

PROJET IA UniHA : Intelligence Artificielle Conversationnelle Hyperspécialisée sur les Marchés UniHA

-

Expression du Besoin

Introduction

1. Contexte du projet

1.1 Rappel sur les outils métier UniHA

Le système d'information actuel d'UniHA s'articule autour de quatre outils métiers majeurs :

- ERP et CRM : basé sur Odoo 19 (en cours de déploiement)
- HERMES : Système de messagerie des marchés UniHA – Suivi des ruptures
- QUANTUM : Collecte des besoins pour les recensements avant publication des marchés
- ESPACE DOCUMENTAIRE : Espace de stockage des documents de marché UniHA

1.2 Volumétrie des Données

Le patrimoine informationnel d'UniHA représente un volume considérable nécessitant une expertise pointue pour son exploitation optimale :

- 12 000 comptes utilisateurs actifs sur le portail adhérents
- 1 400 établissements adhérents avec gestion des droits multi-niveaux
- 4 000 procédures actives (appels d'offres, marchés en cours)
- 15 000 produits référencés avec tarification dynamique
- 300 000 documents de marchés (pièces administratives, cahiers des charges, contrats)
- 1 800 fournisseurs avec coordonnées et historiques contractuels
- Rafraîchissement des données plusieurs fois par jour

1.3 Problématique Identifiée

Les adhérents et collaborateurs UniHA font face à plusieurs difficultés dans l'exploitation quotidienne de ce patrimoine informationnel :

- Complexité d'accès : Recherche manuelle chronophage dans les 4000 procédures et 300 000 documents
- Dispersion des sources : Informations réparties entre SQL Server (marchés), Odoo (CRM), espace documentaire et site web
- Hétérogénéité des profils : Besoins différenciés selon les rôles (acheteurs, comptables, pharmaciens, directeurs)
- Manque de proactivité : Absence de notifications intelligentes sur encours, ruptures ou recensements à venir
- Barrière technique : Nécessité de maîtriser plusieurs interfaces pour accéder aux informations complètes

1.4 Objectifs du projet

Le projet IA UniHA vise à développer une intelligence artificielle conversationnelle hyperspécialisée, accessible en français via interface textuelle et vocale, permettant aux utilisateurs d'interroger l'ensemble du patrimoine informationnel UniHA de manière intuitive, sécurisée et contextualisée selon leurs droits d'accès.

Le projet devra répondre aux exigences suivantes :

- Respect de la réglementation française et européenne (notamment le RGPD et le règlement européen sur l'IA) ;
- Architecture RAG (Retrieval-Augmented Generation) avec enrichissement quotidien automatisé
- Sécurité multi-niveaux basée sur l'authentification OAuth2 Microsoft et les adhésions établissements rattachées aux utilisateurs
- Intégration transparente via iframe sur portail adhérents et site public
- Capacités conversationnelles textuelles et vocales (entrée/sortie)
- Téléchargement intelligent de documents selon les adhésions et droits utilisateurs

2. Cadre Réglementaire

2.1 Conformité au Règlement Européen sur l'intelligence artificielle

Le projet IA UniHA devra respecter les exigences prévues par le règlement européen sur l'intelligence artificielle, selon le calendrier d'entrée en application progressive du règlement.

L'applicabilité et le niveau d'exigence dépendront de la qualification du système au regard des catégories de risque prévues par le règlement, notamment en ce qui concerne une éventuelle classification en système d'IA à haut risque.

Le cas échéant, les obligations suivantes pourront s'appliquer :

- **Mise en place d'un système de gestion de risque** : certaines fonctionnalités du système présentent en effet un risque (accès aux données sensibles fournisseurs et marchés publics).
- **Transparence** : Information claire aux utilisateurs qu'ils interagissent avec une IA
- **Traçabilité** : Logs complets des requêtes et réponses pour audit
- **Supervision humaine** : Mécanisme d'escalade vers collaborateurs UniHA pour validation
- **Qualité des données** : Garantie de véracité et actualité des informations fournies
- **Documentation technique** : Architecture, données d'entraînement, procédures de test
- **Évaluation des risques** : AIPD (Analyse d'Impact Protection des Données) préalable au déploiement

2.2 Sécurité et traçabilité du modèle et des données

En tant qu'entité publique française gérant des données stratégiques (marchés publics, fournisseurs hospitaliers, coordonnées contacts établissements), UniHA privilégie les solutions permettant de privilégier la sécurité et la traçabilité des données et les solutions reposant, lorsque cela est pertinent, sur des technologies open source.

Modèles d'IA :

- Priorité aux LLM open source (ex. solutions Mistral AI, Mixtral)
- Weights du modèle accessible et auditable
- Possibilité de fine-tuning sur corpus UniHA sans transfert vers tiers

Infrastructure d'hébergement :

- Hébergement sur l'infrastructure UniHA (Ajout de VCPU ou serveur dédié)
- Hébergement des données au sein de l'Union européenne

APIs et services :

- Stack technique open source privilégiée (par exemple OpenRAG, LangChain, PGVector...)

- Intégration OAuth2 via Microsoft Graph autorisée (déjà utilisé pour authentification O365 UniHA)
- Chiffrement bout-en-bout des échanges (le cas échéant avec certificats ANSSI)

2.3 Protection des Données Personnelles (RGPD)

Le projet IA UniHA traitant des données personnelles (utilisateurs, contacts fournisseurs) devra respecter le RGPD et notamment les principes suivants :

- **Minimisation** : Accès données strictement limité aux droits utilisateurs
- **Durée conservation** : Logs IA conservés 12 mois, données métiers selon durée légale marchés publics
- **Droits des personnes** : Mécanisme d'accès, rectification, effacement intégré au portail adhérents

3. Organisation du Projet

3.1 Pilotage et Gouvernance

Comité de pilotage (COPIL) :

- Direction des Systèmes d'Information (DSI) : pilotage projet
- Direction des Achats : validation besoins métiers et cas d'usage
- Direction Commerciale : spécifications accès adhérents et non-adhérents
- Délégué à la Protection des Données (DPO) : conformité RGPD
- Direction Générale : arbitrages stratégiques et budgétaires

Fréquence COPIL : Bimensuelle durant phase projet (3 mois), mensuelle en phase run.

Comité technique (COTECH) :

- Chef de projet
- Architecte SI et DSI
- Administrateurs bases de données (SQL Server, Odoo)
- Prestataire retenu développeurs / intégrateurs IA
- Prestataire retenu (RAG/LLM)
- Prestataire retenu : Product Owner
- Représentants métiers (acheteurs, adhérents, pharmaciens ...)

Fréquence COTECH : Hebdomadaire durant développement.

3.2 Phasage du Projet

Phase	Livrables
Phase 1 : Cadrage	Expression besoin, AIPD, choix LLM
Phase 2 : POC	RAG minimal (1000 docs), tests sécurité
Phase 3 : Développement	Intégration complète, iframe, vocal
Phase 4 : Tests	Recette UAT avec 50 utilisateurs pilotes
Phase 5 : Déploiement	Mise en production progressive
Phase 6 : Run et amélioration continue	Monitoring, enrichissement modèle

Table 1: exemple de planning projet IA UniHA

Jalons décisionnels (GO/NO-GO) :

- Fin POC : Validation précision RAG >85% sur jeu test marchés
- Fin développement : Validation tests sécurité intrusion et charge
- Avant production : Homologation RSSI et validation DSI et validation DPO

3.3 Équipe Projet

Ressources internes UniHA :

- Chef de projet
- DSI
- Architecte SI
- Administrateurs BDD/Odoo
- Référents métiers (achats/compta/pharmacie)

Ressources externes (prestataire RAG/LLM) avec les types de profils suivants :

- Architecte IA / Data Scientist
- Développeurs IA (Python, LangChain, APIs)
- Expert sécurité IA
- Product Owner

4. Architecture Technique et Sécurité

4.1 Architecture RAG (Retrieval-Augmented Generation)

L'architecture IA UniHA reposera sur une approche RAG permettant d'ancrer les réponses du LLM dans les données réelles d'UniHA, éliminant les hallucinations et garantissant l'actualité des informations..

Composants de l'architecture RAG :

Données à titre d'exemple, le candidat est libre de proposer sa propre intégration du moment qu'elle respecte les attendus en termes de charge, scalabilité, sécurité, gouvernance et hébergement des données dans l'UE. .

SCHEMA INDICATIF D'UNE ARCHITECTURE CIBLE

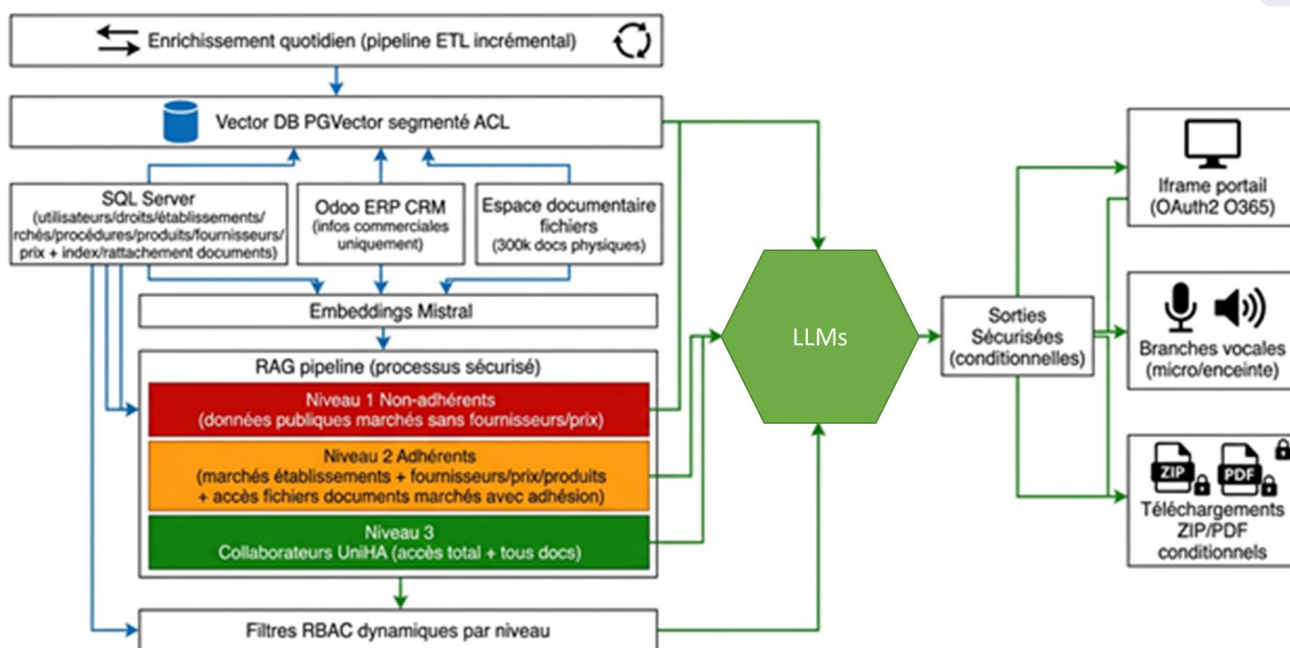


Figure 1: Architecture RAG IA UniHA avec segmentation sécurité

1. Sources de données multiples :

- Base SQL Server : Utilisateurs, établissements, adhésions, marchés (filières, coordinations, procédures, lots, fournisseurs, articles, prix), index documentaire (chemins, noms fichiers, types, rattachements marchés)
- (Voir pertinence) ERP Odoo 19 (AGORA) : CRM fournisseurs, leads commerciaux, factures, stock de produits, disponibilité, états des commandes et livraisons.
- Espace documentaire : 300 000 fichiers PDF/Office (cahiers charges, contrats, pièces administratives)
- Site web UniHA : Informations publiques (organisation, contacts, actualités, guides adhésion)

2. Pipeline ETL et chunking :

- Extraction quotidienne incrémentale (delta uniquement) via connecteurs SQL/Odoo/fichiers
- Chunking intelligent : documents découpés en segments 500-1000 tokens avec overlap 100 tokens
- Métadonnées enrichies : id_marché, id_lot, id_établissement_adhérent, niveau_sécurité, date_validité etc ...

3. Embeddings et base vectorielle :

- Modèle embedding multilingue français

- Base vectorielle: type PGVector (extension PostgreSQL) ou Weaviate auto-hébergé[11]
- Index HNSW pour recherche rapide (<100ms) sur 500k chunks
- Segmentation collections selon niveau sécurité (voir section 4.2)

4. **LLM et génération :**

- Modèle de base : hébergé au sein de l'Union européenne
- Fine-tuning sur 10 000 documents marchés UniHA + FAQ métiers pour jargon spécialisé pour le PC et les tests
- Prompt engineering : Instructions système incluant périmètre UniHA, ton professionnel, limitation réponses aux seules données RAG
- Génération augmentée : Top-5 chunks similaires injectés dans contexte LLM (fenêtre 32k tokens)

5. **Post-traitement et vérification :**

- Filtrage réponses selon droits utilisateur (vérification ACL post-génération)
- Détection hallucinations : Score de confiance RAG, citation sources obligatoire
- Logging complet : Requête, chunks récupérés, réponse, durée, utilisateur (conformité AI Act)

Flux conversationnel type :

1. Utilisateur authentifié pose question vocale/textuelle : "Quels sont les fournisseurs de compresses pour mon établissement ?"
2. Récupération profil utilisateur (niveau sécurité, établissements rattachés, adhésions actives) via API OAuth2
3. Embedding de la question en vecteur
4. Recherche similarité dans base vectorielle avec filtres métadonnées (adhésions , droit utilisateurs utilisateur etc ...)
5. Récupération top-5 chunks pertinents (ex. lots compresses, fournisseurs titulaires, prix, articles ...)
6. Injection chunks + question + instructions système dans prompt LLM
7. Génération réponse par Mistral avec citations sources (id documents/marchés)
8. Post-filtrage : Vérification droits accès sur fournisseurs/prix mentionnés
9. Restitution vocale si demandée ou textuelle avec liens téléchargement documents autorisés
10. Logging requête/réponse pour audit

4.2 Sécurité Multi-Niveaux et Contrôle d'Accès

IA UniHA doit implémenter trois niveaux de sécurité distincts, matérialisés par des collections vectorielles segmentées et des filtres dynamiques sur métadonnées.

Niveau 1 : Non-adhérents (accès public)

- **Périmètre autorisé** : Informations publiques UniHA (organisation, contacts, filières marchés, procédures actives sans détails fournisseurs/prix, pas d'accès aux pièces de marchés)
- **Données accessibles** :
 - Site web UniHA (crawl quotidien)
 - Nomenclature marchés (filières, coordinations, familles produits)
 - Fiches marchés
 - Liste des procédures, de leurs lots et documents publics
 - Guides adhésion et modalités contractuelles
- **Implémentation RAG** : Collection vectorielle "public" sans embeddings prix/fournisseurs
- **Filtrage post-génération** : Suppression automatique de toute mention fournisseur ou tarif ou information de pièces de marchés ou liens vers celles ci si présentes (regex + NER)
- **Exemple requête autorisée** : "Quelles sont les filières marchés disponibles chez UniHA ?"
- **Exemple requête bloquée** : "Quel est le prix des compresses ?" → Réponse : "Information réservée aux adhérents. Contactez contact@uniha.org"

Niveau 2 : Adhérents (accès conditionné par adhésions établissements)

- **Périmètre autorisé** : Ensemble des données marchés auxquels les établissements de rattachement de l'utilisateur ont souscrit une adhésion active
- **Données accessibles** :
 - Marchés adhérents : Procédures, lots, marchés
 - Fournisseurs titulaires et coordonnées complètes
 - Produits et tarifs adhérents (prix négociés)
 - Documents marchés (contrats, avenants, pièces administratives) pour marchés adhérents
 - Encours établissement sur QUANTUM et CA : Recensements à compléter, conventions à signer
 - Ruptures produits et tensions approvisionnement (si profil pharmacien)
 - Factures en attente (si profil comptable)
- **Implémentation RAG** :
 - Collection vectorielle "adhérents" avec métadonnées

- **Téléchargement documents :**
 - Génération liens temporisés (validité 15 minutes) vers API documentaire
 - Vérification droits côté API : Utilisateur → Établissements → Adhésions → Documents autorisés
 - Formats proposés : ZIP (ensemble documents marché) ou fichier individuel PDF/Office
 - Watermarking automatique : Nom utilisateur + établissement + date sur PDF générés
- **Exemple requête autorisée :** "Liste des fournisseurs de médicaments pour l'hôpital de Lyon" (si utilisateur rattaché CHU Lyon avec adhésion marché médicaments)
- **Exemple requête bloquée :** "Prix des PCs portables disponibles pour hôpital de Marseille" (si utilisateur non rattaché à cet établissement ou si pas d'adhésions) → Réponse : "Vous n'avez pas accès aux données de cet établissement ou n'êtes pas adhérent"

Niveau 3 : Collaborateurs UniHA (accès total)

- **Périmètre autorisé :** Accès illimité à l'ensemble des données UniHA (tous marchés, tous fournisseurs, tous établissements, toutes adhésions, CRM complet Odoo)
- **Données accessibles :** Totalité base SQL Server + Odoo + documentaire (300k fichiers)
- **Implémentation RAG :** Collection vectorielle "admin" sans filtres restrictifs
- **Fonctionnalités supplémentaires :**
 - Analytics marchés : Statistiques achats par filière, taux adhésion, CA fournisseurs
 - CRM avancé : Leads commerciaux, relances recensements, KPI adhérents
 - Exports massifs : Possibilité de demander exports CSV/Excel via IA
- **Authentification renforcée :** MFA obligatoire (double authentification) + restriction IP bureau UniHA
- **Exemple requête :** "Liste complète des établissements n'ayant pas adhéré au marché imagerie avec montant potentiel" (pour action commerciale ciblée)

Mécanisme d'authentification OAuth2 et attribution niveau sécurité :

1. Utilisateur accède portail adhérents UniHA ou site public
2. Si non authentifié : Niveau 1 par défaut (non-adhérent)
3. Si authentification OAuth2 O365 réussie :
 - Appel Microsoft Graph API : Récupération email, groupes AD, attributs personnalisés

- Requête BDD UniHA : Vérification existence compte utilisateur (table Users)
 - Si email domaine @uniha.org, non précédé par 'EXT-' et 2FA activé : Attribution Niveau 3 (collaborateur)
 - Sinon : Récupération établissements rattachés (table User_Etablissement) → Attribution Niveau 2 (adhérent) avec liste établissements en contexte session
4. Token JWT signé stocké côté client (cookie httpOnly) avec claims : user_id, niveau_sécurité, établissements_autorisés, expiration (4h)
 5. Chaque requête inclut token JWT validé côté backend avant exécution RAG

4.3 Sécurité Infrastructure et Données

Chiffrement :

- TLS 1.3 pour tous échanges HTTPS (API, iframe)
- Chiffrement base vectorielle au repos (AES-256)
- Secrets (API keys, tokens OAuth2) dans vault chiffré (ex. HashiCorp Vault)

Isolation réseau :

- IA UniHA hébergé sur VLAN dédié sans accès internet sortant direct
- Reverse proxy applicatif délivrant le contenu au iframe publique (protection DDoS, injections)
- Accès bases SQL Server/Odoo via API HTTPS ou VLAN chiffrés

Logs et audit :

- Logs applicatifs : Toutes requêtes utilisateurs avec horodatage, user_id, question, réponse, chunks utilisés, durée (rétention 12 mois)
- Logs sécurité : Tentatives accès non autorisés, échecs authentification, téléchargements documents (rétention 5 ans pour marchés publics)
- SIEM : Remontée alertes temps réel vers SOC UniHA (tentatives énumération fournisseurs, requêtes massives suspectes)

Tests sécurité :

- Pentests externes avant production (injection prompts, énumération données, contournement ACL)
- Red team interne : Tentatives escalade privilèges entre niveaux sécurité
- Fuzzing : Tests robustesse parsing documents (corruption PDF, ZIP bombs)

5. Enrichissement Quotidien du RAG

5.1 Stratégie de Mise à Jour Incrémentale

Les données UniHA étant rafraîchies plusieurs fois par jour (nouveaux marchés, modifications prix, ajouts documents), l'IA UniHA doit synchroniser sa base vectorielle en quasi-temps réel sans réindexations complète (coûteuse en ressources).

Proposition d'architecture pipeline enrichissement, le candidat est libre de proposer sa propre séquence :

1. **Détection changements (CDC - Change Data Capture) :**
 - SQL Server : Triggers sur tables critiques (Marchés, Lots, Fournisseurs, Prix, Documents) loggant modifications dans table audit_changes (id, table, operation, timestamp) ou utilisation des champs 'Date_modification' ou 'date_insertion' des tables UniHA.
 - Odoo : Webhooks configurés sur modèles CRM (leads, factures, recensements) appelant API IA UniHA lors create/update/delete
 - Documentaire : Triggers sur tables critiques (Marchés, Lots, Fournisseurs, Prix, Documents) loggant modifications dans table audit_changes (id, table, operation, timestamp) ou utilisation des champs 'Date_modification' ou 'date_insertion' des tables UniHA
2. **Orchestration ETL temps réel (Apache Airflow ou équivalent) :**
 - DAG "sync_incremental" exécuté toutes les heures (ou trigger événementiel via webhooks)
 - Lecture table audit_changes depuis dernier checkpoint (last_sync_timestamp)
 - Extraction lignes modifiées uniquement via requêtes SQL filtrées (WHERE date_modification > last_sync)
 - Extraction nouveaux fichiers documentaire (liste fichiers avec timestamp > last_sync)
3. **Traitement et chunking différentiel :**
 - Documents modifiés : Rechunking complet du document, suppression anciens chunks en base vectorielle (DELETE WHERE doc_id = X), insertion nouveaux chunks
 - Nouveaux documents : Chunking initial, insertion base vectorielle
 - Métadonnées BDD (marchés, prix) : Update métadonnées chunks existants (ex. mise à jour prix dans metadata sans rechunking si description produit inchangée)
4. **Ré-embedding et indexation :**
 - Génération embeddings pour nouveaux chunks uniquement (économie calcul)

- Insertion base vectorielle PGVector avec métadonnées actualisées
- Rebuild partiel index HNSW si volume insertions >10% base (sinon index incrémental suffisant)

5. **Validation et rollback :**

- Tests automatisés post-sync : Requêtes test sur données modifiées, vérification cohérence réponses
- Si échec validation : Rollback transaction base vectorielle, alerte équipe technique
- Si succès : Update checkpoint last_sync_timestamp, logging statistiques (nb chunks ajoutés/modifiés/supprimés)

5.2 Gestion des Suppressions et Archivages

Suppression de données :

- Documents retirés espace documentaire : Soft delete en base vectorielle (flag is_deleted=true) pour historique audit, puis hard delete après 30 jours
- Marchés clôturés : Maintien dans base vectorielle avec métadonnée statut="clôturé", exclusion recherche sauf requête explicite historique

Archivage :

- Chunks de marchés >5 ans : Migration vers collection vectorielle "archives" interrogeable sur demande explicite (performance optimisée base principale)

5.3 Monitoring et Métriques Qualité

Indicateurs pipeline enrichissement :

- Latence sync : Temps écoulé entre modification source et disponibilité RAG (objectif <1h)
- Taux succès sync : % syncs sans erreur (objectif >99%)
- Volume traité : Nb chunks ajoutés/modifiés par jour (suivi tendance croissance base)

Indicateurs qualité RAG :

- Précision retrieval : % requêtes où top-5 chunks contiennent réponse attendue (évaluation mensuelle sur jeu test 200 questions, objectif >90%)
- Taux citation sources : % réponses incluant références documents/marchés (objectif 100%)
- Feedback utilisateurs : Boutons pouce haut/bas sur réponses, analyse verbatims négatifs

6. Modalités d'Usage et Interfaces

6.1 Intégration via Iframe

L'IA UniHA sera accessible de manière transparente sur deux sites UniHA via encapsulation iframe :

Site public <https://www.uniha.org> :

- Position : Widget chat en bas à droite, icône "Assistance IA UniHA"
- Niveau sécurité : Non-adhérent par défaut (Niveau 1)
- Bouton "Se connecter" redirigeant vers le portail adhérent

Portail adhérents (authentification préalable) :

- Position : widget dédié repliable sur la droite de l'écran "Assistant IA UniHA" + dans menu principal.
- Niveau sécurité : Automatiquement Niveau 2 ou 3 selon profil utilisateur connecté
- Encours personnalisés affichés automatiquement au login (voir section 6.3)

Spécifications techniques iframe :

- URL iframe exemple : <https://iauniha.uniha.org/chat> (sous-domaine dédié)
- Communication parent-iframe via postMessage sécurisé (whitelist domaines autorisés)
- Token JWT (O365) utilisateur transmis via postMessage au chargement iframe (évite authentification double)
- Responsive : Adaptation mobile (plein écran sur smartphone, overlay tablette)
- Accessibilité : Bien prendre en compte les utilisateurs malvoyant ou ayant un handicap

6.2 Fonctionnalité Téléchargement Documents

L'IA UniHA permettra aux utilisateurs autorisés (Niveaux 2 et 3) de demander et télécharger les documents de marchés directement via conversation.

Scénario utilisateur type :

1. **Utilisateur** : "Je voudrais consulter le cahier des charges du marché imagerie médicale"
2. **IA UniHA** : "Le marché imagerie (réf. IMG-2024-001) auquel votre établissement adhère comprend 3 lots. Voici les documents disponibles :
 - Cahier des Clauses Techniques Particulières (CCTP) - 2,3 Mo
 - Règlement de Consultation (RC) - 850 Ko
 - Bordereau de Prix Unitaires (BPU) - 1,1 MoPour les télécharger, cliquez dessus.

3. Souhaitez-vous télécharger un document spécifique ou l'ensemble en ZIP ?"
4. **Utilisateur** : "L'ensemble en ZIP"
5. **IA UniHA** : [Génération ZIP] "Votre archive est prête (4,2 Mo). [Bouton Télécharger]"

Architecture technique téléchargement :

- **Sélection documents** : IA UniHA identifie documents pertinents via recherche RAG (métadonnées id_marché, type_document)
- **Vérification droits** : API backend valide accès utilisateur (Établissements → Adhésions → Marchés → Documents)
- **Génération package** :
 - Document unique : Copie fichier source vers répertoire temporaire sécurisé (TTL 15 min)
 - Archive ZIP : Agrégation dynamique fichiers autorisés avec structure arborescente (dossiers par lots)
 - Watermarking PDF : Application filigrane transparent (nom utilisateur, établissement, date téléchargement, mention "Document confidentiel UniHA") via bibliothèque PyPDF2 ou équivalent
- **URL téléchargement** : Transmettre le ZIP a notre service SMASH via l'API SMASH, qui nous retourne un e URL de téléchargement.
- **Logging** : Traçabilité complète (user_id, établissement, documents téléchargés, timestamp) pour audit et conformité marchés publics

Limitations et protections :

- Taille maximale archive ZIP : 2Go (au-delà, proposition téléchargement séquentiel par lot)
- Throttling : Maximum 10 téléchargements par utilisateur par heure (anti-scraping)
- Détection abus : Alerte RSSI si téléchargement >100 documents sur 24h par un même compte

6.3 Interface Vocale Bidirectionnelle

L'IA UniHA offrira une expérience conversationnelle complète avec entrée vocale (Speech-to-Text) et sortie vocale (Text-to-Speech) en français avec voix féminine et masculine.

Entrée vocale (Speech-to-Text) open source :

- **Technologie** : Solution ou modèle fine-tuné en français
- **Implémentation** :
 - Bouton micro dans interface chat iframe

- Capture audio navigateur via WebRTC (getUserMedia API ou équivalent)
- Envoi stream audio vers backend IA UniHA (WebSocket)
- Transcription temps réel (<2 secondes latence)
- Affichage texte transcrit modifiable avant envoi (correction erreurs)
- **Formats audio** : Opus, MP3, WAV (conversion serveur si nécessaire)
- **Gestion bruit ambiant** : Filtrage passe-haut, détection activité vocale (VAD)

Sortie vocale (Text-to-Speech) :

- **Technologie** : Solution open source française (ex. Coqui TTS) ou API souveraine LinTO TTS[13]
- **Voix** : Voix féminine ou masculine professionnelle française (choix utilisateur)
- **Implémentation** :
 - Synthèse audio serveur (streaming MP3)
 - Lecture automatique dans navigateur (Web Audio API)
 - Contrôles utilisateur : Pause, replay, vitesse lecture (0,75x à 1,5x)
 - Option désactivation : Mode texte seul si souhaité
- **Optimisations** : Cache réponses fréquentes pré-générées, compression audio adaptatif selon bande passante

Cas d'usage vocal privilégiés :

- Utilisateurs en mobilité (pharmaciens en déplacement entre services)
- Situations mains occupées (préparation commandes, vérifications stocks)
- Accessibilité : Utilisateurs malvoyants ou difficultés lecture écran

Performance attendue :

- Latence bout-en-bout (question vocale → réponse vocale) : <8 secondes pour requête simple
- Précision transcription (WER - Word Error Rate) : <5% sur vocabulaire métier UniHA
- Intelligibilité TTS : MOS (Mean Opinion Score) >4/5

6.4 Encours Personnalisés au Login

Au moment de l'authentification sur le portail adhérents, l'IA UniHA affichera proactivement les encours et alertes pertinents selon le profil utilisateur, sans requête explicite.

Logique d'affichage encours :

1. Détection profil utilisateur (via attributs OAuth2 ou table User_Profile BDD) : Comptable, Acheteur, Pharmacien, Directeur
2. Requêtes SQL prédéfinies par profil exécutées au login

3. Génération message synthétique par l' IA UniHA avec données récupérées
4. Affichage carte encours dans interface chat (persistante jusqu'à traitement)

Encours par profil : Ceci est donné à titre d'exemple pour le chiffrage du projet global.

Profil	Encours affichés
Comptable	<ul style="list-style-type: none"> - Factures fournisseurs en attente validation (nb, montant total, ancienneté moyenne) - Factures rejetées nécessitant correction - Échéances paiements à venir (7 jours)
Acheteur	<ul style="list-style-type: none"> - Recensements en cours non finalisés (deadline proche) - Conventions marchés à signer (date limite) - Marchés arrivant à échéance dans 3 mois (renouvellement) - Nouveaux marchés disponibles pour établissement \ Niveau Maxi de commande presque atteint sur le marché au regard du recensement
Pharmacien	<ul style="list-style-type: none"> - Produits en rupture approvisionnement (nb références, fournisseurs concernés) - Produits en tension (stock < seuil alerte) - Alertes sanitaires ANSM sur produits commandés - Modifications tarifaires significatives (>10%) \ Etat des stocks de médicaments commandés \ état des commandes ...
Directeur	<ul style="list-style-type: none"> - KPI synthétiques établissement (taux adhésion marchés, économies réalisées vs tarifs catalogue) - Actions urgentes tous profils (consolidation) - Opportunités nouveaux marchés (estimations économies potentielles)

Table 3: Encours personnalisés par profil utilisateur

Exemple d'affichage au login (profil Acheteur) :

Bonjour Marie, bienvenue sur le portail UniHA. Voici vos encours du jour :

⚠ URGENT - 2 recensements à finaliser avant le 12/03/2026 :

- Marché Dispositifs Médicaux Stériles (lot 3) - CHU Lyon
- Marché Consommables Laboratoire (lot 1 et 2) - CHU Lyon

📄 1 convention à signer :

- Marché Imagerie Coordination Régionale AURA - Échéance 20/03/2026

📅 3 marchés arrivent à échéance (renouvellement à prévoir) :

- Linge Hospitalier (fin 30/06/2026)
- Restauration Patients (fin 31/07/2026)
- Mobilier Médical (fin 15/08/2026)

Comment puis-je vous aider aujourd'hui ?

Actions rapides depuis encours :

- Liens directs vers formulaires recensement (deeplink portail)
- Téléchargement conventions à signer
- Demande informations détaillées sur marché ("Dis-moi en plus sur le marché Linge Hospitalier")

7. Bénéfices Attendus et Indicateurs de Succès

7.1 Bénéfices Utilisateurs

Pour les adhérents :

- **Gain de temps** : Réduction du temps de recherche d'informations marchés
- **Réactivité** : Alertes proactives sur les encours réduisant les risques de ruptures d'approvisionnement liés au non renouvellements de marchés ou dépassement des montants maxi.
- **Autonomie** : Accès 24/7 aux informations sans solliciter le support UniHA

Pour les collaborateurs UniHA :

- **Productivité commerciale** : Ciblage des relances de recensements et propositions d'adhésions via analytics IA
- **Support client** : Réduction des tickets support niveau 1 (questions factuelles traitées par IA)
- **Qualité des données** : Détection d'incohérences marchés (ex. prix aberrants, fournisseurs manquants) via requêtes IA systématiques

Pour UniHA (organisation) :

- **Image innovation et service aux adhérents** : Positionnement de leader digital santé publique, attractivité auprès de nouveaux adhérents
- **Maîtrise complète des technologie (open source) et sécurité des données**
- **Conformité** : Traçabilité renforcée en conformité avec AI Act et RGPD (logs exhaustifs, auditabilité)

8. Risques et Mitigation

Risque	Impact	Mitigation
--------	--------	------------

Hallucinations LLM (infos erronées)	Perte confiance utilisateurs, décisions achats erronées	<ul style="list-style-type: none"> - RAG strict (réponses ancrées données) - Citations sources obligatoires - Supervision humaine - Possibilités escalade - Évaluations qualité trimestrielles
Fuite données sensibles (prix, fournisseurs)	Non-conformité RGPD et secret des affaires.	<ul style="list-style-type: none"> - ACL robustes multi-niveaux - Pentests trimestriels - Chiffrement bout-en-bout - Logs audit exhaustifs
Indisponibilité service	Perte accès informations critiques adhérents	<ul style="list-style-type: none"> - Architecture haute disponibilité (load balancing) - Redondance infrastructure (multi-AZ) - Plan reprise activité <4h
Dérive modèle (performances dégradées)	Baisse satisfaction utilisateurs	<ul style="list-style-type: none"> - Monitoring continu métriques qualité - Ré-entraînement/fine-tuning trimestriel - A/B testing nouvelles versions modèle
Non-conformité AI Act	Sanctions CNIL, blocage service	<ul style="list-style-type: none"> - Audit conformité pré-production (DPO) - Documentation complète (AIPD) - Formation équipes réglementation IA
Adoption insuffisante	ROI non atteint, abandon projet	<ul style="list-style-type: none"> - Co-construction avec utilisateurs pilotes - Formation/webinaires adhérents - Communication proactive bénéfices - Amélioration continue feedback

Table 4: Matrice risques projet IA UniHA

9. Planning de Déploiement et Jalons (à titre indicatif)

Jalon	Livrables clés
-------	----------------

Validation expression besoin	<ul style="list-style-type: none"> - Expression besoin validée COPIL - AIPD validée DPO - Budget alloué
Fin POC	<ul style="list-style-type: none"> - RAG fonctionnel 1000 docs - Tests sécurité 3 niveaux OK - Démo utilisateurs pilotes - GO/NO-GO développement
Fin développement	<ul style="list-style-type: none"> - Intégration complète sources données - Iframe portail fonctionnelle - Vocal bidirectionnel opérationnel - 300k docs indexés
Recette UAT (vérification d'aptitude)	<ul style="list-style-type: none"> - 50 utilisateurs pilotes testeurs - Validation cas usage métiers - Corrections anomalies bloquantes
Homologation sécurité	<ul style="list-style-type: none"> - Pentests externes réussis - Validation RSSI - GO/NO-GO production
Mise en production	<ul style="list-style-type: none"> - Déploiement progressif (10% puis 50% puis 100% utilisateurs) - Formation support - Documentation utilisateur - Communication adhérents
VSR	Un mois
Bilan 6 mois	<ul style="list-style-type: none"> - Évaluation KPI adoption/satisfaction - Mesure ROI - Décision évolutions V2

Table 5: Planning indicatif projet IA UniHA

10. Évolutions Futures Envisagées

Au-delà du périmètre initial (V1), l'IA UniHA pourra évoluer selon les besoins des utilisateurs et la maturité technologique. **Seule la V1 est à chiffrer par le candidat dans son mémoire technique et son BPU.**

Exemples d'évolutions futures possibles (à titre indicatif. Ces évolutions V2, V3 ne doivent pas faire l'objet d'un chiffrage par le candidat)

V2 (2027) - Assistance proactive avancée :

- Notifications push intelligentes (mail/SMS) : Alertes ruptures, deadlines recensements, nouveaux marchés pertinents

- Recommandations personnalisées : Suggestions adhésions marchés basées sur profil consommation établissement
- Intégration outils métiers : Pré-remplissage formulaires recensement depuis données IA, exports directs vers outils achats établissements

V3 (2028) - Analytics et prédictif :

- Prévisions consommations : Anticipation besoins établissements via LLM sur historiques
- Détection anomalies : Alertes automatiques prix aberrants, surconsommations, fraudes potentielles
- Benchmarking adhérents : Comparaisons anonymisées pratiques achats entre établissements similaires

V4 (2029) - Agents autonomes :

- Agent recensement : Pré-remplissage automatique recensements depuis systèmes établissements (avec validation humaine)
- Agent veille : Surveillance automatique évolutions réglementaires/sanitaires impactant marchés
- Agent négociation : Assistance temps réel collaborateurs UniHA lors négociations fournisseurs (suggestions arguments, historiques)