

Référentiels d'activités, de compétences et d'évaluation

Développeur en intelligence artificielle

REFERENTIEL D'ACTIVITES <i>Décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>Identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'EVALUATION <i>Définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITES D'EVALUATION	CRITÈRES D'ÉVALUATION
Bloc de compétences 1 : Réaliser la collecte, le stockage et la mise à disposition des données d'un projet en intelligence artificielle			
<p><u>A1 : Programmer la collecte de données depuis plusieurs sources pour un projet en intelligence artificielle</u></p> <p>- Identification des contraintes techniques propres aux sources de données (consulter la documentation, les règles de confidentialité, etc.).</p> <p>- Rédaction des spécifications techniques pour l'extraction des données.</p>	<p>C1. Automatiser l'extraction de données depuis un service web, une page web (<i>scraping*</i>), un fichier de données, une base de données et un système <i>big data*</i> en programmant le <i>script*</i> adapté afin de pérenniser la collecte des données nécessaires au projet.</p>	<p>E1. Mise en situation (C1, C2, C3, C4, C5)</p> <p>L'évaluation doit se faire dans un contexte de réalisation d'un service numérique réel ou fictif basé sur l'usage de données, à partir du cadrage pour la réalisation d'un service numérique (spécifications fonctionnelles et techniques par exemple).</p> <p>Le projet évalué a pour but d'optimiser, d'automatiser, de pérenniser et de mettre à disposition les flux de données et les données, utiles et nécessaires à la</p>	<p>- La présentation du projet et de son contexte est complète : acteurs, objectifs fonctionnels et techniques, environnements et contraintes techniques, budget, organisation du travail et planification.</p> <p>- Les spécifications techniques précisent : les technologies et outils, les services externes, les exigences de programmation (langages), l'accessibilité (disponibilité, accès).</p> <p>- Le périmètre des spécifications techniques est complet : il couvre</p>

<ul style="list-style-type: none"> - Construction des requêtes HTTP pour la récupération des données depuis un service web (REST*). - Lecture d'un fichier de données dans un script (Python, R, etc.). - Téléchargement de l'HTML d'une ou plusieurs pages web visées par une action de <i>scraping</i>. - Connexion programmatique (Python, R, etc.) à un système de gestion de base de données et à un système big data (Hive, Apache Impala, etc.). - Programmation des filtres/parsing des données utiles dans les résultats obtenus depuis : <ul style="list-style-type: none"> - une API, - depuis la lecture d'un ou de plusieurs fichiers de données, - et depuis l'HTML collecté d'un site web (<i>scraping</i>). - Exécution programmatique des requêtes d'extraction de type SQL*. - Exécution programmatique des requêtes d'extraction depuis un système big data. 		<p>réalisation du service numérique, par les équipes techniques (par exemple en analyse statistique, en business intelligence, en machine learning ou encore en intelligence artificielle).</p> <p><u>Livrable</u> : rapport professionnel individuel</p> <p><u>Évaluation</u> :</p> <ul style="list-style-type: none"> - Correction du rapport professionnel - Soutenance orale individuelle 	<p>l'ensemble des moyens techniques à mettre en œuvre pour l'extraction et l'agrégation des données en un jeu de données brutes final.</p> <ul style="list-style-type: none"> - Le script d'extraction des données est fonctionnel : toutes les données visées sont effectivement récupérées à l'issue de l'exécution du script. - Le script comprend un point de lancement, l'initialisation des dépendances et des connexions externes, les règles logiques de traitement, la gestion des erreurs et des exceptions, la fin du traitement et la sauvegarde des résultats. - Le script d'extraction des données est versionné* et accessible depuis un dépôt Git*. - L'extraction des données est faite depuis un mix entre au moins les sources suivantes : un service web (API REST), un fichier de données, un scraping, une base de données et un système big data.
<ul style="list-style-type: none"> - Ecriture des requêtes d'extraction de type SQL de récupération de données 	<p>C2. Développer des requêtes de type SQL d'extraction des données depuis un système</p>		<ul style="list-style-type: none"> - Les requêtes de type SQL pour la collecte de données sont fonctionnelles : les

<p>stockées en base de données et depuis un système big data (Hive, Spark, etc).</p> <ul style="list-style-type: none"> - Documentation des requêtes d'extraction. 	<p>de gestion de base de données et un système big data en appliquant le langage de requête propre au système afin de préparer la collecte des données nécessaires au projet.</p>		<p>données visées sont effectivement extraites suites à l'exécution des requêtes.</p> <ul style="list-style-type: none"> - La documentation des requêtes met en lumière choix de sélections, filtrages, conditions, jointures, etc., en fonction des objectifs de collecte. - La documentation explicite les optimisations appliquées aux requêtes .
<ul style="list-style-type: none"> - Rédaction des spécifications techniques pour l'agrégation des données. - Programmation des règles d'agrégation des données collectées depuis chaque source en un jeu de données brutes unique. - Programmation de l'identification des entrées corrompues dans le jeu de données (par exemple, données partielles et/ou manquantes). - Programmation de la suppression des entrées corrompues. - Programmation de l'identification des entrées au format non normalisé. - Programmation de la l'homogénéisation des formats de données (par exemple, format des dates, des unités, etc.). 	<p>C3. Développer des règles d'agrégation de données issues de différentes sources en programmant, sous forme de script, la suppression des entrées corrompues et en programmant l'homogénéisation des formats des données afin de préparer le stockage du jeu de données final.</p>		<ul style="list-style-type: none"> - Le script d'agrégation des données est fonctionnel : les données sont effectivement agrégées, nettoyées et normalisées en un seul jeu de données à l'issue de l'exécution du script. - Le script d'agrégation des données est versionné et accessible depuis un dépôt Git. - La documentation du script d'agrégation est complète : dépendances, commandes, les enchaînements logiques de l'algorithme, les choix de nettoyage et d'homogénéisation des formats données.

<ul style="list-style-type: none"> - Versionnement* des scripts avec Git et un dépôt Git en ligne. - Documentation des scripts. 			
<p><u>A2 : Développer la mise à disposition technique des données collectées pour un projet d'intelligence artificielle</u></p> <ul style="list-style-type: none"> - Rédaction des spécifications techniques pour le stockage des données. - Modélisation de la structure des données de la base de données selon la méthode Merise*. - Choix du système de gestion de base de données. - Création de la base de données dans le système de gestion de base de données. - Documentation de la procédure d'installation du système de gestion de base de données. - Rédaction ou mise à jour du registre des traitements de données personnelles en vue de la mise en conformité de la base de données avec le RGPD. - Rédaction des procédures de tri des 	<p>C4. Créer une base de données dans le respect du RGPD en élaborant les modèles conceptuels et physiques des données à partir des données préparées et en programmant leur import afin de stocker le jeu de données du projet.</p>		<ul style="list-style-type: none"> - Les modélisations des données respectent la méthode et le formalisme Merise. - Le modèle physique des données est fonctionnel : il est intégré avec succès lors de la création de la base de données, sans erreur. - La base de données est choisie au regard de la modélisation des données et des contraintes du projet. - La reproduction des procédures d'installation décrites (base de données et API) a pour résultat un système conforme aux objets techniques attendus.. - Le script d'import fourni est fonctionnel : il permet l'insertion des données dans le système mis en place. - La documentation technique du script d'import est versionné à la racine du même dépôt Git que celui utilisé pour le script d'import. - Les documentations techniques des script

<p>données personnelles stockées dans la base de données pour la mise en conformité avec le RGPD (par exemple : détection et suppression des données personnelles inutiles, trop anciennes, etc.).</p> <ul style="list-style-type: none"> - Programmation du script d'import des données en base de données. - Documentation du script d'import. 			<p>couvrent les parties suivantes :</p> <ul style="list-style-type: none"> - les dépendances nécessaires pour la réutilisation des scripts (langages, dépendances externes, etc) - les commandes pour l'exécution des scripts. <p>- Le registre des traitements de données personnelles intègre l'ensemble des traitements de données personnelles impliqués dans la base de données.</p> <p>- Les procédures de tri des données personnelles pour la mise en conformité de la base de données avec le RGPD sont rédigées.</p> <p>- Les procédures de tri détaillent les traitements de conformité (automatisés ou non) à appliquer ainsi que leur fréquence d'exécution.</p>
<ul style="list-style-type: none"> - Rédaction des spécifications techniques des moyens de mise à disposition et d'accès aux données du projet : API (REST) et accès direct à la base de données. - Configuration des accès au jeu de données depuis le serveur de l'API. - Développement de la réception et de la validation des requêtes client (web, mobile...) 	<p>C5. Développer une API mettant à disposition le jeu de données en utilisant l'architecture REST afin de permettre l'exploitation du jeu de données par les autres composants du projet.</p>		<ul style="list-style-type: none"> - La documentation technique de l'API (REST) couvre tous les points de terminaisons. - La documentation technique couvre les règles d'authentification et/ou d'autorisation de l'API. - La documentation technique respecte les standards du modèle choisi (par exemple <i>OpenAPI*</i>).

<ul style="list-style-type: none"> - Développement des requêtes à la base de données en fonction de la mise à disposition des données prévues par les spécifications. - Développement des réponses de l'API au client. - Développement des règles d'autorisation et d'accès aux points de terminaison de l'API REST. - Sécurisation de l'API : Top 10 OWASP API¹ par exemple. - Rédaction de la documentation technique de l'API REST. - Rédaction de la documentation technique d'accès à la base de données adaptée aux environnements techniques du projet. 			<ul style="list-style-type: none"> - L'API REST est fonctionnelle pour l'accès aux données du projet : elle restreint par une autorisation (ou authentification) l'accès aux données, - L'API REST est fonctionnelle pour la mise à disposition : elle permet la récupération de l'ensemble des données nécessaires au projet, comme prévu selon les spécifications données.
Bloc de compétences 2 : Intégrer des modèles et des services d'intelligence artificielle			
<u>A3. Accompagner le choix et l'intégration d'un service d'intelligence artificielle préexistant</u>	C6. Organiser et réaliser une veille technique et réglementaire en animant le travail collectif de sélection des sources, de collecte, de traitement et de partage des informations afin	E2. Cas pratique (C6, C7, C8) L'évaluation doit se faire à partir de l'expression d'un besoin réel ou fictif de	- La thématique de veille choisie porte sur un outil et/ou une réglementation mobilisée dans la mise en situation.

¹ <https://raw.githubusercontent.com/OWASP/API-Security/master/2019/en/dist/owasp-api-security-top-10.pdf>