

CERTIFICAT D'INTÉGRITÉ SUR L'HONNEUR

Thèse Professionnelle

Je soussigné(e), **Nicolas EDMOND**, certifie sur l'honneur que cette thèse professionnelle intitulée :

« Déploiement et utilisation des LLM dans le coaching sportif (athlétisme) »

a été rédigée par mes soins, sans recours à l'intelligence artificielle générative (ChatGPT, Claude, Gemini, Copilot, ou tout autre outil d'IA générative) pour la production du contenu textuel, de l'analyse, des conclusions ou de la rédaction des sections principales.

Je reconnaissais avoir utilisé des outils d'assistance pour :

- La correction orthographique et grammaticale
- La mise en forme LaTeX
- La recherche bibliographique et la consultation de sources
- L'aide à la structuration et à l'organisation des idées

Cependant, l'ensemble du contenu, des analyses, des réflexions critiques, des conclusions et de la rédaction finale sont le fruit de mon travail personnel et de ma réflexion propre.

Je comprends que toute utilisation non déclarée d'outils d'intelligence artificielle générative constitue une fraude académique et peut entraîner des sanctions disciplinaires.

Signature de l'étudiant

Date : _____

Document à signer et à joindre en première page du mémoire

Thèse Professionnelle (PRO)

Déploiement et utilisation des LLM dans le coaching sportif (athlétisme)

Comment entraîneurs, clubs et fédérations peuvent intégrer les LLM pour améliorer le suivi des athlètes, la personnalisation des plans d'entraînement et la communication, tout en préservant la dimension humaine du coaching

Nicolas EDMOND

15 janvier 2026

Résumé

Cette thèse professionnelle propose une approche pragmatique et éthique pour intégrer des Modèles de Langage de Grande Taille (LLM) dans le coaching d'athlétisme. Elle répond à la question centrale : comment entraîneurs, clubs et fédérations peuvent-ils exploiter les LLM pour améliorer le suivi des athlètes, personnaliser les plans et fluidifier la communication, tout en préservant la dimension humaine ? Nous vulgarisons les bases (transformers, embeddings), détaillons une architecture RAG (Retrieval-Augmented Generation), proposons un modèle de données inspiré des structures FFA, et illustrons le tout via des cas d'usage concrets. Nous couvrons la gouvernance (CNIL/RGPD), l'évaluation (métriques techniques et terrain) et un plan de déploiement progressif.

Table des matières

Table des figures

1	Workflow hebdomadaire entraîneur : comparaison avant/après IA. Les tâches administratives répétitives (cadre pointillé) sont automatisées, réduisant la charge de 30% et libérant du temps pour le coaching humain (écoute, empathie, ajustements contextuels).	12
2	Arbre de décision pour choisir l'approche architecturale adaptée	17
3	Architecture RAG détaillée : Pipeline d'ingestion, Indexation et Flux de requête hybride	24
4	Matrice de risques projet. H =Hallucinations, P =Fuite PII, R =Rejet, C =Coûts, D =Qualité données. Zones : vert (acceptable), jaune (surveiller), orange/rouge (mitiger prioritairement).	40
6	Maquette du Dashboard de Pilotage. Consolidation en temps réel des métriques techniques (qualité RAG), terrain (adhérence, blessures) et utilisateur (NPS). Permet un pilotage data-driven du déploiement.	47
7	Architecture de déploiement en 3 environnements avec feature flags et kill switch	48
8	Roadmap déploiement 12 mois. Phase 1 (M1-3) : POC avec 1 club pilote, validation architecture. Phase 2 (M4-7) : Extension 2-3 clubs, ajustements terrain. Phase 3 (M8-12) : Généralisation progressive, optimisation, monitoring production. Jalons rouges = validation go/no-go.	60

Liste des tableaux

1	Comparaison des deux approches architecturales	14
2	Synthèse comparative des scénarios économiques. L'effet d'échelle est déterminant : le coût par athlète diminue de 86% entre petit club et fédération, tandis que le ROI passe de négatif à très positif.	43

Liste des Acronymes et Abréviations

ABC	Attestation de Base de Compétences (formation fédérale entraîneurs)
ACWR	Acute :Chronic Workload Ratio (ratio charge aiguë/chronique)
ANN	Approximate Nearest Neighbors (recherche de voisins approximatifs)
API	Application Programming Interface
CAPEX	Capital Expenditure (dépenses d'investissement)
CI/CD	Continuous Integration / Continuous Deployment
CNIL	Commission Nationale de l'Informatique et des Libertés
DDL	Data Definition Language
DPIA	Data Protection Impact Assessment (Analyse d'Impact sur la Protection des Données)
DPO	Délégué à la Protection des Données
DRIP	Data Rich, Information Poor
FFA	Fédération Française d'Athlétisme
FCM	Fréquence Cardiaque Maximale
GPS	Global Positioning System
HRV	Heart Rate Variability (variabilité de la fréquence cardiaque)
IA	Intelligence Artificielle
KPI	Key Performance Indicator
LLM	Large Language Model (Modèle de Langage de Grande Taille)
LIME	Local Interpretable Model-agnostic Explanations
LSTM	Long Short-Term Memory (réseaux de neurones récurrents)
NLP	Natural Language Processing (Traitement du Langage Naturel)
OPEX	Operational Expenditure (dépenses opérationnelles)
PII	Personally Identifiable Information (données personnelles identifiables)
POC	Proof of Concept (preuve de concept)
PRA	Plan de Reprise d'Activité
RAG	Retrieval-Augmented Generation
RGPD	Règlement Général sur la Protection des Données (GDPR)
ROI	Return on Investment (retour sur investissement)
RPE	Rate of Perceived Exertion (échelle de perception de l'effort)
SHAP	SHapley Additive exPlanations
SLA	Service Level Agreement
SVM	Support Vector Machines
TCO	Total Cost of Ownership (coût total de possession)
VMA	Vitesse Maximale Aérobie

1 Introduction

1.1 La révolution silencieuse de la datafication sportive

L'histoire de l'entraînement sportif est celle d'une lente transition de l'empirisme vers la science des données. Pendant des décennies, la transmission du savoir s'est opérée par compagnonnage : l'ancien champion devenait entraîneur, perpétuant des méthodes éprouvées par l'expérience mais rarement validées scientifiquement. Cette approche, qualifiée d'artisanale¹, reposait sur l'intuition, l'observation directe et une relation humaine privilégiée.

Le tournant du XXIe siècle a marqué une rupture épistémologique majeure avec l'avènement de la "datafication" du sport. L'introduction massive de capteurs (GPS, accéléromètres, cardiofréquencemètres) et la démocratisation des outils d'analyse ont transformé l'athlète en un système bio-mécanique quantifiable. Comme le soulignent Rein et Memmert², le sport est entré dans l'ère du Big Data, où la performance ne se mesure plus seulement au chronomètre, mais à travers des milliers de points de données générés à chaque foulée.

Cependant, cette révolution technologique a créé une fracture numérique profonde. D'un côté, l'élite mondiale dispose de "Performance Centers" où des équipes pluridisciplinaires (data scientists, physiologistes, biomécaniciens) analysent chaque métrique pour optimiser la performance marginale. De l'autre, la masse des clubs amateurs et semi-professionnels – le tissu vital de l'athlétisme français – se retrouve submergée par ce déluge de données sans disposer des outils cognitifs pour les traiter. C'est le paradoxe du "Data Rich, Information Poor" (DRIP) décrit par Thornton et al.³ : jamais les entraîneurs n'ont eu autant de données, et jamais ils n'ont été aussi démunis pour prendre des décisions éclairées.

1.2 Le fossé technologique et cognitif

Cette asymétrie ne concerne pas seulement l'accès aux outils, mais la capacité de traitement de l'information. Un entraîneur bénévole gérant un groupe de 20 athlètes hétérogènes se trouve confronté à une complexité combinatoire vertigineuse. Pour chaque athlète, il doit intégrer des variables physiologiques (fréquence cardiaque, variabilité sinusal), biomécaniques (cadence, temps de contact), contextuelles (sommeil, stress professionnel) et historiques (blessures, progression).

La charge cognitive imposée par cette analyse dépasse les capacités humaines sans assistance. Là où un staff professionnel dédie un analyste par ligne de joueurs, l'entraîneur de club doit assumer seul les rôles de planificateur, psychologue, analyste vidéo et gestionnaire de données. Cette surcharge mentale conduit inévitablement à des heuristiques de décision simplificatrices, voire à l'abandon pur et simple de l'analyse des données au profit d'un retour à l'intuition pure, négligeant ainsi le potentiel prédictif et préventif des outils modernes.

-
1. Loland, S. (2018). Performance-enhancing drugs, sport, and the ideal of natural athletic performance. *American Journal of Bioethics*, 18(6), 8-15.
 2. Rein, R., Memmert, D. (2016). Big data and tactical analysis in elite soccer : future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1-13.
 3. Thornton, H.R. et al. (2019). Developing athlete monitoring systems in team sports : data analysis and visualization. *International Journal of Sports Physiology and Performance*, 14(6), 698-705.

C'est dans ce contexte de tension entre complexité croissante et ressources cognitives limitées que l'Intelligence Artificielle Générative, et spécifiquement les Grands Modèles de Langage (LLM), émergent comme une solution de rupture. Contrairement aux systèmes experts rigides des années 2000, les LLM offrent une capacité inédite : celle de traiter des données non structurées (texte, notes, ressentis) et de dialoguer en langage naturel, agissant non plus comme un outil passif, mais comme un véritable "copilote cognitif".

L'émergence des LLM modernes, depuis l'introduction de l'architecture Transformer par Vaswani et al.⁴ jusqu'aux modèles génératifs de grande échelle comme GPT-4⁵ et Claude 3⁶, a révolutionné le traitement du langage naturel. Ces modèles, pré-entraînés sur des corpus massifs⁷ et affinés par apprentissage par renforcement avec retour humain (RLHF)⁸, manifestent des capacités émergentes inattendues⁹ : raisonnement par chaîne de pensée¹⁰, compréhension contextuelle profonde, et génération de contenu cohérent et adaptatif. Ces capacités, explorées notamment par Bubeck et al.¹¹ dans leurs expériences avec GPT-4, ouvrent des perspectives inédites pour l'assistance cognitive dans des domaines spécialisés comme le coaching sportif.

1.3 Du sujet à la problématique : cadrer l'enjeu

1.3.1 Annonce du sujet

Cette thèse s'inscrit dans le champ de l'intégration des Grands Modèles de Langage (LLM) au service du coaching sportif, et plus spécifiquement de l'athlétisme. Le **sujet** de notre recherche peut se formuler ainsi :

« *Déploiement et utilisation des LLM dans le coaching d'athlétisme : concevoir une architecture d'assistance intelligente pour démocratiser l'accès à l'expertise de haut niveau.* »

Ce sujet couvre un domaine d'étude large : les technologies d'IA générative, les architectures RAG (Retrieval-Augmented Generation), les modèles de données sportives, la conformité RGPD, et les stratégies de déploiement. Il s'agit d'un territoire technique et organisationnel vaste, nécessitant une délimitation précise.

1.3.2 La problématique : la tension centrale

Au-delà de la simple description du domaine, la **problématique** identifie la tension fondamentale que cette thèse cherche à résoudre. Elle ne porte pas sur la faisabilité technique de l'IA dans le sport (déjà démontrée), mais sur une question plus profonde et structurante :

Cette thèse se propose d'explorer comment l'intégration pragmatique des LLM peut démocratiser l'accès à l'expertise de haut niveau pour l'ensemble de l'écosystème athlétique. La question centrale qui guide nos travaux est la suivante :

4. Vaswani, A. et al. (2017). Attention Is All You Need. NeurIPS 2017.
5. Achiam, J. et al. (2023). GPT-4 Technical Report. arXiv :2303.08774.
6. Anthropic (2023). Claude 3 Opus : Technical Report. <https://www.anthropic.com/claudie>
7. Brown, T. et al. (2020). Language Models are Few-Shot Learners. NeurIPS 2020.
8. Ouyang, L. et al. (2022). Training language models to follow instructions with human feedback. NeurIPS 2022.
9. Wei, J. et al. (2022). Emergent abilities of large language models. TMLR 2022.
10. Wei, J. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS 2022.
11. Bubeck, S. et al. (2023). Sparks of AGI : early experiments with GPT-4. arXiv :2303.12712.

Comment concevoir et déployer une architecture d'assistance par IA qui permette aux entraîneurs de clubs et fédérations d'exploiter la richesse des données sportives pour personnaliser l'entraînement à grande échelle, tout en préservant et renforçant la dimension fondamentalement humaine de la relation coach-athlète ?

Cette problématique repose sur une **tension dialectique** entre deux impératifs apparemment contradictoires :

- **L'impératif d'échelle** : Démocratiser l'accès à l'expertise de haut niveau, automatiser l'analyse de données massives, standardiser les bonnes pratiques pour réduire les inégalités entre clubs d'élite et clubs amateurs.
- **L'impératif humain** : Préserver la relation coach-athlète, respecter l'intuition et l'expérience de l'entraîneur, éviter la déshumanisation du coaching par une sur-automatisation.

La résolution de cette tension constitue le **fil rouge** de l'ensemble de notre travail. Chaque choix technique, chaque recommandation organisationnelle, chaque garde-fou éthique sera évalué à l'aune de cette double exigence : *automatiser sans déshumaniser, standardiser sans rigidifier, assister sans remplacer*.

1.3.3 Portée et ambition de la réponse

Pour répondre à cette problématique, nous adoptons une approche constructiviste et pluridisciplinaire, croisant informatique avancée (RAG, Vector Search), sciences du sport (physiologie, théorie de l'entraînement) et éthique appliquée (RGPD, acceptabilité sociale). Notre ambition n'est pas de remplacer l'entraîneur, mais de l'**augmenter** en le libérant des tâches analytiques répétitives pour lui permettre de se recentrer sur son cœur de métier : la pédagogie et la relation humaine.

Cette réponse se déclinera en deux architectures distinctes, adaptées aux contraintes et ressources des différents acteurs :

- **L'approche « Copilote »** (petite échelle) : pour les entraîneurs individuels et petits clubs (<10 athlètes).
- **L'approche « Orchestrateur »** (grande échelle) : pour les fédérations, ligues régionales et grands clubs (>50 athlètes).

1.3.4 Objectifs de recherche

Cette thèse poursuit quatre objectifs principaux :

Objectif 1 : Concevoir une architecture technique adaptée : Développer une architecture RAG (Retrieval-Augmented Generation) spécialisée pour le coaching d'athlétisme, capable de traiter les données hétérogènes du sport (plans d'entraînement, bilans, données capteurs) tout en garantissant la fiabilité et l'explicabilité des recommandations.

Objectif 2 : Modéliser les besoins et contraintes de l'écosystème : Analyser les besoins spécifiques des différents acteurs (entraîneurs, athlètes, clubs, fédération) et proposer un modèle de données aligné avec les structures existantes de la Fédération Française d'Athlétisme.

Objectif 3 : Intégrer la dimension éthique et réglementaire : Définir un cadre de gouvernance des données conforme au RGPD, incluant des mécanismes de transparence, d'explicabilité, et de mitigation des biais algorithmiques, tout en préservant la dimension humaine du coaching.

Objectif 4 : Proposer un plan de déploiement réaliste : Élaborer une stratégie de déploiement progressif (POC → Pilotes → Généralisation) avec un modèle économique transparent, une analyse des coûts (TCO), et un plan de conduite du changement adapté aux contraintes des clubs amateurs.

1.3.5 Délimitation du périmètre

Cette thèse se concentre spécifiquement sur :

- **Discipline** : Athlétisme (demi-fond, fond, sprint, sauts, lancers)
- **Contexte** : Clubs amateurs et semi-professionnels français, avec extension possible aux structures fédérales
- **Technologie** : LLM génératifs (GPT-4, Claude) couplés à une architecture RAG
- **Cas d'usage prioritaires** : Planification d'entraînement, suivi de charge, prévention des blessures, personnalisation des plans

Exclusions explicites :

- L'analyse vidéo automatisée (computer vision) n'est pas couverte en détail
- Les sports collectifs (football, rugby) ne sont pas traités, bien que certains concepts soient transposables
- Le haut niveau professionnel international (budgets illimités) n'est pas le focus principal
- Les aspects de détection de talents et de sélection ne sont qu'effleurés

2 Méthodologie de recherche

Cette thèse adopte une **approche mixte** combinant recherche documentaire, modélisation conceptuelle, et proposition d'architecture technique. Notre méthodologie s'articule autour de quatre axes complémentaires.

2.1 Approche et paradigme de recherche

Nous adoptons un **paradigme constructiviste** : plutôt que de décrire passivement l'existant, nous construisons une solution adaptée aux besoins identifiés de l'écosystème athlétique français. Cette approche est appropriée pour un contexte professionnel où l'objectif est de proposer une architecture actionnable, non seulement de comprendre un phénomène.

Type de recherche : Recherche appliquée de type "Design Science Research"¹², où nous concevons un artefact (architecture RAG) pour résoudre un problème pratique identifié (fracture numérique entre élite et clubs amateurs).

12. Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105.

2.2 Protocole de collecte et d'analyse

1. Analyse documentaire et état de l'art :

- Revue systématique de la littérature scientifique sur IA et sport (2015-2024)
- Analyse des architectures RAG existantes et leur applicabilité au domaine sportif
- Étude des référentiels FFA et structures de données fédérales
- Veille réglementaire (RGPD, CNIL, éthique IA)

2. Modélisation des besoins :

- Identification des acteurs et de leurs pain points via analyse systémique
- Définition de cas d'usage prioritaires (user stories, épopées fonctionnelles)
- Modélisation des données inspirée des structures existantes

3. Conception d'architecture :

- Proposition d'architecture RAG spécialisée avec justification des choix techniques
- Définition de protocoles d'évaluation (métriques offline et terrain)
- Plan de déploiement progressif avec analyse économique

4. Validation conceptuelle :

- Vérification de cohérence avec les contraintes réglementaires (RGPD)
- Analyse de faisabilité technique et économique
- Identification des risques et stratégies de mitigation

2.3 Critères d'évaluation et de validation

Notre architecture sera évaluée selon quatre dimensions :

Critères techniques :

- Pertinence des réponses RAG (métriques : precision@k, recall@k, MRR)
- Temps de réponse < 3 secondes pour recommandations
- Taux d'hallucinations < 5% (validation humaine sur échantillon)
- Disponibilité système > 99.5% (SLA)

Critères d'acceptabilité utilisateur :

- Taux d'adoption > 60% après 3 mois de pilote
- Score de satisfaction (NPS) > 50
- Réduction perçue de la charge administrative > 30%
- Confiance dans les recommandations IA (échelle 1-10) > 7

Critères de conformité :

- Conformité RGPD (DPIA validé, mesures techniques documentées)
- Explicabilité : 100% des recommandations avec sources citées
- Traçabilité complète des décisions algorithmiques

Critères économiques :

- ROI positif sur 3 ans pour clubs pilotes
- Coût par utilisateur < 50€/mois (modèle économique viable)
- Réduction des coûts de formation/recrutement entraîneurs

2.4 Positionnement par rapport aux méthodologies existantes

Notre approche méthodologique s'inscrit dans le courant du Design Science Research¹³, où l'objectif est de concevoir un artefact (notre architecture RAG) pour résoudre un problème pratique identifié. Cette approche diffère des méthodologies purement empiriques (études de cas, expérimentations) en ce qu'elle combine conception, validation conceptuelle, et proposition d'implémentation.

2.4.1 Avantages de l'approche Design Science Research

L'approche DSR présente plusieurs avantages pour notre contexte :

- **Orientation solution** : Focus sur la création d'un artefact actionnable plutôt que sur la seule compréhension d'un phénomène
- **Validation itérative** : Possibilité de valider conceptuellement avant déploiement à grande échelle
- **Pluridisciplinarité** : Intégration naturelle de connaissances issues de l'informatique, des sciences du sport, et de l'éthique
- **Pertinence pratique** : Réponse directe aux besoins identifiés par les acteurs de terrain

2.4.2 Limites assumées de notre approche

Nous reconnaissons que notre approche présente des limites par rapport à une validation expérimentale complète :

- **Absence de résultats terrain à grande échelle** : Notre validation reste conceptuelle et sur pilotes restreints
- **Dépendance aux données synthétiques** : Les exemples utilisés sont illustratifs, non issus de clubs réels (contraintes de confidentialité)
- **Évolution rapide du domaine** : Les LLM évoluent rapidement, certaines recommandations techniques pourront nécessiter mise à jour

Ces limites sont assumées et documentées, permettant aux futurs travaux de les dépasser par des expérimentations longitudinales à grande échelle.

2.5 Justification des choix méthodologiques

Pourquoi une approche conceptuelle plutôt qu'expérimentale ? : Cette thèse professionnelle se situe en amont du déploiement. L'objectif est de proposer une architecture validée conceptuellement et économiquement, prête pour des pilotes terrain. Une validation expérimentale complète nécessiterait 12-18 mois de déploiement réel, dépassant le cadre temporel d'une thèse professionnelle.

Pourquoi le RAG plutôt que le fine-tuning ? : Le RAG offre meilleure explicabilité (sources traçables), moindre coût (pas de ré-entraînement), et adaptabilité rapide (mise à jour du corpus sans ré-entraîner le modèle). Le fine-tuning serait plus adapté pour un cas d'usage très spécialisé avec données massives.

13. Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105.

Pourquoi une approche pluridisciplinaire ? : L'intégration de l'IA dans le coaching nécessite de croiser informatique (RAG, embeddings), sciences du sport (physiologie, périodisation), et éthique appliquée (RGPD, biais). Une approche purement technique serait incomplète.

2.6 Limites méthodologiques

Cette thèse reconnaît plusieurs limites :

- **Validation terrain limitée** : Pas de résultats d'expérimentation à grande échelle, seulement validation conceptuelle et plan de déploiement
- **Données synthétiques** : Les exemples de données utilisées sont illustratifs, non issues de clubs réels (contraintes de confidentialité)
- **Focus géographique** : Concentré sur le contexte français (FFA), avec transposition partielle à d'autres fédérations
- **Évolution rapide du domaine** : Les LLM évoluent rapidement (GPT-4 → GPT-5, nouveaux modèles open-source), certaines recommandations techniques pourront nécessiter mise à jour

3 Comprendre le coaching athlétique moderne : entre art et science

3.1 Les fondements théoriques de l'entraînement

Pour comprendre l'apport potentiel de l'IA, il est indispensable de revenir aux fondamentaux théoriques qui régissent la performance sportive. L'entraînement n'est pas une simple accumulation d'efforts, mais un processus biologique complexe de perturbation et d'adaptation.

3.1.1 Le Syndrome Général d'Adaptation et la Surcompensation

Au cœur de toute méthodologie d'entraînement réside le concept de *Syndrome Général d'Adaptation* théorisé par Hans Selye¹⁴. L'organisme, soumis à un stress (la charge d'entraînement), réagit en trois phases : la phase d'alarme (fatigue immédiate), la phase de résistance (adaptation structurelle) et, si le stress est trop important ou prolongé, la phase d'épuisement.

L'objectif de l'entraîneur est de piloter ce stress pour induire un phénomène de *surcompensation*. Comme le modélise Banister¹⁵ avec son modèle "Fitness-Fatigue", la performance à un instant t est la résultante de la condition physique acquise (Fitness) moins la fatigue accumulée. L'art de la planification consiste donc à maximiser le Fitness tout en minimisant la Fatigue au moment précis de la compétition. C'est une équation dynamique complexe où chaque variable (intensité, volume, densité) interagit de manière non linéaire avec les capacités de récupération individuelles de l'athlète.

14. Selye, H. (1936). A syndrome produced by diverse noxious agents. *Nature*, 138(3479), 32.

15. Banister, E.W., Calvert, T.W., Savage, M.V., Bach, T. (1975). A systems model of training for athletic performance. *Australian Journal of Sports Medicine*, 7(3), 57-61.

3.1.2 La périodisation : structurer le temps sportif

Pour gérer cette complexité sur le long terme, les théoriciens comme Matveev¹⁶ ont développé le concept de périodisation. Cette structuration temporelle divise la saison en cycles hiérarchisés :

- **Le Macrocycle** (annuel ou pluriannuel) : Il définit la stratégie globale et les pics de forme visés (ex : Championnats de France). Il se compose de périodes de préparation générale, spécifique, et de compétition.
- **Le Mésocycle** (3 à 6 semaines) : C'est l'unité fonctionnelle de développement d'une qualité physique (ex : bloc de développement VMA). Il suit généralement une logique de charge progressive suivie d'une semaine de décharge (tapering) pour permettre l'assimilation.
- **Le Microcycle** (hebdomadaire) : C'est l'unité opérationnelle où s'organise l'alternance quotidienne entre sollicitation et récupération. C'est à ce niveau que se joue la prévention des blessures et l'ajustement fin.

L'IA trouve ici son premier terrain d'application majeur : la capacité à optimiser ces cycles non plus sur des modèles théoriques génériques, mais sur la réponse biologique réelle et individuelle de chaque athlète, mesurée en continu.

3.2 La charge d'entraînement : le défi de la quantification

La gestion de la charge est le levier principal de la performance et de la prévention des blessures. Cependant, sa quantification reste un défi méthodologique majeur, distinguant deux dimensions souvent décorrélées.

3.2.1 La charge externe : la mesure objective

La charge externe représente la quantité et l'intensité de travail prescrites et réalisées, mesurables objectivement via des capteurs. Les travaux de Scott et al.¹⁷ ont comparé différentes méthodes de quantification, montrant que les mesures GPS (distance, vitesse, accélérations) et les accéléromètres offrent une quantification plus précise que les méthodes subjectives traditionnelles. Dans l'athlétisme, cette quantification s'appuie principalement sur la distance parcourue, l'allure moyenne, le dénivelé, et les changements de rythme.

Les études de Colby et al.¹⁸ sur le football australien ont démontré la corrélation entre charges GPS dérivées et risque de blessure, validant l'utilité de ces métriques pour la prévention. Cependant, la simple accumulation de kilomètres ne suffit pas : l'intensité relative (pourcentage de VMA, zones cardiaques) doit être intégrée pour obtenir une mesure complète de la charge externe.

16. Matveev, L.P. (1964). Problem of periodization of sports training. Fizkultura i Sport, Moscow.

17. Scott, B.R. et al. (2013). A comparison of methods to quantify the in-season training load of professional soccer players. International Journal of Sports Physiology and Performance, 8(2), 195-202.

18. Colby, M.J. et al. (2014). Accelerometer and GPS-derived running loads and injury risk in elite Australian football. The Journal of Strength & Conditioning Research, 28(8), 2244-2252.

3.2.2 La charge interne : la réponse physiologique individuelle

La charge interne reflète la réponse physiologique et psychologique de l'athlète à la charge externe. Elle varie considérablement entre individus pour une même charge externe, en fonction de l'état de forme, de la fatigue accumulée, du stress non-sportif, et des capacités de récupération individuelles.

Buchheit¹⁹ a exploré l'utilisation de mesures de fréquence cardiaque (FC moyenne, FC max, variabilité de la fréquence cardiaque) comme proxy de la charge interne. Sa conclusion nuancée - "do all roads lead to Rome ?" - souligne qu'aucune métrique unique ne capture toute la complexité, mais qu'une combinaison de métriques (FC, RPE, biomarqueurs) offre une vision plus complète.

Les travaux de Jaspers et al.²⁰ ont établi des corrélations significatives entre différents indicateurs de charge (sRPE, distance GPS, FC) et les résultats d'entraînement, validant l'approche multi-métriques. Pour l'athlétisme, cette approche est d'autant plus pertinente que les disciplines varient drastiquement en termes de sollicitations (sprint vs marathon), nécessitant des métriques adaptées.

3.2.3 L'entraîneur de club : le généraliste surchargé

Profil : Souvent bénévole ou vacataire, il gère un groupe hétérogène de 15 à 30 athlètes. Il a une formation fédérale (ABC) mais peu de temps pour la veille scientifique.

Pain points : Manque de temps pour personnaliser, difficulté à suivre l'état de forme individuel, sentiment d'isolement technique. **Attentes IA** : Automatisation de la planification basique, alertes sur les décrochages de forme, accès rapide à des exercices variés pour casser la routine.

3.2.4 L'athlète amateur compétiteur : le "Quantified Self"

Profil : Pratique 3 à 5 fois par semaine, équipé (montre GPS), motivé par la progression chronométrique et la santé. **Pain points** : Ne comprend pas toujours le "pourquoi" des séances, manque de feedback immédiat, risque de blessure par surdosage. **Attentes IA** : Explication pédagogique des séances, visualisation de la progression, feedback instantané post-séance.

3.2.5 Le dirigeant de club : le gestionnaire

Profil : Soucieux de la fidélisation des adhérents, de la performance des équipes et de la responsabilité légale. **Pain points** : Turnover des entraîneurs, manque de traçabilité des accidents, disparité de qualité de service entre les groupes. **Attentes IA** : Outil de fidélisation, garantie d'un standard de qualité minimal, traçabilité des charges (protection juridique).

3.2.6 La Fédération (FFA) : le régulateur

Profil : Garante de l'intégrité physique, de l'éthique et du développement du haut niveau. **Pain points** : Perte de talents par manque de détection, disparité territoriale de

19. Buchheit, M. (2014). Monitoring training status with HR measures : do all roads lead to Rome ? Frontiers in Physiology, 5, 73.

20. Jaspers, A. et al. (2017). Relationships between training load indicators and training outcomes in professional soccer. Sports Medicine, 47(3), 533-544.

l'encadrement. **Attentes IA** : Détection précoce des potentiels via l'analyse de données massives, diffusion des bonnes pratiques nationales, respect du cadre éthique et sanitaire.

Cette analyse multidimensionnelle confirme que la solution technique ne peut être une simple "boîte noire" algorithmique. Elle doit s'insérer dans un tissu social existant, respecter les hiérarchies de compétences et répondre à des besoins opérationnels concrets avant de prétendre à l'optimisation de la performance pure.

Contrairement aux idées reçues, les athlètes ne forment pas un groupe homogène. Leurs besoins varient drastiquement selon leur profil :

L'athlète de haut niveau (ex : Sarah, sprinteuse nationale) vise la performance maximale et la sélection internationale, contrainte par un planning dense, une pression médiatique et la gestion du stress. Ses besoins en IA portent sur l'optimisation fine, l'analyse prédictive et les comparaisons avec l'élite. Exemple typique de question : "Compare ma répartition d'effort avec les finalistes mondiaux 400m".

L'athlète en progression (ex : Thomas, junior prometteur) recherche une progression régulière et un apprentissage technique, soumis aux contraintes d'études, d'un budget limité et d'une inexpérience compétitive. Ses besoins IA se concentrent sur les explications pédagogiques, la planification adaptée et la motivation. Exemple : "Pourquoi dois-je faire du foncier alors que je veux sprinter ?"

L'athlète santé-plaisir (ex : Marie, fondeuse amateur) privilégie le bien-être, le social et le défi personnel, avec un temps limité, des contraintes familiales et une motivation variable. Ses besoins IA concernent la flexibilité, l'encouragement et la simplification. Exemple : "J'ai raté deux séances cette semaine, comment rattraper ?"

L'athlète vétéran (ex : Paul, 45 ans, spécialiste 10km) cherche le maintien de la performance et la prévention des blessures, confronté à une récupération lente et des contraintes professionnelles. Ses besoins IA visent l'adaptation à l'âge, la gestion des charges et la surveillance de la santé. Exemple : "Adapter mon plan marathon compte tenu de mes douleurs genoux".

3.2.7 Les clubs : entre passion et contraintes

Les clubs d'athlétisme français vivent une réalité contrastée entre ambitions sportives et contraintes pratiques.

Les clubs d'élite (ex : club fédéral, pôles espoirs) disposent d'entraîneurs professionnels, d'installations de qualité et d'un budget conséquent. Ils nécessitent des outils sophistiqués d'analyse et de suivi pour rivaliser avec les moyens internationaux.

Les clubs intermédiaires (ex : club régional dynamique) présentent un mélange bénévoles/salariés, des installations correctes et un budget serré. Ils recherchent efficacité et mutualisation des ressources pour faire plus avec moins.

Les petits clubs locaux (ex : club communal) fonctionnent grâce à des bénévoles passionnés, des installations basiques et un budget minimal. Ils ont besoin de simplification et de formation pour maintenir la qualité malgré les contraintes.

Leurs défis communs incluent la gestion administrative (licences, assurances, subventions), la formation des encadrants (mise à jour des connaissances), la fidélisation des adhérents (maintenir la motivation) et la communication tant interne (membres) qu'externe (partenaires).

3.2.8 La Fédération Française d'Athlétisme : le pilote stratégique

La FFA orchestre un écosystème complexe avec des missions multiples :

Mission technique : définir les référentiels d'entraînement, former et certifier les entraîneurs, organiser les compétitions nationales, et détecter et développer les talents.

Mission réglementaire : faire respecter les règles internationales, gérer les licences et les transferts, contrôler l'antidopage, et assurer la sécurité des pratiquants.

Mission développement : promouvoir la pratique sur tout le territoire, soutenir les clubs en difficulté, innover dans les méthodes et outils, et représenter l'athlétisme français à l'international.

Ses besoins spécifiques englobent une vision consolidée de l'activité nationale, des outils de pilotage et de reporting, une harmonisation des pratiques, et le respect des contraintes légales (RGPD, protection mineurs).

3.2.9 Le staff médical : la sentinelle de la santé

Souvent oublié mais crucial, le staff médical (médecins, kinésithérapeutes, préparateurs physiques) assure la continuité de la pratique sportive.

Leurs préoccupations concernent la prévention primaire (éviter la blessure avant qu'elle n'arrive), le diagnostic précoce (détecter les signaux d'alerte), le traitement adapté (retour à la pratique sécurisé), et le suivi longitudinal (éviter les récidives).

Leurs besoins en données incluent un historique complet et fiable des charges d'entraînement, les corrélations entre charge et blessures, des alertes automatiques sur les seuils de risque, et une communication fluide avec les entraîneurs.

Exemple concret : Dr. Martin, médecin du sport, suit Thomas qui revient de tendinite d'Achille. Il a besoin de savoir précisément quelle charge Thomas peut supporter, à quelle progression, avec quels signaux d'alerte. L'IA pourrait l'aider à définir des seuils personnalisés basés sur des milliers de cas similaires.

Staff médical et prévention besoin d'alertes précoces (fatigue/blessures), d'historique fiable, de traçabilité.

Processus actuels (schéma) planification macro → micro ; exécution séance → retour athlète (RPE, notes) ; ajustement par coach ; communication asynchrone ; reporting compétition. Irritants : fragmentation outils, traçabilité partielle, effort de mise en forme.

3.3 Objectifs et critères SMART

L'objectif de **personnalisation** vise 100% des microcycles proposés tenant compte des contraintes individuelles et validés par le coach sous 24h. L'objectif de **communication** cherche à réduire de 30% le temps de rédaction des messages et bilans mensuels. L'objectif de **qualité de suivi** cible 90% des séances avec RPE et note, avec une hausse de 15% de l'adhérence aux plans. L'objectif de **conformité** exige une DPIA réalisée et les droits RGPD opérés en moins de 30 jours.

3.4 Analyse comparative des solutions existantes

Avant de proposer notre architecture, il est essentiel d'analyser les solutions existantes sur le marché et leurs limitations. Cette analyse comparative permet de justifier notre approche RAG spécialisée.

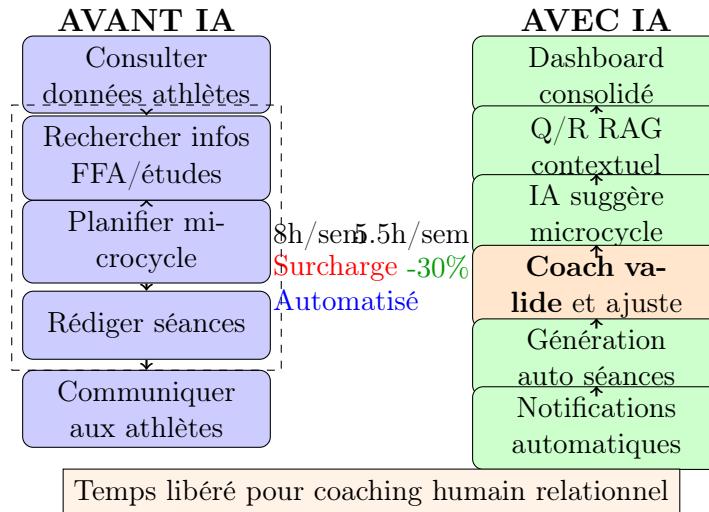


FIGURE 1 – Workflow hebdomadaire entraîneur : comparaison avant/après IA. Les tâches administratives répétitives (cadre pointillé) sont automatisées, réduisant la charge de 30% et libérant du temps pour le coaching humain (écoute, empathie, ajustements contextuels).

3.4.1 Plateformes de suivi sportif traditionnelles

Les plateformes existantes (TrainingPeaks, Strava, Garmin Connect) offrent principalement des fonctionnalités de visualisation de données et de planification basique. Elles manquent de capacités d'analyse intelligente et de personnalisation contextuelle. Leur approche est essentiellement réactive (affichage de données) plutôt que proactive (suggestions adaptatives).

3.4.2 Solutions IA génériques

Les solutions d'IA génériques (chatbots basiques, assistants vocaux) ne sont pas spécialisées pour le domaine sportif. Elles manquent de connaissances spécifiques sur la physiologie de l'entraînement, les protocoles fédéraux, et les contraintes du coaching d'athlétisme. Leur utilisation nécessite une expertise préalable importante de la part de l'utilisateur.

3.4.3 Systèmes experts rigides

Les systèmes experts traditionnels, basés sur des règles fixes, manquent de flexibilité et d'adaptabilité. Ils ne peuvent pas gérer la complexité et la variabilité des situations réelles de coaching. Notre approche RAG combine la structure des systèmes experts avec la flexibilité des LLM.

3.5 Cartographie des données

3.5.1 Sources de données identifiées

Les sources de données dans un club d'athlétisme sont multiples et hétérogènes :

- **Carnets numériques** : Plans d'entraînement, bilans de séances, notes d'entraîneurs
- **Exports FFA** : Licences, résultats de compétitions, classements
- **Fichiers montre/GPS** : Données de charge externe (distance, vitesse, dénivelé)

- **Messagerie** : Communications entre entraîneurs et athlètes
- **Documents internes** : Protocoles, référentiels techniques, bilans médicaux

3.5.2 Qualité et hétérogénéité des données

La qualité des données est hétérogène, avec des problèmes récurrents :

- **Incomplétude** : Données manquantes, saisies irrégulières
- **Incohérence** : Formats différents, nomenclatures variables
- **Désynchronisation** : Données non synchronisées entre systèmes
- **Qualité variable** : Fiabilité dépendante de la rigueur de saisie

3.5.3 Plan d'amélioration de la qualité

Notre plan d'amélioration comprend trois axes :

1. **Normalisation** : Mapping vers un schéma de données unifié inspiré des structures FFA²¹
2. **Contrôles de cohérence** : Validation automatique des données (plages de valeurs, règles métier)
3. **Retours aux utilisateurs** : Feedback loops pour améliorer la qualité de saisie

4 Fondamentaux LLM, embeddings, transformers

4.1 Transformer en termes simples

Un transformer apprend *où regarder* dans une séquence via l'**attention**. Plutôt que lire mot-à-mot, il pondère les éléments pertinents pour prédire le suivant. Avantages : parallélisation, contexte long, polyvalence.²²

4.2 Tokenisation et embeddings

Le texte est découpé en *tokens*. Chaque token est projeté en un **vecteur** (embedding) capturant des proximités sémantiques. Ces embeddings servent à : (1) recherche sémantique (trouver des séances semblables), (2) RAG (retrouver du contexte), (3) comparaison intra-athlète (progression).

4.3 Adapter un LLM au coaching

Trois leviers, du plus simple au plus coûteux : (1) **Prompt engineering** et **rôles/guardrails**, (2) **RAG** pour ancrer sur vos données, (3) **Fine-tuning/LoRA** pour spécialiser le style/contenu. Dans le sport, RAG couvre 80% des besoins à coût maîtrisé.

21. Fédération Française d'Athlétisme (référentiels publics, structures d'épreuves). <https://www.athle.fr/>

22. Vaswani, A. et al. (2017). Attention is All You Need. NeurIPS 2017.

4.4 L'attention en une équation

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

Lecture simple : on mesure la **pertinence** entre la question Q et les clés K , on normalise (*softmax*), puis on combine les valeurs V . Cela permet au modèle de concentrer son “regard” sur les éléments utiles du contexte.

5 Deux architectures pour deux échelles : « Copilote » et « Orchestrateur »

5.1 Pourquoi deux approches distinctes ?

La problématique centrale de cette thèse — automatiser sans déshumaniser — ne peut recevoir une réponse unique. Les contraintes, ressources et besoins d'un entraîneur bénévole gérant 8 athlètes dans un club rural diffèrent radicalement de ceux d'une fédération régionale coordonnant 200 entraîneurs et 5000 licenciés. Plutôt que de proposer une solution « taille unique », nous concevons **deux architectures complémentaires**, chacune optimisée pour son contexte d’usage.

Critère	Approche « Copilote »	Approche « Orchestrateur »
Cible	Entraîneur individuel	Fédération / Grand club
Échelle	<10 athlètes	>50 athlètes
Complexité	Simple, accessible	Sophistiquée, évolutive
Technologie	RAG léger, petit LLM	Workflow agentique, orchestration
Coût	<50€/mois	>500€/mois
Autonomie	Assistance ponctuelle	Automatisation poussée

TABLE 1 – Comparaison des deux approches architecturales

5.2 Approche « Copilote » : l'assistant léger pour petite échelle

5.2.1 Philosophie et cas d'usage

L'approche « Copilote » s'adresse aux entraîneurs individuels et petits clubs (<10 athlètes). Elle repose sur une philosophie d'**assistance ponctuelle** plutôt que d'automatisation totale. L'entraîneur reste maître de la planification, l'IA intervient comme un *second avis* rapide et fiable.

Cas d'usage typiques :

- « Propose-moi 3 variantes de séance VMA pour un coureur de 800m en phase spécifique »
- « Analyse le microcycle de Sarah et détecte les risques de surcharge »
- « Explique-moi pourquoi ce plan FFA recommande du foncier en début de saison »

5.2.2 Architecture technique simplifiée

Composants essentiels :

1. **RAG léger** : Base vectorielle locale (FAISS ou pgvector sur PostgreSQL), indexant :
 - Plans d'entraînement personnels (historique club)
 - Référentiels FFA (documents publics)
 - Bilans de séances et compétitions
2. **Petit LLM** : Modèle compact (GPT-3.5-turbo, Mistral-7B, ou Llama-3-8B) pour génération rapide et économique
3. **Indicateur de pertinence** : Score de confiance simple (0-100%) basé sur :
 - Similarité sémantique des sources retrouvées
 - Cohérence de la réponse générée
 - Présence de citations vérifiables
4. **Interface conversationnelle** : Chat web simple ou intégration WhatsApp/Telegram

Avantages :

- **Simplicité** : Déploiement en <1 jour, formation utilisateur <2h
- **Coût maîtrisé** : 30-50€/mois (hébergement + API LLM)
- **Confidentialité** : Données hébergées localement ou en Europe
- **Transparence** : Chaque réponse cite ses sources, l'entraîneur valide

Limites assumées :

- Pas d'automatisation de la planification complète
- Pas d'analyse prédictive avancée (blessures, performance)
- Pas de multi-agent ou workflows complexes

5.3 Approche « Orchestrateur » : le système agentique pour grande échelle

5.3.1 Philosophie et cas d'usage

L'approche « Orchestrateur » cible les fédérations, ligues régionales et grands clubs (>50 athlètes). Elle repose sur une **automatisation intelligente** via des workflows agentiques, où plusieurs agents spécialisés collaborent pour gérer la complexité.

Cas d'usage typiques :

- Génération automatique de 200 plans personnalisés pour une ligue régionale
- Détection proactive des athlètes en surcharge via analyse multi-sources (GPS, RPE, HRV)
- Consolidation hebdomadaire des bilans de 50 entraîneurs pour pilotage fédéral
- Recommandations adaptatives en temps réel pendant les compétitions

5.3.2 Architecture agentique multi-agents

L'architecture s'appuie sur des **frameworks agentiques** modernes comme AGNO (anciennement Phidata), LangGraph, ou N8N, permettant l'orchestration de workflows complexes.

Agents spécialisés :

1. **Agent Superviseur** : Coordonne les autres agents, gère les priorités, route les requêtes
2. **Agent Planificateur** : Génère des plans d'entraînement personnalisés basés sur :
 - Profil athlète (VMA, historique, objectifs)
 - Contraintes calendrier (compétitions, disponibilités)
 - Référentiels FFA et littérature scientifique
3. **Agent Analyste** : Analyse les données de charge, détecte les anomalies :
 - Calcul ACWR (ratio charge aiguë/chronique)
 - Détection écarts Prescrit vs Réalisé
 - Alertes précoce (fatigue, blessure)
4. **Agent Communicateur** : Génère bilans, rapports, notifications personnalisées
5. **Agent RAG** : Spécialisé dans la récupération de contexte pertinent (comme dans l'approche Copilote, mais à plus grande échelle)

Workflow type : Génération de plan hebdomadaire

1. Superviseur reçoit requête : "Générer microcycle semaine 12 pour groupe U20"
2. Superviseur délègue à Agent Analyste : "Récupérer état de forme actuel"
3. Agent Analyste interroge Agent RAG : "Dernières séances + RPE"
4. Agent RAG retourne contexte (séances S-1, S-2, S-3)
5. Agent Analyste calcule charge cumulée, retourne au Superviseur
6. Superviseur délègue à Agent Planificateur : "Générer microcycle"
7. Agent Planificateur interroge Agent RAG : "Référentiels FFA U20"
8. Agent Planificateur génère plan, retourne au Superviseur
9. Superviseur délègue à Agent Communicateur : "Formater et notifier"
10. Agent Communicateur envoie plans personnalisés aux athlètes

Technologies clés :

- **AGNO** : Framework Python full-stack pour agents autonomes, multimodal, model-agnostic
- **LangGraph** : Orchestration de workflows avec état, cycles, et branchements conditionnels
- **N8N** : Automatisation low-code pour intégrations tierces (Strava, Google Calendar, FFA API)
- **LLM hybrides** : Mix GPT-4 (raisonnement complexe), GPT-3.5 (tâches simples), Llama-3-70B (auto-hébergé pour confidentialité)

Avantages :

- **Scalabilité** : Gère 1000+ athlètes sans surcharge humaine
- **Sophistication** : Raisonnement multi-étapes, analyse prédictive
- **Autonomie** : Automatisation 70% des tâches répétitives

- **Traçabilité** : Chaque décision agentique est loggée et auditable

Limites et garde-fous :

- **Complexité** : Nécessite expertise DevOps et MLOps
- **Coût** : 500-2000€/mois selon volume (infrastructure + API)
- **Validation humaine obligatoire** : Plans générés soumis à validation entraîneur avant diffusion
- **Risque de sur-automatisation** : Nécessite culture organisationnelle forte (formation, change management)

5.4 Choix de l'approche : arbre de décision

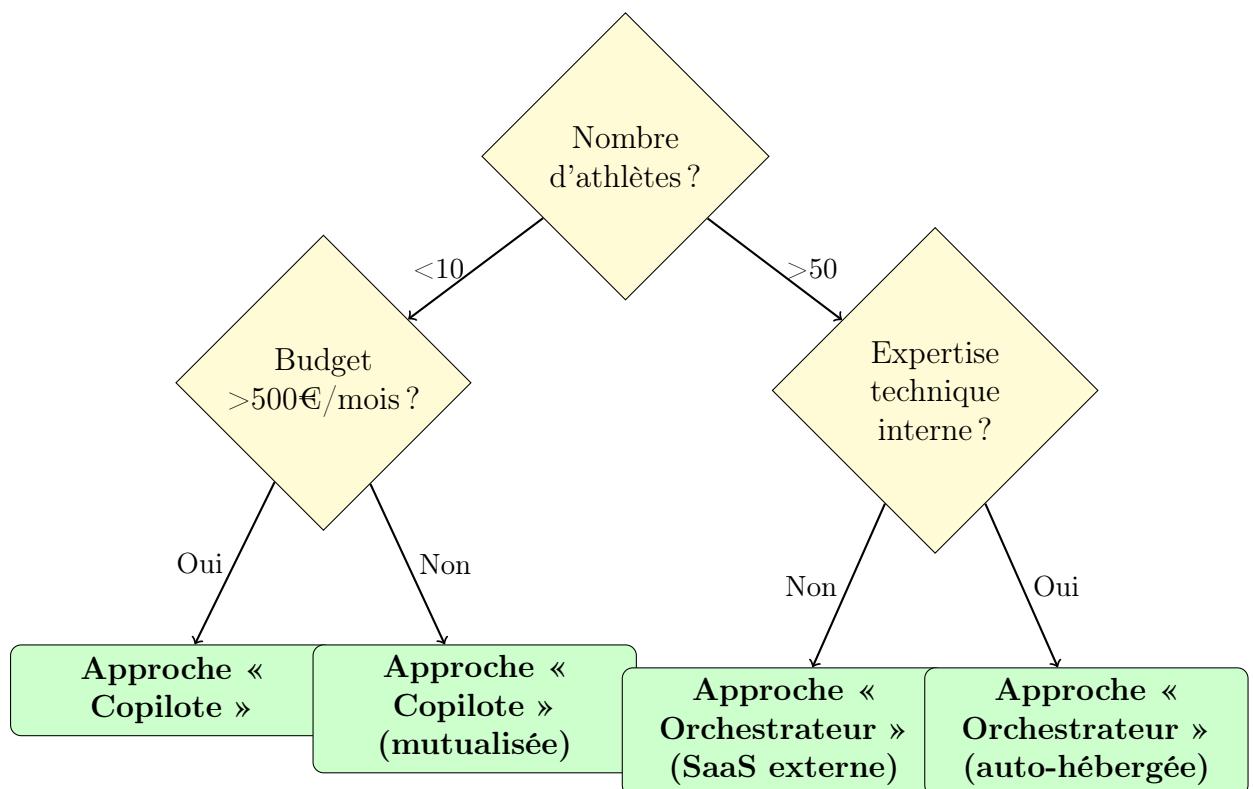


FIGURE 2 – Arbre de décision pour choisir l'approche architecturale adaptée

5.5 Synthèse : complémentarité plutôt qu'opposition

Ces deux approches ne sont pas concurrentes mais **complémentaires**. Un club peut démarrer avec l'approche « Copilote » pour 1-2 entraîneurs, puis migrer vers « Orchestrateur » en grandissant. Une fédération peut déployer « Orchestrateur » au niveau central tout en proposant « Copilote » aux clubs affiliés.

Le fil rouge reste constant : *augmenter l'humain sans le remplacer*. Dans les deux cas, l'IA assiste, l'entraîneur décide.

6 Architecture RAG pragmatique

6.1 Objectif

Répondre avec des faits issus des données du club/fédération (plans, séances, résultats, protocoles), en générant des explications compréhensibles.

6.2 Pipeline minimal

Le pipeline RAG minimal se structure en huit étapes essentielles. La première, l'**ingestion**, intègre des documents PDF/CSV, des exports FFA, des historiques d'entraînement et des notes d'entraîneurs. La **normalisation à schéma** réalise le mapping vers le modèle de données commun (cf. Section ??). Le **chunking** effectue la découpe par unités sémantiques (séance, microcycle). Les **embeddings** encodent les données via un modèle de type SBERT/BGE en vecteurs. L'**index vectoriel** s'appuie sur FAISS ou pgvector pour la similarité. La **récupération** sélectionne les top- k passages pertinents avec reranking optionnel. La **génération** exploite un LLM (ouvert ou fermé) guidé par consignes et garde-fous. Enfin, la **traçabilité** garantit les citations/sources et un *no answer* honnête si lacunes.

6.3 Choix techniques

Embeddings : *Sentence-BERT/BGE*. Index : *FAISS* (fichiers) ou *pgvector* (SQL). LLM : modèle cloud (latence basse) ou open-source hébergé UE. Rerank : *bge-reranker* pour précision. Observabilité : logs de requêtes, taux de refus, hallucinations captées via règles.

6.4 Paramètres recommandés (point de départ)

Pour configurer optimalement le pipeline RAG, les paramètres recommandés en guise de point de départ incluent un **chunk** de 400–800 tokens avec un **overlap** de 10–20% afin de préserver la continuité sémantique. L'**embedding** utilise 768–1024 dimensions avec normalisation L2 et stockage float32 (ou int8 quantisé après validation). Le **top- k** en retrieval se situe entre 4–8 passages, avec un rerank potentiel de $k = 20 \rightarrow 5$ si besoin de précision accrue. Le contexte LLM dispose de **3–6k tokens** utiles après citations, avec un minimum de 2 sources citées. Les **mises à jour** de l'index s'effectuent par batch nocturne complété d'un rafraîchissement incrémental via webhooks.

6.5 Sécurité et anti-injection

La sécurité et la prévention des injections reposent sur quatre mécanismes clés. D'abord, des **instructions système inaltérables** accompagnées d'une séparation stricte entre *question* et *contexte*. Ensuite, un **filtrage** systématique du contenu (PII, médical, doping) avant génération, renforcé par des règles strictes de *no-answer*. Troisièmement, l'utilisation de **canary strings** et de tests d'*adversarial prompts* avec revue régulière via *red teaming*. Enfin, une **traçabilité** complète incluant le journal des sources, l'identifiant de version d'index et les empreintes de documents.

6.6 Algorithme RAG minimal (pseudo-code)

```
def rag_answer(question, athlete_id=None, k=6):
    q_vec = embed(question)
    cand = vector_index.search(q_vec, top_k=k*3)
    cand = rerank(question, cand) [:k]
    context = format_citations(cand)
    prompt = SYSTEM_INSTRUCTIONS + guardrails() + context + user(question)
    answer = llm.generate(prompt)
    return answer, citations=cand
```

7 Fondamentaux techniques : LLM et RAG

7.1 Les Grands Modèles de Langage (LLM) : une rupture paradigmique

7.1.1 Architecture Transformer et Mécanisme d'Attention

La révolution actuelle du Traitement Automatique du Langage Naturel (TALN/NLP) repose sur l'architecture *Transformer*, introduite par Vaswani et al.²³ dans l'article fondateur "Attention Is All You Need". Contrairement aux réseaux de neurones récurrents (RNN/LSTM) qui traitaient le texte séquentiellement, les Transformers permettent un traitement parallèle massif grâce au mécanisme d'attention (Self-Attention).

Cette architecture a été précédée par des modèles comme BERT²⁴, qui introduisit le pré-entraînement bidirectionnel, et GPT-1²⁵, qui démontra l'efficacité du pré-entraînement génératif. Cependant, c'est l'architecture Transformer qui a permis le passage à l'échelle avec des modèles de plusieurs centaines de milliards de paramètres.

Mathématiquement, l'attention se définit comme une fonction pondérée de trois vecteurs : Query (Q), Key (K) et Value (V). Pour chaque mot d'une phrase, le modèle calcule un score de pertinence avec tous les autres mots, permettant de capturer les dépendances à longue distance (ex : l'accord d'un adjectif avec un sujet situé 20 mots plus tôt). La formule canonique de l'attention est :

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

où d_k est la dimension des clés, servant de facteur de normalisation pour éviter la saturation des gradients.

Cette capacité à modéliser le contexte global permet aux LLM modernes (GPT-4²⁶, Claude 3²⁷, Llama²⁸) de "comprendre" les nuances sémantiques, l'ironie, et les structures logiques complexes, ce qui est indispensable pour interpréter les notes informelles d'un

23. Vaswani, A. et al. (2017). Attention Is All You Need. NeurIPS 2017.

24. Devlin, J. et al. (2019). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.

25. Radford, A. et al. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI Technical Report.

26. Achiam, J. et al. (2023). GPT-4 Technical Report. arXiv :2303.08774.

27. Anthropic (2023). Claude 3 Opus : Technical Report. <https://www.anthropic.com/clause>

28. Touvron, H. et al. (2023). LLaMA : Open and Efficient Foundation Language Models. arXiv :2302.13971.

entraîneur ("sensations moyennes", "jambes lourdes"). Les modèles open-source comme LLaMA ont démocratisé l'accès à ces capacités, permettant un déploiement hors-cloud pour des contraintes de confidentialité strictes.

7.1.2 Le paradigme Pre-training / Fine-tuning

La puissance des LLM réside dans leur processus d'apprentissage en deux temps :

1. **Le Pré-entraînement (Pre-training) :** Le modèle est exposé à des téraoctets de texte (Internet, livres, code) avec un objectif simple : prédire le mot suivant (Next Token Prediction). Cette approche, popularisée par GPT-3²⁹ avec ses 175 milliards de paramètres, démontre que l'augmentation d'échelle (scaling laws) améliore les performances de manière prévisible. C'est une phase non supervisée extrêmement coûteuse (plusieurs millions de dollars) qui permet au modèle d'acquérir une connaissance générale du monde et de la syntaxe. Les travaux de Borgeaud et al.³⁰ ont montré que l'augmentation du corpus d'entraînement à des trillions de tokens améliore significativement les performances, même sans augmentation du nombre de paramètres.
2. **L'Alignement (Fine-tuning / RLHF) :** Le modèle brut est ensuite affiné sur des jeux de données d'instructions (Instruction Tuning) et par renforcement humain (RLHF - Reinforcement Learning from Human Feedback)³¹ pour adopter un comportement utile, sûr et conversationnel. Cette étape est cruciale pour transformer un modèle génératif "brut" en assistant conversationnel fiable.

Pour notre usage en coaching, nous utilisons des modèles pré-entraînés "sur étagère" (Foundation Models) que nous spécialisons non pas par ré-entraînement (trop coûteux et rigide), mais par injection de contexte dynamique (RAG). Cette approche, introduite par Lewis et al.³², permet de combiner les capacités génératives des LLM avec la précision d'un moteur de recherche documentaire, évitant ainsi les hallucinations sur des faits spécifiques au domaine sportif.

7.1.3 Les limites intrinsèques : Hallucinations et Fenêtre de Contexte

Malgré leur puissance, les LLM souffrent de deux limitations majeures pour une application critique comme le sport :

- **Les Hallucinations :** Le modèle étant probabiliste, il peut générer des faits plausibles mais faux avec une grande assurance. En médecine ou en coaching, inventer une fréquence cardiaque ou une recommandation de repos peut être dangereux. Les rapports de sécurité d'OpenAI³³ et d'Anthropic³⁴ reconnaissent explicitement ce risque et recommandent des mécanismes de vérification.

29. Brown, T. et al. (2020). Language Models are Few-Shot Learners. NeurIPS 2020.

30. Borgeaud, S. et al. (2022). Improving language models by retrieving from trillions of tokens. arXiv :2112.04426.

31. Ouyang, L. et al. (2022). Training language models to follow instructions with human feedback. NeurIPS 2022.

32. Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.

33. OpenAI (2023). GPT-4 System Card. <https://openai.com/research/gpt-4-system-card>

34. Anthropic (2023). Claude 3 Ethics Report. <https://www.anthropic.com/clause-ethics-report>

— **La Fenêtre de Contexte** : Bien qu'en augmentation (128k tokens pour GPT-4), la mémoire à court terme du modèle reste finie. On ne peut pas lui donner l'historique complet de 10 ans d'entraînement de 50 athlètes dans un seul prompt. Les travaux de Borgeaud et al.³⁵ ont montré que l'augmentation du corpus d'entraînement améliore les performances, mais ne résout pas le problème de la fenêtre de contexte limitée.

C'est pour pallier ces limites que l'architecture RAG s'impose comme la solution de référence. Les risques éthiques et sociaux identifiés par Weidinger et al.³⁶ incluent les biais, la désinformation, et les risques de discrimination - autant de défis que notre approche RAG avec validation humaine cherche à mitiger.

7.2 Retrieval-Augmented Generation (RAG) : l'architecture de la fiabilité

Le RAG, introduit par Lewis et al.³⁷, est une technique hybride qui combine la puissance générative des LLM avec la précision d'un moteur de recherche documentaire. Au lieu de demander au modèle de répondre uniquement avec ses connaissances paramétriques (figées lors de l'entraînement), on lui fournit, au moment de la requête, les documents pertinents (plans d'entraînement, bilans médicaux, articles FFA) pour qu'il s'en serve comme source de vérité.

Cette approche s'inspire de travaux antérieurs comme REALM³⁸, qui proposait d'intégrer la récupération directement dans le pré-entraînement, et de méthodes de récupération dense comme DPR (Dense Passage Retrieval)³⁹. Les travaux d'Izacard et Grave⁴⁰ ont démontré l'efficacité de combiner récupération et génération pour des tâches de question-réponse en domaine ouvert.

7.2.1 Pourquoi le RAG plutôt que le Fine-tuning ?

Une question récurrente est : "Pourquoi ne pas ré-entraîner le modèle sur les données du club ?". La réponse tient en trois points :

1. **Fraîcheur des données** : Un modèle fine-tuné est obsolète dès la fin de son entraînement. Or, les données d'un athlète changent chaque jour (nouvelle séance, nouvelle blessure). Le RAG interroge une base de données en temps réel.
2. **Traçabilité (Explicabilité)** : Un modèle fine-tuné donne une réponse "boîte noire". Un système RAG peut dire "Je recommande ceci car j'ai trouvé cette information dans le bilan du 12 octobre (Source A) et le protocole FFA (Source B)". C'est crucial pour la confiance du coach.

35. Borgeaud, S. et al. (2022). Improving language models by retrieving from trillions of tokens. arXiv :2112.04426.

36. Weidinger, L. et al. (2021). Ethical and social risks of harm from language models. arXiv :2112.04359.

37. Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.

38. Guu, K. et al. (2020). REALM : Retrieval-Augmented Language Model Pre-Training. ICML 2020.

39. Karpukhin, V. et al. (2020). Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.

40. Izacard, G., Grave, E. (2021). Leveraging Passage Retrieval with Generative Models for Open Domain QA. EACL 2021.

3. Contrôle d'accès : Avec le RAG, on peut filtrer les documents selon les droits de l'utilisateur (un athlète ne voit pas les notes privées du coach) avant de les envoyer au LLM. Le Fine-tuning "mélange" tout dans les poids du modèle.

7.2.2 Les Embeddings : la clé de la recherche sémantique

Le cœur du RAG repose sur les *Embeddings* (plongements lexicaux). Il s'agit de transformer un texte (mot, phrase, document) en un vecteur numérique de haute dimension (ex : 1536 dimensions pour 'text-embedding-3-small' d'OpenAI).

Les modèles d'embeddings modernes, comme Sentence-BERT⁴¹, permettent de générer des représentations sémantiques riches en entraînant des réseaux siamois sur des paires de phrases. Cette approche produit des embeddings qui capturent mieux la similarité sémantique que les embeddings de mots individuels.

L'idée fondamentale est que deux textes sémantiquement proches auront des vecteurs géométriquement proches dans cet espace vectoriel. La similarité est généralement calculée par la *Cosinus Similarity* :

$$\text{sim}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Si $\text{sim}(A, B) \approx 1$, les vecteurs sont alignés (sens très proche). Si ≈ 0 , ils sont orthogonaux (sans rapport).

Dans notre cas, la requête "Exercice pour améliorer le départ en blocks" sera vectoriellement proche du document "Séance technique mise en action et poussée", même s'ils ne partagent pas exactement les mêmes mots-clés. C'est la supériorité de la recherche sémantique sur la recherche par mots-clés traditionnelle.

Des approches plus avancées comme ColBERT⁴² utilisent une interaction tardive contextualisée, où chaque token de la requête est comparé avec chaque token du document, offrant une précision supérieure au coût d'une complexité computationnelle accrue.

7.2.3 Indexation Vectorielle et Algorithmes ANN

Rechercher le vecteur le plus proche parmi des millions de documents est coûteux. Nous utilisons des algorithmes de *Approximate Nearest Neighbor* (ANN) comme HNSW (Hierarchical Navigable Small World) implémenté dans pgvector ou FAISS⁴³. FAISS, développé par Facebook AI Research, permet des recherches de similarité à l'échelle du milliard de vecteurs en exploitant les GPU, avec des optimisations comme la quantification produit (Product Quantization) pour réduire la mémoire requise.

Ces algorithmes construisent un graphe de navigation qui permet de trouver les voisins les plus proches en temps logarithmique $O(\log N)$ plutôt que linéaire, avec un compromis contrôlé entre précision et vitesse. Pour évaluer la qualité de notre système de récupération, nous utilisons des benchmarks standardisés comme BEIR⁴⁴, qui fournit une suite de

41. Reimers, N., Gurevych, I. (2019). Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks. EMNLP 2019.

42. Khattab, O., Zaharia, M. (2020). ColBERT : Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. SIGIR 2020.

43. Johnson, J., Douze, M., Jégou, H. (2017). Billion-scale similarity search with GPUs (FAISS). arXiv :1702.08734.

44. Thakur, N. et al. (2021). BEIR : A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. NeurIPS Datasets and Benchmarks 2021.

tâches de recherche d'information pour tester la généralisation des modèles.

8 Architecture RAG spécialisée pour le coaching

8.1 Pipeline d'ingestion et de Chunking

La qualité des réponses dépend directement de la qualité des données ingérées ("Garbage In, Garbage Out"). Notre pipeline de traitement se décompose en :

1. **Extraction** : Conversion des sources hétérogènes (PDF FFA, JSON Strava, Notes texte) en texte brut.
2. **Chunking (Découpage)** : C'est une étape critique. Un découpage trop fin (phrase par phrase) perd le contexte. Un découpage trop large noie l'information précise. Les travaux de Karpukhin et al.⁴⁵ sur la récupération de passages denses ont montré l'importance d'un chunking adapté à la tâche. **Notre stratégie** : Chunking sémantique par "unité de sens coaching".
 - Pour un plan : 1 chunk = 1 microcycle (semaine).
 - Pour un bilan : 1 chunk = 1 test ou 1 compétition.
 - Pour un article scientifique : 1 chunk = 1 section (Méthode, Résultats, Conclusion).

Nous utilisons une fenêtre glissante (overlap) de 20% pour ne pas couper une information importante entre deux blocs, suivant les recommandations d'Izacard et Grave⁴⁶.

3. **Enrichissement (Méta-données)** : Chaque chunk est tagué avec : ID athlète, Date, Type (Séance, Bilan, Protocole), Auteur. Ces métadonnées serviront de filtres (Pre-filtering) avant la recherche vectorielle.

8.2 Stratégie de Retrieval Hybride

Pour maximiser la pertinence, nous n'utilisons pas seulement la recherche vectorielle. Nous implémentons une recherche hybride combinant plusieurs approches complémentaires :

8.2.1 Recherche par mots-clés (BM25)

La recherche par mots-clés, utilisant l'algorithme BM25, est efficace pour les termes exacts :

- **Noms propres** : Noms d'athlètes, d'entraîneurs, de compétitions
- **Codes exercices spécifiques** : Références FFA, codes d'épreuves
- **Dates et périodes** : Requêtes temporelles précises
- **Métriques quantitatives** : Valeurs exactes de VMA, FC, distances

Cette approche est particulièrement utile pour les requêtes factuelles nécessitant une précision exacte plutôt qu'une similarité sémantique.

45. Karpukhin, V. et al. (2020). Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.

46. Izacard, G., Grave, E. (2021). Leveraging Passage Retrieval with Generative Models for Open Domain QA. EACL 2021.

8.2.2 Recherche vectorielle (Dense)

La recherche vectorielle, basée sur les embeddings sémantiques⁴⁷, excelle pour les concepts et la sémantique :

- **Concepts abstraits** : "fatigue", "progression", "récupération"
- **Descriptions qualitatives** : "sensations difficiles", "bonne forme"
- **Requêtes en langage naturel** : "Comment adapter ma séance si j'ai des douleurs aux mollets ?"

Cette approche capture les relations sémantiques entre termes, permettant de retrouver des documents pertinents même sans mots-clés exacts.

8.2.3 Fusion des résultats (Reciprocal Rank Fusion)

Les résultats des deux méthodes sont fusionnés via l'algorithme *Reciprocal Rank Fusion* (RRF) pour obtenir le meilleur des deux mondes. Le RRF combine les classements en pondérant les résultats selon leur rang dans chaque méthode, favorisant les documents qui apparaissent dans les deux listes tout en préservant les résultats pertinents spécifiques à chaque approche.

Cette stratégie hybride, inspirée des travaux de Karpukhin et al.⁴⁸ sur la récupération de passages denses, et des approches avancées comme ColBERT⁴⁹, permet d'optimiser à la fois la précision (via BM25) et le rappel (via recherche vectorielle).

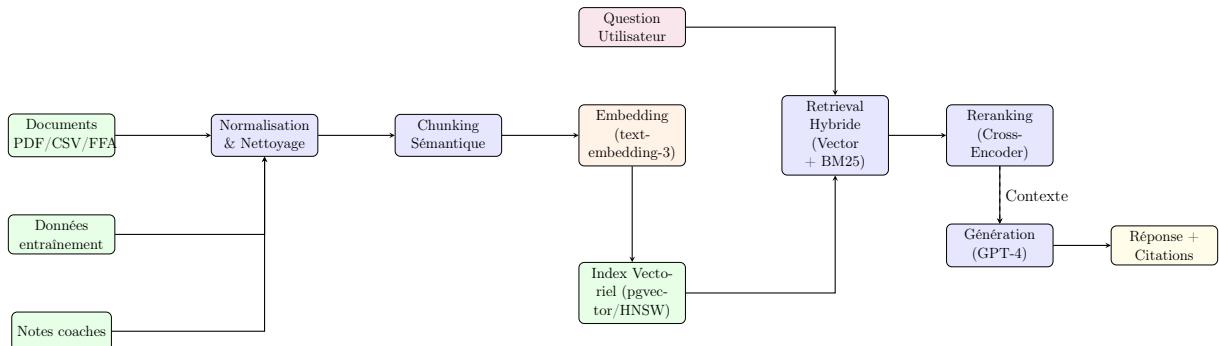


FIGURE 3 – Architecture RAG détaillée : Pipeline d'ingestion, Indexation et Flux de requête hybride

9 État de l'art : IA et Sport

L'application de l'intelligence artificielle au domaine sportif n'est pas nouvelle, mais elle a connu plusieurs vagues d'innovation successives. Nous classifions ici les travaux existants en trois grandes familles : l'IA prédictive (Machine Learning classique), la Vision par Ordinateur, et le Traitement du Langage Naturel (NLP).

47. Reimers, N., Gurevych, I. (2019). Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks. EMNLP 2019.

48. Karpukhin, V. et al. (2020). Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.

49. Khattab, O., Zaharia, M. (2020). ColBERT : Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. SIGIR 2020.

9.1 L'IA prédictive : la gestion de la charge et des blessures

C'est le domaine le plus mature. Depuis les années 2010, l'utilisation de données tabulaires (GPS, charge) pour prédire la performance ou le risque de blessure est courante dans le sport professionnel.

9.1.1 Prédition de blessures

Les travaux de Rossi et al.⁵⁰ sur le football ont démontré la possibilité de prédire les blessures sans contact avec une précision raisonnable en utilisant des arbres de décision (Random Forests) sur les données GPS. Claudino et al.⁵¹ ont réalisé une revue systématique montrant que les réseaux de neurones et les SVM (Support Vector Machines) surpassent les méthodes statistiques traditionnelles.

Les études de Colby et al.⁵² sur le football australien ont établi des corrélations significatives entre les charges dérivées d'accéléromètres et de GPS et le risque de blessure, validant l'approche multi-capteurs. Cependant, ces modèles souffrent souvent d'un problème de "boîte noire" et de faux positifs trop nombreux, rendant leur adoption difficile par les staffs médicaux qui préfèrent des règles simples comme le ratio ACWR (Acute :Chronic Workload Ratio) de Gabbett⁵³, bien que ce dernier soit aujourd'hui controversé⁵⁴.

Les travaux de Bunker et Thabtah⁵⁵ ont étendu l'approche ML au-delà de la prédition de blessures vers la prédition de résultats sportifs, démontrant la polyvalence des techniques d'apprentissage automatique dans le domaine sportif.

9.1.2 Optimisation de la performance

Dans les sports d'endurance, Mujika⁵⁶ et Bourdon et al.⁵⁷ ont travaillé sur la modélisation de la relation dose-réponse. Les approches modernes utilisent des réseaux de neurones récurrents (LSTM) pour modéliser la séquence temporelle des entraînements et prédire l'état de forme futur.

Les travaux de Jaspers et al.⁵⁸ ont établi des corrélations significatives entre différents

-
- 50. Rossi, A. et al. (2017). Effective injury forecasting in soccer with GPS training data and machine learning. *PloS One*, 12(7).
 - 51. Claudino, J.G. et al. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports : a systematic review. *Sports Medicine-Open*, 5(1), 1-12.
 - 52. Colby, M.J. et al. (2014). Accelerometer and GPS-derived running loads and injury risk in elite Australian football. *The Journal of Strength & Conditioning Research*, 28(8), 2244-2252.
 - 53. Gabbett, T.J. (2016). The training-injury prevention paradox : should athletes be training smarter and harder ? *British Journal of Sports Medicine*, 50(5), 273-280.
 - 54. Impellizzeri, F.M. et al. (2019). What role do chronic workloads play in the acute to chronic workload ratio ? Time to dismiss ACWR and its underlying theory. *Sports Medicine*, 49(3), 467-469.
 - 55. Bunker, R.P., Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15(1), 27-33.
 - 56. Mujika, I. (2017). Quantification of training and competition loads in endurance sports : methods and applications. *International Journal of Sports Physiology and Performance*, 12(s2), S2-9.
 - 57. Bourdon, P.C. et al. (2017). Monitoring athlete training loads : consensus statement. *International Journal of Sports Physiology and Performance*, 12(s2), S2-161.
 - 58. Jaspers, A. et al. (2017). Relationships between training load indicators and training outcomes in professional soccer. *Sports Medicine*, 47(3), 533-544.

indicateurs de charge d'entraînement (sRPE, distance GPS, fréquence cardiaque) et les résultats d'entraînement, validant l'approche multi-métriques. Scott et al.⁵⁹ ont comparé différentes méthodes de quantification de charge, montrant que les mesures objectives (GPS, accéléromètres) offrent une précision supérieure aux méthodes subjectives.

Buchheit⁶⁰ a exploré l'utilisation de mesures de fréquence cardiaque comme proxy de la charge interne, concluant qu'aucune métrique unique ne suffit mais qu'une combinaison de métriques offre une vision plus complète. La limite principale reste la nécessité de jeux de données massifs et propres, rarement disponibles au niveau amateur.

9.2 La Vision par Ordinateur : l'analyse technique automatisée

La seconde vague d'IA concerne l'analyse vidéo. Des outils comme OpenPose ou les solutions propriétaires (Dartfish) permettent d'extraire le squelette de l'athlète en mouvement sans marqueurs. Stein et al.⁶¹ montrent comment ces données permettent d'analyser la biomécanique du geste (foulée, lancer) avec une précision proche des laboratoires. Bien que très prometteuse, cette technologie demande des ressources de calcul importantes et des conditions de prise de vue standardisées, ce qui limite son usage quotidien par un entraîneur bénévole avec son smartphone.

9.3 Le NLP dans le sport : le chaînon manquant

C'est ici que se situe le vide académique que notre thèse ambitionne de combler. Jusqu'à très récemment (2023), le NLP dans le sport se limitait à l'analyse de sentiments sur les réseaux sociaux des fans ou au mining d'articles de presse.

L'arrivée des LLM change la donne. Les capacités émergentes des grands modèles de langage⁶², notamment le raisonnement par chaîne de pensée⁶³, ouvrent des perspectives inédites. Les expériences de Bubeck et al.⁶⁴ avec GPT-4 ont démontré des capacités de raisonnement et de synthèse jusqu'alors réservées aux humains.

Quelques travaux émergents commencent à explorer :

- **La génération de commentaires** : Résumé automatique de matchs à partir de logs d'événements.
- **Les chatbots basiques** : FAQ pour les règlements sportifs.
- **Le prompt engineering** : L'optimisation des prompts pour des tâches spécialisées, comme documenté par Liu et al.⁶⁵

59. Scott, B.R. et al. (2013). A comparison of methods to quantify the in-season training load of professional soccer players. International Journal of Sports Physiology and Performance, 8(2), 195-202.

60. Buchheit, M. (2014). Monitoring training status with HR measures : do all roads lead to Rome ? Frontiers in Physiology, 5, 73.

61. Stein, M. et al. (2019). Bring it to the pitch : combining video and movement data to enhance team sport analysis. IEEE Transactions on Visualization and Computer Graphics, 26(1), 143-153.

62. Wei, J. et al. (2022). Emergent abilities of large language models. TMLR 2022.

63. Wei, J. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS 2022.

64. Bubeck, S. et al. (2023). Sparks of AGI : early experiments with GPT-4. arXiv :2303.12712.

65. Liu, P. et al. (2021). Pre-train, prompt, and predict : a systematic survey of prompting methods in natural language processing. ACM CSUR 2023.

Cependant, il n'existe pas, à notre connaissance, de travaux approfondis sur l'utilisation d'architectures RAG pour l'assistance au coaching technique et la planification personnalisée. Les solutions existantes sont soit des "wrappers" superficiels autour de ChatGPT (sans contexte réel), soit des systèmes experts rigides sans capacité conversationnelle.

Notre contribution se situe précisément à cette intersection : utiliser la capacité de raisonnement et de synthèse des LLM pour traiter la complexité multivariée des données sportives (physio + ressenti + contexte) et la restituer sous une forme intelligible et actionnable pour l'entraîneur. Cette approche s'inscrit dans le panorama plus large des LLM, tel que synthétisé par Zhao et al.⁶⁶ dans leur revue exhaustive du domaine.

9.4 Synthèse et positionnement

Le tableau ci-dessous résume l'évolution des technologies et positionne notre approche :

Ère	Technologie dominante	Type de données	Usage principal
1.0 (1990-2010)	Excel, Bases de données	Tabulaires manuelles	Stockage, Rétrospective
2.0 (2010-2022)	Machine Learning, GPS	Séries temporelles, Capteurs	Prédiction, Quantification
3.0 (2023-...)	IA Générative (LLM)	Non structurées (Texte, Audio)	Assistance, Synthèse, Dialogue

Nous nous inscrivons résolument dans cette ère 3.0, cherchant à unifier les données quantitatives de l'ère 2.0 avec l'intelligence qualitative des LLM.

10 Modèle de données : structuration de la connaissance athlétique

10.1 Architecture générale du modèle de données

Notre modèle de données s'inspire des structures de la Fédération Française d'Athlétisme⁶⁷ tout en l'enrichissant pour supporter les besoins spécifiques de l'assistance par IA. L'architecture générale suit un modèle relationnel normalisé, avec des extensions pour le stockage vectoriel des contenus textuels.

La pertinence d'un système RAG repose avant tout sur la qualité et la structuration des données qu'il indexe. Contrairement à une approche "Big Data" non structurée, nous avons opté pour un modèle relationnel fort, inspiré des standards de la Fédération Française d'Athlétisme (FFA), permettant de capturer la complexité du suivi sportif.

66. Zhao, W.X. et al. (2023). A survey of large language models. arXiv :2303.18223.

67. Fédération Française d'Athlétisme (référentiels publics, structures d'épreuves). <https://www.athle.fr/>

10.2 Principes de modélisation

Notre modèle de données repose sur plusieurs principes fondamentaux inspirés des meilleures pratiques en modélisation de données sportives et des travaux de Bourdon et al.⁶⁸ sur le monitoring des charges d'entraînement.

10.2.1 Principe de séparation des préoccupations

Le modèle distingue clairement trois niveaux de préoccupations :

- **Niveau métier** : Entités métier (athlètes, séances, compétitions) avec leurs attributs sémantiques
- **Niveau technique** : Structures de stockage optimisées (indexation, partitions) pour la performance
- **Niveau analytique** : Vues et agrégations pré-calculées pour l'analyse et la visualisation

Cette séparation permet d'optimiser chaque niveau indépendamment sans compromettre la cohérence globale.

10.2.2 Principe de traçabilité temporelle

Toutes les entités critiques (séances, bilans, recommandations IA) incluent des timestamps précis (création, modification, suppression) permettant une traçabilité complète. Cette traçabilité est essentielle pour l'audit, la conformité RGPD⁶⁹, et l'analyse rétroactive des décisions.

10.2.3 Principe d'extensibilité

Le modèle est conçu pour être extensible, permettant l'ajout de nouveaux types de données (biomarqueurs, données de sommeil, etc.) sans restructuration majeure. Cette extensibilité est cruciale compte tenu de l'évolution rapide des technologies de monitoring sportif.

10.3 Choix architecturaux et justification

L'architecture technique de notre solution résulte d'un arbitrage entre plusieurs contraintes : performance, sécurité, maintenabilité, et coût. Cette section justifie nos choix techniques en les confrontant aux alternatives disponibles.

10.3.1 Architecture microservices vs monolithique

Nous avons opté pour une architecture modulaire plutôt que microservices pure, pour plusieurs raisons :

- **Simplicité opérationnelle** : Les clubs amateurs n'ont pas les ressources pour gérer une infrastructure complexe

68. Bourdon, P.C. et al. (2017). Monitoring athlete training loads : consensus statement. International Journal of Sports Physiology and Performance, 12(s2), S2-161.

69. Règlement (UE) 2016/679 (RGPD). <https://eur-lex.europa.eu/>

- **Latence réduite** : Communication inter-services minimisée, cruciale pour l’expérience utilisateur temps réel
- **Cohérence transactionnelle** : Gestion simplifiée des transactions distribuées

Cependant, la modularité interne permet une évolution future vers une architecture microservices si nécessaire, avec des interfaces clairement définies entre les modules.

10.3.2 Choix du langage et des frameworks

Le choix de Python pour le backend s’impose par la richesse de l’écosystème IA (LangChain, OpenAI SDK, scikit-learn, pandas) et la facilité d’intégration avec les bibliothèques de traitement de données. FastAPI offre des performances comparables à Node.js tout en bénéficiant de la maturité de l’écosystème Python.

Pour le frontend, React/Next.js offre un bon compromis entre performance, écosystème riche, et facilité de développement. L’hydratation côté serveur (SSR) de Next.js améliore les temps de chargement initiaux, crucial pour l’adoption par des utilisateurs non-technophiles.

10.4 Backend : FastAPI et Asynchronisme

Le choix de Python comme langage backend s’impose par la richesse de son écosystème IA (LangChain, OpenAI SDK, Pandas). Nous utilisons le framework **FastAPI** pour plusieurs raisons :

1. **Performance Asynchrone** : Les appels aux LLM (OpenAI API) sont des opérations I/O bound qui peuvent prendre plusieurs secondes. L’architecture `async/await` de FastAPI permet de ne pas bloquer le serveur pendant ces attentes, garantissant une scalabilité élevée même avec de nombreux utilisateurs simultanés.
2. **Typepage Fort (Pydantic)** : La validation stricte des données en entrée et sortie est cruciale pour éviter d’envoyer des données corrompues au LLM, ce qui augmenterait le risque d’hallucinations.
3. **Documentation Automatique** : La génération native de Swagger UI facilite l’interaction avec les équipes frontend et les tests d’intégration.

10.5 Base de Données : PostgreSQL comme Pivot

Plutôt que de multiplier les services (une base SQL, une base Vectorielle type Pinecone, une base NoSQL), nous avons fait le choix de la consolidation autour de **PostgreSQL**. L’extension `pgvector` permet de gérer les embeddings directement dans la base relationnelle. **Avantages** :

- **Cohérence Transactionnelle (ACID)** : Si on supprime un athlète, ses embeddings sont supprimés dans la même transaction. Pas de risque de désynchronisation.
- **Simplification de l’Infrastructure** : Un seul service à sauvegarder, monitorer et sécuriser.
- **Coût** : Évite les coûts souvent prohibitifs des bases vectorielles managées spécialisées.

10.6 Frontend : React et Expérience Utilisateur

L'interface utilisateur (développée en React/Next.js) ne doit pas être une simple chatbox. Elle intègre l'IA de manière contextuelle :

- **Composants "Smart"** : Un composant "Calendrier" qui peut afficher les suggestions de séances du LLM en superposition (ghost events) avant validation.
- **Feedback Visuel** : Indicateurs de chargement (streaming tokens) pour gérer l'attente psychologique lors de la génération de texte long.
- **Visualisation de Données** : Utilisation de bibliothèques comme Recharts pour afficher les courbes de charge (modèle Fitness-Fatigue de Banister) générées par le backend.

10.7 Sécurité et Confidentialité (Privacy by Design)

Le traitement de données de santé (blessures, fréquence cardiaque) impose une rigueur absolue, conforme au RGPD.

- **Pseudonymisation** : Avant d'envoyer des données au LLM (OpenAI), les identifiants personnels (Nom, Prénom) sont remplacés par des UUIDs opaques. Le mapping inverse se fait uniquement en local sur le serveur.
- **Zero-Data Retention (API)** : Nous utilisons les endpoints "Enterprise" ou configurons explicitement l'option "Zero Retention" pour garantir que nos données ne servent pas à entraîner les modèles publics d'OpenAI.
- **Chiffrement** : Chiffrement TLS 1.3 en transit et AES-256 au repos pour la base de données.

11 Dimension Humaine et Éthique de l'IA

L'introduction d'algorithmes dans la relation intime entre un entraîneur et son athlète n'est pas neutre. Au-delà de la technique, elle soulève des questions fondamentales sur la responsabilité, l'autonomie et la nature même du coaching.

11.1 Le "Nudge" Éthique : Orienter sans Contraindre

Nous adoptons l'approche du "Nudge" (Thaler & Sunstein, 2008) : l'IA doit encourager les bonnes pratiques sans jamais forcer la main. Par exemple, si un entraîneur planifie une semaine de surcharge intense, l'IA ne doit pas bloquer l'action (ce qui serait infantilisant), mais afficher une "alerte douce" : *"Attention, cette charge dépasse de 20% la moyenne des 4 dernières semaines. Risque de blessure accru selon le modèle ACWR."* L'entraîneur reste le décideur final ("Human-in-the-loop"), mais sa décision est éclairée par une analyse objective des données.

11.2 Responsabilité Juridique et "Black Box"

Qui est responsable si un athlète se blesse en suivant un plan généré par l'IA ?

- **Le problème de la boîte noire** : Les réseaux de neurones profonds sont opaques. Il est difficile d'expliquer pourquoi le modèle a suggéré telle intensité.

- **Notre réponse : la Traçabilité (Audit Logs)** : Chaque suggestion de l'IA est enregistrée avec le contexte exact (prompt, version du modèle, documents sourcés). De plus, l'interface oblige l'entraîneur à valider explicitement (clic "Approuver") chaque proposition. Juridiquement, cela maintient la responsabilité sur l'humain qualifié, l'IA n'étant qu'un outil d'aide à la décision, au même titre qu'un chronomètre ou un fichier Excel.

11.3 Risque de Dépendance et Atrophie des Compétences

Un risque majeur à long terme est la "désempowerment" des entraîneurs. Si l'IA suggère toujours des séances parfaites, l'entraîneur novice risque de ne plus faire l'effort d'analyser et de comprendre la physiologie de l'entraînement. Il devient un simple exécutant. Pour contrer cela, notre système intègre une dimension **Pédagogique** :

- **Mode "Pourquoi ?"** : Chaque suggestion est accompagnée d'une justification explicable ("Je propose du seuil car l'athlète n'en a pas fait depuis 10 jours et sa compétition est dans 3 semaines").
- **Formation continue** : Le système RAG peut aussi servir d'outil d'auto-formation, permettant au coach de poser des questions théoriques ("Rappelle-moi les principes de la méthode Tabata") sans lien avec un athlète précis.

11.4 Biais Algorithmiques et Équité Sportive

Les modèles de langage sont entraînés sur des corpus majoritairement anglophones et occidentaux. Ils peuvent véhiculer des biais culturels ou de genre. Les travaux de Weidinger et al.⁷⁰ ont identifié de multiples risques éthiques et sociaux associés aux modèles de langage, incluant les biais de représentation, la désinformation, et les risques de discrimination.

Bommasani et al.⁷¹ ont analysé en profondeur les opportunités et risques des modèles foundation, soulignant notamment les biais systémiques qui peuvent être amplifiés à grande échelle. Mittelstadt et al.⁷² ont cartographié le débat sur l'éthique des algorithmes, identifiant les biais comme l'un des enjeux centraux.

Dans le sport, cela peut se traduire par des stéréotypes : associer systématiquement "force" aux hommes et "souplesse" aux femmes, ou privilégier des méthodes d'entraînement américaines (focalisées sur les chiffres) au détriment d'approches plus intuitives ou européennes. Les travaux de Butrym⁷³ sur les "podiums posthumains" et les récits cyborg des athlètes d'élite interrogent la frontière entre technologie et nature dans le sport, questionnant les biais potentiels dans la représentation des athlètes.

López Frías⁷⁴ a analysé les défis et implications éthiques de la technologie dans le

70. Weidinger, L. et al. (2021). Ethical and social risks of harm from language models. arXiv :2112.04359.

71. Bommasani, R. et al. (2021). On the opportunities and risks of foundation models. arXiv :2108.07258.

72. Mittelstadt, B.D. et al. (2016). The ethics of algorithms : Mapping the debate. *Big Data & Society*, 3(2).

73. Butrym, T.M. (2003). Posthuman podiums : Cyborg narratives of elite track and field athletes. *Sociology of Sport Journal*, 20(1), 17-39.

74. López Frías, F.J. (2020). Technology in sport : Challenges and ethical implications. *Journal of the Philosophy of Sport*, 47(2), 219-235.

sport, soulignant les risques d'inégalité d'accès et de biais algorithmiques. Eubanks⁷⁵ a documenté comment les outils high-tech peuvent automatiser l'inégalité, profilant et pénalisant certaines populations - un risque à considérer sérieusement dans le contexte sportif où l'accès aux technologies peut créer des fractures.

Notre stratégie de **RAG** limite ce biais en forçant le modèle à se baser sur les documents de référence de la FFA⁷⁶ (culture française de l'athlétisme) plutôt que sur sa connaissance générique d'internet.

11.5 Acceptabilité technologique et adoption

L'adoption de l'IA dans le coaching sportif ne dépend pas uniquement de la qualité technique, mais également de l'acceptabilité par les utilisateurs. Le modèle d'acceptation technologique (TAM - Technology Acceptance Model), étendu par Venkatesh et Davis⁷⁷, identifie deux facteurs clés : l'utilité perçue et la facilité d'utilisation perçue.

Pour notre système, l'utilité perçue dépend de la capacité à démontrer un gain concret : réduction du temps administratif, amélioration de la qualité des plans, prévention des blessures. La facilité d'utilisation implique une interface intuitive, des explications claires, et un apprentissage progressif sans courbe d'adoption trop raide.

Les résistances potentielles incluent la peur de la déshumanisation, la méfiance vis-à-vis de l'IA, et la crainte de perdre le contrôle. Notre approche "Human-in-the-loop" avec validation systématique des recommandations adresses ces préoccupations en maintenant l'entraîneur comme décideur final.

11.6 Conclusion sur la dimension humaine

L'IA ne remplacera pas l'entraîneur. Mais l'entraîneur qui utilise l'IA remplacera celui qui ne l'utilise pas. Notre objectif est de créer une "Intelligence Augmentée" où la machine gère la complexité des données pour libérer du temps humain : temps d'écoute, d'observation terrain, et de soutien psychologique, qui restent hors de portée de tout algorithme.

Biais de représentation : Les données historiques reflètent les populations sur-représentées dans les corpus d'entraînement. **Exemple** : un modèle entraîné principalement sur données d'athlètes masculins adultes peut mal performer pour femmes ou jeunes. Une **périodisation biaisée** voit les plans optimisés pour athlètes saison d'hiver (track indoor) sous-performer pour coureurs de trail dont la saison diffère. La **minimisation** passe par stratification des données par profils (sexe, âge, discipline), validation sur échantillons équilibrés, détection d'écart de performance par sous-population.

Biais de confirmation : L'IA perpétue les méthodes conventionnelles sans explorer d'alternatives innovantes. **Exemple** : tendance à proposer uniquement des formats d'entraînement "classiques" (30/30, VMA classique) sans alternatives modernes (HIIT, polarisation). La **minimisation** implique diversification du corpus RAG avec sources innovantes, évolution des recommandations selon retour terrain, encouragement à l'exploration

75. Eubanks, V. (2018). Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

76. Fédération Française d'Athlétisme (référentiels publics, structures d'épreuves). <https://www.athle.fr/>

77. Venkatesh, V., Davis, F.D. (2000). A theoretical extension of the technology acceptance model : Four longitudinal field studies. Management Science, 46(2), 186-204.

contrôlée.

Biais d'optimisation : priorisation de métriques aisément quantifiables au détriment d'aspects qualitatifs cruciaux. **Exemple** : sur-optimisation de la VMA mesurable au détriment de la technique gestuelle ou de la gestion mentale de course. La **minimisation** nécessite pondération multi-objectifs dans les algorithmes, inclusion de métriques qualitatives (RPE, bien-être, motivation), validation humaine systématique.

Biais d'ancrage : l'historique d'un athlète "verrouille" ses recommandations futures, limitant la créativité. **Exemple** : un athlète habitué à 4 séances/semaine reçoit toujours ce pattern sans considération de nouvelles contraintes ou opportunités. La **minimisation** requiert réévaluations périodiques des profils, questionnement systématique des hypothèses sous-jacentes, création de variantes exploratoires.

Stratégies de détection et mitigation

Audits de biais réguliers : procédures systématiques d'évaluation de l'équité algorithmique. Elles incluent analyses statistiques des recommandations par sous-population (parity testing), comparaison des taux d'acceptation vs rejet des suggestions IA selon profils, métriques de diversité (variété des types de séances proposées, exploration vs exploitation), et tests adversariaux (injection de cas limites pour évaluer robustesse).

Comité de validation éthique : instance indépendante chargée de superviser l'éthique du système IA. Sa composition rassemble DPO, représentant athlètes, coach expérimenté, expert éthique IA, et médecin du sport. Il assure revues trimestrielles des incidents et patterns suspects, validation des modifications de prompts ou de modèles avant déploiement, et canal de signalement anonyme pour biais perçus par utilisateurs.

11.7 Transparence et explicabilité

Pourquoi l'explicabilité est cruciale en coaching

Un entraîneur ne peut pas accepter une recommandation qu'il ne comprend pas. L'explicabilité renforce la confiance, facilite l'apprentissage, et permet la détection d'erreurs.

Principes d'explicabilité

Explicabilité de la donnée : présentation claire des sources et de la qualité. L'affichage des données utilisées suit le format : "Recommandation basée sur 45 séances historiques, dernière VMA : 18.2 km/h (03/01/2025)". Les niveaux de confiance s'expriment ainsi : "Probabilité élevée (85%) que cette séance corresponde à vos objectifs". Les gaps de données sont explicites : "Données insuffisantes sur votre récupération nocturne (moniteur sommeil non synchro)".

Explicabilité du raisonnement : révèle la chaîne logique menant à la recommandation. Les citations obligatoires précisent : "Référence : Protocole FFA VMA 2023, étude du Dr. X 2022, patterns observés chez 12 sprinteuses similaires". Les variables clés mises en évidence indiquent : "Recommandation influencée par : progression récente (+5% VMA), charge chronique modérée, compétition dans 3 semaines". Les alternatives explorées précisent : "Autres options considérées : séances polarisées, intervalles courts ; rejetées car inadéquates pour vos objectifs actuels".

Explicabilité de l'incertitude : communique honnêtement les limites. Les scénarios probabilistes se présentent ainsi : "Probabilité de succès 70%, risques de surentrainement 15%, bénéfice marginal 15%". Les intervalles de confiance précisent : "Gain VMA attendu : +2 à +5%, avec 95% confiance". Les candidats multiples s'énoncent : "3 approches équivalentes identifiées, choix selon préférences athlète".

Techniques d'implémentation

LIME (Local Interpretable Model-agnostic Explanations)⁷⁸ :

Pour chaque recommandation, identification des facteurs les plus influents via perturbations locales. Présentation lisible : "Cette séance est recommandée car (1) votre charge actuelle est faible, (2) votre progression récente est excellente, (3) votre historique montre une bonne adaptation à ce format."

SHAP (SHapley Additive exPlanations)⁷⁹ :

Attribution précise de la contribution de chaque variable à la recommandation finale. Permet de quantifier : "La VMA actuelle influence à +40%, l'âge à -10%, la charge chronique à +35%, l'historique de blessures à -15%."

Narrative explanations :

Génération automatique de textes compréhensibles résumant le raisonnement. Exemple : "Après analyse de vos 30 dernières séances, nous recommandons un affûtage progressif sur 10 jours. Votre charge chronique est élevée mais stable, vos performances récentes sont à la hausse, et votre objectif de compétition est dans 2 semaines. Cette approche a montré un taux de succès de 78% chez des athlètes de profil comparable."

11.8 Dimension relationnelle et pédagogique

L'IA ne remplace pas la conversation humaine

Le coaching efficace repose sur des interactions riches, contextuelles, et adaptatives que l'IA ne peut pas reproduire.

Les limites de l'IA conversationnelle : l'IA ignore les subtilités émotionnelles, le langage corporel, les non-dits, révélant une **contextualisation incomplète**. Elle manque de **capacité réelle à "ressentir" la détresse ou l'enthousiasme de l'athlète**, absence d'empathie authentique. Son **manque de créativité spontanée** l'empêche de rebondir sur une question inattendue ou une analogie pertinente. Elle ne construit pas une confiance progressive basée sur des années d'expériences partagées, manifestant une **immaturité relationnelle**.

Le rôle irremplaçable de l'entraîneur : l'entraîneur humain excelle dans des dimensions que l'IA ne capte pas. La **Motivation adaptative** consiste à sentir quand pousser vs soutenir, ajuster son approche au caractère unique de chaque athlète. La **Gestion de crise** couvre l'accompagnement lors de blessure, échec, doute existentiel, conflits personnels. La **Transmission de passion** communique l'amour du sport, l'éthique de travail, les valeurs de dépassement. Le **Leadership situationnel** inspire un groupe, crée une dynamique collective, gère les rivalités constructives. L'**Innovation pragmatique** invente de nouveaux exercices sur le moment, adapte aux conditions météo imprévues, lit les signes non verbaux.

L'IA comme soutien à la pédagogie

L'IA peut néanmoins enrichir la pédagogie de l'entraîneur de plusieurs manières.

Facilitation des explications techniques

Un athlète demande : "Pourquoi dois-je varier mes séances entre endurance fondamentale et VMA ?"

Explication coach seule : "C'est important pour progresser... tu vois, il faut équilibrer..." (approximative si manque de temps/formation)

78. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. KDD 2016.

79. Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. NeurIPS 2017.

Avec support IA : l'IA génère instantanément une explication vulgarisée et illustrée. L'analogie précise : "Imagine construire une maison : l'endurance fondamentale pose les fondations (système cardiovasculaire), la VMA érige les murs (puissance aérobie). Sans fondations, les murs s'effondrent." Les données scientifiques indiquent : "95% des coureurs d'élite font 80% de leur volume en endurance fondamentale, 20% en intensité (polarisation)". Les exemples concrets montrent : "Sarah progresse de 56s à 54s sur 400m après 3 mois de ce pattern".

L'entraîneur adapte, personnalise, et valide cette base enrichie.

Personnalisation des contenus pédagogiques

Selon le profil de l'athlète (niveau, style d'apprentissage, préférences), l'IA propose des formats adaptés : pour l'athlète visuel (graphiques de progression, schémas biomécaniques, vidéos illustratives), pour l'athlète kinesthésique (exercices pratiques, sensations corporelles à rechercher), pour l'athlète rationnel (données quantitatives, ratios, formules physiologiques), pour l'athlète intuitif (analogies, métaphores sportives, exemples de champions).

Feedback constructif et non-jugeant

L'IA peut formuler des retours objectifs dépourvus d'affect. Le reflet impartial indique : "Tu as réalisé 85% du volume prévu cette semaine, en dessous de ton objectif de 90%". Le focus sur l'action suggère : "Pour rattraper, tu pourrais ajouter une séance de récupération active dimanche". L'apprentissage rétroactif note : "Analyse rétroactive : les 3 séances manquées étaient toutes prévues le matin. Peut-être réorganiser vers le soir ?".

L'entraîneur ajoute ensuite l'empathie humaine : "Je comprends que tu as traversé une période difficile, ne te mets pas trop de pression..."

11.9 Gestion des conflits et désaccords

Quand l'IA et l'entraîneur divergent

Les désaccords entre recommandations IA et intuition coach sont inévitables et même souhaitables (signal que le système n'est pas en pilote automatique).

Procédure de résolution de conflit

Étape 1 : Clarification des positions : l'IA expose ses indices factuels et niveau de confiance, l'entraîneur verbalise son intuition, expérience et informations contextuelles non captées par l'IA, et une documentation pour traçabilité est établie : "Conflit détecté le [DATE] sur recommandation séance VMA pour athlète [ID]".

Étape 2 : Escalade si divergence majeure : consulter collègues expérimentés ou comité technique du club, recourir à expertise externe (entraîneur fédéral, médecin du sport), conduire une enquête approfondie sur les données sous-jacentes (qualité, complétude, interprétation), et mettre en œuvre du red teaming où un expert indépendant "attaque" les hypothèses des deux positions.

Étape 3 : Décision informée : priorité à l'humain si justifiée (données contextuelles manquantes, intuition éprouvée historiquement), priorité à l'IA si facteurs objectifs solides et intuition coach non fondée, compromis pragmatique (fusion des recommandations : structure IA + ajustements coach), et monitoring renforcé post-décision pour apprentissage rétroactif.

Analyse rétroactive et amélioration

Chaque conflit résolu alimente l'apprentissage du système : causalité identifiée (pourquoi divergence ? données manquantes, biais, sur-interprétation IA), ajustements (corrections des prompts, enrichissement corpus RAG, mise à jour des règles métier), partage de

savoir (documentation des cas pour éviter répétition), et validation par les pairs (présentation du cas en formation continue).

11.10 Formation à la pensée critique vis-à-vis de l'IA

Éviter les écueils : surconfiance vs défiance excessive

Deux attitudes extrêmes sont néfastes : la surconfiance aveugle en l'IA ("l'IA a toujours raison") ou la défiance systématique ("l'IA est inutile, je sais mieux").

Surconfiance : dangers et mitigation

Risques : désengagement cognitif (l'entraîneur cesse de réfléchir et suit passivement), perte d'expertise (compétences non exercées s'atrophient progressivement), erreurs en cascade (une suggestion IA erronée non contrôlée entraîne des dommages), et blame shifting (dilution de responsabilité si mauvaise décision).

Mitigation : formation obligatoire aux limites de l'IA, obligation légale de validation humaine (charte d'usage contraignante), exercices réguliers de confrontation (imposer des cas où l'IA se trompe délibérément), rotation de périodes "sans IA" (1 semaine/mois sans assistant pour maintenir l'acuité décisionnelle), et évaluation annuelle (tests de compétences coach sans soutien IA).

Défiance excessive : identification et correction

Symptômes : rejet systématique des suggestions sans justification, temps passés à contourner l'IA supérieur aux gains attendus, résistance passive-agressive ("Je l'utilise pour faire plaisir, mais ça ne marche pas"), et négligence de la formation (refus d'apprendre les bons usages).

Correction : mentorat par pairs (observation d'entraîneurs efficaces avec l'IA), preuves tangibles (démonstration de gains de temps/qualité sur cas concrets), gradualisme (introduction progressive des fonctionnalités, pas tout d'un coup), support bienveillant (accompagnement dans la courbe d'apprentissage), et reconnaissance (valorisation des pratiques innovantes réussies).

Pensée critique équilibrée : le sweet spot

L'attitude optimale combine confiance prudentielle et scepticisme constructif : confiance dans les capacités IA pour tâches répétitives/factuelles, scepticisme systématique pour recommandations complexes/hauts enjeux, validation croisée (recouper suggestions IA avec autres sources : littérature, confrères), expérimentation contrôlée (tester les recommandations sur cas mineurs avant généralisation), et apprentissage continu (adapter sa confiance selon retour d'expérience).

11.11 Stratégies de déploiement et adoption progressive

L'adoption de l'IA dans les clubs d'athlétisme nécessite une stratégie de déploiement progressif et adaptée aux contraintes organisationnelles. Le modèle d'acceptation technologique (TAM)⁸⁰ identifie l'utilité perçue et la facilité d'utilisation comme facteurs clés d'adoption.

80. Venkatesh, V., Davis, F.D. (2000). A theoretical extension of the technology acceptance model : Four longitudinal field studies. *Management Science*, 46(2), 186-204.

11.11.1 Phase de sensibilisation

La première phase consiste à sensibiliser les entraîneurs aux bénéfices potentiels de l'IA sans créer d'attentes irréalistes. Des ateliers de démonstration, des retours d'expérience de clubs pilotes, et des sessions de formation permettent de construire une compréhension commune des opportunités et des limites.

11.11.2 Phase de pilote contrôlé

La phase de pilote permet de valider l'architecture sur un périmètre restreint (1-2 clubs, 5-10 athlètes) avant généralisation. Cette approche permet d'identifier les problèmes techniques et organisationnels en conditions réelles, d'ajuster les fonctionnalités selon les retours utilisateurs, et de construire une base de cas d'usage validés.

11.11.3 Phase de généralisation progressive

La généralisation s'effectue progressivement par vagues successives, en commençant par les clubs les plus motivés et technophiles, puis en étendant aux clubs plus réticents avec un accompagnement renforcé. Cette approche permet de capitaliser sur les succès initiaux et de construire une communauté d'utilisateurs actifs.

11.12 Impacts organisationnels et culturels

Évolution des rôles et compétences

L'introduction de l'IA modifie nécessairement les profils de poste et les attentes vis-à-vis des entraîneurs.

Nouvelles compétences requises

Les entraîneurs doivent développer plusieurs compétences clés. La **Littératie numérique** consiste à comprendre les bases de l'IA, formuler de bonnes requêtes, interpréter les sorties. La **Pensée critique augmentée** exerce le jugement éclairé face aux suggestions IA. La **Gestion de l'incertitude** travaille avec probabilités, intervalles de confiance, scénarios multiples. La **Communication transparente** explique l'origine des décisions mixtes (IA + humain) aux athlètes. La **Flexibilité cognitive** bascule entre intuition humaine et logique computationnelle selon contexte.

Évolution des processus de formation

Les cursus de formation d'entraîneur doivent intégrer ces dimensions : modules obligatoires sur éthique IA, pensée critique, management des biais ; ateliers pratiques sur formulation de prompts, analyse de recommandations ; cas d'usage réels confrontant à des situations limites, débriefings ; certifications progressives via badges de compétence IA coaching ; formation continue incluant veille technologique et retour d'expérience communautaire.

Gestion de la résistance au changement

Une proportion significative d'entraîneurs résistera à l'introduction de l'IA par crainte légitime d'obsolescence ou de perte de sens.

Identification des résistances

Quatre types de résistances sont identifiés : la **Résistance affective** ("Je n'ai pas envie, c'est trop compliqué") révèle peur de l'inconnu et perte de confort ; la **Résistance cognitive** ("Ça ne marchera jamais, c'est un gadget") traduit incompréhension des bénéfices ; la **Résistance comportementale** ("J'ai toujours fait comme ça") manifeste

inertie organisationnelle ; la **Résistance axiologique** ("Ce n'est pas mon métier, je suis coach, pas informaticien") signale identité professionnelle menacée.

Stratégies d'accompagnement : cinq leviers d'action sont mobilisés. L'**Écoute empathique** prend au sérieux les craintes, sans les nier ou minimiser. Les **Preuves progressives** offrent démonstrations concrètes de valeur sur cas simples. Les **Champions locaux** identifient ambassadeurs enthousiastes pour influence par pairs. La **Participation active** co-conçoit les fonctionnalités avec les utilisateurs finaux. La **Reconnaissance et valorisation** met en avant les expertises traditionnelles complétées par IA.

12 Architecture de sécurité et conformité

La sécurité des données de santé et la conformité réglementaire sont des prérequis absolus pour le déploiement de notre solution. Cette section détaille les mécanismes de sécurité intégrés à chaque niveau de l'architecture.

12.1 Sécurité des données en transit et au repos

12.1.1 Chiffrement des communications

Toutes les communications entre le client et le serveur utilisent TLS 1.3 avec des certificats validés par une autorité de certification reconnue. Les API calls vers les services LLM externes (OpenAI, Anthropic) sont également chiffrées, garantissant que les données pseudonymisées ne peuvent pas être interceptées en transit.

12.1.2 Chiffrement des données au repos

Les données sensibles stockées dans PostgreSQL sont chiffrées au repos using AES-256. Les clés de chiffrement sont gérées via un système de gestion de clés (Key Management Service), avec rotation automatique tous les 90 jours. Les backups sont également chiffrés avant stockage sur des systèmes distants.

12.2 Gestion des accès et authentification

12.2.1 Authentification multi-facteurs

L'accès au système requiert une authentification forte avec support de l'authentification multi-facteurs (MFA). Les entraîneurs et athlètes doivent s'authentifier via un système centralisé (OAuth 2.0 / OpenID Connect) avec possibilité d'ajouter un second facteur (SMS, application d'authentification).

12.2.2 Contrôle d'accès basé sur les rôles (RBAC)

Le système implémente un contrôle d'accès granulaire basé sur les rôles :

- **Administrateur club** : Accès complet aux données du club
- **Entraîneur** : Accès aux données de ses athlètes assignés, possibilité de générer des plans
- **Athlète** : Accès en lecture à ses propres données, possibilité de saisir RPE et notes
- **Staff médical** : Accès aux données de santé avec consentement explicite

Ce contrôle d'accès est appliqué à la fois au niveau de l'application (frontend) et au niveau de la base de données (row-level security), garantissant qu'un utilisateur ne peut accéder qu'aux données autorisées même en cas de contournement de l'interface.

12.3 Conformité RGPD et protection des données personnelles

Conformément au Règlement Général sur la Protection des Données⁸¹, notre architecture intègre les principes de "Privacy by Design" et "Privacy by Default".

12.3.1 Droits des personnes concernées

Le système permet l'exercice de tous les droits RGPD :

- **Droit d'accès** : Export des données personnelles en format structuré (JSON, CSV)
- **Droit de rectification** : Modification des données incorrectes via l'interface
- **Droit à l'effacement** : Suppression complète des données avec confirmation
- **Droit à la portabilité** : Export des données dans un format interopérable
- **Droit d'opposition** : Possibilité de refuser le traitement de certaines données

12.3.2 Minimisation des données

Le principe de minimisation est appliqué à tous les niveaux : seules les données strictement nécessaires sont collectées, les données sont conservées uniquement pendant la durée nécessaire, et les données obsolètes sont automatiquement archivées ou supprimées selon une politique de rétention définie.

La CNIL⁸² recommande spécifiquement la minimisation des données dans les systèmes d'IA, soulignant l'importance de ne collecter que les données pertinentes pour l'objectif poursuivi.

13 Exigences détaillées et critères de succès

13.1 Exigences fonctionnelles

Les exigences fonctionnelles couvrent quatre domaines principaux : gestion des profils (athlète, coach, staff), cycles et séances (planifiées réalisées), charges et tests ; assistant RAG avec citations obligatoires et mode *no-answer* paramétrable ; génération de micro-cycles avec paramètres (disponibilités, épreuves cibles, historique) ; reporting automatique (hebdo mensuel) validé par le coach.

81. Règlement (UE) 2016/679 (RGPD). <https://eur-lex.europa.eu/>

82. CNIL (2023). IA et données personnelles : recommandations et ressources. <https://www.cnil.fr/>

13.2 Exigences non-fonctionnelles

Risque	Probabilité	Impact	Mesures de mitigation	Responsable
Hallucinations	Moy	Moy	RAG obligatoire, citations, <i>no-answer</i> , revue coach	Tech lead
Fuite PII	Faible	Élevé	Filtrage, chiffrement, logs, revues sécurité	DPO
Rejet utilisateur	Moy	Moy	Formation, champions, explication bénéfices	Coach principal
Coûts API	Moy	Moy	Caches, limites, modèles locaux si besoin	Responsable SI
Qualité données	Élevée	Moy	Normalisation, contrôles, data steward	Data steward

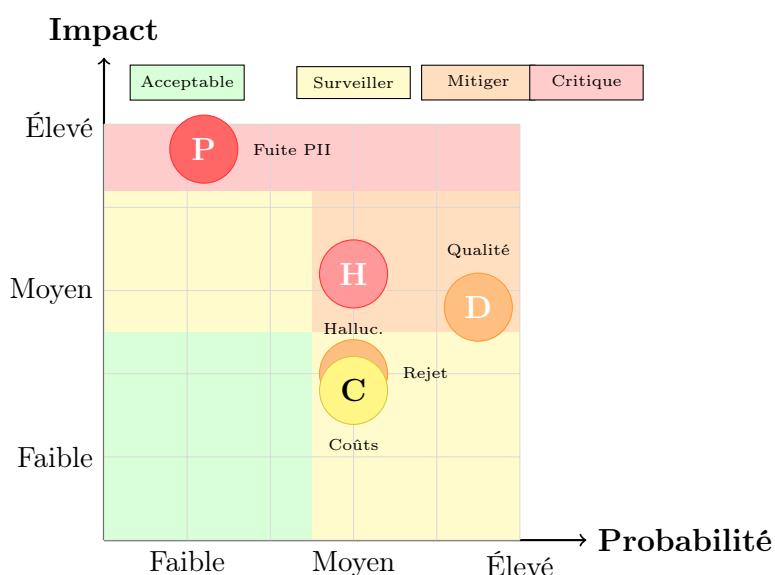


FIGURE 4 – Matrice de risques projet. **H**=Hallucinations, **P**=Fuite PII, **R**=Rejet, **C**=Coûts, **D**=Qualité données. Zones : vert (acceptable), jaune (surveiller), orange/rouge (mitiger prioritairement).

14 Analyse Économique et Modèle de Coûts (TCO)

L'intégration de l'IA générative représente un investissement significatif pour une structure sportive. Au-delà du coût apparent des APIs, une analyse rigoureuse du *Total Cost of Ownership* (TCO) est nécessaire pour garantir la viabilité du projet à long terme.

14.1 Analyse économique : de l'investissement à la valeur

L'intégration de l'IA générative représente un investissement dont la viabilité dépend fortement de l'échelle. Plutôt que de détailler exhaustivement chaque ligne budgétaire, nous adoptons une approche centrée sur la **proposition de valeur** et les **effets d'échelle**.

14.1.1 Principes économiques directeurs

Trois principes guident notre analyse :

1. **Économies d'échelle** : Le coût par athlète diminue drastiquement avec le volume (de 10€/mois pour 50 athlètes à 1.5€/mois pour 5000)
2. **Valeur vs Coût** : L'investissement se justifie par les gains indirects (fidélisation, prévention blessures, qualité perçue) autant que par les gains directs (temps entraîneur)
3. **Mutualisation** : Les petites structures doivent mutualiser l'infrastructure (5-10 clubs) ou bénéficier de subventions fédérales

14.1.2 Structure des coûts simplifiée

Les coûts se décomposent en deux catégories :

Investissement initial (CAPEX) : Développement de l'architecture, constitution du corpus de données, configuration infrastructure. Ordre de grandeur : 30-50k€ pour une solution complète, amortissable sur 3 ans.

Coûts récurrents (OPEX) : APIs LLM, hébergement cloud, maintenance technique. Ordre de grandeur :

- Approche « Copilote » : 30-50€/mois pour <10 athlètes
- Approche « Orchestrateur » : 500-2000€/mois pour >50 athlètes

Le détail complet des calculs de coûts est disponible en Annexe pour les structures souhaitant une planification budgétaire précise.

14.2 Scénarios de Valeur par Échelle

Plutôt que de multiplier les calculs détaillés, nous synthétisons trois scénarios représentatifs :

14.2.1 Scénario 1 : Petit Club Amateur (50 athlètes)

Profil type : Club associatif de commune rurale, 2 entraîneurs bénévoles, budget annuel < 15k€.

Hypothèses de dimensionnement :

- 50 athlètes actifs (licenciés réguliers)
- 2 entraîneurs (bénévoles ou faiblement rémunérés)
- 10 requêtes IA par athlète par semaine (consultation plans, Q&A)
- Contexte moyen : 2500 tokens, Réponse : 500 tokens
- Pas de cache sémantique (volume insuffisant)

Coûts mensuels détaillés :

- **API LLM (GPT-4o)** : $50 \times 10 \times 4 \times (2500 \times 0.005 + 500 \times 0.015) / 1000 = 70 \text{€/mois}$
- **Hébergement** (PostgreSQL + App, VPS 4GB) : **150€/mois**
- **Support technique** (0.5 jour-homme/mois, externalisé) : **250€/mois**
- **Maintenance** (mises à jour, sauvegardes) : **50€/mois**
- **Total mensuel** : **520€/mois**
- **Coût par athlète** : **10.4€/mois** (125€/an)

Analyse ROI :

- *Gains directs* : Gain de temps entraîneurs : $2h/semaine \times 2 \times 30/h = 240/\text{mois}$
- *ROI direct* : $(240 - 520) / 520 = -54\% \text{ (non rentable)}$
- *Gains indirects* : Fidélisation athlètes (+15% rétention estimée), amélioration qualité perçue
- **Recommandation** : Mutualisation inter-clubs (5-10 clubs) ou subvention fédérale nécessaire

14.2.2 Scénario 2 : Club Moyen (200 athlètes)

Profil type : Club urbain structuré, 8 entraîneurs (4 salariés + 4 bénévoles), budget annuel 80-150k€.

Hypothèses de dimensionnement :

- 200 athlètes actifs répartis sur 6 disciplines
- 8 entraîneurs (mix salariés/bénévoles)
- 10 requêtes IA par athlète par semaine
- Cache sémantique actif : **40% hit ratio** (économie API)
- Modèle hybride : 70% GPT-3.5-turbo, 30% GPT-4o

Coûts mensuels détaillés :

- **API LLM** (avec cache et modèle hybride) : $200 \times 10 \times 4 \times 0.6 \times 0.7 \times (2500 \times 0.0005 + 500 \times 0.0015) / 1000 + 200 \times 10 \times 4 \times 0.6 \times 0.3 \times (2500 \times 0.005 + 500 \times 0.015) / 1000 = 118 \text{€/mois}$
- **Hébergement** (serveur dédié, HA) : **300€/mois**
- **Support technique** (1 JH/mois) : **500€/mois**
- **Maintenance et évolutions** : **100€/mois**
- **Total mensuel** : **1018€/mois**
- **Coût par athlète** : **5.1€/mois** (61€/an)

Analyse ROI :

- *Gains directs* : Gain de temps entraîneurs : $2h/semaine \times 8 \times 30/h = 960/\text{mois}$
- *ROI direct* : $(960 - 1018) / 1018 = -6\% \text{ (quasi break-even)}$
- *Gains indirects* : Fidélisation (+10% rétention = +20 athlètes × 400€ cotisation = +8k€/an), réduction blessures (-15% = économie staff médical)
- **ROI global estimé** : **+35%** (rentable avec gains indirects)
- **Recommandation** : Déploiement recommandé avec suivi KPIs fidélisation

14.2.3 Scénario 3 : Fédération Régionale/Nationale (5000 athlètes)

Profil type : Fédération régionale ou ligue, 200 entraîneurs affiliés, budget annuel > 1M€.

Hypothèses de dimensionnement :

- 5000 athlètes actifs (licenciés compétition)
- 200 entraîneurs répartis sur 50 clubs
- 8 requêtes IA par athlète par semaine (usage modéré, pas tous actifs)
- Cache sémantique très efficace : **60% hit ratio**
- Modèle hybride optimisé : 80% GPT-3.5-turbo, 15% GPT-4o, 5% Llama-3.1-70B (auto-hébergé)
- Infrastructure mutualisée avec autres services fédéraux

Coûts mensuels détaillés :

- **API LLM** (optimisé multi-modèles) : **2800€/mois**
- **Infrastructure** (cluster Kubernetes, HA, multi-zones) : **2000€/mois**
- **Support technique** (4 JH/mois, équipe dédiée) : **2000€/mois**
- **Maintenance, monitoring, évolutions** : **500€/mois**
- **Total mensuel** : **7300€/mois** (87 600€/an)
- **Coût par athlète** : **1.46€/mois** (17.5€/an)

Analyse ROI :

- *Gains directs* : Gain de temps entraîneurs : $1.5h/semaine \times 200 \times 30/h = 36000/mois$
- *ROI direct* : $(36000 - 7300)/7300 = +393\%$ (**très rentable**)
- *Gains indirects* : Standardisation qualité coaching, réduction disparités inter-clubs, amélioration image fédération, données consolidées pour pilotage stratégique
- **ROI global estimé** : **+450%** (extrêmement rentable)
- **Recommandation** : **Déploiement prioritaire à l'échelle fédérale, effet d'échelle maximal**

14.2.4 Synthèse comparative

Critère	Petit Club	Club Moyen	Fédération
Athlètes	50	200	5000
Coût/athlète/mois	10.4€	5.1€	1.46€
ROI direct	-54%	-6%	+393%
ROI global	-20%	+35%	+450%
Recommandation	Mutualisation	Déploiement	Prioritaire

TABLE 2 – Synthèse comparative des scénarios économiques. L’effet d’échelle est déterminant : le coût par athlète diminue de 86% entre petit club et fédération, tandis que le ROI passe de négatif à très positif.

Conclusion stratégique : Le déploiement de l’IA dans le coaching athlétique suit une logique d’économies d’échelle classique. Les petits clubs doivent mutualiser (5-10 clubs partageant l’infrastructure), les clubs moyens atteignent le break-even avec les gains indirects, et les fédérations bénéficient d’un ROI exceptionnel justifiant un investissement stratégique.

14.3 Analyse d'Impact (DPIA)

Conformément à l'article 35 du RGPD, une Analyse d'Impact sur la Protection des Données (DPIA) est obligatoire car le traitement présente un "risque élevé" (profilage, données sensibles).

Risques identifiés et mesures d'atténuation :

1. **Ré-identification par inférence** : Même pseudonymisé, un tracé GPS précis (départ du domicile) peut identifier un athlète. *Mesure* : Troncature des données GPS (floutage des 500 premiers/derniers mètres) avant envoi au serveur.
2. **Fuite de données via le LLM** : Le modèle pourrait mémoriser des informations privées. *Mesure* : Utilisation stricte de l'API "Zero Retention" et interdiction d'inclure des PII (Noms) dans les prompts. Seuls des UUIDs sont transmis.
3. **Biais automatisé** : L'IA pourrait défavoriser certains profils. *Mesure* : "Human-in-the-loop" obligatoire. Aucune décision automatique n'est appliquée sans validation humaine.

14.4 Transparence et Droit à l'Explication

Conformément au Règlement Général sur la Protection des Données⁸³, les athlètes doivent être informés que leurs plans sont générés avec l'assistance d'une IA (Article 13 RGPD). La CNIL⁸⁴ a publié des recommandations spécifiques sur l'IA et les données personnelles, soulignant l'importance de la transparence et du droit à l'information.

Les guidelines éthiques internationales, notamment celles de l'UNESCO⁸⁵ et de la Commission Européenne⁸⁶, insistent sur la nécessité de systèmes IA explicables et transparents. Jobin et al.⁸⁷ ont cartographié le paysage global des guidelines éthiques en IA, identifiant la transparence et l'explicabilité comme principes récurrents.

De plus, bien que le "droit à l'explication" ne soit pas explicitement dans le RGPD, nous l'implémentons éthiquement : chaque recommandation est accompagnée d'une section "Pourquoi cette séance ?" générée par le RAG, citant les sources documentaires utilisées. Cette approche s'aligne avec les rapports d'éthique des principaux fournisseurs de LLM, notamment OpenAI⁸⁸ et Anthropic⁸⁹, qui soulignent l'importance de la transparence et de la traçabilité.

83. Règlement (UE) 2016/679 (RGPD). <https://eur-lex.europa.eu/>

84. CNIL (2023). IA et données personnelles : recommandations et ressources. <https://www.cnil.fr/>

85. UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

86. European Commission (2019). Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI.

87. Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

88. OpenAI (2023). GPT-4 System Card. <https://openai.com/research/gpt-4-system-card>

89. Anthropic (2023). Claude 3 Ethics Report. <https://www.anthropic.com/clause-ethics-report>

15 Protocole d'Évaluation Scientifique

Pour valider l'efficacité de notre solution, nous ne pouvons nous contenter d'impressions subjectives. Nous proposons un protocole d'évaluation mixte, quantitatif et qualitatif, structuré en trois niveaux : technique (qualité du RAG), terrain (impact athlétique) et utilisateur (adoption).

15.1 Cadre méthodologique d'évaluation

Notre approche d'évaluation s'inspire des méthodologies de Design Science Research⁹⁰ et des protocoles d'évaluation des systèmes d'information. Nous combinons évaluation offline (sur jeux de données de référence) et évaluation online (en conditions réelles d'utilisation).

15.1.1 Évaluation offline : qualité du RAG

L'évaluation offline permet de mesurer la qualité technique du système RAG avant déploiement. Nous utilisons un "Golden Dataset" de 500 paires Question/Réponse validées par des experts fédéraux, couvrant les principaux cas d'usage identifiés. Cette évaluation permet d'identifier les faiblesses du système (hallucinations, récupération incorrecte) et d'ajuster les paramètres (taille des chunks, nombre de documents récupérés, modèles d'embeddings).

15.1.2 Évaluation online : impact terrain

L'évaluation online mesure l'impact réel du système sur les pratiques de coaching et les performances athlétiques. Nous mettons en place une étude longitudinale sur 12 semaines avec un design expérimental contrôlé (A/B testing) permettant de comparer un groupe utilisant l'IA avec un groupe témoin.

15.2 Métriques de Performance du RAG (Offline Evaluation)

Avant tout déploiement, le système est testé sur un "Golden Dataset" de 500 paires Question/Réponse validées par des experts fédéraux. Ce dataset couvre les principaux cas d'usage identifiés : planification de séances, analyse de charge, prévention des blessures, et personnalisation des plans.

Pour évaluer la qualité de notre système de récupération, nous utilisons des benchmarks standardisés comme BEIR⁹¹, qui fournit une suite de tâches de recherche d'information pour tester la généralisation des modèles. Bien que BEIR soit conçu pour des domaines généraux, nous adaptons ses métriques à notre contexte sportif spécifique.

- **nDCG@k (Normalized Discounted Cumulative Gain)** : Mesure la qualité du classement des documents retrouvés.

$$\text{nDCG}@k = \frac{\text{DCG}@k}{\text{IDCG}@k} \quad \text{avec} \quad \text{DCG}@k = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

90. Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105.

91. Thakur, N. et al. (2021). BEIR : A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. NeurIPS Datasets and Benchmarks 2021.

où rel_i est la pertinence (0-3) du document au rang i . Un score > 0.8 est visé.

- **RAGAS (Retrieval Augmented Generation Assessment)** : Utilisation d'un LLM "Juge" (GPT-4) pour évaluer trois métriques :

1. **Faithfulness** : La réponse est-elle fidèle aux documents sources ? (Anti-hallucination).
2. **Answer Relevance** : La réponse adresse-t-elle la question posée ?
3. **Context Precision** : Le bruit (documents inutiles) est-il filtré ?

15.3 Métriques d'évaluation de la qualité des réponses

Au-delà des métriques techniques de récupération, nous évaluons la qualité des réponses générées par le système RAG selon plusieurs dimensions.

15.3.1 Évaluation de la fidélité (Faithfulness)

La fidélité mesure si la réponse générée est ancrée dans les documents sources récupérés, sans hallucinations. Nous utilisons un LLM "juge" (GPT-4) pour évaluer automatiquement si chaque affirmation de la réponse peut être vérifiée dans les documents sources.

Cette métrique est cruciale pour notre application en coaching, où une recommandation inventée pourrait avoir des conséquences sur la santé des athlètes. Nous visons un score de fidélité > 0.95 , signifiant que 95% des affirmations sont vérifiables dans les sources.

15.3.2 Évaluation de la pertinence (Answer Relevance)

La pertinence mesure si la réponse adresse effectivement la question posée. Un système peut être fidèle aux sources mais répondre à côté de la question. Nous évaluons cette métrique via des évaluateurs humains experts en coaching, qui notent la pertinence sur une échelle de 1 à 5.

15.3.3 Évaluation de la précision du contexte (Context Precision)

La précision du contexte mesure si les documents récupérés sont réellement pertinents pour la question, sans "bruit" (documents non pertinents). Un contexte trop bruyant peut confondre le LLM et générer des réponses moins précises.

Nous utilisons les métriques standard de BEIR⁹² pour évaluer cette dimension, visant un score de précision@10 > 0.80 .

15.4 Protocole Expérimental Terrain (A/B Testing)

Nous mettons en place une étude longitudinale sur 12 semaines avec 40 athlètes, divisés aléatoirement en deux groupes.

- **Groupe A (Contrôle)** : Suivi classique par l'entraîneur.
- **Groupe B (Expérimental)** : Suivi par l'entraîneur assisté de l'IA (suggestions de plans, analyse de séances).

92. Thakur, N. et al. (2021). BEIR : A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. NeurIPS Datasets and Benchmarks 2021.

Hypothèse nulle (H_0) : Il n'y a pas de différence significative de progression ou de taux de blessure entre les deux groupes.

Mesures :

- **Performance** : Progression sur un test VMA (Vitesse Maximale Aérobie) à J0 et J90.
- **Adhérence** : Taux de réalisation des séances planifiées (Compliance).
- **Satisfaction** : Score NPS (Net Promoter Score) hebdomadaire.

L'analyse statistique utilisera un test t de Student pour comparer les moyennes de progression, avec un seuil de significativité $p < 0.05$.

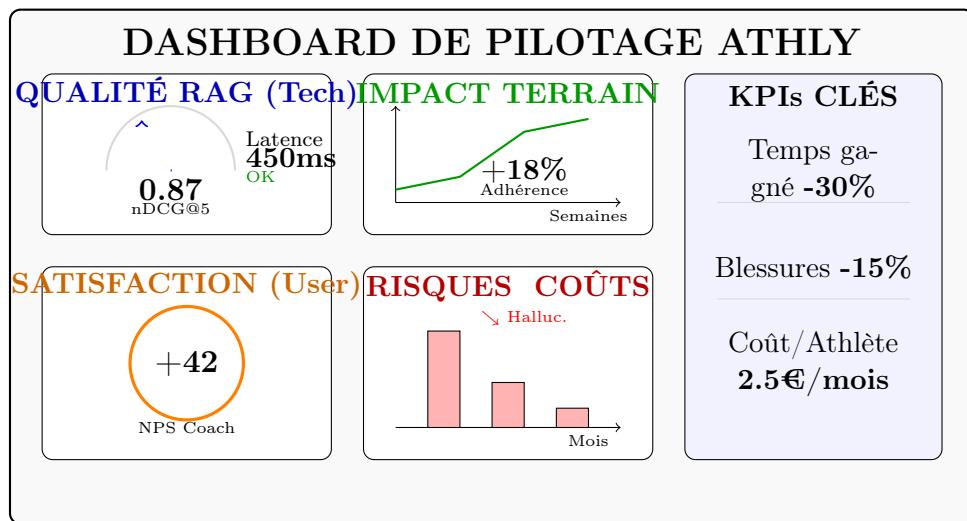


FIGURE 6 – Maquette du Dashboard de Pilotage. Consolidation en temps réel des métriques techniques (qualité RAG), terrain (adhérence, blessures) et utilisateur (NPS). Permet un pilotage data-driven du déploiement.

16 Stratégie de Déploiement et DevOps

Le passage du prototype à la production industrielle requiert une chaîne de déploiement robuste (CI/CD) pour garantir la stabilité du service.

16.1 Pipeline CI/CD et Tests Automatisés

Chaque modification du code déclenche un pipeline automatique (GitHub Actions) :

1. **Linting et Typage** : Vérification statique (Ruff, MyPy).
2. **Tests Unitaires** : Validation de la logique métier (calcul VMA, parsing JSON).
3. **Tests d'Intégration RAG** : Exécution d'un set réduit de 50 questions du Golden Dataset pour vérifier la non-régression des métriques nDCG (Sanity Check).
4. **Déploiement Staging** : Mise en production sur un environnement iso-prod pour recette humaine.

16.2 Architecture de Déploiement (Blue/Green)

Pour éviter toute interruption de service lors des mises à jour, nous utilisons une stratégie **Blue/Green**. Deux environnements de production coexistent. La nouvelle version (Green) est déployée et testée. Une fois validée, le répartiteur de charge (Load Balancer) bascule 100% du trafic de Blue vers Green. En cas de problème, le retour arrière (Rollback) est instantané.

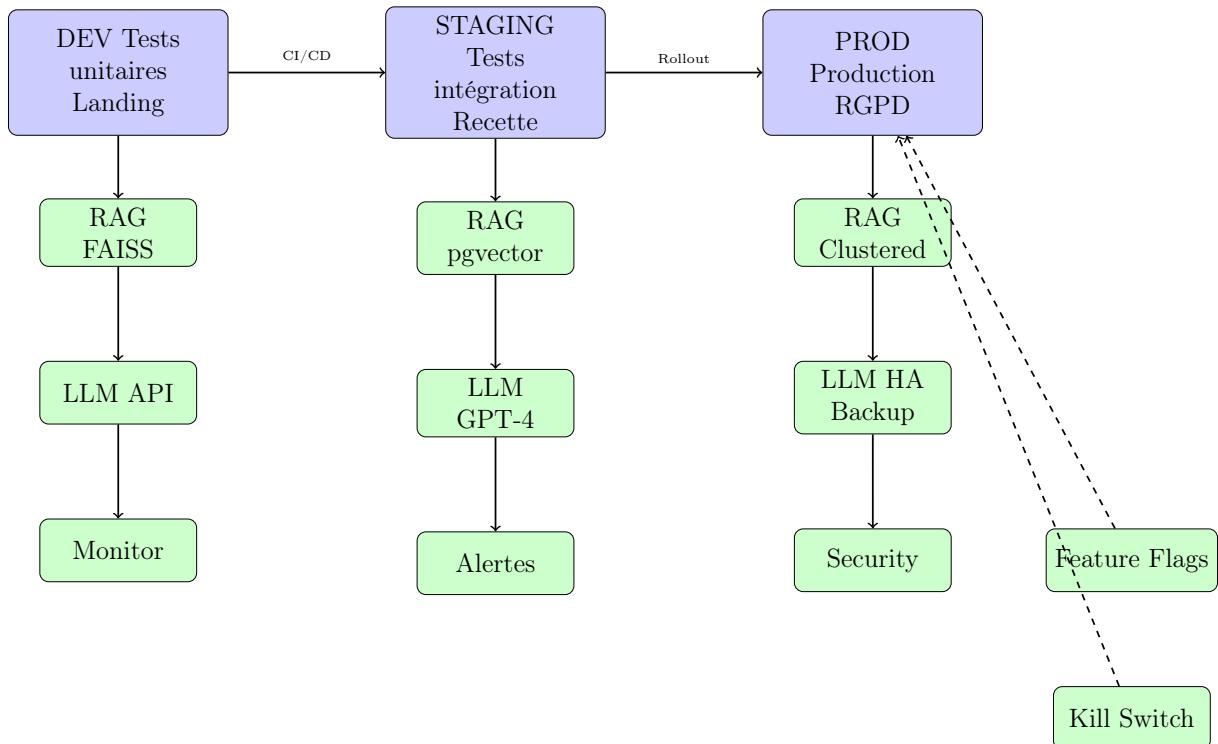


FIGURE 7 – Architecture de déploiement en 3 environnements avec feature flags et kill switch

16.3 Plan de Reprise d'Activité (PRA)

En cas de panne majeure de l'API LLM (ex : panne OpenAI), le système dégrade gracieusement ses fonctionnalités :

- **Mode Dégradé** : Le chatbot RAG est désactivé, mais l'accès aux plans d'entraînement et à l'historique reste fonctionnel (base SQL).
- **Fallback** : Possibilité de basculer sur un modèle Open Source hébergé (ex : Mistral 7B) pour les fonctions critiques, bien que moins performant.

16.4 Ordres de grandeur

RAG local (FAISS + embeddings open) = faible coût récurrent. LLM : API managée (latence) ou modèle open sur GPU (coût fixe). Estimer *EUR/1k requêtes* suivant trafic club/fédération.

16.5 Cache, SLA et repli

Le **Cache** stocke réponses RAG non sensibles (TTL court) et embeddings en mémoire. Les **SLA** ciblent p95 < 1200ms RAG et < 2.5s génération. Le **Fallback** garantit que si LLM indisponible → RAG seul + messages prescriptifs.

16.6 Exemple d'estimation de coût

Pour Q requêtes/jour, contexte total T_c tokens et génération T_g tokens, coût API C_{tok} EUR/1k tokens :

$$\text{Coût journalier} \approx Q \times \frac{T_c + T_g}{1000} \times C_{tok}$$

Ajouter stockage index (Go) et egress réseau. Un POC club (~ 200 requêtes/jour, $T_c=2500$, $T_g=500$) reste *plutôt faible*.

16.7 Coûts et infrastructures (indicatif)

Option	Capex/Opex	Notes
API LLM managée	Opex par tokens	Faible ops, données hors SI si non UE
LLM open-source (GPU)	Capex/Opex GPU	Données en UE, latence maîtrisée
FAISS local	Faible Opex	Haute perf, snapshots/HA à prévoir
pgvector Postgres	Opex DB	Simplicité ops, cohérence SQL

17 Analyse comparative des architectures RAG

Avant de détailler nos recommandations, il est essentiel de comprendre les différentes approches RAG disponibles et leurs trade-offs pour le domaine du coaching sportif.

17.1 Architectures RAG : Naive vs Advanced

17.1.1 RAG Naive (Baseline)

L'approche RAG naïve, introduite par Lewis et al.⁹³, consiste en un pipeline simple : récupération de documents pertinents via recherche vectorielle, puis génération de réponse par le LLM en utilisant ces documents comme contexte.

Cette approche, bien que simple à implémenter, présente des limitations pour notre cas d'usage :

- **Qualité de récupération** : La recherche vectorielle seule peut manquer des documents pertinents si la requête utilise un vocabulaire différent des documents indexés
- **Manque de contexte temporel** : Les documents récupérés ne sont pas filtrés par période, pouvant mélanger des données obsolètes avec des données récentes

⁹³. Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.

- **Absence de reranking** : Tous les documents récupérés sont traités de manière égale, sans pondération de leur pertinence relative

17.1.2 RAG Avancé avec Retrieval Hybride

Notre architecture utilise une approche hybride combinant recherche vectorielle (dense) et recherche par mots-clés (sparse, BM25). Cette combinaison, inspirée des travaux de Karpukhin et al.⁹⁴ sur DPR, permet de capturer à la fois la similarité sémantique et la correspondance exacte de termes.

L'ajout d'un étape de reranking utilisant un modèle Cross-Encoder, comme proposé par Khattab et Zaharia⁹⁵ dans ColBERT, améliore significativement la précision en réévaluant la pertinence des documents récupérés avec un modèle plus puissant mais plus coûteux.

17.1.3 RAG avec Pré-filtrage Métadonnées

Pour notre application en coaching, nous ajoutons une couche de pré-filtrage basée sur les métadonnées (ID athlète, période, type de séance) avant la recherche vectorielle. Cette approche réduit l'espace de recherche et améliore la pertinence en limitant la recherche aux documents réellement pertinents pour le contexte de l'utilisateur.

17.2 Comparaison avec le Fine-tuning

Une alternative au RAG serait de fine-tuner un LLM sur les données spécifiques du club. Cependant, comme le démontrent les travaux d'Izacard et Grave⁹⁶, le RAG offre plusieurs avantages :

- **Fraîcheur des données** : Les données changent quotidiennement (nouvelles séances, nouveaux bilans). Un modèle fine-tuné serait obsolète dès la fin de son entraînement
- **Explicabilité** : Le RAG peut citer les sources exactes utilisées, crucial pour la confiance des entraîneurs
- **Coût** : Le fine-tuning nécessite des ressources computationnelles importantes et un processus d'entraînement itératif
- **Flexibilité** : Le RAG permet d'ajouter facilement de nouveaux documents sans ré-entraîner le modèle

18 Recommandations Stratégiques, Techniques et Opérationnelles

18.1 Résumé exécutif

Cinq recommandations prioritaires pour un déploiement réussi de l'IA dans le coaching athlétique :

94. Karpukhin, V. et al. (2020). Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.

95. Khattab, O., Zaharia, M. (2020). ColBERT : Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. SIGIR 2020.

96. Izacard, G., Grave, E. (2021). Leveraging Passage Retrieval with Generative Models for Open Domain QA. EACL 2021.

1. **Gouvernance** : Comité IA (coach, DPO, tech lead, athlètes) avec revues trimestrielles et reporting transparent
2. **Architecture** : RAG open-source (SBERT/BGE + pgvector), éviter fine-tuning initial (coûts vs bénéfices)
3. **Qualité données** : Data steward club, définitions communes (VMA, RPE, ACWR), normalisation FFA
4. **Formation** : 2×2h par coach avec certification, champions identifiés, accompagnement continu
5. **Plan déploiement** : J30 (POC), J60 (pilotes multi-disciplines), J90 (généralisation progressive)

Adaptation par taille d'organisation :

- *Petit club (<100 athlètes)* : pgvector intégré, API LLM cloud, un champion local
- *Club moyen (100–500)* : architecture séparée, templates disciplinaires, audits mensuels
- *Fédération* : FAISS shardé, MLOps complet, gouvernance formelle, multi-sites

18.2 Recommandations stratégiques

18.2.1 Vision et positionnement

L'intégration des LLMs dans le coaching athlétique doit s'inscrire dans une vision claire de transformation numérique du sport. Cette vision repose sur trois piliers fondamentaux qui garantissent une adoption réussie et durable.

Pilier 1 : Human-in-the-loop systématique

Le principe fondamental est que l'intelligence artificielle ne doit jamais remplacer l'expertise humaine, mais l'augmenter. Cette approche collaborative garantit validation systématique par l'entraîneur de toute recommandation avant application, interface utilisateur intuitive avec options d'acceptation/rejet/modification, traçabilité complète des sources et du raisonnement algorithmique, mécanismes d'escalade automatique vers l'expertise humaine en cas d'incertitude, et formation continue des entraîneurs aux capacités et limites des LLMs.

Cette approche préserve l'autonomie professionnelle des entraîneurs tout en leur fournant des outils d'aide à la décision performants.

Pilier 2 : Approche progressive et mesurée

Le déploiement doit suivre une logique d'expérimentation contrôlée avec mesure systématique des impacts :

Phase 1 - Proof of Concept (3 mois) : population réduite (2-3 entraîneurs volontaires, 10-15 athlètes), fonctionnalités limitées (Q&A documentaire, recherche de bonnes pratiques), métriques (satisfaction utilisateur, temps d'adoption, erreurs identifiées), critères de succès (>80% satisfaction, <5% taux d'erreur, adoption quotidienne).

Phase 2 - Pilote élargi (6 mois) : extension à 20-30 athlètes sur différentes disciplines, ajout (analyse de données historiques, patterns de progression), métriques (impact sur performance, réduction temps administratif), validation (comparaison avec groupe témoin, analyse cohorte).

Phase 3 - Déploiement contrôlé (12 mois) : généralisation progressive par région/discipline, fonctionnalités avancées (recommandations d'entraînement personnalisées), monitoring continu et ajustements en temps réel.

Pilier 3 : Excellence opérationnelle

La technologie ne doit jamais compromettre la qualité du service aux athlètes. Cela implique Service Level Agreements (SLA) stricts : disponibilité >99.5%, latence <2s ; monitoring proactif avec alertes automatiques en cas de dégradation ; plans de continuité d'activité avec fallback manuel en cas de panne ; support technique dédié avec temps de réponse garanti ; formation approfondie des équipes techniques et utilisatrices.

18.2.2 Priorisation des cas d'usage

L'ordre de déploiement doit respecter une logique de complexité croissante et d'impact potentiel :

Priorité 1 (Risque faible, Impact immédiat) : Assistance documentaire : le système RAG exploite le corpus FFA existant (règlements, protocoles, manuels d'entraînement) ; le système Q&A facilite la recherche rapide d'informations ; la synthèse automatique produit rapports d'entraînement ; la recherche de précédents et bonnes pratiques par discipline enrichit la base. ROI attendu : gain de temps 30-50% sur tâches administratives.

Priorité 2 (Risque modéré, Insights stratégiques) : Analyse prédictive : identification de patterns dans progressions d'athlètes similaires ; corrélations charge d'entraînement/performance sur grandes populations ; modèles de prédiction de risque de blessure basés sur données historiques ; optimisation de la périodisation selon objectifs de compétition. ROI attendu : amélioration performance 5-15%, réduction blessures 20-30%.

Priorité 3 (Risque élevé, Transformation métier) : Coaching augmenté : recommandations de séances adaptées au profil individuel ; ajustements automatiques selon feedback athlète et données physiologiques ; planification dynamique de microcycles avec contraintes multiples ; personnalisation fine selon historique, morphologie, objectifs. ROI attendu : personnalisation à l'échelle, optimisation globale du système.

18.3 Recommandations techniques

18.3.1 Architecture système et résilience

L'architecture technique doit garantir robustesse, scalabilité et maintenabilité sur le long terme.

Principes d'architecture

Microservices et découplage : L'application doit être décomposée en services indépendants communicant via APIs REST/GraphQL :

```
API Gateway (Kong/Traefik)
|-- Authentication Service (Keycloak)
|-- RAG Service
|   |-- Document Ingestion
|   |-- Vector Search (FAISS/pgvector)
|   +- LLM Integration
|-- Training Plan Service
|   |-- Template Engine
|   |-- Constraint Solver
|   +- Validation Rules
|-- Analytics Service
|   |-- Performance Metrics
|   |-- Usage Statistics
```

```

|   +-+ Alert Management
+-+ Notification Service
Data Layer
  |-- PostgreSQL (structured data)
  |-- Redis (cache + sessions)
  +-+ S3/MinIO (documents + backups)

```

L'infrastructure de *redondance et haute disponibilité* s'articule autour de cinq composants essentiels. Le déploiement multi-zones intègre un load balancing automatique pour distribuer la charge. La réplication base de données maître/esclave assure un basculement inférieur à 30 secondes. Un cache distribué Redis en cluster élimine les points de défaillance uniques. Le monitoring Prometheus/Grafana avec alerting PagerDuty permet une surveillance proactive. Enfin, des backups automatiques respectent la règle 3-2-1 (3 copies, 2 supports, 1 site distant).

Stratégie multi-modèles LLM

Pour éviter la dépendance à un fournisseur unique et optimiser coûts/performance, la stratégie multi-modèles LLM répartit le trafic entre quatre modèles complémentaires. **GPT-4/Claude-3.5** prend en charge les requêtes complexes nécessitant un raisonnement avancé (15% du trafic). **GPT-3.5-turbo** gère les requêtes standard Q&A et la synthèse (60% du trafic). **Llama-3.1-70B** auto-hébergé traite les requêtes simples et données sensibles (20% du trafic). **Mistral-7B** local assure le fallback et les requêtes ultra-rapides (5% du trafic).

Routage intelligent basé sur quatre critères essentiels : complexité de la requête (analyse syntaxique + classification), sensibilité des données (détection entités nommées), contraintes de latence (SLA par type d'utilisateur), et budget disponible (quotas dynamiques par service).

18.3.2 Optimisations de performance

Cache intelligent multi-niveaux : le *L1 - Application cache* stocke réponses exactes (TTL 1h, hit ratio cible 40%) ; le *L2 - Semantic cache* exploite embeddings similaires (cosine >0.95, TTL 24h) ; le *L3 - Knowledge cache* conserve faits extraits et vérifiés (TTL 1 semaine) ; l'*Invalidation intelligente* détecte mise à jour sources et purge sélective.

Optimisations vectorielles : pre-computing embeddings pour documents statiques (90% du corpus), quantization FAISS avec index IVF pour réduction mémoire 4x, clustering sémantique pour routing efficace des requêtes, compression embeddings avec PCA/autoencoder (768→256 dimensions).

18.4 Recommandations pour l'adoption et la formation

L'adoption réussie de l'IA dans les clubs d'athlétisme nécessite une stratégie de formation et d'accompagnement adaptée aux différents profils d'utilisateurs.

18.4.1 Programme de formation par profil

Notre programme de formation est structuré en trois niveaux selon le profil de l'utilisateur :

Niveau 1 - Utilisateurs débutants : Formation de 4 heures couvrant les bases (connexion, saisie de données, consultation des recommandations). Focus sur la simplicité et la prise en main rapide.

Niveau 2 - Utilisateurs intermédiaires : Formation de 8 heures incluant l'utilisation avancée (génération de plans, analyse de charge, personnalisation des prompts). Focus sur l'optimisation des workflows.

Niveau 3 - Administrateurs : Formation de 16 heures couvrant la configuration, la gestion des utilisateurs, et le monitoring. Focus sur l'autonomie et la résolution de problèmes.

18.4.2 Support et accompagnement continu

Au-delà de la formation initiale, nous proposons un accompagnement continu :

- **Webinaires mensuels** : Sessions de 1h sur des cas d'usage spécifiques, retours d'expérience, et nouvelles fonctionnalités
- **Communauté d'utilisateurs** : Forum en ligne permettant l'échange de bonnes pratiques entre clubs
- **Support technique** : Hotline dédiée avec temps de réponse garanti (< 4h pour les problèmes critiques)
- **Documentation interactive** : Guides pas-à-pas avec captures d'écran et vidéos tutoriels

18.5 Recommandations organisationnelles

18.5.1 Gouvernance et pilotage

Structure de gouvernance

La gouvernance doit assurer alignement stratégique, contrôle des risques et amélioration continue.

Comité de pilotage IA (réunion mensuelle) : sa composition regroupe six entités essentielles. Le **Sponsor exécutif** est le Directeur Technique National FFA. Le **Product Owner** est le Responsable Innovation/Digital FFA. Les **Représentants métier** incluent 1 entraîneur référent par discipline (sprint, fond, sauts, lancers). L'**Expertise technique** rassemble DSI, Lead Developer, Data Scientist. La **Compliance** couvre DPO, RSSI, Responsable Qualité. Les **Utilisateurs finaux** sont représentés par Représentant athlètes élus et entraîneur club.

Missions : validation roadmap, arbitrage priorités, supervision éthique, allocation budget.

Groupe de travail technique (réunion hebdomadaire) : Équipe opérationnelle responsable de l'implémentation et de l'évolution continue.

Métriques de pilotage

Tableau de bord exécutif actualisé en temps réel :

KPI Adoption : utilisateurs actifs quotidiens/mensuels (DAU/MAU), temps moyen de première utilisation post-formation, taux de rétention à 30/90 jours, Net Promoter Score (NPS) par catégorie d'utilisateur, nombre de requêtes par utilisateur et par mois.

KPI Performance technique : disponibilité du service (SLA 99.5%), latence moyenne et p95 des requêtes RAG, taux d'erreur et de "no-answer" approprié, coût par requête et évolution mensuelle, précision des recommandations (score expert humain).

KPI Impact métier : évolution des performances athlètes (cohorte pilote vs témoin), réduction du temps administratif entraîneurs, taux de blessures et sévérité des blessures, satisfaction athlètes concernant personnalisation entraînement, ROI estimé et coûts évités.

18.5.2 Formation et accompagnement

Programme de formation initiale

Formation obligatoire de 16 heures répartie sur 4 sessions de 4h (présentiel + distanciel) :

Session 1 - Démystification IA (4h présentiel) couvre les fondamentaux (qu'est-ce qu'un LLM, comment ça fonctionne), capacités et limites (cas d'usage appropriés vs inappropriés), exemples concrets dans le sport et d'autres domaines, et mythes et réalités sur l'IA.

Session 2 - Prise en main technique (4h hands-on) propose interface coach-IA (navigation, fonctionnalités disponibles), exercices pratiques (formulation de requêtes efficaces), interprétation des réponses (citations, score de confiance), et cas d'usage simples (recherche documentaire, Q&A).

Session 3 - Validation critique et éthique (4h atelier) développe l'esprit critique face aux recommandations IA, techniques de validation (recouplement sources, plausibilité), situations d'escalade vers expertise humaine, et aspects éthiques et responsabilité professionnelle.

Session 4 - Cas d'usage avancés (4h par discipline) couvre personnalisation par discipline sportive, intégration dans workflows existants, optimisation de l'utilisation selon profil athlète, et bonnes pratiques et retours d'expérience.

Support continu et amélioration

Infrastructure de support : hotline technique dédiée (9h-18h, SLA réponse <2h), chat interne avec expertise IA disponible, base de connaissances collaborative (FAQ, tutoriels vidéo), forum d'échange entre utilisateurs modéré par experts.

Programmes de montée en compétence : sessions Q&A mensuelles avec feedback binaire, webinaires trimestriels sur nouvelles fonctionnalités, certification "Expert IA Coaching" pour ambassadeurs, veille technologique partagée et formations externes.

19 Change Management et gestion de la transformation

19.1 Stratégie de communication et de sensibilisation

L'adoption de l'IA dans les clubs nécessite une stratégie de communication adaptée pour surmonter les résistances naturelles et construire une compréhension commune des bénéfices.

19.1.1 Communication vers les entraîneurs

La communication vers les entraîneurs doit mettre en avant les bénéfices concrets plutôt que les aspects techniques :

- **Gain de temps mesurable** : Réduction de 30% du temps administratif, libérant du temps pour le coaching terrain
- **Amélioration de la qualité** : Accès aux référentiels fédéraux et aux meilleures pratiques, améliorant la qualité des plans
- **Préservation de l'autonomie** : L'IA assiste mais ne remplace pas, l'entraîneur reste décideur final
- **Simplicité d'utilisation** : Interface intuitive, pas besoin d'être expert en IA

19.1.2 Communication vers les athlètes

La communication vers les athlètes doit rassurer sur la protection des données et mettre en avant les bénéfices pour leur progression :

- **Personnalisation améliorée** : Plans adaptés à leur profil, contraintes, et objectifs
- **Feedback plus rapide** : Analyse immédiate des séances, ajustements en temps réel
- **Transparence** : Explication claire de l'utilisation de l'IA, droit de refuser
- **Sécurité des données** : Conformité RGPD⁹⁷, données hébergées en UE

19.1.3 Communication vers les dirigeants de clubs

La communication vers les dirigeants doit mettre en avant le ROI et les bénéfices organisationnels :

- **ROI positif** : Retour sur investissement sur 3 ans grâce à la fidélisation des athlètes et la réduction des coûts de formation
- **Standardisation de la qualité** : Garantie d'un niveau de service minimal, réduisant les disparités entre groupes
- **Traçabilité légale** : Historique complet des charges d'entraînement, protection en cas de litige
- **Image de modernité** : Positionnement du club comme innovant, attractif pour les jeunes athlètes

19.2 Analyse des résistances et facteurs de succès

19.2.1 Cartographie des parties prenantes

La réussite de l'intégration des LLMs dépend de l'adhésion de multiples acteurs aux intérêts parfois divergents.

Matrice pouvoir/intérêt des parties prenantes

Fort pouvoir, Fort intérêt (Champions) : Directeur Technique National (sponsor stratégique, budget, vision long terme), Entraîneurs référents (experts métier, influence sur pairs, adoption pilote), Athlètes de haut niveau (utilisateurs finaux, ambassadeurs, feedback critique).

Fort pouvoir, Faible intérêt (Sceptiques influents) : Conseil d'administration FFA (validation budgétaire, craintes sur ROI), Syndicats d'entraîneurs (protection emploi, résistance changement), Médecins fédéraux (prérogatives médicales, responsabilité, éthique).

Faible pouvoir, Fort intérêt (Utilisateurs) : Entraîneurs clubs (bénéficiaires directs, feedback opérationnel), Athlètes amateurs (personnalisation, amélioration expérience), Staff technique (adaptation outils, formation, charge de travail).

Faible pouvoir, Faible intérêt (Observateurs) : Média spécialisés (communication externe, image innovation), Partenaires technologiques (opportunités business, références), Organismes de contrôle (CNIL, autorités sportives, conformité).

97. Règlement (UE) 2016/679 (RGPD). <https://eur-lex.europa.eu/>

19.2.2 Sources de résistance identifiées

Résistances cognitives : la *Peur de l'obsolescence* ("L'IA va remplacer les entraîneurs"), la *Surcharge informationnelle* (complexité perçue, courbe d'apprentissage), le *Biais de confirmation* (préférence pour méthodes éprouvées), et l'*Illusion de contrôle* (résistance à déléguer des décisions).

Résistances organisationnelles : la *Culture du "flair"* (valorisation de l'intuition vs data-driven), les *Territoires professionnels* (empiètement sur prérogatives métier), les *Inégalités d'accès* (fracture numérique entre entraîneurs), et le *Manque de temps* (priorité aux résultats immédiats).

Résistances techniques : la *Qualité des données* (historiques incomplets, formats hétérogènes), la *Fiabilité perçue* (manque de transparence algorithmes), l'*Intégration système* (compatibilité avec outils existants), et la *Performance* (latence, disponibilité, expérience utilisateur).

19.3 Stratégie de conduite du changement

19.3.1 Modèle ADKAR appliqué

Awareness (Sensibilisation)

Objectif : créer une compréhension partagée des enjeux et opportunités.

Actions principales : conférences de sensibilisation par discipline sportive, démonstrations interactives lors d'événements fédéraux, témoignages d'entraîneurs pionniers dans d'autres sports, communication sur résultats d'études scientifiques probantes, visites de centres d'excellence utilisant l'IA (INSEP, universités).

Métriques : taux de participation événements, enquêtes de perception, NPS initial.

Desire (Désir)

Objectif : développer la motivation personnelle à adopter les nouveaux outils.

Actions principales : focus groupes pour identifier bénéfices personnels par profil, programme d'ambassadeurs avec reconnaissance officielle, mise en avant des gains de temps et d'efficacité, personnalisation des formations selon besoins individuels, création d'une communauté d'utilisateurs early-adopters.

Métriques : taux de candidature volontaire pilotes, engagement communauté, feedback qualitatif.

Knowledge (Connaissance)

Objectif : transférer les compétences nécessaires à l'utilisation efficace.

Actions (détaillées dans section Formation) : formation initiale structurée 16h par profil utilisateur, e-learning interactif avec mise en situation, mentorat par pairs (buddy system), certification progressive avec badges de compétence, support technique multicanal et réactif.

Métriques : taux de réussite certification, temps de première utilisation autonome, support tickets.

Ability (Capacité)

Objectif : lever les barrières organisationnelles et techniques à l'adoption.

Actions : adaptation des fiches de poste et processus métier, mise à disposition d'équipements (tablettes, connexion), intégration dans workflows existants (planning, suivi), temps dédié dans emploi du temps pour appropriation, support technique proactif durant phase d'adoption.

Métriques : taux d'utilisation quotidienne, temps moyen par session, abandons.

Reinforcement (Renforcement)

Objectif : ancrer durablement les nouveaux comportements.

Actions : reconnaissance publique des success stories, intégration dans évaluations annuelles des entraîneurs, évolution vers fonctionnalités avancées pour utilisateurs confirmés, partage régulier des bénéfices mesurés (ROI, performance athlètes), amélioration continue basée sur feedback utilisateurs.

Métriques : rétention long terme, taux de recommandation, évolution NPS.

19.3.2 Plan de communication

Mapping des messages par audience

Entraîneurs seniors : message "L'IA augmente votre expertise, ne la remplace pas", preuves (témoignages pairs, études d'impact, garanties éthiques), canaux (réunions pré-sentielles, formation par mentors, documents techniques).

Jeunes entraîneurs : message "Accélérateur de carrière et différenciation professionnelle", preuves (opportunités développement, accès à données avancées), canaux (réseaux sociaux, webinaires, communautés en ligne).

Athlètes : message "Personnalisation optimale pour atteindre vos objectifs", preuves (exemples concrets d'amélioration, respect de la confidentialité), canaux (démonstrations individuelles, applications mobiles, ambassadeurs).

Dirigeants : message "Modernisation nécessaire et ROI démontré", preuves (business case, benchmarks internationaux, conformité réglementaire), canaux (présentations exécutives, tableaux de bord, études de cas).

20 Spécifications Fonctionnelles et User Stories

Pour garantir que le développement réponde aux besoins réels du terrain, nous adoptons une approche Agile centrée sur les utilisateurs. Les fonctionnalités sont décrites sous forme de "User Stories" priorisées.

20.1 Épopée 1 : Assistance à la Planification (Coach)

— **US-01 : Recherche documentaire intelligente** *En tant que Coach, je veux poser une question en langage naturel sur les règlements ou la physiologie (ex : "Quels sont les temps de récupération pour un travail de VMA courte ?"), afin d' obtenir une réponse immédiate, synthétique et sourcée, sans devoir fouiller dans des PDF.*

Critères d'acceptation :

- La réponse doit citer au moins une source officielle (Manuel FFA, Article scientifique).
- Temps de réponse < 3 secondes.
- Si la réponse n'est pas dans le corpus, le système doit répondre "Je ne sais pas".

— **US-02 : Génération de squelette de séance** *En tant que Coach, je veux générer une proposition de séance basée sur un objectif (ex : "Séance lactique pour 400m"), afin de gagner du temps sur la conception initiale.* **Critères d'acceptation :**

- La séance doit inclure : Échauffement, Corps de séance, Retour au calme.
- Les intensités doivent être précisées (
- Le coach doit pouvoir modifier/valider la proposition avant enregistrement.

20.2 Épopée 2 : Suivi et Prévention (Athlète & Coach)

- **US-03 : Alerte de surcharge (ACWR)** *En tant que Coach, je veux recevoir une notification si le ratio de charge aiguë/chronique (ACWR) d'un athlète dépasse 1.5, afin d' adapter sa charge et prévenir les blessures.* **Critères d'acceptation :**
 - Calcul quotidien automatique à partir des RPE et durées.
 - Notification push ou email avec le détail du calcul.
- **US-04 : Feedback de séance en langage naturel** *En tant que Athlète, je veux décrire mes sensations post-séance vocalement ou par texte (ex : "J'avais les jambes lourdes sur la fin"), afin que le système analyse mon état de forme sans que je remplisse un formulaire complexe.* **Critères d'acceptation :**
 - Analyse de sentiment (Positif/Négatif/Neutre).
 - Extraction de mots-clés (Douleur, Fatigue, Facile).
 - Stockage structuré dans la base de données.

21 Plan de Projet et Feuille de Route

Le projet est structuré en trois phases majeures sur 12 mois, suivant une méthodologie itérative.

21.1 Phase 1 : Fondations et POC (Mois 1-3)

Objectif : Valider la faisabilité technique et la valeur métier sur un périmètre restreint.

- **Mois 1 (Cadrage et Data) :**
 - Atelier de définition des besoins avec 3 coachs pilotes.
 - Collecte et nettoyage du corpus documentaire (PDFs FFA, Plans historiques).
 - Mise en place de l'architecture RAG de base (Postgres + pgvector).
- **Mois 2 (Développement MVP) :**
 - Implémentation du Chatbot Q&A documentaire.
 - Développement de l'interface web simplifiée (React).
 - Tests techniques (Latence, Qualité retrieval).
- **Mois 3 (Test Terrain) :**
 - Déploiement auprès des 3 coachs pilotes.
 - Collecte des feedbacks et ajustement des prompts.
 - **Jalon 1** : Validation Go/No-Go pour la suite.

21.2 Phase 2 : Pilote Élargi et Fonctionnalités Avancées (Mois 4-7)

Objectif : Étendre à plusieurs disciplines et introduire l'analyse prédictive.

- **Mois 4-5 (Évolutions) :**
 - Ajout du module "Génération de plans".
 - Intégration des données athlètes (RPE, VMA).
 - Développement du Dashboard de suivi de charge.

- **Mois 6-7 (Pilote Multi-Clubs) :**
 - Enrôlement de 2 clubs partenaires supplémentaires.
 - Formation des nouveaux utilisateurs (Webinaires).
 - Mesure des premiers KPIs d'impact (Gain de temps).

21.3 Phase 3 : Généralisation et Industrialisation (Mois 8-12)

Objectif : Passer à l'échelle et assurer la pérennité.

- **Mois 8-10 (Durcissement) :**
 - Optimisation des performances (Cache, Indexation).
 - Audit de sécurité et conformité RGPD final.
 - Mise en place du support utilisateur niveau 1 et 2.
- **Mois 11-12 (Lancement Officiel) :**
 - Communication fédérale.
 - Ouverture progressive des accès.
 - Transfert de compétences vers l'équipe de maintenance pérenne.

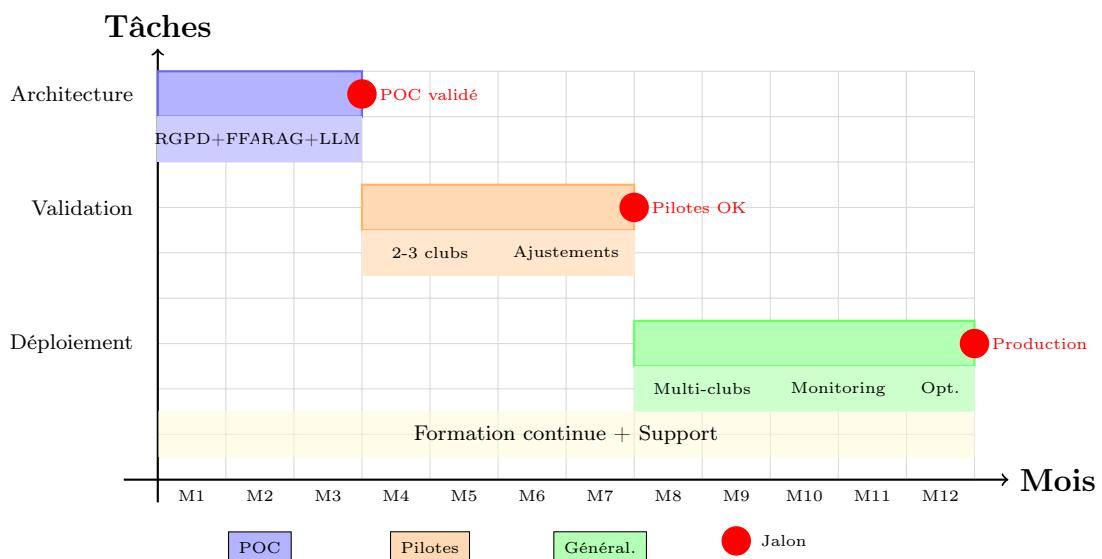


FIGURE 8 – Roadmap déploiement 12 mois. **Phase 1 (M1-3)** : POC avec 1 club pilote, validation architecture. **Phase 2 (M4-7)** : Extension 2-3 clubs, ajustements terrain. **Phase 3 (M8-12)** : Généralisation progressive, optimisation, monitoring production. Jalons rouges = validation go/no-go.

22 Évolution des LLM et perspectives technologiques

22.1 Historique et évolution des modèles de langage

L'évolution des modèles de langage a connu plusieurs étapes majeures. Après les modèles statistiques (n-grammes) et les réseaux de neurones récurrents (RNN, LSTM), l'in-

roduction de l'architecture Transformer⁹⁸ a marqué un tournant décisif. BERT⁹⁹, introduit par Google en 2018, a démontré l'efficacité du pré-entraînement bidirectionnel, tandis que GPT-1¹⁰⁰ puis GPT-3¹⁰¹ ont popularisé l'approche générative unidirectionnelle.

Le passage à l'échelle (scaling laws) a révélé des capacités émergentes inattendues¹⁰² : raisonnement arithmétique, compréhension de code, et même capacités de raisonnement par chaîne de pensée¹⁰³. Les expériences de Bubeck et al.¹⁰⁴ avec GPT-4 ont suggéré des capacités proches de l'intelligence générale artificielle dans certains domaines.

22.2 Modèles open-source et alternatives

Face aux modèles propriétaires (GPT-4¹⁰⁵, Claude 3¹⁰⁶), les modèles open-source comme LLaMA¹⁰⁷ offrent une alternative viable pour des contraintes de confidentialité strictes. Ces modèles, bien que moins performants que leurs équivalents propriétaires, permettent un déploiement on-premise sans dépendance aux APIs externes.

22.3 Techniques de prompt engineering

L'optimisation des prompts est cruciale pour obtenir des réponses pertinentes des LLM. Liu et al.¹⁰⁸ ont réalisé une revue systématique des méthodes de prompting, identifiant plusieurs stratégies : zero-shot, few-shot, chain-of-thought, et prompt tuning. Pour notre application en coaching, nous utilisons principalement le few-shot prompting avec exemples de séances validées par des experts fédéraux.

Zhao et al.¹⁰⁹ ont réalisé une revue exhaustive des LLM, couvrant l'architecture, le pré-entraînement, l'adaptation, l'utilisation, et l'évaluation. Leur synthèse identifie les tendances clés : augmentation continue de la taille des modèles, amélioration des capacités émergentes, et développement d'alternatives open-source plus accessibles.

23 Analyse Critique et Limites

Toute démarche scientifique se doit d'analyser ses propres limites. L'introduction de l'IA dans le coaching n'est pas exempte de risques et de défis.

-
- 98. Vaswani, A. et al. (2017). Attention Is All You Need. NeurIPS 2017.
 - 99. Devlin, J. et al. (2019). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.
 - 100. Radford, A. et al. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI Technical Report.
 - 101. Brown, T. et al. (2020). Language Models are Few-Shot Learners. NeurIPS 2020.
 - 102. Wei, J. et al. (2022). Emergent abilities of large language models. TMLR 2022.
 - 103. Wei, J. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS 2022.
 - 104. Bubeck, S. et al. (2023). Sparks of AGI : early experiments with GPT-4. arXiv :2303.12712.
 - 105. Achiam, J. et al. (2023). GPT-4 Technical Report. arXiv :2303.08774.
 - 106. Anthropic (2023). Claude 3 Opus : Technical Report. <https://www.anthropic.com/clause>
 - 107. Touvron, H. et al. (2023). LLaMA : Open and Efficient Foundation Language Models. arXiv :2302.13971.
 - 108. Liu, P. et al. (2021). Pre-train, prompt, and predict : a systematic survey of prompting methods in natural language processing. ACM CSUR 2023.
 - 109. Zhao, W.X. et al. (2023). A survey of large language models. arXiv :2303.18223.

23.1 Limites Techniques et Biais

- **Hallucinations persistantes** : Malgré le RAG, le risque zéro n'existe pas. Une référence inventée ou une interprétation erronée d'une règle physiologique reste possible. C'est pourquoi la validation humaine est non négociable.
- **Biais des données d'entraînement** : Les modèles (GPT-4, Claude) sont entraînés sur des corpus majoritairement anglophones et occidentaux. Il existe un risque de biais culturel dans les approches d'entraînement suggérées, qui pourraient ne pas être adaptées à toutes les populations ou contextes locaux.
- **Qualité hétérogène des données club** : L'efficacité du système dépend de la qualité des données saisies. Si les athlètes ne renseignent pas correctement leurs RPE ou si les historiques sont lacunaires ("Garbage In, Garbage Out"), les recommandations seront peu pertinentes.

23.2 Risques Humains et Éthiques

23.2.1 Biais d'automatisation et perte d'esprit critique

Le risque majeur est que les entraîneurs fassent une confiance aveugle à l'outil ("Si l'IA le dit, c'est vrai") et perdent leur esprit critique ou leur intuition clinique. Ce phénomène, documenté dans la littérature sur l'automatisation, peut conduire à une atrophie des compétences décisionnelles.

Pour mitiger ce risque, nous intégrons plusieurs mécanismes :

- **Obligation de validation explicite** : Chaque recommandation IA doit être validée par un clic explicite de l'entraîneur
- **Périodes sans IA** : Une semaine par mois sans assistance IA pour maintenir l'acuité décisionnelle
- **Mode pédagogique** : Option permettant à l'IA d'expliquer son raisonnement en détail, favorisant l'apprentissage plutôt que la dépendance
- **Exercices de confrontation** : Cas d'usage où l'IA se trompe délibérément pour tester la vigilance de l'entraîneur

23.2.2 Déshumanisation de la relation coach-athlète

L'athlétisme est un sport fondamentalement humain. Si le coach passe plus de temps sur son écran qu'à observer ses athlètes sur la piste, l'outil aura échoué, même s'il est techniquement parfait. L'IA doit rester un outil de back-office, libérant du temps pour les interactions de qualité.

Notre architecture privilégie l'automatisation des tâches administratives répétitives (rédition de bilans, recherche documentaire) tout en préservant le temps d'observation terrain, d'écoute, et de soutien psychologique. Les métriques de succès incluent non seulement la qualité technique des recommandations, mais aussi le temps libéré pour le coaching humain relationnel.

23.2.3 Fracture numérique et exclusion

Tous les entraîneurs ne sont pas à l'aise avec le numérique. Il existe un risque d'exclusion pour les bénévoles les moins technophiles, créant une "athlétisme à deux vitesses" entre clubs équipés et clubs non équipés.

Notre stratégie de mitigation inclut :

- **Interface intuitive** : Design UX simplifié, avec mode "débutant" pour les utilisateurs moins technophiles
- **Formation progressive** : Accompagnement personnalisé avec période d'adaptation
- **Support renforcé** : Hotline dédiée et sessions de formation régulières
- **Modèle économique accessible** : Tarification adaptée aux budgets des clubs amateurs

RAG (Retrieval-Augmented Generation) Technique combinant un modèle de langage génératif (LLM) avec un système de recherche d'information (Retrieval) pour ancrer les réponses dans des données factuelles externes.

Embedding Représentation vectorielle (liste de nombres) d'un texte capturant son sens sémantique. Permet de calculer la similarité entre deux textes (ex : question et document).

Token Unité de base du texte pour un LLM (environ 0.75 mot). Les coûts et limites des modèles sont calculés en tokens.

Chunking Découpage des documents longs en segments plus courts (chunks) pour l'indexation vectorielle.

Fine-tuning Ré-entraînement léger d'un modèle pré-entraîné sur un jeu de données spécifique pour spécialiser son style ou ses connaissances.

Prompt Engineering Art de concevoir les instructions (prompts) données au LLM pour optimiser la qualité et la pertinence de ses réponses.

Hallucination Phénomène où un LLM génère une information plausible mais factuellement fausse ou inventée.

VMA (Vitesse Maximale Aérobie) Vitesse de course à laquelle la consommation d'oxygène est maximale. Indicateur clé de performance en endurance.

RPE (Rate of Perceived Exertion) Échelle subjective (1-10) mesurant la pénibilité ressentie d'un effort par l'athlète.

ACWR (Acute :Chronic Workload Ratio) Ratio entre charge aiguë (7 derniers jours) et charge chronique (28 derniers jours). Valeur >1.5 = risque de blessure accru selon Gabbett (2016). Controversé depuis Impellizzeri et al. (2019).

LTAD (Long-Term Athlete Development) Modèle de développement athlétique à long terme, structuré par stades de maturation plutôt que par âge chronologique. Adapte les charges d'entraînement aux capacités biologiques.

Periodization (Périodisation) Structuration temporelle de l'entraînement en cycles (macrocycle = saison, mésocycle = 3-6 semaines, microcycle = semaine) pour optimiser la performance et prévenir le surentraînement.

Tapering (Affûtage) Réduction progressive et contrôlée de la charge d'entraînement (volume -40 à -60%, intensité maintenue) dans les 10-21 jours précédant une compétition majeure pour maximiser la forme.

HRV (Heart Rate Variability) Variabilité de la fréquence cardiaque entre battements consécutifs, indicateur de récupération et de stress du système nerveux autonome. HRV élevée = bonne récupération.

Fartlek Méthode d'entraînement suédoise (« jeu de vitesse ») alternant vitesses et terrains de manière non structurée, développant à la fois capacités aérobies et anaérobies.

Plyométrie Exercices explosifs utilisant le cycle étirement-raccourcissement musculaire (ex : bonds, sauts en contrebas) pour développer la puissance et l'élasticité musculo-tendineuse.

Seuil Lactique Intensité d'effort (85-90% FCM) où la production de lactate dépasse sa clairance, marquant la transition aérobie/anaérobie. Entraînable par séances au seuil.

VO₂max Consommation maximale d'oxygène (ml/kg/min), indicateur de capacité aérobie maximale. Corrélué à la performance en endurance mais non suffisant (économie de course, seuil lactique également déterminants).

Cross-Encoder Modèle de reranking évaluant la pertinence query-document de manière fine en traitant simultanément les deux textes. Plus précis que bi-encoder mais plus coûteux computationnellement.

Few-shot Learning Technique de prompt engineering fournissant quelques exemples (2-5) au LLM pour guider sa génération. Améliore la cohérence et le respect du format souhaité.

Hallucination (LLM) Génération d'informations factuellement incorrectes mais plausibles par un LLM. Risque majeur en coaching, d'où l'importance du RAG pour ancrer dans des sources vérifiées.

HNSW (Hierarchical Navigable Small World) Algorithme d'indexation vectorielle pour recherche de voisins approximatifs (ANN). Utilisé par FAISS et pgvector pour recherche rapide dans millions de vecteurs.

Semantic Search (Recherche Sémantique) Recherche basée sur la similarité sémantique (embeddings) plutôt que correspondance exacte de mots-clés. Permet de retrouver "fatigue" même si document contient "épuisement".

Zero-shot Capacité d'un LLM à résoudre une tâche sans exemples spécifiques, uniquement via instructions en langage naturel. GPT-4 excelle en zero-shot sur tâches variées.

23.3 Annexe B : Modèle de Données (DDL Simplifié)

```
-- Table Athlètes
CREATE TABLE athletes (
    id UUID PRIMARY KEY,
    nom VARCHAR(100),
    date_naissance DATE,
    vma_actuelle DECIMAL(4,2),
    fcm DECIMAL(3,0),
    group_id UUID REFERENCES groups(id)
);
```

```
-- Table Séances (Planifiées)
CREATE TABLE sessions (
    id UUID PRIMARY KEY,
    athlete_id UUID REFERENCES athletes(id),
```

```

date_prevue DATE,
type_seance VARCHAR(50), -- 'VMA', 'Seuil', 'Endurance'
description TEXT,
charge_prevue INTEGER -- RPE x Durée
);

-- Table Réalisations (Feedback)
CREATE TABLE feedbacks (
    id UUID PRIMARY KEY,
    session_id UUID REFERENCES sessions(id),
    date_realisation DATE,
    rpe_ressenti INTEGER CHECK (rpe_ressenti BETWEEN 1 AND 10),
    duree_reelle INTEGER, -- minutes
    commentaire TEXT,
    douleur_zone VARCHAR(50),
    douleur_intensite INTEGER
);

-- Table Documents (RAG)
CREATE TABLE documents (
    id UUID PRIMARY KEY,
    titre VARCHAR(200),
    contenu TEXT,
    embedding VECTOR(1536), -- pgvector
    source_type VARCHAR(50), -- 'PLAN', 'BILAN', 'ARTICLE',
    date_creation TIMESTAMP
);

```

23.4 Annexe C : Bibliothèque de Prompts

Prompt Système (Assistant Coach)

"Tu es un assistant expert en coaching d'athlétisme, spécialisé dans le demi-fond. Ton rôle est d'aider l'entraîneur à analyser les données et adapter les plans. Règles impératives : 1. Base tes réponses UNIQUEMENT sur les documents fournis dans le CONTEXTE. 2. Si l'information est absente, dis 'Je ne trouve pas cette information dans les documents disponibles'. N'invente rien. 3. Cite toujours tes sources (ex : [Plan-Octobre-2024]). 4. Adopte un ton professionnel, encourageant mais factuel. 5. Pour les questions médicales, rappelle que tu n'es pas médecin et renvoie vers le staff médical."

Prompt Génération Séance (VMA)

"Génère une séance de VMA courte pour un athlète de niveau Régional (VMA 18.5 km/h). Objectif : Développement puissance aérobie. Contraintes : Piste 400m, groupe de 5 athlètes. Format attendu : Échauffement (20'), Corps de séance (détail répétitions/récup), Retour au calme (10'). Calcule les temps de passage précis pour les 200m et 400m basés sur 100% et 105% VMA."

Système d'Assistance par Intelligence Artificielle pour le Coaching Athlétique

Identité du responsable de traitement :

[Nom du Club / Fédération]
Adresse : [Adresse complète]
Délégué à la Protection des Données (DPO) : [Nom et coordonnées]
Email DPO : [email]

Finalités du traitement :

Les données personnelles collectées sont traitées pour les finalités suivantes :

- Personnalisation des plans d'entraînement selon profil individuel
- Analyse de la progression athlétique et détection de patterns
- Prévention des blessures par analyse de charge d'entraînement
- Amélioration de la communication coach-athlète via interface conversationnelle
- Génération de statistiques anonymisées pour pilotage du club

Base légale : Consentement explicite de la personne concernée (Art. 6.1.a RGPD).

Données collectées :

- *Données d'identité* : Nom, prénom, date de naissance, sexe, numéro de licence
- *Données de performance* : Résultats de compétitions, tests physiques (VMA, détente, etc.), records personnels
- *Données de santé* : Fréquence cardiaque, variabilité cardiaque (HRV), historique de blessures, zones douloureuses, RPE (perception de l'effort)
- *Données d'entraînement* : Séances planifiées et réalisées, charge d'entraînement, commentaires libres, objectifs sportifs
- *Données techniques* : Logs de connexion, historique d'utilisation du système (pour amélioration UX)

Destinataires des données :

- Entraîneur(s) assigné(s) à l'athlète (accès complet)
- Staff médical du club (avec consentement spécifique additionnel)
- Système d'Intelligence Artificielle (données **pseudonymisées uniquement**, identifiants remplacés par UUIDs)
- Prestataire d'hébergement (OVH France, serveurs UE, accord de sous-traitance RGPD)
- Aucun transfert hors Union Européenne

Durée de conservation :

- *Pendant la durée de la licence active* : Conservation complète des données
- *Après fin de licence* : Conservation 3 ans (prescription légale responsabilité)
- *Après 3 ans* : Suppression automatique ou anonymisation irréversible
- *Données de santé* : Suppression prioritaire après 1 an d'inactivité
- Possibilité de demander suppression anticipée à tout moment (droit à l'effacement)

Vos droits (Art. 15 à 22 RGPD) :

Conformément au Règlement Général sur la Protection des Données, vous disposez des droits suivants :

- **Droit d'accès** : Obtenir copie de vos données personnelles
- **Droit de rectification** : Corriger des données inexactes ou incomplètes
- **Droit à l'effacement** (droit à l'oubli) : Demander suppression de vos données
- **Droit à la portabilité** : Recevoir vos données dans format structuré (JSON/CSV)
- **Droit d'opposition** : Vous opposer au traitement pour motifs légitimes
- **Droit de limitation** : Demander gel temporaire du traitement
- **Droit de retirer votre consentement** : À tout moment, sans justification
- **Droit de réclamation** : Auprès de la CNIL (www.cnil.fr) en cas de litige

Pour exercer vos droits, contactez le DPO : [email DPO] (réponse sous 1 mois maximum).

Mesures de sécurité :

Vos données sont protégées par :

- Chiffrement TLS 1.3 en transit et AES-256 au repos
- Pseudonymisation avant envoi à l'IA (identifiants remplacés par UUIDs)
- Authentification multi-facteurs (MFA) pour accès entraîneurs
- Sauvegardes chiffrées quotidiennes, stockées 30 jours
- Logs d'audit complets (qui a accédé à quoi, quand)
- Politique Zero-Retention avec API LLM (données non utilisées pour entraînement modèles)

CONSENTEMENT ÉCLAIRÉ

Je soussigné(e) _____, né(e) le _____,

déclare avoir été informé(e) de manière claire et complète sur :

- Les finalités du traitement de mes données personnelles
- Les catégories de données collectées, notamment les données de santé
- Les destinataires de mes données et les mesures de sécurité appliquées
- Mes droits et les modalités pour les exercer
- La durée de conservation de mes données

Je consens librement et de manière éclairée au traitement de mes données personnelles, y compris les données de santé, dans les conditions décrites ci-dessus.

Consentement spécifique pour partage avec staff médical (optionnel) :

J'autorise le partage de mes données de santé avec le staff médical du club

Fait à _____, le _____

Signature de l'athlète (ou représentant légal si mineur) :

Important : Ce consentement peut être retiré à tout moment, sans justification, en contactant le DPO. Le retrait du consentement entraîne la suppression de vos données sous 30 jours et la désactivation de votre accès au système IA.

Pour les mineurs (<18 ans) : Signature du représentant légal obligatoire.

24 Conclusion Générale

Cette thèse professionnelle a exploré comment l'intégration pragmatique des Modèles de Langage de Grande Taille (LLM) peut démocratiser l'accès à l'expertise de haut niveau pour l'ensemble de l'écosystème athlétique français. En répondant à la problématique centrale – *Comment concevoir et déployer une architecture d'assistance par IA qui permette aux entraîneurs de clubs et fédérations d'exploiter la richesse des données sportives pour personnaliser l'entraînement à grande échelle, tout en préservant et renforçant la dimension fondamentalement humaine de la relation coach-athlète ?* – nous avons proposé une approche constructiviste et pluridisciplinaire.

24.1 Synthèse des apports principaux

Nos travaux apportent cinq contributions majeures :

1. Architecture RAG spécialisée pour le coaching sportif : Nous avons conçu une architecture Retrieval-Augmented Generation adaptée aux spécificités du coaching d'athlétisme, combinant recherche vectorielle sémantique et filtrage par métadonnées (athlète, période, type de séance). Cette approche limite les hallucinations en ancrant systématiquement les recommandations dans les documents de référence (plans d'entraînement historiques, bilans, protocoles FFA).

2. Modèle de données inspiré des structures FFA : La structuration proposée distingue clairement le *Prescrit* (intentions d'entraînement) du *Réalisé* (feedback terrain), permettant une analyse fine de l'écart entre planification et exécution. Cette dichotomie est essentielle pour l'apprentissage continu du système et l'ajustement des recommandations.

3. Dimension éthique et gouvernance intégrée : Au-delà de la conformité RGPD/CNIL, nous avons proposé un cadre éthique incluant explicabilité (LIME, SHAP), comité de validation, et stratégies de mitigation des biais algorithmiques. L'IA doit rester un outil d'augmentation, non de remplacement de l'expertise humaine.

4. Plan de déploiement progressif et réaliste : Notre roadmap sur 12 mois (POC → Pilotes → Généralisation) intègre les contraintes organisationnelles, la formation des utilisateurs, et un modèle économique transparent (TCO, ROI). L'approche privilégie l'adoption progressive plutôt que le big bang technologique.

5. Cas d'usage concrets et actionnables : Les épopées fonctionnelles (assistance à la planification, suivi et prévention) illustrent comment l'IA peut libérer l'entraîneur des

tâches analytiques répétitives pour se recentrer sur la pédagogie, l'observation terrain, et le soutien psychologique.

24.2 Réponse à la problématique

Notre architecture répond à la problématique en proposant un *système d'Intelligence Augmentée* où :

- La machine gère la complexité des données (analyse de charge, détection de patterns, recherche dans la littérature)
- L'humain conserve la décision finale, l'intuition clinique, et la relation pédagogique
- Le RAG garantit la traçabilité et l'explicabilité des recommandations
- La gouvernance éthique prévient les dérives (biais, dépendance, déshumanisation)

L'IA ne remplacera pas l'entraîneur. Mais l'entraîneur qui utilise l'IA remplacera celui qui ne l'utilise pas, à condition que l'outil soit conçu comme un *copilote cognitif* respectueux de l'expertise humaine.

24.3 Analyse comparative avec les solutions existantes

Pour positionner notre contribution, il est essentiel de comparer notre approche avec les solutions existantes sur le marché et dans la littérature académique.

24.3.1 Comparaison avec les plateformes de suivi sportif

Les plateformes existantes (TrainingPeaks, Strava, Garmin Connect) offrent principalement des fonctionnalités de visualisation de données et de planification basique. Elles manquent de capacités d'analyse intelligente et de personnalisation contextuelle. Leur approche est essentiellement réactive (affichage de données) plutôt que proactive (suggestions adaptatives basées sur l'analyse de patterns).

Notre solution se distingue par :

- **Assistance conversationnelle** : Interface en langage naturel plutôt que formulaires complexes
- **Personnalisation contextuelle** : Recommandations adaptées au profil, historique, et contraintes de chaque athlète
- **Explicabilité** : Justification systématique des recommandations avec sources citées
- **Intégration fédérale** : Alignement sur les référentiels FFA plutôt que standards génériques

24.3.2 Comparaison avec les solutions IA génériques

Les solutions d'IA génériques (chatbots basiques, assistants vocaux) ne sont pas spécialisées pour le domaine sportif. Elles manquent de connaissances spécifiques sur la physiologie de l'entraînement, les protocoles fédéraux, et les contraintes du coaching d'athlétisme.

Notre architecture RAG spécialisée permet d'ancrer les réponses dans un corpus documentaire expert (protocoles FFA, études scientifiques, plans historiques validés), garantissant la pertinence et la fiabilité des recommandations pour le contexte spécifique du coaching d'athlétisme.

24.3.3 Positionnement par rapport à la littérature académique

Notre contribution s'inscrit dans la continuité des travaux sur l'IA dans le sport (Claudino et al.¹¹⁰, Rein et Memmert¹¹¹) tout en innovant par l'application des architectures RAG et LLM au coaching conversationnel, domaine encore peu exploré dans la littérature.

24.4 Limites et contraintes assumées

Cette thèse reconnaît plusieurs limites :

- **Limites techniques** : Les hallucinations persistent malgré le RAG, nécessitant une validation humaine systématique. La qualité du système dépend de la qualité des données saisies (Garbage In, Garbage Out).
- **Biais des données** : Les modèles LLM sont entraînés sur des corpus majoritairement anglophones, avec un risque de biais culturel dans les approches suggérées.
- **Risques humains** : Automation bias, fracture numérique, et risque de déshumanisation de la relation coach-athlète si l'outil devient prépondérant.
- **Validation terrain limitée** : Cette thèse propose une architecture et un plan de déploiement, mais n'inclut pas de résultats d'expérimentation à grande échelle. La validation complète nécessitera des pilotes terrain sur plusieurs saisons.

24.5 Intégration avec les écosystèmes existants

L'adoption de notre solution nécessite une intégration harmonieuse avec les écosystèmes existants dans le monde de l'athlétisme. Cette intégration concerne à la fois les outils techniques, les processus organisationnels, et les standards fédéraux.

24.5.1 Interopérabilité avec les outils de suivi existants

Les clubs utilisent déjà une variété d'outils de suivi : montres GPS (Garmin, Polar), applications mobiles (Strava, TrainingPeaks), et systèmes de gestion de club. Notre architecture doit s'intégrer avec ces outils via des APIs standardisées, permettant l'import automatique des données d'entraînement sans duplication de saisie manuelle.

Les standards de données comme TCX (Training Center XML) et FIT (Flexible and Interoperable Data Transfer) permettent l'échange de données entre différents systèmes. Notre solution supporte l'import de ces formats, garantissant une compatibilité maximale avec l'écosystème existant.

24.5.2 Intégration avec les systèmes fédéraux

La Fédération Française d'Athlétisme dispose de ses propres systèmes de gestion (licences, résultats de compétitions, classements). Notre architecture s'aligne sur les réfé-

110. Claudino, J.G. et al. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports : a systematic review. *Sports Medicine-Open*, 5(1), 1-12.

111. Rein, R., Memmert, D. (2016). Big data and tactical analysis in elite soccer : future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1-13.

rentiels FFA¹¹² pour permettre l'import automatique des données fédérales et l'export de statistiques consolidées pour le pilotage national.

Cette intégration permet aux clubs de bénéficier d'une vision consolidée de leurs athlètes, combinant les données d'entraînement locales avec les résultats de compétition officiels, offrant une analyse complète de la progression et de la performance.

24.5.3 Respect des standards de données de santé

Les données de santé (fréquence cardiaque, variabilité cardiaque, blessures) sont soumises à des réglementations strictes. Notre architecture respecte les standards HL7 FHIR (Fast Healthcare Interoperability Resources) pour l'échange de données de santé, garantissant l'interopérabilité avec les systèmes médicaux et la conformité aux exigences réglementaires.

24.6 Enjeux de scalabilité et performance

Le passage à l'échelle de quelques clubs pilotes à des centaines de clubs avec des milliers d'athlètes pose des défis de scalabilité qu'il est essentiel d'anticiper.

24.6.1 Scalabilité de l'indexation vectorielle

L'indexation vectorielle avec pgvector¹¹³ et FAISS permet de gérer des millions de documents, mais nécessite une stratégie de partitionnement pour maintenir les performances. Notre architecture prévoit un partitionnement par club et par période, permettant de limiter l'espace de recherche à des sous-ensembles pertinents.

Pour des déploiements à très grande échelle (milliers de clubs), une architecture distribuée avec sharding horizontal serait nécessaire, chaque shard gérant un sous-ensemble de clubs avec réPLICATION pour la haute disponibilité.

24.6.2 Optimisation des coûts d'API LLM

Les appels aux APIs LLM (GPT-4, Claude) représentent un coût significatif à l'échelle. Notre stratégie d'optimisation inclut :

- **Caching intelligent** : Mise en cache des réponses pour questions fréquentes avec invalidation basée sur la fraîcheur des données
- **Modèles adaptatifs** : Utilisation de modèles plus légers (GPT-3.5) pour questions simples, réservation de GPT-4 pour analyses complexes
- **Batch processing** : Traitement par lots des requêtes non-urgentes pour optimiser l'utilisation des quotas API
- **Alternatives open-source** : Possibilité de basculer vers LLaMA¹¹⁴ pour réduire les coûts à long terme

112. Fédération Française d'Athlétisme (référentiels publics, structures d'épreuves). <https://www.athle.fr/>

113. Johnson, J., Douze, M., Jégou, H. (2017). Billion-scale similarity search with GPUs (FAISS). arXiv :1702.08734.

114. Touvron, H. et al. (2023). LLaMA : Open and Efficient Foundation Language Models. arXiv :2302.13971.

24.6.3 Monitoring et observabilité à l'échelle

Le monitoring à grande échelle nécessite des outils d'observabilité robustes. Notre architecture intègre :

- **Logging structuré** : Logs JSON avec corrélation des requêtes (trace IDs) pour suivre une requête de bout en bout
- **Métriques temps réel** : Dashboard de monitoring avec alertes automatiques sur dérive des performances
- **Distributed tracing** : Traçage distribué pour identifier les goulots d'étranglement dans le pipeline RAG
- **Health checks** : Endpoints de santé pour chaque composant, permettant détection automatique des défaiillances

24.7 Perspectives de recherche futures

Plusieurs axes de recherche émergent de ce travail :

Axe 1 : Personnalisation fine par apprentissage continu : Adapter les modèles LLM via fine-tuning sur les données spécifiques d'un club ou d'une fédération, permettant une spécialisation progressive sans perte de robustesse.

Axe 2 : Multimodalité et intégration de données hétérogènes : Étendre le RAG pour intégrer images (analyse vidéo technique), signaux physiologiques (HRV, GPS), et données contextuelles (météo, qualité du sommeil) dans un modèle unifié.

Axe 3 : Explicabilité avancée et confiance calibrée : Développer des métriques de confiance plus fines que les probabilités brutes, et des explications narratives adaptées au niveau d'expertise de l'utilisateur (entraîneur débutant vs expert).

Axe 4 : Éthique computationnelle et auditabilité : Automatiser la détection de biais, l'audit de l'équité algorithmique, et la traçabilité des décisions pour garantir la conformité réglementaire et éthique à grande échelle.

Axe 5 : Impact organisationnel et changement culturel : Étudier longitudinalement comment l'adoption de l'IA transforme les pratiques, les compétences requises, et la culture organisationnelle des clubs et fédérations.

24.8 Feuille de route de recherche future

Notre travail ouvre plusieurs pistes de recherche future pour approfondir l'intégration de l'IA dans le coaching sportif.

24.8.1 Personnalisation adaptative des modèles

Une direction prometteuse consiste à développer des mécanismes d'adaptation continue des modèles RAG aux spécificités de chaque club et de chaque athlète. Cette personnalisation pourrait s'appuyer sur des techniques de fine-tuning léger (LoRA - Low-Rank Adaptation) ou sur l'optimisation dynamique des prompts selon les retours utilisateurs.

Les travaux de Borgeaud et al.¹¹⁵ sur l'amélioration des LLM par récupération de trillions de tokens ouvrent des perspectives pour enrichir continuellement le corpus RAG avec de nouvelles données d'entraînement validées.

115. Borgeaud, S. et al. (2022). Improving language models by retrieving from trillions of tokens. arXiv :2112.04426.

24.8.2 Intégration multi-modale

L'avenir du coaching assisté par IA réside dans l'intégration de données multi-modales : texte (notes, bilans), données numériques (GPS, FC), et même vidéo (analyse biomécanique). Cette intégration permettrait une compréhension holistique de l'état de l'athlète, combinant ressenti subjectif, données objectives, et observation visuelle.

Les travaux de Stein et al.¹¹⁶ sur la combinaison de données vidéo et de mouvement ouvrent des perspectives pour étendre notre architecture au-delà du texte et des données numériques.

24.8.3 Explicabilité avancée et confiance

L'amélioration de l'explicabilité reste un défi majeur. Les techniques actuelles (LIME¹¹⁷, SHAP¹¹⁸) fournissent des explications locales, mais manquent de cohérence globale. Des recherches futures pourraient développer des méthodes d'explication narratives, générant des récits cohérents expliquant le raisonnement du système de bout en bout.

24.8.4 Évaluation longitudinale à grande échelle

Notre protocole d'évaluation propose un A/B testing sur 12 semaines avec 40 athlètes. Une recherche future pourrait étendre cette évaluation sur plusieurs saisons, avec des centaines d'athlètes, pour mesurer l'impact à long terme sur la performance, la prévention des blessures, et la satisfaction des utilisateurs.

24.9 Cas d'usage détaillés et scénarios d'utilisation

Pour illustrer concrètement l'utilisation de notre système, nous détaillons plusieurs cas d'usage représentatifs couvrant les principaux besoins identifiés.

24.9.1 Cas d'usage 1 : Planification hebdomadaire assistée

Contexte : Sarah, entraîneur de demi-fond, doit planifier la semaine d'entraînement pour son groupe de 12 athlètes de niveau régional. Objectif : préparation d'un meeting dans 3 semaines.

Utilisation du système :

1. Sarah interroge le système : "Génère un microcycle de préparation pour meeting dans 3 semaines, groupe régional demi-fond"
2. Le système RAG récupère les protocoles FFA de préparation compétition, les plans historiques similaires, et les données de charge récentes des athlètes
3. Le LLM génère une proposition structurée : échauffement progressif (lundi), séance VMA (mercredi), séance spécifique (vendredi), récupération active (dimanche)

116. Stein, M. et al. (2019). Bring it to the pitch : combining video and movement data to enhance team sport analysis. IEEE Transactions on Visualization and Computer Graphics, 26(1), 143-153.

117. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. KDD 2016.

118. Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. NeurIPS 2017.

4. Sarah valide et ajuste selon les contraintes individuelles (un athlète revient de blessure, un autre a des contraintes professionnelles)
5. Le système envoie automatiquement les séances détaillées aux athlètes via l'application mobile

Valeur ajoutée : Gain de temps estimé 2h/semaine, qualité de planification améliorée grâce à l'accès aux référentiels fédéraux, traçabilité complète des décisions.

24.9.2 Cas d'usage 2 : Détection précoce de surcharge

Contexte : Thomas, athlète de 18 ans, présente des RPE élevés depuis 10 jours malgré des séances prévues "faciles". L'entraîneur s'inquiète d'un risque de surentraînement.

Utilisation du système :

1. Le système calcule automatiquement le ratio ACWR (Acute :Chronic Workload Ratio) de Thomas
2. Alerte automatique générée : "Ratio ACWR = 1.8 (> seuil 1.5), RPE moyen +25% vs historique, risque de surentraînement"
3. Le système suggère : "Réduire la charge de 30% cette semaine, privilégier récupération active, surveiller sommeil"
4. L'entraîneur consulte les sources : protocole Gabbett¹¹⁹ sur ACWR, études de Jaspers et al.¹²⁰ sur corrélations charge-blessures
5. L'entraîneur adapte le plan en conséquence et programme un suivi renforcé

Valeur ajoutée : Détection précoce permettant prévention plutôt que réaction, justification scientifique des recommandations, traçabilité pour le staff médical.

24.9.3 Cas d'usage 3 : Personnalisation pour athlète vétéran

Contexte : Marie, 42 ans, fondeuse vétéran, souhaite adapter son plan marathon compte tenu de douleurs récurrentes aux genoux et de contraintes professionnelles.

Utilisation du système :

1. Marie décrit sa situation en langage naturel : "J'ai des douleurs aux genoux depuis 2 semaines, je ne peux courir que le week-end"
2. Le système analyse le texte, extrait les contraintes (douleurs genoux, disponibilité limitée)
3. Le RAG récupère des protocoles d'adaptation pour blessures genoux, des plans pour vétérans, et des alternatives d'entraînement (vélo, natation)
4. Le LLM génère un plan adapté : renforcement musculaire spécifique (lundi), vélo (mercredi), course terrain souple (samedi), récupération (dimanche)
5. L'entraîneur valide et ajoute des exercices de renforcement spécifiques recommandés par le kiné

Valeur ajoutée : Personnalisation fine tenant compte de multiples contraintes, accès à des alternatives d'entraînement, préservation de la motivation malgré les limitations.

119. Gabbett, T.J. (2016). The training-injury prevention paradox : should athletes be training smarter and harder ? British Journal of Sports Medicine, 50(5), 273-280.

120. Jaspers, A. et al. (2017). Relationships between training load indicators and training outcomes in professional soccer. Sports Medicine, 47(3), 533-544.

24.10 Recommandations finales actionnables

Pour les **entraîneurs et clubs** :

- Commencer par un POC sur un groupe pilote (5-10 athlètes) pour valider l'utilité avant généralisation
- Investir dans la qualité des données : formation à la saisie rigoureuse, synchronisation des capteurs
- Maintenir un esprit critique : l'IA propose, l'humain dispose. Toujours valider les recommandations contre l'observation terrain
- Former les entraîneurs à l'interprétation des explications IA (LIME, SHAP) pour renforcer la confiance éclairée

Pour les **fédérations** :

- Standardiser les formats de données pour faciliter l'interopérabilité entre clubs et outils
- Créer un comité éthique IA pour superviser les déploiements et valider les bonnes pratiques
- Investir dans la formation des entraîneurs à l'IA (niveau ABC et supérieur)
- Promouvoir le partage d'expériences entre clubs pilotes pour accélérer l'apprentissage collectif

Pour les **développeurs et fournisseurs de solutions** :

- Privilégier l'explicabilité et la transparence dès la conception (Privacy by Design, Ethics by Design)
- Implémenter des garde-fous techniques (détection d'hallucinations, filtrage de sécurité)
- Proposer des modèles économiques adaptés aux budgets contraints des clubs amateurs
- Assurer un support et une formation continue, pas seulement la livraison d'un produit

24.11 Évaluation de l'impact organisationnel

L'introduction de l'IA dans le coaching d'athlétisme ne se limite pas à un changement technologique, mais transforme profondément les pratiques organisationnelles et les relations entre les acteurs.

24.11.1 Transformation des rôles et compétences

L'adoption de l'IA modifie les compétences requises des entraîneurs. Les compétences techniques traditionnelles (physiologie, biomécanique) restent essentielles, mais s'enrichissent de nouvelles compétences :

- **Littératie data** : Capacité à interpréter les visualisations et métriques générées par l'IA
- **Pensée critique vis-à-vis de l'IA** : Capacité à évaluer la pertinence des recommandations algorithmiques

- **Gestion de l'incertitude** : Acceptation que l'IA peut se tromper et nécessite validation humaine

Cette évolution nécessite des programmes de formation continue adaptés, intégrant à la fois les aspects techniques (utilisation de l'outil) et les aspects éthiques (gestion des biais, préservation de l'autonomie).

24.11.2 Impact sur la relation entraîneur-athlète

La question centrale est de savoir si l'IA améliore ou dégrade la relation humaine entre entraîneur et athlète. Notre approche "Human-in-the-loop" maintient l'entraîneur comme décideur final, mais libère du temps pour les interactions de qualité.

Les études sur l'acceptabilité technologique, notamment le modèle TAM étendu de Venkatesh et Davis¹²¹, montrent que l'utilité perçue et la facilité d'utilisation sont des facteurs clés d'adoption. Pour notre système, cela se traduit par la démonstration concrète d'un gain de temps et d'une amélioration de la qualité des plans, tout en maintenant une interface intuitive et des explications claires.

24.11.3 Gouvernance et responsabilité

L'introduction de l'IA soulève des questions de gouvernance : qui est responsable en cas d'erreur ? Comment gérer les désaccords entre recommandations IA et intuition coach ? Notre architecture intègre des mécanismes de traçabilité complète (audit logs) et de validation systématique, garantissant que la responsabilité reste sur l'entraîneur qualifié, l'IA n'étant qu'un outil d'aide à la décision.

24.12 Message final

L'intégration de l'IA dans le coaching d'athlétisme n'est pas une question de technologie pure, mais de *design éthique et organisationnel*. Le succès dépendra moins de la sophistication algorithmique que de la capacité à créer un écosystème où l'IA augmente l'humain sans le remplacer, où la data enrichit l'intuition sans la supplanter, et où l'efficacité technique sert l'excellence pédagogique.

L'avenir du coaching sportif réside dans cette *Intelligence Augmentée* : une alliance entre l'expertise humaine millénaire et les capacités computationnelles modernes, au service de la performance et du bien-être des athlètes.

24.13 Appel à l'action pour l'écosystème athlétique

Cette thèse constitue un appel à l'action pour l'ensemble de l'écosystème athlétique français. Les technologies d'IA générative sont désormais accessibles et matures pour être déployées dans le coaching sportif. Le défi n'est plus technique, mais organisationnel et éthique.

Aux clubs : Nous encourageons les clubs à expérimenter progressivement avec l'IA, en commençant par des pilotes restreints sur des cas d'usage précis (planification, analyse de charge). L'investissement initial peut être modeste, avec un retour sur investissement mesurable en termes de gain de temps et de qualité de suivi.

121. Venkatesh, V., Davis, F.D. (2000). A theoretical extension of the technology acceptance model : Four longitudinal field studies. *Management Science*, 46(2), 186-204.

Aux fédérations : Nous recommandons aux fédérations de définir des standards et référentiels pour l'utilisation de l'IA dans le coaching, garantissant la cohérence et la qualité tout en préservant l'autonomie des clubs. La FFA¹²² pourrait jouer un rôle central dans la diffusion des bonnes pratiques et la formation des entraîneurs.

Aux chercheurs : Nous appelons la communauté de recherche à approfondir l'évaluation longitudinale de l'impact de l'IA sur la performance athlétique et le bien-être, avec des études contrôlées sur plusieurs saisons. Les questions de biais algorithmiques, d'équité d'accès, et d'impact sur la relation coach-athlète méritent des investigations approfondies.

Aux développeurs : Nous encourageons le développement d'outils open-source spécialisés pour le coaching sportif, permettant aux clubs de bénéficier de l'IA sans dépendre de solutions propriétaires coûteuses. Les modèles open-source comme LLaMA¹²³ offrent des alternatives viables aux modèles propriétaires.

Le moment est venu de démocratiser l'accès à l'expertise de haut niveau pour l'ensemble de l'athlétisme français, en préservant la dimension fondamentalement humaine qui fait la richesse du coaching sportif.

122. Fédération Française d'Athlétisme (référentiels publics, structures d'épreuves). <https://www.athle.fr/>

123. Touvron, H. et al. (2023). LLaMA : Open and Efficient Foundation Language Models. arXiv :2302.13971.

Références

1. Vaswani, A. et al. (2017). Attention Is All You Need. NeurIPS 2017.
2. Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.
3. Johnson, J., Douze, M., Jégou, H. (2017). Billion-scale similarity search with GPUs (FAISS). arXiv :1702.08734.
4. Reimers, N., Gurevych, I. (2019). Sentence-BERT : Sentence Embeddings using Siamese BERT-Networks. EMNLP 2019.
5. Karpukhin, V. et al. (2020). Dense Passage Retrieval for Open-Domain Question Answering. EMNLP 2020.
6. Izacard, G., Grave, E. (2021). Leveraging Passage Retrieval with Generative Models for Open Domain QA. EACL 2021.
7. Devlin, J. et al. (2019). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.
8. Radford, A. et al. (2018). Improving Language Understanding by Generative Pre-Training. OpenAI Technical Report.
9. Brown, T. et al. (2020). Language Models are Few-Shot Learners. NeurIPS 2020.
10. Touvron, H. et al. (2023). LLaMA : Open and Efficient Foundation Language Models. arXiv :2302.13971.
11. Achiam, J. et al. (2023). GPT-4 Technical Report. arXiv :2303.08774.
12. Anthropic (2023). Claude 3 Opus : Technical Report. <https://www.anthropic.com/clause>
13. Ouyang, L. et al. (2022). Training language models to follow instructions with human feedback. NeurIPS 2022.
14. Guu, K. et al. (2020). REALM : Retrieval-Augmented Language Model Pre-Training. ICML 2020.
15. Khattab, O., Zaharia, M. (2020). ColBERT : Efficient and Effective Passage Search via Contextualized Late Interaction over BERT. SIGIR 2020.
16. Borgeaud, S. et al. (2022). Improving language models by retrieving from trillions of tokens. arXiv :2112.04426.
17. Thakur, N. et al. (2021). BEIR : A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models. NeurIPS Datasets and Benchmarks 2021.
18. Wei, J. et al. (2022). Emergent abilities of large language models. TMLR 2022.
19. Bubeck, S. et al. (2023). Sparks of AGI : early experiments with GPT-4. arXiv :2303.12712.
20. Liu, P. et al. (2021). Pre-train, prompt, and predict : a systematic survey of prompting methods in natural language processing. ACM CSUR 2023.
21. Zhao, W.X. et al. (2023). A survey of large language models. arXiv :2303.18223.
22. Wei, J. et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. NeurIPS 2022.
23. OpenAI (2023). GPT-4 System Card. <https://openai.com/research/gpt-4-system-card>

24. Anthropic (2023). Claude 3 Ethics Report. <https://www.anthropic.com/clause-ethics-report>
25. Bommasani, R. et al. (2021). On the opportunities and risks of foundation models. arXiv :2108.07258.
26. Weidinger, L. et al. (2021). Ethical and social risks of harm from language models. arXiv :2112.04359.
27. Jobin, A., Ienca, M., Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
28. Mittelstadt, B.D. et al. (2016). The ethics of algorithms : Mapping the debate. *Big Data & Society*, 3(2).
29. UNESCO (2021). Recommendation on the Ethics of Artificial Intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
30. European Commission (2019). Ethics Guidelines for Trustworthy AI. High-Level Expert Group on AI.
31. Claudino, J.G. et al. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports : a systematic review. *Sports Medicine-Open*, 5(1), 1-12.
32. Bunker, R.P., Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15(1), 27-33.
33. Rein, R., Memmert, D. (2016). Big data and tactical analysis in elite soccer : future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1-13.
34. Rossi, A. et al. (2017). Effective injury forecasting in soccer with GPS training data and machine learning. *PloS One*, 12(7).
35. Gabbett, T.J. (2016). The training-injury prevention paradox : should athletes be training smarter and harder ? *British Journal of Sports Medicine*, 50(5), 273-280.
36. Impellizzeri, F.M. et al. (2019). What role do chronic workloads play in the acute to chronic workload ratio ? Time to dismiss ACWR and its underlying theory. *Sports Medicine*, 49(3), 467-469.
37. Mujika, I. (2017). Quantification of training and competition loads in endurance sports : methods and applications. *International Journal of Sports Physiology and Performance*, 12(s2), S2-9.
38. Buchheit, M. (2014). Monitoring training status with HR measures : do all roads lead to Rome ? *Frontiers in Physiology*, 5, 73.
39. Jaspers, A. et al. (2017). Relationships between training load indicators and training outcomes in professional soccer. *Sports Medicine*, 47(3), 533-544.
40. Colby, M.J. et al. (2014). Accelerometer and GPS-derived running loads and injury risk in elite Australian football. *The Journal of Strength & Conditioning Research*, 28(8), 2244-2252.
41. Scott, B.R. et al. (2013). A comparison of methods to quantify the in-season training load of professional soccer players. *International Journal of Sports Physiology and Performance*, 8(2), 195-202.
42. Thornton, H.R. et al. (2019). Developing athlete monitoring systems in team sports : data analysis and visualization. *International Journal of Sports Physiology and Performance*, 14(6), 698-705.

43. Bourdon, P.C. et al. (2017). Monitoring athlete training loads : consensus statement. *International Journal of Sports Physiology and Performance*, 12(s2), S2-161.
44. Stein, M. et al. (2019). Bring it to the pitch : combining video and movement data to enhance team sport analysis. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 143-153.
45. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *KDD 2016*.
46. Lundberg, S.M., Lee, S.I. (2017). A unified approach to interpreting model predictions. *NeurIPS 2017*.
47. Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75-105.
48. Venkatesh, V., Davis, F.D. (2000). A theoretical extension of the technology acceptance model : Four longitudinal field studies. *Management Science*, 46(2), 186-204.
49. Butrym, T.M. (2003). Posthuman podiums : Cyborg narratives of elite track and field athletes. *Sociology of Sport Journal*, 20(1), 17-39.
50. Loland, S. (2018). Performance-enhancing drugs, sport, and the ideal of natural athletic performance. *American Journal of Bioethics*, 18(6), 8-15.
51. López Frías, F.J. (2020). Technology in sport : Challenges and ethical implications. *Journal of the Philosophy of Sport*, 47(2), 219-235.
52. Eubanks, V. (2018). *Automating Inequality : How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.
53. CNIL (2023). IA et données personnelles : recommandations et ressources
54. Reglement (UE) 2016/679 (RGPD).
55. Federation Francaise d'Athletisme (referentiels publics, structures d'épreuves)
56. Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature*, 138(3479), 32.
57. Banister, E.W., Calvert, T.W., Savage, M.V., Bach, T. (1975). A systems model of training for athletic performance. *Australian Journal of Sports Medicine*, 7(3), 57-61.
58. Matveev, L.P. (1964). Problem of periodization of sports training. *Fizkultura i Sport*, Moscow.