

CAHIER DES CHARGES

HACKBYIFRI 2026

Thème : « INTÉGRATION EFFICACE DU NUMÉRIQUE DANS L'APPRENTISSAGE »

NEXUS

Assistant IA Tout-en-Un pour Étudiants IFRI

1. Nom de l'équipe :

FOUR

2. Présentation des membres :

Membre 1 (Chef de groupe) :

- Nom : ADEBOYE-SENAN
- Prénom(s) : Christ-Emile
- École : IFRI
- Filière : SeLot
- Année : Licence 1
- Téléphone : 01 92 58 03 07
- Email : emileadeboye2@gmail.com

Membre 2 :

- Nom : HOUNSOU
- Prénom(s) : Tchègnon Benny Samuel
- École : IFRI
- Filière : SI
- Année : Licence 1
- Téléphone : 01 92 62 54 31
- Email : hounsoutchegnon@gmail.com

Membre 3 :

- Nom : GBESSEMEHLAN
- Prénom(s) : Sènam Léoncelle
- École : IFRI
- Filière : SI
- Année : Licence 1
- Téléphone : 01 92 70 89 13
- Email : leoncelle.gb@gmail.com

Membre 4 :

- Nom : HOUNSOUNNOU
- Prénom(s) : Mahouti Ruth
- École : IFRI
- Filière : GL
- Année : Licence 1
- Téléphone : 01 44 69 76 65
- Email : hounsounnouruth@gmail.com

3. Problèmes identifiés :

Problème principal :

Les étudiants de l'IFRI font face à plusieurs défis majeurs dans leur apprentissage qui freinent leur progression académique et leur épanouissement personnel :

- Programme académique dense et exigeant : Les étudiants doivent maîtriser des concepts complexes (algorithmes, réseaux, mathématiques) dans des délais courts, avec peu de temps pour assimiler et approfondir chaque matière et composer .
- Parfois manque de connexions pour faire des recherches.
- Stress académique sans soutien psychologique : Pression des examens, charge de travail intense, mais aucun espace dédié au bien-être mental et à l'écoute empathique accessible 24/7.
- Absence de vitrine professionnelle : Difficile de valoriser ses compétences, projets et certificats auprès des recruteurs pour décrocher stages et emplois.
- Manque d'espace d'échange sécurisé : Aucune plateforme de confiance pour vendre/acheter du matériel informatique (PC, périphériques) entre étudiants.

Contexte du problème :

À l'IFRI, les étudiants suivent des formations intensives en informatique, réseaux et programmation. Les cours sont denses, techniques, et nécessitent une compréhension approfondie de concepts abstraits.

Cependant, plusieurs facteurs limitent l'efficacité de l'apprentissage :

- Classes surchargées (+250 étudiants par promotion)
- Ratio professeur/étudiant très faible (1 professeur pour ~70 étudiants)
- Manque de temps pour un accompagnement individuel personnalisé
- Absence d'outils pédagogiques intelligents adaptés au contexte IFRI

De plus, la pression académique génère du stress et de l'anxiété, mais aucune plateforme de soutien psychologique adaptée aux étudiants n'existe.

Catégories d'étudiants touchées :

- Étudiants en difficulté : ceux qui peinent à suivre le rythme des cours et ont besoin d'explications supplémentaires accessibles à tout moment.
- Étudiants autodidactes : ceux qui veulent approfondir au-delà du programme officiel avec un tuteur IA disponible 24/7.
- Étudiants en révision : ceux qui préparent examens et projets sans accès immédiat aux enseignants.
- Étudiants sous pression : ceux qui ont besoin d'un espace d'écoute confidentiel pour gérer stress et anxiété.
- Étudiants cherchant du matériel : ceux qui veulent acheter/vendre ordinateurs et périphériques de confiance.

4. Description de la solution :

Nom de la solution :

NEXUS

(Assistant IA Tout-en-Un pour Étudiants)

Description :

Nexus est un assistant pédagogique intelligent alimenté par l'IA, spécialement conçu pour répondre aux besoins des étudiants de l'IFRI.

La solution regroupe CINQ modules complémentaires qui s'intègrent harmonieusement pour offrir une expérience d'apprentissage complète, personnalisée et accessible 24/7 :

Module	Description
Brain	Tuteur conversationnel intelligent capable de répondre aux questions, expliquer des concepts techniques (réseaux, bases de données, programmation, maths, etc), générer du code, et s'adapter au niveau de l'étudiant. Utilise les cours IFRI comme base de connaissances via RAG (Retrieval-Augmented Generation). <u>Note</u> : 100 % offline
Care	Assistant de bien-être mental offrant écoute empathique, conseils personnalisés pour gérer stress et anxiété académique. Confidentialité totale garantie (données locales uniquement). <u>Note</u> : 100 % offline
Hub Skills	Portfolio professionnel en ligne permettant de créer son profil avec compétences, projets GitHub, certificats. URL unique partageable sur LinkedIn/CV. Export PDF automatique (à venir).
Messages	Système de communication temps réel entre étudiants avec partage de fichiers, notifications desktop, et latence ultra-faible (<100ms via WebSocket). Utile pour la marketplace et le hub entre les étudiants.
Marketplace	Plateforme d'achat/vente de matériel informatique (PC, souris, claviers, écrans) entre étudiants IFRI de confiance. Système de notation, paiement Mobile Money. La vente de cours/exercices est interdite (moralement incorrect).

Comment Nexus répond au problème ?

- Disponibilité 24/7 : Les étudiants peuvent poser leurs questions à tout moment, sans dépendre des horaires des professeurs.
- Recherche intelligente : Trouve les informations pertinentes parmi les centaines de PDFs de cours en quelques secondes grâce au RAG.
- Explications personnalisées : Nexus adapte ses réponses au niveau de l'étudiant et reformule jusqu'à compréhension complète.
- Base de connaissances locale : Utilise les cours IFRI réels pour des réponses contextualisées au programme spécifique.
- Support émotionnel : Module Care pour gérer le stress académique, prévenir le burnout et offrir un espace d'écoute confidentiel.
- Vitrine professionnelle : Hub Skills pour valoriser compétences et projets auprès des recruteurs.
- Échange sécurisé : Marketplace pour acheter/vendre matériel entre étudiants de confiance avec système de notation.
- 100% hors ligne : Fonctionne sans connexion internet (données privées, pas de cloud).

Valeur ajoutée :

Contrairement aux solutions IA génériques (ChatGPT, Claude, Gemini), Nexus est spécifiquement calibré pour l'IFRI avec :

- Connaissance des cours officiels IFRI (via RAG avec index L1 pré-construit)
- Ton pédagogique adapté aux étudiants béninois
- Multi-modèles IA pour différents usages (explication pédagogique, génération de code, bien-être empathique)
- Fonctionnement 100% local (pas de dépendance internet, données privées garanties)
- Marketplace éthique (matériel seulement, pas de vente de cours)
- Modules intégrés (Brain + Care + Hub + Messages + Marketplace) en une seule plateforme.

5. Exigences fonctionnelles :

Module Brain

- Authentification utilisateur sécurisée (JWT + bcrypt + salt)
- Chat conversationnel multi-tours avec historique persistant
- Système de Retrieval-Augmented Generation (RAG) pour interroger les cours IFRI
- Génération de code (Python, C, JavaScript, SQL) avec explications détaillées
- Explication de concepts techniques (réseaux, maths, algorithmes, bases de données, etc)
- Adaptation du ton selon le contexte (pédagogue, expert, coach)
- Gestion multi-conversations (création, suppression, basculement)
- Streaming temps réel des réponses (WebSocket pour latence minimale)
- Toggle utilisation du RAG (avec/sans cours selon besoin)

- Sélection de modèle IA selon la tâche (Qwen: conversationnel, Phi: recherche approfondie, Gemma: créatif/empathique)
- Disponibilité 24/7 pour soutien en cas de besoin urgent
- Fonctionne sans connection internet

Module Care

- Conversation empathique et confidentielle avec IA bienveillante
- Tracker d'humeur rapide avec émojis ()
- Détection de signes de stress/anxiété/burnout dans les conversations
- Suggestions d'activités de bien-être personnalisées (respiration, pause, exercice)
- Confidentialité absolue garantie (données stockées localement uniquement)
- Disponibilité 24/7 pour soutien en cas de besoin urgent
- Fonctionne sans connection internet

Module Hub Skills

- Création profil complet (photo, bio, compétences techniques)
- Gestion de certificats avec upload images/PDFs
- Portfolio projets avec liens GitHub, démo live, descriptions
- Système de compétences avec niveaux de maîtrise
- Design professionnel glassmorphism moderne
- Export CV PDF automatique (fonctionnalité à venir)

Module Messages

- Chat temps réel via WebSocket (latence < 100ms)
- Partage de fichiers (PDF, images, code)
- Notifications desktop natives
- Historique complet des conversations
- Recherche d'utilisateurs par nom/pseudo

Module Marketplace

- Upload et vente de matériel informatique (PC, souris, claviers, écrans, composants)
- Système de notation et avis (étoiles + commentaires)
- Catégorisation (PC portables, périphériques, composants, etc.)
- Paiement Mobile Money intégré (MTN, Moov, Flooz)
- Interdiction stricte de vente de cours/exercices (contraire à l'éthique académique)

Fonctionnalités transversales

- Interface utilisateur moderne avec thème sombre/clair (glassmorphism)
- Support Markdown pour formatage riche des réponses
- Typing indicator animé (●●●) pendant génération IA
- Responsive design (fonctionne sur desktop et mobile)

6. Technologies utilisées :

Backend

- Python 3.11 : Langage principal pour l'orchestration IA et le traitement de données
- FastAPI : Framework web asynchrone pour API REST et WebSockets (hautes performances, documentation auto-générée)
- llama-cpp-python : Exécution locale des LLMs quantifiés (pas de GPU nécessaire, coût zéro)
- Sentence-Transformers : Génération d'embeddings pour RAG (recherche sémantique dans les cours)
- FAISS : Base de données vectorielle pour stockage et recherche rapide d'embeddings
- PyJWT + bcrypt : Authentification sécurisée avec tokens JWT et hachage bcrypt des mots de passe
- DiskCache : Système de cache persistant pour historiques et sessions

Frontend

- React 18 : Framework JavaScript pour interface utilisateur réactive et moderne
- React Router : Navigation multi-pages fluide (login, chat, profil, hub, marketplace)
- CSS3 custom : Design glassmorphism avec thème sombre/clair, animations fluides
- WebSocket client : Communication temps réel bidirectionnelle

Intelligence Artificielle

- Qwen2.5-3B-Instruct (Q4_K_M) : Modèle principal pour chat conversationnel (1.93GB, rapide, multilingue)
- Phi-3.5-mini-instruct (Q4_K_M) : Modèle spécialisé raisonnement complexe et recherche approfondie (2.39GB)
- Gemma-2-2B-it (Q6_K) : Modèle pour contenu créatif et empathique (Care) (1.6GB)
- all-MiniLM-L6-v2 : Modèle d'embeddings pour RAG (léger, efficace, 786 dimensions)
- Tiny (Gemma 700MB) : Modèle de démonstration inclus par défaut

Base de données & Stockage

- DiskCache : Cache persistant pour historiques de conversations et sessions utilisateurs
- FAISS : Base vectorielle pour index RAG
- Système de fichiers : Upload utilisateurs (images, certificats, produits marketplace)

Infrastructure & Déploiement

- Docker : Conteneurisation avec images multi-stage (optimisation taille)
- Docker Compose : Orchestration des services avec volumes persistants
- Volumes Docker : Persistance données (cache, modèles, uploads, index)

7. Modèle économique :

Stratégie de monétisation

NEXUS adopte un modèle hybride adapté au contexte étudiant béninois :

1. Version actuelle (v1.0) : 100% gratuite pour tous les étudiants IFRI
 - Accès illimité à tous les modules (Brain, Care, Hub, Messages, Marketplace)
 - Pas de limitation de messages ou de fonctionnalités
 - Objectif : adoption massive et retours utilisateurs
2. Version future (post-hackathon) :
 - Freemium : Module Brain gratuit avec limitation (50 messages/jour)
 - Abonnement étudiant : 2000 FCFA/mois pour accès illimité Brain + Care
 - Commission Marketplace : 10% sur chaque vente de matériel (vendeur garde 90%)
 - Licence institutionnelle : Partenariat avec IFRI pour accès financé par l'école

Public payant (futur)

- Étudiants IFRI (1000+ utilisateurs potentiels : L1, L2, L3 toutes filières)
- Universités partenaires au Bénin (UAC, École Polytechnique d'Abomey-Calavi, et autres)
- Vendeurs de matériel sur Marketplace (commission 10%)
- Partenariats entreprises tech pour stages/emplois (recommandations IA)

8. Perspectives et évolutions futures :

Court terme (3-6 mois)

- Finalisation complète module Marketplace avec recherche avancée
- Système de gamification (badges, points XP pour motivation)
- Export CV PDF automatique depuis Hub Skills
- Blocage de vente de cours automatique
- Amélioration du design

Moyen terme (6-12 mois)

- Génération automatique de QCM et exercices personnalisés basés sur les cours
- Mode révision avec répétition espacée (algorithme Anki)
- Intégration vocale (speech-to-text pour questions orales)
- Collaboration temps réel (étude de groupe avec Nexus)
- Dashboard analytique pour professeurs (suivi progression classe)
- Appels audio/vidéo dans Messages

Long terme (1-2 ans)

- Expansion vers autres écoles/universités au Bénin et en Afrique de l'Ouest

- Support multilingue (anglais, yoruba, fon pour inclusivité)
- Modèles IA fine-tunés spécifiquement sur corpus IFRI complet
- Intégration calendrier académique IFRI (emplois du temps, examens)
- Partenariats entreprises tech pour stages/emplois (recommandations IA)
- API publique pour développeurs tiers (écosystème d'extensions)

9. Architecture et faisabilité technique :

Architecture globale

Nexus suit une architecture client-serveur moderne avec séparation claire des responsabilités pour optimiser performances et maintenabilité :

Couche	Technologies
Frontend	React 18 + React Router + WebSocket client + CSS3 glassmorphism
API Gateway	FastAPI (REST + WebSocket) + JWT auth + Chiffrement Fernet pour confidentialité
Intelligence IA	llama-cpp (Qwen, Phi, Gemma) + Lazy loading multi-modèles
RAG Engine	Sentence-Transformers + FAISS + Chunking intelligent (500 tokens/chunk)
Storage	DiskCache (sessions/conversations) + FAISS (vecteurs) + Filesystem (uploads)

Flux de fonctionnement

- 1. Authentification : Utilisateur se connecte → Backend vérifie credentials → Génère token JWT → Stocke session chiffrée dans DiskCache
- 2. Connexion WebSocket : Frontend ouvre WebSocket avec token → Backend valide → Établit canal bidirectionnel persistant
- 3. Question utilisateur : User tape question → Envoyée via WebSocket → Backend route vers module approprié (Brain/Care/Hub/Messages/Marketplace)
- 4. RAG (si activé) : Question → Embedding (all-MiniLM) → Recherche similarité FAISS → Top-n chunks pertinents → Injection dans contexte LLM
- 5. Génération IA : Prompt construit (système + historique + contexte RAG + question) → LLM génère tokens → Streaming via WebSocket
- 6. Affichage temps réel : Frontend reçoit tokens progressivement → Affichage automatique → Markdown render + auto-scroll
- 7. Sauvegarde : Réponse complète → Stockée dans DiskCache avec métadonnées (timestamp, conv_id, titre auto-généré)

Contraintes techniques

- RAM : Minimum 4GB recommandé (modèles IA ~2GB chacun + OS + FAISS). Solution : Lazy loading (charge modèle uniquement quand utilisé, libère mémoire après)

- CPU : Génération sur CPU peut être lente (8-15 tokens/sec selon modèle). Optimisé avec quantification Q4/Q6 et contexte limité à 4096 tokens
- Stockage : ~8GB pour modèles + embeddings + base vectorielle. Stockage utilisateurs (Hub, Marketplace) peut croître de quelque Ko

Risques techniques et mitigations

- Hallucinations IA : LLMs peuvent générer fausses informations. Mitigation : RAG avec cours officiels, citations sources obligatoires, feedback utilisateur
- Scalabilité : Serveur unique limite à ~50 utilisateurs simultanés pour hub et marketplace (brain et care sont 100 % offline). Solution future : Load balancing + scaling horizontal avec Kubernetes
- Qualité RAG : Dépend de la qualité des embeddings, chunking et quantité de cours. Mitigation : Fine-tuning embeddings sur corpus IFRI.
- Sécurité données : Historiques contiennent données sensibles. Mitigation : Chiffrement Fernet en base, HTTPS obligatoire, pas de stockage cloud