Projet d’assistant virtuel en français-écriture

Démarche de déploiement d'un assistant virtuel IA en contexte éducatif (Moodle)

Projet : Assistant virtuel pour le développement des compétences en écriture (Français)

Contexte : Intégration Moodle sur serveur dédié (Centre de services scolaire)

Destinataire : Ministère de la Cybersécurité et du Numérique / Direction des ressources informationnelles du CSS

Synthèse exécutive

Ce document détaille la démarche de conformité en 9 points pour le déploiement d'un outil d'intelligence artificielle (IA) générative visant à offrir une rétroaction formative aux élèves. L’IAg se comporte comme l'assistant de l'enseignant et elle lui soumet des propositions de rétroaction à fournir à l'élève.

L'architecture repose sur un principe de minimisation des données (Privacy by Design) : l'IA est hébergée sur un serveur dédié et n'a accès qu'aux textes soumis, sans lien direct avec le dossier scolaire de l'élève.

Point 1 : Description du besoin et valeur ajoutée pédagogique

Objectif : Démontrer que le recours à l'IA répond à un besoin spécifique que les outils traditionnels ne comblent pas.

- **Besoin identifié :** Les élèves ont besoin de rétroactions fréquentes et immédiates sur la structure et la syntaxe de leurs textes pour progresser. Les délais de correction humaine limitent la fréquence des exercices d'écriture formative.
- **Valeur ajoutée de l'IA :** L'assistant permet à l’enseignant de fournir une rétroaction plus rapide sur des critères précis (orthographe, cohérence textuelle, vocabulaire) et d’encourager l'autorégulation de l'élève. Il permet d’augmenter la qualité de la rétroaction en passant d'un système par codification à des commentaires écrits qui sont plus parlants pour les élèves.

- **Alignement :** Le projet s'aligne avec la Stratégie numérique en éducation en soutenant le développement de la compétence à écrire (programme de français).

Point 2 : Classification et inventaire des données

Objectif : Identifier la sensibilité des informations traitées conformément à la *Directive sur la sécurité de l'information*.

- **Nature des données :**
  - **Données structurelles :** Textes rédigés par les élèves (productions écrites).
  - **Métadonnées :** Horodatage de la soumission.
  - **Exclusions strictes :** Aucun accès aux bases de données administratives (GPI, Charlemagne, etc.), aucun accès aux noms, adresses, ou codes permanents via l'assistant.
- **Niveau de classification :**
  - Bien que l'assistant ne collecte pas d'identifiants directs, les textes eux-mêmes peuvent contenir des renseignements personnels (RP) involontaires (ex: un élève signe son texte ou raconte une expérience personnelle).
  - **Classification retenue : Intermédiaire (Protégé B)**. Les productions des élèves sont considérées comme des renseignements confidentiels et une propriété intellectuelle.

Point 3 : Architecture technologique et souveraineté des données

Objectif : Garantir que l'infrastructure respecte les exigences de résidence des données du Québec.

- **Hébergement :** Serveur dédié situé physiquement au Québec (ou dans une zone infonuagique validée par le MCN avec chiffrement géré par le client).
- **Cloisonnement (Sandboxing) :**
  - L'IA opère dans un conteneur isolé.
  - Les échanges se font via une API interne sécurisée entre Moodle et le serveur IA.
  - Aucune donnée ne transite par des serveurs publics d'IA (ex: pas d'appel API vers OpenAI public aux États-Unis sans contrat d'entreprise strict assurant la non-utilisation des données pour l'entraînement).

Point 4 : Gestion des identités et des accès (GIA)

Objectif : S'assurer que seules les personnes autorisées accèdent au service.

- **Authentification unifiée (SSO) :** L'accès passe exclusivement par le module d'authentification de Moodle (déjà sécurisé par le CSS). L'IA n'a pas de gestion d'utilisateurs propre.
- **Anonymisation des requêtes :**
    - Lorsque Moodle envoie le texte à l'IA, il utilise un jeton de session temporaire (ID de transaction) et non le nom de l'élève.
    - L'assistant "voit" le texte, l'analyse, et renvoie le résultat au jeton associé. La réassociation avec l'élève se fait uniquement côté Moodle (serveur sécurisé).

Point 5 : Évaluation des facteurs relatifs à la vie privée (EFVP)

Objectif : Se conformer à la *Loi 25* (Loi modernisant des dispositions législatives en matière de protection des renseignements personnels).

- **Analyse de proportionnalité :** La collecte est limitée au strict minimum (le texte à analyser).
- **Risque identifié :** Présence de RP dans le corps du texte (ex: "Je m'appelle Julie et j'habite rue des Érables").
- **Mesure d'atténuation (Mitigation) :**
    - Mise en place d'un mécanisme de **dé-identification automatisé** (filtres Regex et NER) visant à réduire la présence de renseignements personnels avant l'analyse. Bien que ce filtrage ne puisse garantir une anonymisation absolue (risque résiduel), il est complété par l'architecture 'stateless' (sans mémoire) qui assure qu'aucune donnée n'est conservée par le modèle après la génération de la rétroaction.

Point 6 : Mesures de sécurité de l'information (InfoSec)

Objectif : Protéger l'intégrité et la confidentialité des échanges.

- **Chiffrement :**
    - En transit : Protocole TLS 1.3 obligatoire pour toutes les communications API.
    - Au repos : Disques du serveur dédié chiffrés (AES-256).
- **Journalisation (Logs) :**
    - Les journaux d'accès sont conservés pour audit de sécurité, mais ne contiennent pas le contenu des textes (seulement métadonnées : qui, quand, volume de données).
- **Gestion des vulnérabilités :** Mises à jour de sécurité automatisées du système d'exploitation du serveur dédié et du plugin Moodle.

Point 7 : Gouvernance de l'IA (Éthique, Biais et Fiabilité)

Objectif : Assurer que l'IA ne génère pas de contenus préjudiciables ou pédagogiquement faux.

- **Encadrement du "Prompt System" :** Le système utilise des instructions strictes (System Prompts) pour interdire à l'IA de rédiger le texte à la place de l'élève ou de tenir des propos discriminatoires.
- **Supervision humaine (Human-in-the-loop) :**
  - Aucune interaction directe entre l'IA et l'élève sans l'intervention de l'enseignant. C'est l'enseignant qui fait la requête à l'IA et qui reçoit la réponse.
- **Transparence :** L'interface indique clairement ce qui a été créé par l'IAg : le dispositif mis en place présente une icône IA à côté de chaque commentaire généré par l'IA de manière à différencier ce qui est généré par l'IA et par l'enseignant. Lorsque l'enseignant édite le commentaire, l'icône est retiré (contrôle et validation par l'humain).

Point 8 : Consentement et transparence

Objectif : Informer les usagers (élèves et parents pour les mineurs) de l'utilisation de l'IA.

- **Communication :** Une fiche d'information est envoyée aux parents expliquant que l'outil est local, sécurisé et non relié aux données administratives.
- **Consentement :** Selon la politique du CSS, l'utilisation peut être rendue obligatoire si jugée essentielle au cours, car l'EFVP démontre un risque minimal, ou optionnelle avec une alternative offerte (correction par les pairs) pour ceux qui refusent.

Point 9 : Plan de contingence et réversibilité

Objectif : Prévoir la fin du projet ou une panne majeure.

- **Continuité des affaires :** Si le serveur dédié tombe en panne, le module Moodle se désactive automatiquement et le cours reprend son mode normal (correction enseignante différée) sans perte de données pour Moodle.
- **Stratégie de sortie (Exit Strategy) :**
  - À la fin du semestre ou du contrat, toutes les données temporaires sur le serveur dédié sont purgées de manière sécurisée (effacement cryptographique).
  - Aucune donnée ne reste captive d'un fournisseur tiers propriétaire.

Statut du dossier : Prêt pour révision par le Responsable de la sécurité de l'information (RSI) et le Comité d'éthique.