

## **PROGRAMME DE FORMATION CONCEPTEUR DÉVELOPPEUR D'APPLICATIONS OU CERTIFICATION PROFESSIONNELLE DÉVELOPPEUR EN INTELLIGENCE ARTIFICIELLE ET DATA SCIENCE RNCP36581**

Bloc E6.1 – Créer un modèle de données d'une solution I.A en utilisant des méthodes de Data science

Cahier des Charges de la MSPR « Développement et déploiement d'une application dans le respect du cahier des charges Client // Création d'un backend métier permettant le nettoyage et la visualisation des données »

### **COMPÉTENCES ÉVALUÉES :**

- Définir les sources et les outils nécessaires pour permettre de collecter les données.
- Recueillir de manière sécurisée les informations à partir de sources adaptées (sources hétérogènes, internes fournies par le client ou externes accessibles en Open Data) permettant de définir les données à collecter pour réaliser l'architecture de données.
- Paramétrer les outils afin d'importer les données de manière automatisée et sécurisée.
- Analyser, nettoyer, trier et s'assurer de la qualité des données afin de les rendre exploitables pour la solution I.A, en utilisant des outils d'analyse et de visualisation des données et se basant sur des approches de la Data science.
- Construire la structure de stockage des données (modèle de données) qui répond au mieux au besoin d'analyse.
- Représenter graphiquement les relations entre les données afin de les visualiser en créant des tableaux de bord accessibles à tout public garantissant ainsi l'accessibilité numérique.
- Exploiter de manière automatisée et analyser les informations recueillies dans les structures de stockage des données (requêtage ou interrogation) afin de répondre aux exigences de la solution IA défini dans le cahier des charges.

*Compétences supplémentaires évaluées pour les apprenants suivant le programme de formation Concepteur Développeur d'applications (Livret ECF) :*

- Installer et configurer son environnement de travail en fonction du projet.
- Développer des interfaces utilisateur
- Développer des composants métiers

### **PHASE 1 : PRÉPARATION DE CETTE MISE EN SITUATION PROFESSIONNELLE RECONSTITUÉE**

- **Durée de préparation** : 19 heures
- **Mise en œuvre** : Travail d'équipe constituée de 4 apprenants (5 maximum si groupe impair)

### **PHASE 2 : PRÉSENTATION ORALE COLLECTIVE + ENTRETIEN COLLECTIF**

- **Objectif** : mettre en avant et démontrer que les compétences visées par ce bloc sont bien acquises.
- **Moyen** : L'équipe utilise un support de présentation
- **Durée totale par groupe** : 50 mn se décomposant comme suit :
  - 20 mn de soutenance orale par l'équipe.
  - 30 mn d'entretien collectif avec le jury (questionnement complémentaire).
- **Jury d'évaluation** : 2 personnes (binôme d'évaluateurs) par jury – Ces évaluateurs ne sont pas intervenus durant la période de formation et ne connaissent pas les apprenants à évaluer.

## I – CONTEXTE

HealthAI Coach est une jeune startup française positionnée sur le marché en pleine expansion de la santé connectée et du coaching personnalisé. Son ambition est de proposer une plateforme digitale intégrant à la fois un suivi nutritionnel, un accompagnement sportif et une surveillance de certains indicateurs de santé, afin de répondre à une demande croissante de solutions de bien-être basées sur la donnée.

### 1. Business model

L'entreprise se positionne sur un marché en pleine croissance, où les consommateurs recherchent des solutions simples, personnalisées et accessibles pour gérer leur santé au quotidien. Pour répondre à cette demande, HealthAI Coach a adopté un **modèle économique hybride** :

- **Freemium** : accès gratuit donnant droit aux fonctionnalités de base (journal alimentaire, suivi d'activité, calcul d'IMC, tableaux de progression simples).
- **Premium – 9,99 €/mois** : accès à des recommandations personnalisées générées par IA, à des plans nutritionnels et sportifs détaillés, ainsi qu'à un suivi plus fin des objectifs.
- **Premium+ – 19,99 €/mois** : fonctionnalités avancées incluant l'intégration de données biométriques (fréquence cardiaque, sommeil, poids) via objets connectés, et la possibilité de bénéficier de consultations en ligne avec des nutritionnistes partenaires.
- **B2B – Distribution en marque blanche** : offre destinée aux salles de sport, mutuelles et entreprises, permettant d'intégrer la plateforme sous leur propre marque afin d'enrichir leur offre de services auprès de leurs adhérents, collaborateurs ou clients.

### 2. Cibles principales

La plateforme s'adresse prioritairement à :

- les **Millennials et la Génération Z (25–35 ans)**, soucieux de leur santé et habitués aux outils numériques,
- les **urbains actifs** disposant de peu de temps pour un suivi personnalisé mais souhaitant optimiser leur hygiène de vie,
- les **débutants en nutrition et en sport** qui recherchent des conseils fiables et structurés,
- les **personnes ayant des objectifs spécifiques** : perte de poids, renforcement musculaire, amélioration du sommeil ou maintien de la santé générale.

### 3. Enjeux du marché et concurrence

Le marché mondial des applications de santé et de fitness connaît une croissance soutenue (+20 %/an), stimulée par l'adoption massive des objets connectés (montres, bracelets, balances intelligentes) et par une sensibilisation accrue aux problématiques de santé préventive.

Cependant, ce **marché est fortement** concurrentiel, avec des acteurs établis comme **MyFitnessPal, Yazio ou Fitbit Premium**, qui proposent déjà des solutions de suivi alimentaire et d'activité. Ces applications se concentrent principalement sur le suivi quantitatif (calories, pas, minutes actives).

HealthAI Coach se différencie par :

- l'intégration **d'IA générative** et prédictive pour formuler des recommandations personnalisées,
- un **suivi global et holistique** (nutrition, sport, sommeil, biométrie),
- une **approche accessible à tous**, grâce à un modèle freemium inclusif,
- et une **stratégie B2B** en marque blanche, permettant aux entreprises et clubs de sport d'offrir un service différenciant à leurs utilisateurs.

#### 4.Écosystème technologique

Pour soutenir ce business model et répondre aux attentes de ses cibles, HealthAI Coach développe un **écosystème complet** :

une application mobile iOS/Android pour l'utilisateur final,

- un backend métier central chargé de collecter, nettoyer et stocker les données,
- une API sécurisée permettant d'exposer les informations aux applications front-end et aux partenaires,
- des tableaux de bord analytiques pour la direction et les clients B2B,
- et, à terme, des modules IA de recommandation et de prédiction intégrés dans la plateforme.

Le succès de la startup repose sur sa capacité à mettre en place un socle technique fiable, évolutif et industrialisable, capable d'absorber une volumétrie croissante de données tout en garantissant qualité, sécurité et accessibilité.

## **II – CAHIER DES CHARGES**

HealthAI Coach souhaite mettre en place une infrastructure technique robuste afin de collecter des données hétérogènes (issues d'APIs publiques, de jeux de données open data et de fichiers simulant les données biométriques des utilisateurs), de les transformer et de les stocker dans une base relationnelle unique.

L'objectif est double :

- D'une part, disposer d'un **référentiel fiable et de qualité** pour préparer les travaux futurs en intelligence artificielle (recommandations personnalisées),
- et d'autre part, fournir aux équipes produit un **tableau de bord interactif** permettant de visualiser les indicateurs essentiels (progression des utilisateurs, habitudes nutritionnelles, suivi d'activité).

L'entreprise attend que ce socle technique soit **automatisé, sécurisé et reproductible**, afin de réduire au maximum les interventions manuelles et de faciliter son déploiement dans différents environnements. L'ensemble doit être conçu pour évoluer : il servira de base à l'ajout ultérieur de micro-services (notamment des modules IA de prédiction et de personnalisation).

La mission confiée consiste à **concevoir, développer et livrer le backend métier** de la future plateforme HealthAI Coach. Concrètement, il s'agit :

- de mettre en œuvre un système de collecte automatisée capable d'intégrer différentes sources de données, en respectant les contraintes de sécurité et de fiabilité,
- de développer un processus de transformation et de nettoyage garantissant l'exploitabilité des données (qualité, cohérence, complétude),
- de concevoir et implémenter une base de données relationnelle adaptée aux besoins de l'entreprise, accompagnée d'une documentation claire et de scripts de migration,
- de proposer une API REST permettant à la fois aux applications front-end (mobile/web) et aux équipes internes de consulter et d'exploiter les données consolidées,
- et enfin, de livrer une interface de visualisation accessible permettant de suivre les indicateurs clés pour la direction et les partenaires de HealthAI Coach.

Ces éléments doivent être conçus dans une logique industrielle : l'entreprise n'attend pas une démonstration technique isolée, mais bien un **prototype structuré**, respectant les bonnes pratiques de développement et prêt à être intégré dans son écosystème global.

### III – LES BESOINS EXPRIMÉ PAR HEALTHAI COACH

La startup centralise des données provenant de multiples sources pour alimenter ses algorithmes de recommandation et ses outils d'analyse métier. Elle a besoin d'une plateforme backend complète capable d'ingérer, nettoyer et analyser ces données hétérogènes. Cette solution doit permettre aux data scientists d'accéder à une donnée exploitable pour leurs modèles, et aux product managers de monitorer la qualité des flux tout en extrayant des insights business critiques.

#### 1. Ingestion et traitement de données

La plateforme doit intégrer un pipeline d'ingestion automatisé capable :

- d'importer régulièrement des datasets externes (formats CSV, JSON, XLSX),
- de valider automatiquement la structure et la cohérence des données,
- et de gérer les erreurs dans les sources de données.

Les données concernées incluent :

- les **profils utilisateurs** (âge, objectifs, contraintes spécifiques),
- une **base nutritionnelle** (aliments, macronutriments, recettes),
- un **catalogue d'exercices** (types, niveaux de difficulté, équipements requis),
- et des **métriques de performance** (progression dans le temps, données biométriques simulées comme poids, sommeil, fréquence cardiaque).

#### 2. Interface d'administration et API de gestion

Une interface web d'administration doit permettre aux équipes internes :

- de consulter un **dashboard de pilotage en temps réel** avec des métriques de qualité,
- d'utiliser des **outils de nettoyage interactifs** pour corriger manuellement des anomalies,
- de gérer un **workflow de validation et d'approbation** avant mise en production des données,
- d'exporter les données nettoyées vers des services en aval (formats JSON ou CSV).

Cette interface reposera sur une **API REST de gestion flexible**, permettant la manipulation programmatique des données (CRUD utilisateurs, alimentation, exercices, progression). L'API doit être sécurisée, documentée via OpenAPI, et conçue pour évoluer avec de futurs modules IA et front-end mobile.

#### 3. Analytics et visualisation business

La solution doit inclure un module analytique générant :

- des **métriques utilisateurs** (répartition par âge, objectifs, taux de progression),
- des **analyses nutritionnelles** (tendances alimentaires, déficits/excès par profil),
- des **statistiques fitness** (exercices les plus pratiqués, niveaux d'intensité),
- et des **KPIs business** (engagement, conversion premium, satisfaction).

Ces indicateurs devront être accessibles via un tableau de bord interactif conforme aux standards d'accessibilité (RGAA niveau AA), afin d'être compris aussi bien par les data scientists que par les décideurs non techniques.

#### 4. Exigences complémentaires

Les données traitées devront être disponibles au format **JSON ou CSV**.

Les équipes devront **fournir une justification du choix des datasets** utilisés.

Une solution générique et extensible est attendue, mais le **minimum requis** est une chaîne complète de traitement fonctionnelle sur au **moins deux sources de données**, sélectionnées et justifiées par l'équipe projet.

Le livrable doit être conçu dans une **logique industrielle**, prêt à être intégré dans l'écosystème HealthAI Coach, et non comme une simple démonstration technique isolée.

### IV – LES LIVRABLES

À l'issue de la mission, HealthAI Coach attend une série de livrables clairement identifiés, qui permettront à la fois de vérifier la qualité du travail accompli et de préparer l'intégration dans l'écosystème global de la startup. Ces livrables doivent être produits dans une logique industrielle, c'est-à-dire documentés, reproductibles et exploitables par d'autres équipes (développeurs, data scientists, product managers).

#### 1. Documentation des données et flux

Un **rapport d'inventaire des sources de données** devra recenser toutes les sources utilisées, internes comme externes, en précisant leur origine, leur format, leur fréquence de mise à jour et les règles appliquées pour en assurer la qualité. Ce rapport sera accompagné d'un **diagramme des flux de données**, permettant de visualiser le cheminement complet entre la collecte brute, les traitements de nettoyage, le stockage en base relationnelle et l'exposition via API.

#### 2. Pipelines ETL opérationnels

Les équipes devront livrer un **pipeline d'ingestion et de transformation** automatisé, incluant :

- le code source complet, versionné et commenté,
- les scripts de planification ou de configuration (cron, Airflow, ou équivalent),
- et un système de gestion des erreurs et des logs retraçant chaque exécution.

#### 3. Jeux de données nettoyés et exploitables

Un **jeu de données consolidé** et débarrassé de ses anomalies devra être produit. Ce dataset servira de référence pour l'évaluation de la qualité des pipelines et constituera la base de travail des futurs modules d'intelligence artificielle.

#### 4. Base de données relationnelle et scripts associés

Un **modèle de données relationnel** documenté au format Merise(MCD/MLD/MPD) ou UML, devra être livré. Ce schéma sera accompagné des **scripts SQL de création et de migration** permettant de déployer la base de manière automatisée. L'objectif est de fournir une structure pérenne, versionnée et reproductible dans n'importe quel environnement.

#### 5. API REST documentée

Une **API REST** fonctionnelle devra être mise à disposition, offrant les principales opérations de consultation et de gestion des données (CRUD utilisateurs, alimentation, exercices, métriques). Cette API devra être sécurisée, testée et fournie avec une **documentation complète au format OpenAPI** afin de faciliter son adoption par les équipes front-end et partenaires externes.

## **6. Interface web et tableau de bord interactif**

Une interface d'administration accessible par navigateur devra permettre aux équipes internes de visualiser les flux de données, de valider ou corriger certaines anomalies, et d'exporter les données nettoyées. Elle devra intégrer un tableau de bord interactif présentant des indicateurs clairs (qualité des données, progression utilisateurs, tendances nutrition et activité, KPIs business). Le respect des standards d'accessibilité numérique (RGAA niveau AA) sera un critère déterminant.

## **7. Rapport technique et guide de déploiement**

Un rapport technique (5 à 8 pages) devra retracer l'ensemble de la démarche : contexte, choix technologiques, résultats obtenus, difficultés rencontrées et perspectives d'évolution. Ce rapport constituera la mémoire du projet et facilitera sa reprise par les équipes internes. En complément, un guide de déploiement détaillant la procédure de mise en œuvre (Docker/Docker Compose, variables d'environnement, prérequis logiciels) devra permettre à toute équipe technique de reproduire l'environnement et de lancer la solution en moins de trente minutes.

## **8. Support de soutenance**

En complément des livrables, l'équipe projet devra préparer un support de présentation destiné à la soutenance finale devant le client (public technique). Ce support devra synthétiser les principaux éléments du travail réalisé : démarche suivie, difficultés rencontrées, solutions mises en place, résultats obtenus et perspectives.

Il est important de souligner que l'évaluation de cette MSPR repose **sur la combinaison des trois éléments** suivants :

- la **qualité du travail réalisé** au cours du projet,
- la **pertinence et l'exhaustivité** des livrables remis
- et la capacité de l'équipe à **présenter, justifier et valoriser** ce travail lors de la soutenance orale.

**Les équipes devront donc s'assurer que la soutenance reflète bien l'ensemble des compétences attendues**, en démontrant à la fois la maîtrise technique et la capacité à communiquer efficacement auprès d'un client professionnel.

## V – RESSOURCES FOURNIES

Afin de permettre à l'équipe projet de mener à bien la mission, HealthAI Coach mettra à disposition un ensemble de ressources techniques et documentaires. Ces éléments constituent la base de travail commune et devront être exploités et enrichis par les apprenants dans le cadre de la mise en œuvre du backend métier.

### 1. Jeux de données de référence

Un premier socle de données sera remis aux équipes, composé de fichiers structurés (CSV et JSON) simulant différents volets du domaine d'activité de HealthAI Coach :

- une base nutritionnelle contenant des aliments, leurs apports énergétiques et leurs macronutriments associés ;
  - **Daily Food & Nutrition Dataset**  
<https://www.kaggle.com/datasets/adilshamim8/daily-food-and-nutrition-dataset>  
Données : apports quotidiens, valeurs nutritionnelles, tracking santé
  - **Diet Recommendations Dataset**  
<https://www.kaggle.com/datasets/ziya07/diet-recommendations-dataset>  
Données : profils santé, besoins diététiques, recommandations IA
- un catalogue d'exercices sportifs décrivant pour chaque activité son type, son niveau d'intensité et les équipements nécessaires ;
  - **ExerciseDB API Repository (1300+ exercices)**  
<https://github.com/ExerciseDB/exercisedb-api/tree/main>  
Recommandation : Fork du repository sur votre compte GitHub personnel  
Données : nom, type, muscle groups, équipement, niveau, images, instructions
- un jeu de profils utilisateurs (fictifs) comprenant des informations démographiques (âge, sexe, poids, taille) et des objectifs personnalisés (perte de poids, prise de masse, amélioration du sommeil, maintien de la forme) ;
  - **Gym Members Exercise Dataset (973 échantillons)**  
<https://www.kaggle.com/datasets/valakhorasani/gym-members-exercise-dataset>  
Données : âge, genre, poids, taille, BPM max/moyen, calories, BMI, body fat %
  - **Fitness Tracker Dataset (données d'activité quotidienne)**  
<https://www.kaggle.com/datasets/nadeemajeedch/fitness-tracker-dataset>  
Données : steps, calories burn, minutes d'activité, profils diversifiés

### 2. Assistance et périmètre

Dans le cadre de ce projet pédagogique, l'équipe projet n'aura aucun contact direct avec HealthAI Coach. Le cahier des charges constitue la seule expression officielle du besoin. Toute demande de clarification devra être traitée avec l'encadrant pédagogique, jouant le rôle du client.

### 3. Webographie

- Documentation Pandas : <https://pandas.pydata.org/pandas-docs/stable/>
- Documentation PostgreSQL : <https://www.postgresql.org/docs/>
- Documentation Power BI : <https://docs.microsoft.com/fr-fr/power-bi/>
- Documentation Apache Superset : <https://superset.apache.org/docs/intro>
- Datasets Kaggle : <https://www.kaggle.com/datasets/>
- Guide ETL Apache Hop : <https://hop.apache.org/manual/latest/getting-started/>
- Guide ETL Talend : <https://www.talend.com/fr/resources/guide-etl/>
- Documentation Grafana : <https://grafana.com/docs/>
- Guide Metabase : <https://www.metabase.com/docs/latest/>