Plateforme d'apprentissage « Secure by Design »

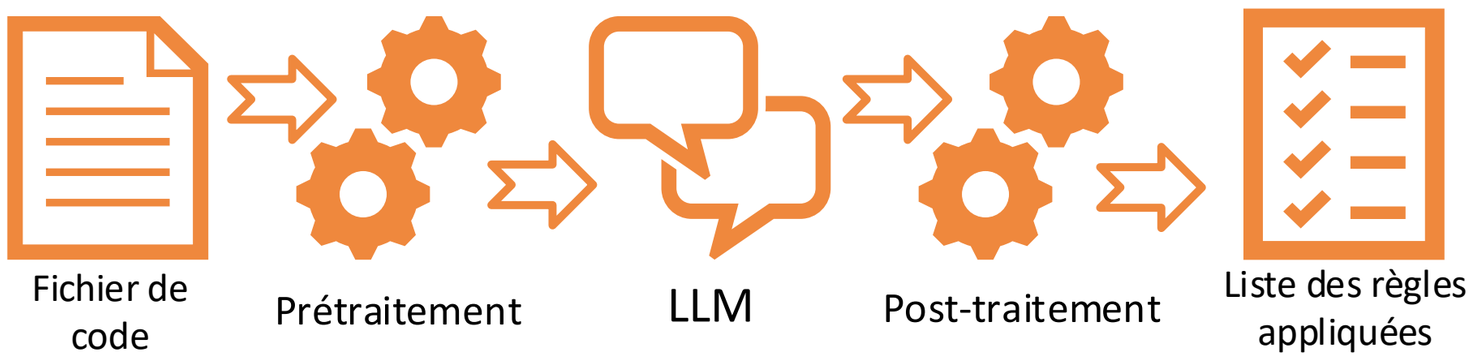
Auteurs : COCAGNE Mathis – KONE Zie Ismaël – LAMBERT Alexis – LECOCQ Enzo

CONTEXTE ET OBJECTIF

Ce projet industriel est une collaboration avec la startup Cybalgoris, spécialiste du Secure by Design. Son objectif est d'accompagner les développeurs, de la formation à la revue de code, en passant par un outil de gestion de projet : CYBERNOE. L'objectif principal de notre projet industriel consiste à étudier la faisabilité et à concevoir un outil de revue automatique de code, visant à vérifier si des règles de Secure by Design sont appliquées dans un fichier de code source.

MÉTHODES ET DÉVELOPPEMENTS

Nous avons segmenté notre projet en trois phases. Dans un premier temps, nous avons conçu des exemples de code se rapportant à la formation Secure by Design, afin de l'illustrer et de la rendre plus accessible. Dans un second temps, nous avons élaboré un mapping entre la formation et les différentes phases de développement d'une application. Enfin, dans un troisième temps, nous avons étudié la faisabilité et conçu le prototype d'un outil de revue automatique de code. Pour ce faire, nous avons utilisé le langage Python et le framework Langchain, grâce auquel nous avons pu interagir avec des LLMs. Nous avons utilisé le modèle de langage GPT fourni par Microsoft Azure. Le fonctionnement de l'outil consiste en trois phases principales. On lui fournit tout d'abord un fichier de code source auquel on applique un ou plusieurs pré-traitements, que nous avons nous-mêmes mis en place au nombre de vingt. Cette étape retourne des segments de code, qui feront l'objet d'une ou plusieurs requêtes au modèle d'IA. Enfin, lors du post-traitement, l'outil va interpréter automatiquement chaque réponse. Nous avons en sortie une liste exhaustive des règles de Secure by Design qui sont présentes dans le fichier, avec des justifications et/ou explications pour chacune des règles.



RÉSULTATS ET CONCLUSION

Ce projet a démontré la faisabilité de l'utilisation de modèles de langage large (LLMs) pour retrouver une règle de Secure by Design dans un fichier de code source. Cet outil peut analyser 200 fichiers de code de 1500 lignes en moins de 20 minutes et détecter 30 règles différentes. Notre preuve de concept atteint un taux de réussite de 91 % sur notre base de tests.

MOTS-CLÉS : Cybersécurité, Secure by Design, code review, assistant virtuel, IA, LLM