Ecole Supérieure d'Informatique et du Numérique
COLLEGE OF ENGINEERING & ARCHITECTURE

RAPPORT DU PROJET

4-ème année

CYBERSECURITY

Dans le cadre du cours :

Artificial Intelligence

Sous le thème :

PhishGuard AI – Cyber Assistant

Réalisé par : Ayman Bouissehak
Akram Mokhtari

Encadré par : Hamza Gamouh
Hakim Hafidi

Année Universitaire : 2025-2026

Table des matières

1. Introduction et contexte du projet

1.1. Contexte du projet

La transformation numérique et l'usage massif des services en ligne ont entraîné une augmentation significative des cybermenaces, en particulier les attaques de phishing. Ces attaques exploitent la confiance des utilisateurs à travers des messages frauduleux imitant des entités légitimes (banques, plateformes de paiement, services administratifs, etc.). Face à cette évolution, l'intelligence artificielle s'impose comme un outil efficace pour analyser de grandes quantités de données textuelles et détecter des schémas malveillant.

Dans le cadre du module d'intelligence artificielle, ce projet vise à mettre en pratique les concepts étudiés en développant une application concrète combinant IA et cybersécurité.

1.2. Introduction

Le projet **PhishGuard AI – Cyber Assistant** est une application web intelligente destinée à assister les utilisateurs dans la détection des tentatives de phishing et à améliorer leur sensibilisation en cybersécurité. L'application propose deux fonctionnalités principales : l'analyse automatique de messages ou d'URLs suspects et un chatbot intelligent capable de répondre aux questions liées au phishing et aux bonnes pratiques de sécurité.

Ce projet s'inscrit dans une démarche pédagogique et pratique, mettant l'accent sur l'intégration de modèles d'IA, le développement applicatif et le travail en équipe.

2. Définition du problème

2.1 Problématique identifiée

Le phishing constitue aujourd'hui l'une des cybermenaces les plus répandues et les plus dangereuses. Il s'agit d'attaques basées sur l'ingénierie sociale visant à tromper les utilisateurs afin de leur soutirer des informations sensibles telles que des mots de passe, des données bancaires ou des informations personnelles. Ces attaques prennent généralement la forme de messages ou de liens imitant des services légitimes, ce qui les rend difficiles à identifier pour des utilisateurs non experts.

Malgré l'existence de solutions de sécurité traditionnelles (filtres anti-spam, antivirus, navigateurs sécurisés), de nombreux utilisateurs continuent de tomber victimes de phishing, principalement par manque de sensibilisation et de compréhension des mécanismes de ces attaques.

2.2 Objectifs du projet

L'objectif principal de ce projet est de développer une application intelligente capable d'assister les utilisateurs dans la détection des tentatives de phishing. Les objectifs spécifiques sont les suivants :

- Analyser automatiquement des messages ou des URLs suspects à l'aide de techniques basées sur l'intelligence artificielle.

- Fournir à l'utilisateur un retour clair et compréhensible sur le niveau de risque détecté.
- Sensibiliser les utilisateurs aux bonnes pratiques de cybersécurité à travers un assistant conversationnel intelligent .

2.3 Solution proposée

Pour répondre à cette problématique, le projet **PhishGuard AI – Cyber Assistant** propose une application web intégrant des modèles d'intelligence artificielle dédiés à la détection du phishing. L'utilisateur peut soumettre un message ou une URL afin d'obtenir une analyse automatique indiquant s'il s'agit d'un contenu légitime ou potentiellement malveillant.

En complément, un chatbot de cybersécurité est intégré à l'application afin de répondre aux questions des utilisateurs, expliquer les résultats obtenus et renforcer leur compréhension des risques liés au phishing. Cette approche combine détection automatique et sensibilisation, offrant ainsi une solution à la fois technique et pédagogique.

3. Conception et architecture

3.1 Architecture globale

L'application **PhishGuard AI – Cyber Assistant** repose sur une architecture client–serveur composée de trois couches principales : le frontend, le backend et les services d'intelligence artificielle. Cette architecture modulaire permet une bonne séparation des responsabilités et facilite l'évolution future de l'application.

- **Frontend** : interface web permettant aux utilisateurs de soumettre des messages ou des URLs suspects et d'interagir avec le chatbot.
- **Backend** : serveur applicatif chargé de recevoir les requêtes, de traiter les données et de communiquer avec les modèles d'IA.
- **Services IA** : modèles d'intelligence artificielle utilisés pour l'analyse du phishing et la génération de réponses du chatbot.

3.2 Fonctionnalités basées sur l'intelligence artificielle

L'application intègre deux fonctionnalités principales basées sur l'IA :

- **Détection du phishing** : L'utilisateur peut saisir un message ou une URL suspecte. Le système analyse le contenu à l'aide d'un modèle d'IA afin d'estimer s'il s'agit d'un phishing probable ou d'un contenu légitime.

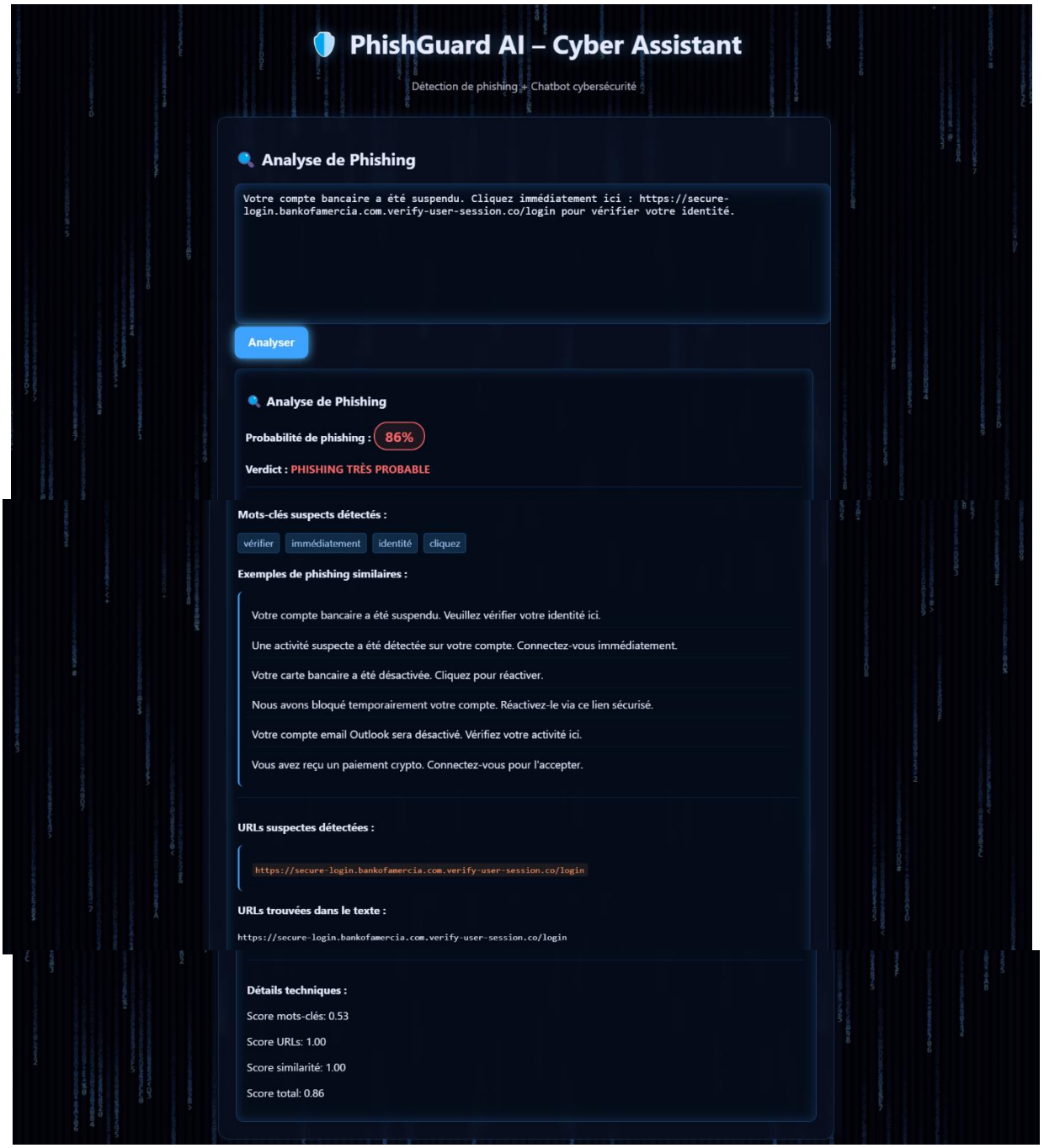

Figure 1: Exemple d'analyse d'un message et Url classé comme *Phishing probable*.

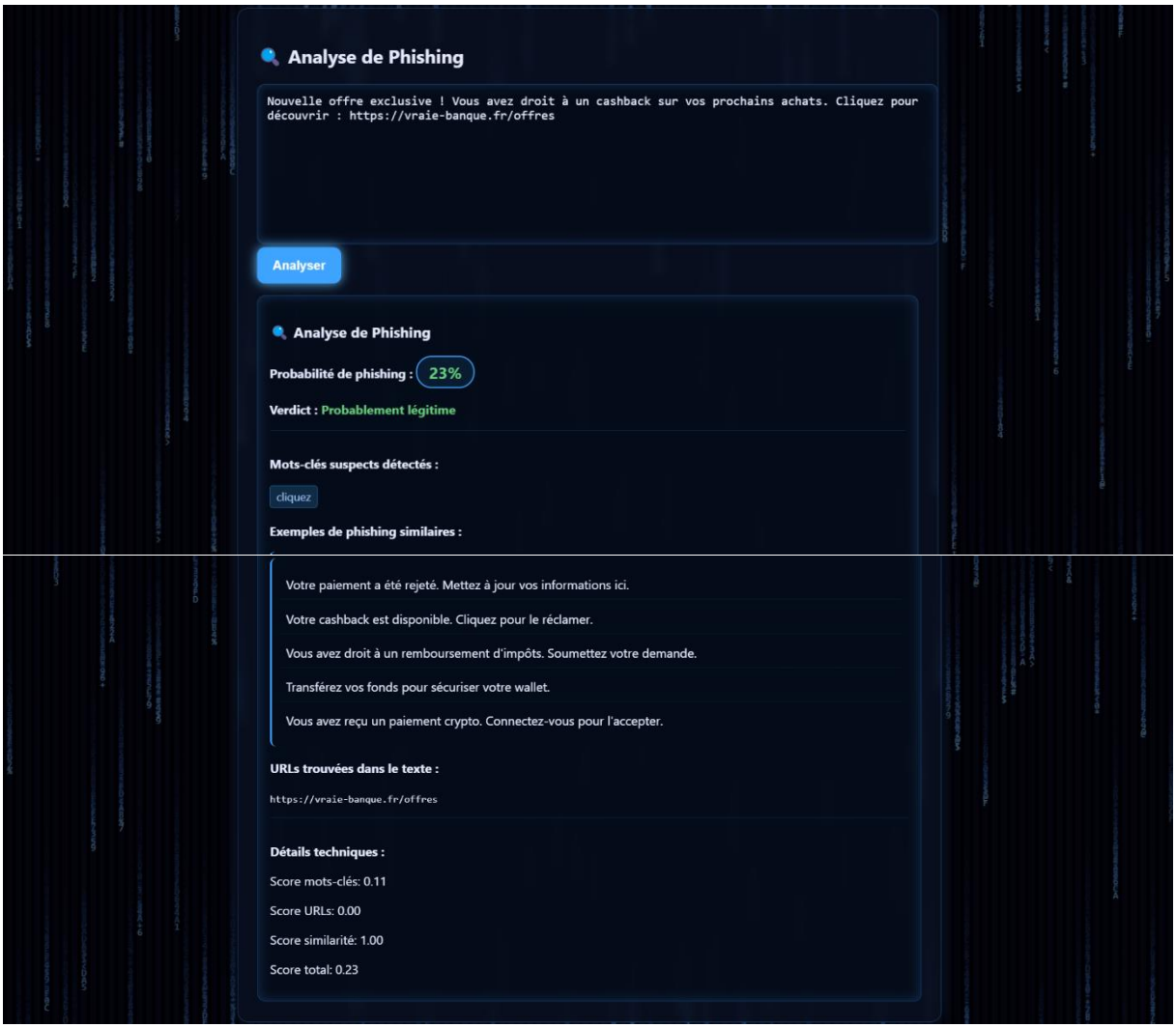

Figure 2 : Exemple d'analyse d'un message ou URL classé comme *Légitime*

Chatbot de cybersécurité : Un assistant conversationnel intelligent est intégré afin de répondre aux questions des utilisateurs concernant le phishing, d'expliquer les résultats fournis par le système et de proposer des conseils de sécurité.

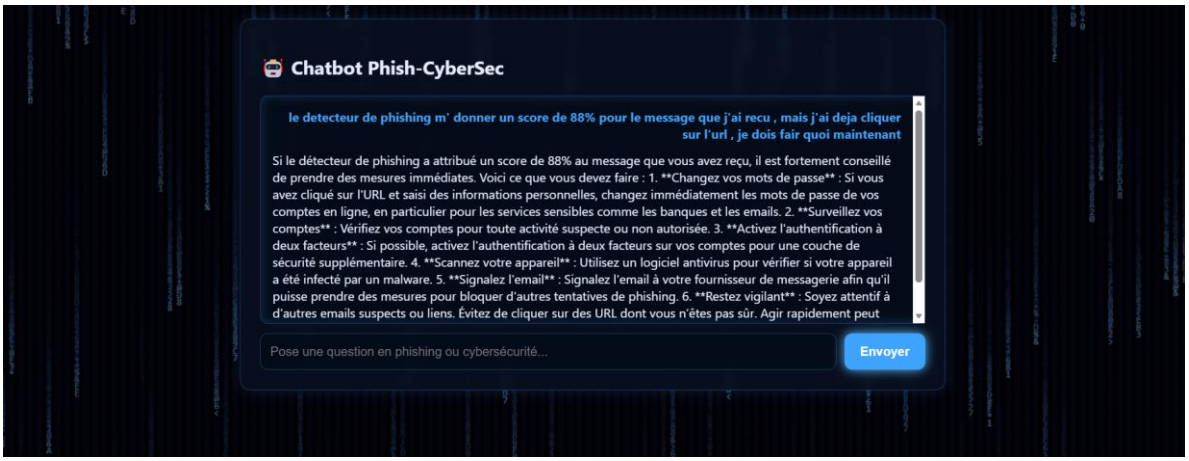

Figure 4 : Interaction avec le chatbot de cybersécurité.

3.3 Rôle des LLMs et approche RAG

3.3.1 Rôle des modèles de langage (LLMs)

Les modèles de langage de grande taille (Large Language Models – LLMs) jouent un rôle central dans le projet **PhishGuard AI – Cyber Assistant**. Ces modèles sont capables de comprendre et d'analyser le langage naturel, ce qui les rend particulièrement adaptés à l'analyse de messages textuels utilisés dans les attaques de phishing.

Dans ce projet, les LLMs sont utilisés pour :

- Analyser le contenu sémantique des messages et des URLs soumis par l'utilisateur.
- Identifier des indices linguistiques et contextuels caractéristiques des tentatives de phishing (urgence, incitation à l'action, demandes d'informations sensibles, etc.).
- Générer des réponses claires et pédagogiques à destination de l'utilisateur.

Grâce à ces capacités, les LLMs permettent une analyse plus flexible et intelligente que les approches classiques basées uniquement sur des règles statiques.

3.3.2 Approche RAG (Retrieval-Augmented Generation)

Afin d'améliorer la fiabilité et la pertinence des réponses fournies par le chatbot, une approche **RAG (Retrieval-Augmented Generation)** peut être intégrée au projet. Cette approche combine la génération de texte par un LLM avec la récupération d'informations issues d'une base de connaissances.

Le principe du RAG dans le cadre de PhishGuard AI est le suivant :

Les questions de l'utilisateur sont analysées par le système.

Les informations pertinentes (exemples de phishing, règles de cybersécurité, bonnes pratiques) sont récupérées depuis une base de données ou un ensemble de documents.

Ces informations sont fournies au LLM comme contexte supplémentaire.

Le LLM génère alors une réponse enrichie, plus précise et mieux contextualisée.

Cette approche permet de limiter les réponses approximatives et d'assurer une meilleure cohérence avec des connaissances validées en cybersécurité.

4. Conclusion et perspectives

Le projet **PhishGuard AI – Cyber Assistant** a permis de développer une application web intégrant des modèles d'intelligence artificielle pour la détection de phishing et l'accompagnement des utilisateurs dans la cybersécurité.

Grâce à l'utilisation de **LLMs** et de l'approche **RAG**, l'application est capable d'analyser des messages et des URLs suspects, de fournir un retour clair sur le niveau de risque et de répondre de manière pédagogique aux questions des utilisateurs. Cette combinaison entre détection automatique et sensibilisation représente une approche innovante et éducative face aux cybermenaces.

Ce projet met en évidence la **synergie entre intelligence artificielle et cybersécurité**. En combinant les capacités d'analyse des LLMs avec la détection automatisée de menaces, l'application ne se contente pas de signaler les contenus suspects, mais accompagne également l'utilisateur en expliquant le raisonnement derrière chaque analyse. Cette approche renforce la **prévention et la sensibilisation** aux risques numériques, démontrant que l'IA peut être un outil complémentaire et pédagogique dans la lutte contre le phishing.