Cyber Mirror

L'IA qui révèle les failles numériques

Introduction

L'essor des technologies numériques et de l'intelligence artificielle s'accompagne d'une collecte massive de données personnelles. Les usages quotidiens du web, des réseaux sociaux et des services en ligne exposent les utilisateurs à des traitements de données parfois opaques, pouvant entraîner des atteintes à la vie privée et des risques en matière de cybersécurité.

Dans ce contexte, le Règlement Général sur la Protection des Données (RGPD), entré en application en 2018 au sein de l'Union européenne, vise à encadrer le traitement des données personnelles et à renforcer les droits des utilisateurs. Il impose notamment des principes fondamentaux tels que la minimisation des données, la transparence des traitements, la limitation des finalités et la protection des données dès la conception (*privacy by design*).

Le projet **Cyber Mirror** s'inscrit pleinement dans cette démarche. Il a été conçu comme un dispositif de sensibilisation qui illustre les enjeux du RGPD tout en respectant strictement ses principes. Aucune donnée personnelle réelle n'est collectée, stockée ou exploitée au sein de l'application. Les informations utilisées lors du parcours utilisateur sont entièrement simulées et ont pour unique finalité la démonstration pédagogique des risques liés aux comportements numériques.

Objectifs

Vous devez concevoir une application web interactive ayant pour objectif de sensibiliser les utilisateurs aux enjeux de la cybersécurité à travers l'utilisation de l'intelligence artificielle. Votre projet devra adopter une approche pédagogique et immersive afin de rendre des notions techniques accessibles à un public non spécialiste.

Problématique

Avec la généralisation des usages numériques, les utilisateurs exposent quotidiennement leurs données personnelles à divers risques (phishing, mots de passe faibles, tracking, ingénierie sociale). Cependant, ces menaces restent souvent abstraites et mal comprises. La problématique de ce projet est donc la suivante :

> *Comment utiliser l'intelligence artificielle pour sensibiliser efficacement les utilisateurs aux risques de cybersécurité, tout en proposant une expérience interactive, compréhensible et éthique ?*

1 Concept et fonctionnement

Cyber Mirror propose à l'utilisateur un parcours interactif basé sur des données simulées. L'utilisateur est invité à répondre à un questionnaire portant sur des habitudes numériques fictives telles que l'utilisation de mots de passe, la navigation sur internet ou les comportements face aux courriels suspects.

À partir de ces réponses, une intelligence artificielle analyse les comportements déclarés et génère un profil de risque. L'application fournit ensuite une restitution visuelle et pédagogique mettant en évidence les failles potentielles, accompagnée d'explications compréhensibles sur les mécanismes exploités en cybersécurité.

Aucune donnée personnelle réelle n'est collectée ou stockée, ce qui permet de respecter les principes du RGPD et de garantir une approche éthique du projet.

L'accent est mis sur l'explicabilité de l'IA afin d'éviter l'effet de « boîte noire » et de favoriser la compréhension des résultats par l'utilisateur ;

2 Scoring et classification des habitudes numériques

2.1 Objectif

Le projet **Cyber Mirror** inclut un questionnaire interactif sur les habitudes numériques fictives. L'objectif est de calculer un *score de risque cyber* pour chaque utilisateur simulé, puis de le classer dans un *niveau de risque* :

- Faible
- Modéré
- Élevé

Ce système de scoring n'est pas délégué à l'IA, mais par le code de l'application pour garantir la transparence (RGPD).

Une fois les scores calculés vous devrez intégrer un LLM pour générer la synthèse pédagogique.

2.2 Moteur de Calcul (Le "Cerveau" Déterministe)

Chaque réponse du questionnaire se voit attribuer un nombre de points. Le **score total** est obtenu par la somme des points, pondérés par thématique :

Thématique	**Poids (%)**
Mots de passe	40
Navigation web	30
Emails / phishing	20
Réseaux sociaux	10

TABLE 1 – Pondération des thématiques pour le calcul du score

2.2.1 Exemple de scoring par question

Mots de passe

- Q1 - Type de mot de passe :
  - Mot de passe unique et complexe : 0 point
  - Variations du même mot de passe : 5 points
  - Mot de passe simple : 10 points
  - Ne sait pas : 8 points
- Q2 - Gestion des mots de passe :
  - Gestionnaire de mots de passe : 0 point
  - Mémorisation personnelle : 4 points
  - Notes non sécurisées : 10 points
  - Navigateur : 6 points
- Q3 - Authentification à deux facteurs :
  - Toujours : 0 point
  - Parfois : 4 points
  - Rarement : 7 points
  - Jamais : 10 points

Navigation web

- Q4 - Vérification HTTPS :
  - Toujours : 0 point
  - Souvent : 3 points
  - Rarement : 7 points
  - Jamais : 10 points
- Q6 - Gestion des cookies :
  - Personnalisation : 0 point
  - Cookies nécessaires : 2 points
  - Tout accepter : 7 points
  - Ne sait pas : 8 points

Emails suspects
- Q7 - Email urgent :
  - Vérification avant d'agir : 0 point
  - Cliquer sur le lien : 8 points
  - Répondre : 10 points
  - Ignorer : 3 points
- Q9 - Reconnaissance phishing :
  - Oui, facilement : 0 point
  - Oui, mais pas toujours : 4 points
  - Non : 8 points
  - Incertain : 6 points

Réseaux sociaux
- Q10 - Partage d'informations personnelles :
  - Rarement : 0 point
  - Parfois : 3 points
  - Souvent : 7 points
  - Ne fait pas attention : 10 points

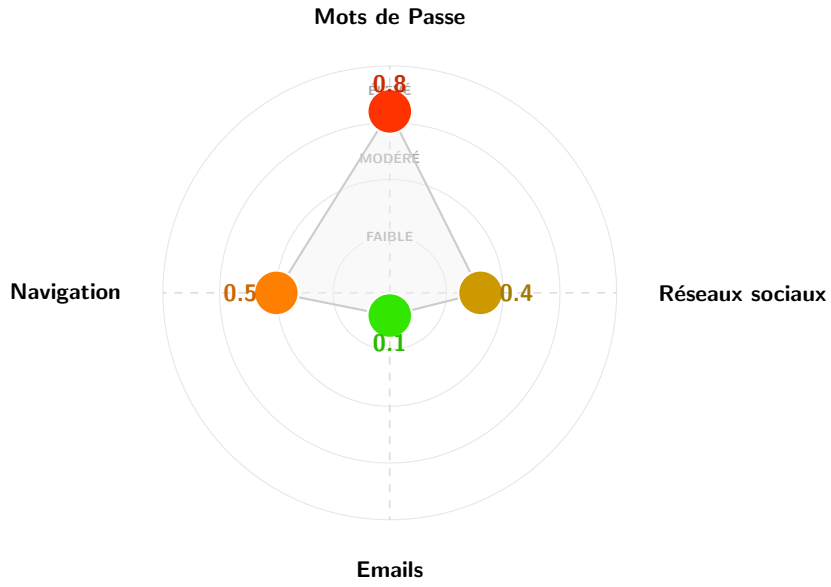

FIGURE 1 – Exemple de graphique radar représentant le niveau de risque par thématique pour un utilisateur fictif. Les cercles concentriques indiquent les seuils Faible / Modéré / Élevé. Les points sont colorés du vert (risque faible) au rouge (risque élevé) et annotés avec la valeur exacte.

2.2.2 Calcul du score global

Le **score total** est obtenu en additionnant les points de toutes les réponses :

$$\text{Score total} = \sum \text{Points par réponse} \tag{1}$$

Le score minimum est 0 et le score maximum est approximativement 100. La figure 1 donne un exemple de restitution des résultats sous forme de radar.

2.3 Intégration du LLM : Restitution Augmentée

Pour rappel, vous devez utiliser un LLM pour générer la synthèse pédagogique. Son rôle dans Cyber Mirror est de transformer des données (scores et statistiques) en une expérience pédagogique engageante. Vous devrez

mettre en place une chaîne de traitement où l'IA agit sous la supervision directe du moteur de calcul. Afin d'éviter les hallucinations (cas où l'IA inventerait des scores non calculés), vous devez définir un Prompt de Système strict. Ce prompt doit forcer l'IA à respecter la structure suivante :

- **Entrée :** Le LLM reçoit en entrée les scores bruts et le profil de l'utilisateur.
- **Contrainte :** L'IA doit obligatoirement baser ses critiques sur les points de la grille d'évaluation ayant généré le plus de risques. Elle ne doit pas inventer les scores.
- **Sortie** : Une restitution en langage naturel, adoptant un ton de "Coach Cyber", expliquant de manière contextuelle les vulnérabilités détectées et les mesures correctives prioritaires à appliquer.

Deux options d'intégration de LLM sont possible [1] :

1. Option Cloud (OpenAI) : Utilisation de l'API GPT-4o/GPT-3.5 via des requêtes HTTP POST sécurisées.
2. Option Locale (Ollama via Docker) : Utilisation de l'image Docker `ollama/ollama`. Le service doit être déclaré dans le fichier `docker-compose.yml` et le modèle (ex : `mistral` ou `llama3`) doit être pré-chargé.

Cette approche permet de rendre le calcul transparent et compréhensible, conformément aux principes de *l'IA explicable* et du RGPD.

3 Travail à rendre

1. **Dépôt Git complet** : Incluant le code source (Frontend et Backend). Un fichier `README.md` clair expliquant comment déployer le projet et, le cas échéant, comment charger le modèle dans Ollama via la commande `docker exec`.
2. **Rapport de "Privacy by Design"**. Une note d'une page expliquant comment le projet respecte le RGPD (absence de stockage de données réelles, anonymisation avant envoi au LLM).
3. **Dossier d'Explicabilité (XAI)** Une démonstration de la fidélité de l'IA. Les étudiants doivent montrer, via 2 ou 3 cas types (ex : un profil "imprudent" vs un profil "expert"), que l'IA génère bien une synthèse cohérente avec les scores du radar.

4 Barème de notation

Critère	Points	Indicateur de réussite
Moteur de Règles	5 pts	"Le calcul du score est précis, transparent et sans erreur."
Intégration LLM	5 pts	L'IA est connectée (API/Ollama) et suit strictement le prompt system.
Visualisation (Radar)	4 pts	Le graphique est dynamique et réagit aux réponses de l'utilisateur.
Architecture Docker	3 pts	Le projet se lance intégralement avec une seule commande.
Conformité, Éthique	3 pts	Documentation claire sur le RGPD et l'explicabilité des résultats.

1. Si vous voulez être 100% local (éthique +++), faites tourner un modèle dans un conteneur.