d’identité tout en garantissant la précision des informations extraites et la conformité aux exigences réglementaires.

Le rapport a présenté de manière structurée les différentes étapes du projet :

* L’analyse du besoin et l’étude de l’état de l’art,
* La modélisation fonctionnelle et l’architecture technique du système,
* L’implémentation des modules principaux (OCR, reconnaissance faciale, vérification IA),
* Et enfin, les résultats issus des tests fonctionnels et de performance.

Les résultats obtenus montrent que le système est capable de traiter efficacement des documents variés, avec un **taux de réussite global de 94%** sur les scénarios standards. La combinaison de Tesseract OCR, des techniques de traitement d’image, et d’un modèle de reconnaissance faciale comme DeepFace permet d’atteindre un bon compromis entre rapidité, précision et sécurité.

Toutefois, le système reste perfectible sur certains aspects, notamment la gestion des documents de mauvaise qualité, la détection de falsification avancée, et la robustesse dans des conditions réelles d’utilisation. Ces points constituent des perspectives d’amélioration et d’extension future du projet.

En conclusion, ce projet illustre concrètement comment les technologies d’IA peuvent transformer les processus KYC traditionnels, en les rendant plus **automatisés**, **scalables** et **fiables**, tout en respectant les principes de sécurité et de confidentialité des données personnelles.

# **Annexes**

## Annexe A : Exemple de rapport généré par le système

Nom : DUPONT

Prénom : Alice

Date de naissance : 12/03/1994

Type de document : Passeport

N° de document : XX1234567

Score de reconnaissance faciale : 95.3%

Résultat : ✅ Vérification validée

Commentaire : Document authentique, photo conforme

Annexe B : Extrait de code - Traitement OCR

import pytesseract

import cv2

image = cv2.imread('document.jpg')

gray = cv2.cvtColor(image, cv2.COLOR\_BGR2GRAY)

text = pytesseract.image\_to\_string(gray, lang='fra+eng')

print(text)

## Annexe C : Données de test (50 documents anonymisés)

* 30 cartes nationales d’identité
* 10 passeports
* 10 permis de conduire
* 50 selfies associés
* Documents en français et en anglais

# **Bibliographie**

1. **Tesseract OCR** – <https://github.com/tesseract-ocr/tesseract>
2. **DeepFace** – <https://github.com/serengil/deepface>
3. KYC & Compliance: A practical guide – Finextra Research, 2020
4. ISO/IEC 19794-5:2011 – Biometric data interchange formats – Face image data
5. OpenCV Documentation – https://docs.opencv.org
6. RGPD (Règlement Général sur la Protection des Données) – <https://eur-lex.europa.eu>
7. Goodfellow et al., Deep Learning, MIT Press, 2016
8. Scikit-learn Documentation – <https://scikit-learn.org>

<https://www.kaggle.com/datasets/ealaxi/paysim1?resource=download>

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/data>