

Référentiels d'activités, de compétences et d'évaluation

Expert en ingénierie et science des données

Niveau 7

Évaluation des candidats en situation de handicap

Avec l'accord du candidat, l'identification d'une situation de handicap peut être déclarée dans le dossier académique destiné au jury final qui le prendra en compte dans son évaluation. Tout candidat en situation de handicap peut également bénéficier d'un aménagement de ses modalités d'évaluation au cours de sa formation (notamment un tiers temps pour les soutenances).

Référentiels d'activités, de compétences et d'évaluation

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>BLOC 1. Manager et piloter un projet Data et le développement d'une solution</p> <p>Dans un contexte de plus en plus digitalisé, la gestion de projets Data prend une place centrale au sein des organisations cherchant à valoriser leurs données. Le pilotage d'un projet Data inclut la gestion complète du cycle de vie du projet, de la phase de cadrage à la mise en production de la solution. L'objectif est de répondre aux besoins métiers, tout en intégrant les contraintes techniques, juridiques et réglementaires.</p> <p>Dans le cadre de la gestion et du pilotage d'un projet Data, le titulaire de la certification est responsable de définir le périmètre du projet en s'appuyant sur le recueil des besoins métiers. Il réalise une analyse contextuelle approfondie, incluant une cartographie des données existantes et une évaluation des besoins en matière de stratégie Data. Il organise le projet en termes d'objectifs, d'étapes clés, d'indicateurs de suivi de performance, et d'assignation des ressources ainsi que la coordination de l'équipe projet. Il garantit également le respect des délais, des coûts et de la qualité des livrables. Une partie essentielle de son activité consiste à réaliser une veille technologique et réglementaire (Big Data, Cloud, RGPD, etc.), à auditer les données existantes et à évaluer les risques pour identifier la solution technique la plus adaptée. Il procède, si nécessaire, à un prototypage pour valider la solution avec les parties prenantes et l'ajuster, si nécessaire.</p> <p>Selon le contexte, le titulaire travaille en autonomie ou en collaboration avec des équipes internes / externes, garantissant la performance du projet tout en respectant les bonnes pratiques en matière de gestion des données, d'accessibilité et de réglementation en vigueur.</p>			
<p>A.1.1 Gestion et coordination du projet</p>			
<p>Cadrage et définition du projet</p> <ul style="list-style-type: none"> - Recueil des besoins en matière de data 	<p>C1.1.1 Collecter les besoins métiers et analyser le contexte de l'organisation (dont la gestion des données) afin d'identifier et</p>	<p>L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en</p>	<p>CE1 - Choix et conformité méthodologiques (pour le recueil / analyse des besoins)</p> <ul style="list-style-type: none"> - Une méthodologie de recueil et d'analyse du besoin est définie.

<p><i>(expression des besoins métiers (ex : via entretiens, via atelier) ; état des lieux de contexte de l'organisation / du département ou service concerné ; état des lieux en matière de data (ex : stratégie data, cartographie des données).</i></p>	<p>de formaliser les objectifs / les attendus en matière de data et les éventuelles contraintes associées (organisationnelles, techniques, budgétaires, réglementaires).</p>	<p>situation professionnelle reconstituée.</p> <p>À partir d'un contexte donné, le candidat doit analyser un besoin data puis cadrer et conduire un projet en réponse à ce besoin :</p> <ul style="list-style-type: none"> - Analyser et reformuler les besoins métiers. - Rechercher, compiler et analyser des informations en lien avec le besoin data. - Définir des indicateurs à suivre - Cadrer, mettre en œuvre et évaluer le projet. - Effectuer un retour d'expérience. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<ul style="list-style-type: none"> - <i>Les choix méthodologiques (ex : sélection des sources d'informations, choix des techniques et outils de recueil type animation d'un atelier ou entretiens) prennent en compte les objectifs visés et le contexte de l'organisation.</i> - <i>La méthodologie est correctement appliquée (ex : utilisation / application correcte des outils de recueil).</i> <p>CE2 - Pertinence / complétude du recueil / analyse des besoins</p> <ul style="list-style-type: none"> - <i>Les éléments organisationnels, techniques, budgétaires et réglementaires sont correctement identifiés et pris en compte.</i> - <i>Les contraintes spécifiques à l'organisation sont collectées, décrites et intégrées dans l'analyse.</i> - <i>Le mode de fonctionnement de l'organisation (gestion des données) est pris en compte.</i> <p>CE3 - Identification et formulation correcte des besoin</p> <ul style="list-style-type: none"> - <i>Les problématiques métiers sont entièrement identifiées et retranscrites.</i> - <i>Les besoins des services et directions métiers sont correctement énoncés et justifiés.</i>
<ul style="list-style-type: none"> - <i>Organisation du projet (formalisation des objectifs, définition des étapes de réalisation, assignation des tâches, élaboration du calendrier prévisionnel, définition des modalités de suivi de projet / des indicateurs).</i> 	<p>C1.1.2. Définir les modalités de réalisation et de suivi du projet data et le planifier afin de s'assurer de sa bonne mise en œuvre et de sa conformité avec les objectifs visés :</p> <ul style="list-style-type: none"> - <i>planifier les étapes de réalisation</i> - <i>élaborer le calendrier prévisionnel</i> - <i>définir les modalités de suivi du projet avec des indicateurs appropriés.</i> 		<p>CE1 - Qualité de l'organisation et de la planification du projet</p> <ul style="list-style-type: none"> - <i>Le plan de projet est complet, détaillé et aligné avec les objectifs visés.</i> - <i>Les étapes du projet, les jalons et les livrables sont clairement définis.</i> - <i>Les responsabilités de chaque membre de l'équipe sont bien réparties et documentées.</i> - <i>Les ressources nécessaires (humaines, financières, matérielles) sont correctement identifiées et allouées.</i>

<p>Veille technologique, juridique / réglementaire (dans le champ de la data. Ex : Big data, Cloud, Protection des données) :</p> <ul style="list-style-type: none"> - <i>Mise en place d'un système de veille permanent (sur une ou des thématiques en lien avec son métier / son organisation).</i> - <i>Recherche ponctuelle sur une problématique donnée.</i> 	<p>C1.1.3. Rechercher, collecter, compiler et analyser des informations clés. Identifier de nouvelles opportunités, solutions ou pratiques (tendances, technologies émergentes, etc.), afin de bénéficier de connaissances techniques, juridiques ou réglementaires actualisées dans le champ de la data et de les diffuser en vue de leur partage (notamment en tenant en compte les Personnes en situation de handicap (PSH)).</p>		<p>CE1 - Complétude et cohérence du périmètre de la veille</p> <ul style="list-style-type: none"> - <i>Les objectifs de la veille sont fixés.</i> - <i>Les thématiques / mots clés à suivre sont identifiés / formalisés et en cohérence avec les objectifs visés.</i> - <i>Le choix des outils de curation retenus sont justifiés et en cohérence avec les objectifs visés.</i> - <i>Le recueil des données est effectué à partir de sources fiables.</i> - <i>La veille est réalisée en adéquation aux besoins métiers.</i> <p>CE2 - Pertinence de l'analyse des résultats de la veille.</p> <ul style="list-style-type: none"> - <i>Les données collectées permettent de répondre aux objectifs visés.</i> - <i>Les impacts en termes d'opportunités / de nouvelles pratiques sont identifiés et justifiés.</i> - <i>Le support de présentation respecte les règles d'accessibilité.</i>
<p>Coordination / suivi de projet</p> <ul style="list-style-type: none"> - <i>Gestion / animation des équipes / parties prenantes.</i> - <i>Gestion des délais / coûts / livrables.</i> - <i>Mesure de la performance du projet.</i> 	<p>C1.1.4. Conduire les actions et les échanges (ex : réunion, état d'avancement) entre les différentes parties prenantes (y compris avec les PSH) du projet data afin de s'assurer de sa bonne mise en œuvre et de créer une synergie optimale.</p>		<p>CE1 - Qualité de la coordination</p> <ul style="list-style-type: none"> - <i>Les attentes des différentes parties prenantes sont identifiées et correctement gérées.</i> - <i>La coordination entre les différentes parties prenantes est documentée de manière efficace.</i> <p>CE2 - Qualité du suivi de projet</p> <ul style="list-style-type: none"> - <i>Les modalités de gestion / suivi des délais / coûts sont définies / outillées.</i> - <i>Les différentes avancées du projet sont communiquées en respect des bonnes pratiques aux parties prenantes.</i> - <i>Les livrables finaux sont conformes aux critères fixés.</i>
	<p>C1.1.5. Suivre, contrôler et analyser le projet data en termes de délais, de coûts, de livrables et de performance :</p>		

	<ul style="list-style-type: none"> - S'assurer de leur conformité aux critères et indicateurs définis, y compris l'efficacité et la qualité du projet. - Ajuster le projet si nécessaire et rendre compte aux différentes parties prenantes, notamment aux personnes en situation de handicap. 		<ul style="list-style-type: none"> - Le support de présentation respecte les règles d'accessibilité. <p>CE3 - Qualité de l'analyse de la performance</p> <ul style="list-style-type: none"> - Les indicateurs de performance sont suivis / correctement évalués / analysés selon une fréquence adaptée. Des actions correctives sont définies en cas de besoin.
Accompagnement des parties prenantes	C1.1.6. Conseiller et apporter un appui stratégique et méthodologique en matière de data tout au long du projet, en accompagnant les parties prenantes pour faciliter la prise de décision, notamment en intégrant les problématiques liées au handicap.		<p>CE1 - Pertinence / qualité des conseils / appui</p> <ul style="list-style-type: none"> - Les conseils / appuis prennent en compte la demande initiale. <ul style="list-style-type: none"> - Ils sont adaptés au projet, au contexte. - Ils sont expliqués / justifiés / objectivés. <p>CE2 - Posture adaptée</p> <ul style="list-style-type: none"> - Une posture de conseil est adaptée (écoute, reformulation). - Le ton / le vocabulaire sont adaptés au public. <p>CE3 - Respect de la législation en vigueur et des bonnes pratiques de la profession.</p> <ul style="list-style-type: none"> - La législation et réglementation en vigueur en matière de data est respectée. - Les bonnes pratiques en matière de gestion de projet sont appliquées au fil de l'eau (ex : documentation du projet, traçabilité). - Les bonnes pratiques en matière d'accessibilité sont prises en compte.
A.1.2. Définition et mise en oeuvre de la solution Data			
Audit global de la data	C1.2.1. Auditer la solution data en analysant la donnée, tant sur son historique que sur sa qualité, afin d'évaluer sa disponibilité et son adéquation par rapport aux besoins identifiés :	L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.	<p>CE1 - Choix et conformité de la méthodologie de l'audit :</p> <ul style="list-style-type: none"> - Une méthodologie d'audit est définie. - Les choix méthodologiques prennent en compte les objectifs visés et le contexte de l'organisation.

	<ul style="list-style-type: none"> - Évaluer l'historique de la donnée, - Vérifier la qualité et la disponibilité des données, - Comparer leur adéquation avec les besoins identifiés. 	<p>À partir d'un contexte donné, le candidat doit proposer une solution data et en évaluer la faisabilité :</p> <ul style="list-style-type: none"> - Cartographier les données. - Identifier et évaluer les risques de la solution. - Rechercher une solution technique compatible. 	<ul style="list-style-type: none"> - La méthodologie est correctement appliquée. <p>CE2 - Qualité de l'analyse des données :</p> <ul style="list-style-type: none"> - Le cycle de vie de la donnée au sein de l'organisation est pris en compte. - L'historique de la donnée est pris en compte. - La disponibilité et l'adéquation des données sont correctement établies. - Les éventuelles problématiques ou contraintes liées aux données (ex : accessibilité, architecture) sont identifiées et décrites.
<p>Analyse des risques de la solution</p>	<p>C1.2.2. Identifier et évaluer les risques de la solution data en matière d'accessibilité, de sécurité et de développement durable afin de répondre aux normes / réglementation en vigueur.</p>	<ul style="list-style-type: none"> - Lister, évaluer, décrire, et prioriser les cas d'usage. - Prototyper la solution. - Proposer une solution technique. - Justifier ses choix techniques / financiers. - Présenter et valider la solution aux parties prenantes. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>CE1 - Choix et conformité de la méthodologie d'identification et d'évaluation des risques :</p> <ul style="list-style-type: none"> - Une méthodologie est définie. - Les choix méthodologiques prennent en compte les objectifs visés et le contexte de l'organisation. - La méthodologie est correctement appliquée (ex : les risques sont classifiés par catégories par ex accessibilité, développement durable, etc.). <p>CE2 - Exhaustivité des risques identifiés :</p> <ul style="list-style-type: none"> - L'ensemble des risques potentiels liés au projet data sont identifiés et documentés. - L'ensemble des domaines à risque (accessibilité, sécurité, développement durable) ont été analysés. <p>CE3 - Pertinence et justesse de l'évaluation des risques identifiés :</p> <ul style="list-style-type: none"> - Les risques sont correctement établis et hiérarchisés. - Chaque risque identifié est accompagné des mesures d'atténuation correspondantes. - Les risques sont classifiés par catégories (ex:

			<p><i>accessibilité, développement durable, etc.).</i></p> <ul style="list-style-type: none"> - <i>Les risques de sécurité des données sont identifiés et évalués selon les pratiques et réglementations en matière de sécurité (ex: RGPD, ISO 27001).</i> - <i>Les risques liés au développement durable sont identifiés et évalués en tenant compte des critères environnementaux et sociétaux.</i>
<p>Définition d'une solution (technologies, modèles etc.) :</p> <ul style="list-style-type: none"> - <i>Recherche d'une ou de plusieurs solutions techniques.</i> - <i>Recensement, évaluation (voire priorisation) et description des cas d'usage / des tâches à résoudre (via un cahier des charges, un product backlog, une fiche de qualification d'un cas d'usage etc.).</i> - <i>Évaluation de la compatibilité de la (des) solution(s) technique(s).</i> 	<p>C1.2.3. Identifier une solution technique (interne ou de marché) compatible aux contraintes éventuelles (métiers / techniques dont SI) afin de répondre aux besoins identifiés :</p> <ul style="list-style-type: none"> - <i>Lister et évaluer la faisabilité technique, financière et opérationnelle des cas d'usage.</i> - <i>Formaliser les cas d'usage.</i> - <i>Partager et valider les cas d'usage auprès des parties prenantes.</i> - <i>Tenir en compte les personnes en situation de handicap, dans la conception des supports de présentation.</i> 		<p>CE1 - Compatibilité de la solution technique :</p> <ul style="list-style-type: none"> - <i>La solution est compatible avec le projet data, ses contraintes (ex : budgétaires, techniques, etc.)</i> - <i>La solution respecte la demande initiale, est réaliste et fonctionnelle.</i> - <i>Le choix de la solution technique est justifié.</i> - <p>CE2 - Choix et conformité de la méthodologie d'identification, d'évaluation et de priorisation des cas d'usage :</p> <ul style="list-style-type: none"> - <i>Une méthodologie est définie.</i> - <i>Les choix méthodologiques prennent en compte les objectifs visés et le contexte de l'organisation.</i> - <i>La méthodologie est correctement appliquée.</i> <p>CE3 - Identification correcte des cas d'usage :</p> <ul style="list-style-type: none"> - <i>Les cas d'usage traduisent correctement les besoins (dans leur ensemble).</i> - <i>L'ensemble des ressources humaines, techniques / matérielles nécessaires aux cas d'usage sont identifiées.</i> <p>CE4 - Évaluation / priorisation correcte des des cas d'usage :</p> <ul style="list-style-type: none"> - <i>Les cas d'usage sont évalués de manière correcte, selon des critères définis (ex :</i>

			<p><i>compétences, technologies, données etc.)</i></p> <ul style="list-style-type: none"> - <i>Les cas d'usage sont priorisés au besoin, de manière correcte, selon des critères préalablement définis.</i> <p>CE5 - Description correcte / juste des cas d'usage :</p> <ul style="list-style-type: none"> - <i>Les cas d'usage sont correctement décrits (en termes d'objectifs, attendus, livrables, efforts, contraintes).</i> - <i>Les éléments sont justes / justifiés.</i> - <i>Les critères d'accessibilité des supports de présentation respectent les principes d'accessibilité :</i> <ul style="list-style-type: none"> - <i>Contenu non textuel.</i> - <i>Utilisation de la couleur.</i> - <i>Contraste (minimum).</i> - <i>Redimensionnement du texte.</i> - <i>Titre de page.</i>
<p><i>Au besoin, prototypage de la solution. Le prototypage n'est pas systématiquement réalisé, il dépend des besoins / du contexte. Il peut également prendre une multiplicité de formes.</i></p>	<p>C1.2.4. Effectuer - en cas de besoin - un prototype de la solution afin d'en confirmer la faisabilité technique.</p>		<p>CE1 - Choix et conformité méthodologique :</p> <ul style="list-style-type: none"> - <i>Une méthodologie est définie.</i> - <i>Les choix méthodologiques prennent en compte les objectifs visés et le contexte de l'organisation.</i> - <i>La méthodologie est correctement appliquée.</i> <p>CE2 - Qualité du prototype :</p> <ul style="list-style-type: none"> - <i>Le prototype permet de s'assurer de la conformité technique de la solution proposée.</i> - <i>Le prototype permet de mesurer la plus value de la solution proposée.</i>
<p>Présentation de la solution data aux parties prenantes (product owner, utilisateurs finaux, experts techniques, sponsors etc.).</p>	<p>C1.2.5. Présenter la solution data et expliquer ses choix auprès des parties prenantes (interlocuteurs technique / métier) afin d'en démontrer la pertinence et</p>		<p>CE1 - Qualité de la présentation écrite et orale :</p> <ul style="list-style-type: none"> - <i>La communication est adaptée au public cible (niveau de détail, terminologie, etc.).</i> - <i>Les supports de présentation sont adaptés à tout public en particulier aux PSH.</i> - <i>Les choix sont présentés de manière claire et</i>

	l'adéquation aux besoins.		<p><i>justifiée.</i></p> <p>CE2 - Respect de la législation en vigueur et des bonnes pratiques de la profession :</p> <ul style="list-style-type: none">- <i>La législation et réglementation en vigueur en matière de data est respectée.</i>- <i>La solution est documentée au fil de l'eau.</i>
--	---------------------------	--	--

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION

BLOC 2. Concevoir et déployer une infrastructure complète de gestion des données (collecte, traitement et stockage)

Dans le cadre de la conception et du déploiement d'une infrastructure complète de gestion des données, le titulaire de la certification est responsable de l'installation et du paramétrage de l'environnement de travail en fonction des besoins spécifiques du projet. Il élabore les procédures de collecte, de traitement et de stockage des données en choisissant les méthodes et outils adaptés, tels que les solutions Python, DLT ou le traitement en mode Batch ou Stream. Le choix du système de stockage est également une étape clé pour assurer la performance et la disponibilité des données.

En collaboration avec la DSI et le RSSI, il garantit la sécurité et la confidentialité des données (en collaboration avec le RSSI), via la gestion des accès, des identités, et des autorisations. Il conçoit et organise les bases de données, en mettant en place des systèmes distribués ou en réseau (AWS, Azure, pgAdmin) pour faciliter l'échange fluide des données. Le titulaire supervise également la définition et le déploiement de l'infrastructure, en sélectionnant les composants critiques et en veillant à leur installation. Il effectue enfin les tests nécessaires pour s'assurer que la solution répond aux exigences fonctionnelles et techniques de l'organisation.

A.2.1 Définition des processus de collecte et gestion des accès

Choix, installation et paramétrage de l'environnement de travail	C2.1.1 Configurer l'environnement de travail (dont ressources : machines physiques ou virtuelles, nécessaires à la gestion des données (collecte - traitement - stockage)).	L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée. À partir d'un contexte donné, le candidat doit concevoir et déployer une infrastructure de gestion des données et sécuriser leur accès : <ul style="list-style-type: none"> - Paramétrer 	CE1 - Pertinence de la configuration de l'environnement de travail : <ul style="list-style-type: none"> - La configuration de l'environnement est adaptée aux besoins spécifiques de la gestion des données (collecte, traitement, stockage). - La configuration tient compte des contraintes et opportunités techniques de l'infrastructure existante. - Le choix de machines physiques et virtuelles sont justifiés en fonction des exigences de performance et de capacité.
---	---	--	--

		<p>l'environnement de travail</p> <ul style="list-style-type: none"> - Définir un processus du cycle de vie de la donnée. - Implémenter un système d'authentification pour l'accès aux données. - Rédiger la documentation associée. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>CE2 - Conformité de la configuration de l'environnement de travail :</p> <ul style="list-style-type: none"> - <i>La configuration de l'environnement assure la conformité avec les normes de sécurité des données et les réglementations en vigueur (ex. RGPD).</i> - <i>Les mesures de sécurité sont adéquates et permettent de protéger les données tout au long de leur cycle de vie (collecte, traitement, stockage).</i> - <i>Les mécanismes de protections des accès et des permissions sont définis et correctement configurés.</i> <p>CE3 - Qualité et efficacité de la configuration de l'environnement de travail :</p> <ul style="list-style-type: none"> - <i>La configuration couvre tous les aspects nécessaires et assure une gestion optimale des données à chaque étape (collecte, traitement, stockage).</i> - <i>Les configurations assurent une gestion fluide et performante des données.</i> - <i>La documentation des configurations est claire et complète.</i>
<p>Elaboration d'une solution de collecte et de stockage des données</p> <ul style="list-style-type: none"> - <i>Définition des procédures de collecte / de traitement et de stockage.</i> - <i>Choix des méthodes / des outils de collecte / de traitement des données (Build from scratch, Aribyte, DLT, Python, Batch versus Stream, web</i> 	<p>C2.1.2. Définir les processus (procédures, méthodes / outils) de collecte, de traitement et de stockage des données en cohérence avec les besoins et l'environnement technique afin de proposer un système de gestion des données adapté (aux besoins), fonctionnel, performant et sécurisé.</p>		<p>CE1 - Pertinence du processus de gestion des données :</p> <ul style="list-style-type: none"> - <i>Le processus de gestion des données est adapté aux besoins spécifiques du projet.</i> - <i>Le processus de gestion des données tient compte des contraintes et opportunités techniques de l'environnement SI existant (il est compatible avec le SI existant).</i> - <i>Le processus de gestion des données est aligné avec les données.</i> - <i>Les procédures, méthodes et outils choisis pour chaque étape sont pertinents et justifiés.</i>

<ul style="list-style-type: none"> - <i>scraping, etc.)</i> - <i>Choix du système de stockage (ex : stockage cloud, stockage en réseau (NAS/SAN) etc.).</i> 			<ul style="list-style-type: none"> - <i>Les technologies, méthodes et outils choisis sont compatibles avec l'infrastructure technique en place.</i> <p>CE2 - Conformité du processus de gestion des données :</p> <ul style="list-style-type: none"> - <i>Le processus de gestion des données assure la conformité et la qualité des données.</i> - <i>Le processus de gestion des données inclut des mesures de sécurité adéquates pour protéger les données à chaque étape.</i> - <i>Le processus de gestion des données respecte les normes et réglementations en vigueur (ex: RGPD).</i> - <i>Les mécanismes de protection des données (ex. contrôle d'accès) sont définis.</i> <p>CE3 - Qualité et clarté du processus de gestion des données :</p> <ul style="list-style-type: none"> - <i>Les processus couvrent tous les aspects nécessaires pour répondre aux besoins identifiés.</i> - <i>Les processus sont optimisés : ils permettent une gestion fluide et rapide des données à toutes les étapes (collecte, traitement, stockage).</i>
<ul style="list-style-type: none"> - Gestion de l'accès aux données - <i>Définition ou application des politiques d'accès</i> - <i>Gestion des identités et des authentifications</i> - <i>Gestion des</i> 	<p>C2.1.3. Mettre en place un système d'authentification conforme aux procédures internes (en lien avec le RSSI ou la DSI) et à la réglementation en termes de protection des données et de sécurité en</p>		<p>CE1 Choix du système d'authentification des données :</p> <ul style="list-style-type: none"> - <i>Les méthodes d'authentification choisies sont appropriées et justifiées.</i> <p>CE2 - Opérationnalité et conformité du système</p>

<p><i>autorisations</i></p>	<p>vigueur afin de permettre un accès sécurisé aux données.</p>		<p>d'authentification des données :</p> <ul style="list-style-type: none"> - <i>Le système d'authentification est correctement installé et configuré selon les spécifications définies.</i> - <i>Le système d'authentification fonctionne : il protège les données contre les accès non autorisés.</i> <p>CE3 - Respect de la législation en vigueur et des bonnes pratiques de la profession :</p> <ul style="list-style-type: none"> - <i>Le système de gestion des données (dont système d'authentification) respecte la législation et la réglementation en vigueur, notamment la politique DLP, en cohérence avec la politique de sécurité SI définie par le RSSI :</i> <ul style="list-style-type: none"> - <i>les règles de classement pour la protection des informations confidentielles et sensibles ;</i> - <i>l'octroi des droits d'accès ;</i> - <i>la protection des informations stockées dans les serveurs ou cloud ;</i> - <i>les mesures de prévention des pertes de données.</i> - <i>Le système de gestion des données (dont système d'authentification) respecte les règles de l'art (ex : principe de conception type MVC (modèle-vue-contrôleur), et design patterns (factory, etc.)).</i> - <i>Le système de gestion des données (dont système d'authentification) est documenté.</i>
-----------------------------	---	--	---

A.2.2 Conception du système de stockage

<p>Mise en place de la solution de collecte et de stockage des données :</p> <ul style="list-style-type: none"> - Conception / structuration / organisation des BD : <ul style="list-style-type: none"> - Caractérisation des données (typologie, etc.) - Élaboration du schéma ou du modèle de données. - Mise en place d'un système de fichiers distribués ou en réseau. - Mise en place d'un système de gestion de base de données (ex : pgAdmin, services cloud : AWS, Azure) - Mise en place d'un système d'échange de données. 	<p>C2.2.1. Structurer l'architecture des données et concevoir des BDD relationnelles ou non (SQL / noSQL), respectant la politique de sécurité définie par le RSSI, afin de permettre l'exploitation des données par le SGBD :</p> <ul style="list-style-type: none"> - Définir / structurer / organiser les données et leurs relations en cohérence avec leurs caractéristiques. - Créer des bases de données relationnelles ou non (SQL / noSQL) afin de contenir les données. - Organiser les données par fichier, distribué ou en réseau, afin de disposer d'un accès direct et unifié au SBD depuis n'importe quel appareil / réseau. 	<p>L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.</p> <p>À partir d'un contexte donné, le candidat doit élaborer et mettre en place un système de gestion des données :</p> <ul style="list-style-type: none"> - Modéliser une architecture de données (schéma de base de données / modèles / tables de relations) - Installer et paramétrer un système de gestion de base de données. - Concevoir des BDD relationnelles ou non. - Installer et paramétrer un outil d'extraction et d'intégration des données. - Concevoir et exécuter un plan de test (aspects critiques - fonctionnalités à tester - scénarios - critères d'acceptation). - Produire un reporting des tests. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>CE1 - Pertinence de l'architecture des données (de la modélisation / structuration) :</p> <ul style="list-style-type: none"> - L'architecture des données répond aux besoins identifiés. - L'architecture des données permet un accès fluide aux données. - L'architecture des données est cohérente avec les caractéristiques des données (ex : typologie). <p>CE2 - Qualité de l'architecture des données (de la modélisation / structuration) :</p> <ul style="list-style-type: none"> - Les modèles de données (conceptuels, logiques et physiques) sont complets et précis. - Les données et leurs relations sont clairement définies et justifiées. - Les données sont clairement définies et organisées de manière cohérente avec la structure NoSQL utilisée, et les éventuelles relations (lorsqu'elles existent) sont justifiées par des choix de conception spécifiques au modèle NoSQL. - L'architecture est modélisée à l'aide d'un outil adapté. <p>CE3 - Pertinence et opérationnalité des BDD :</p> <ul style="list-style-type: none"> - Les bases de données sont fonctionnelles et alignées avec les exigences du projet. - Les tables et structures relationnelles ou non relationnelles (SQL / noSQL) sont correctement définies et implémentées. - Les choix entre les bases de données relationnelles et non relationnelles sont justifiés par les besoins spécifiques des données. <p>CE4 - Conformité des BDD :</p>
--	---	---	--

			<ul style="list-style-type: none"> - Les scripts / requêtes sont correctement réalisés et exécutés (ex : choix du langage, structure du code). - Des tests sont effectués pour valider la fonctionnalité, la performance et la sécurité des bases de données.
	<p>C2.2.2. Installer et paramétrer un système de gestion de base de données et un outil d'extraction (ex. : Airbyte, DBT), en collaboration avec la DSI, afin de permettre des opérations sur les fichiers composant les BDD, de diriger l'accès aux données et de fluidifier l'intégration des données dans le système d'échange.</p>		<p>CE1 - Conformité de l'installation du SGBD et de l'outil d'extraction :</p> <ul style="list-style-type: none"> - Le SGBD et l'outil d'extraction sont correctement installés sur les machines physiques ou virtuelles prévues. - Les étapes d'installation suivent les normes professionnelles (ex : les exigences minimales de hardware et software pour l'installation de l'outil d'extraction sont respectées). <p>CE2 - Pertinence du paramétrage du SGBD et de l'outil d'extraction :</p> <ul style="list-style-type: none"> - Le SGBD et l'outil d'extraction sont paramétrés pour optimiser les performances en fonction des besoins spécifiques du projet. - Les paramètres de configuration sont ajustés pour maximiser la sécurité, la performance et la fiabilité. - La configuration tient compte des spécificités du système d'exploitation et de l'environnement technique. <p>CE3 - Opérationnalité des processus d'extraction et d'intégration des données :</p> <ul style="list-style-type: none"> - Les données sont extraites et intégrées de manière fluide et efficace. - Les processus d'extraction et d'intégration sont robustes et minimisent les erreurs et les pertes

			de données.
Test de la solution de collecte et de stockage des données.	C2.2.3. Établir et exécuter un processus de test afin de s'assurer de l'opérationnalité du SGBD, de la disponibilité et de la qualité des données ainsi que la performance des requêtes.		<p>CE1 - Conformité et complétude de la méthodologie de tests :</p> <ul style="list-style-type: none"> - <i>Un plan de test complet est élaboré, incluant tous les aspects critiques du système de gestion des données.</i> - <i>Les scénarios de test couvrent les différentes fonctionnalités du système, y compris la collecte, le traitement et le stockage des données.</i> - <i>Les critères d'acceptation sont clairement définis et alignés avec les objectifs du projet.</i> - <i>Les tests sont exécutés de manière rigoureuse.</i> - <i>Les outils de test appropriés sont utilisés.</i> <p>CE2 - Opérationnalité des tests :</p> <ul style="list-style-type: none"> - <i>Les tests permettent de vérifier que les performances du système (vitesse, capacité, fiabilité) répondent aux exigences du projet.</i> - <i>Les tests de charge et de stress montrent que le système peut gérer les volumes de données prévus sans dégradation des performances.</i> <p>CE3 - Respect de la législation en vigueur et des bonnes pratiques de la profession :</p> <ul style="list-style-type: none"> - <i>L'ensemble des actions sont effectuées dans le respect de la législation / réglementation en vigueur.</i> - <i>L'ensemble des actions sont effectuées dans les règles de l'art.</i> - <i>L'ensemble des actions sont documentées.</i>
A.2.3 Création et intégration d'une infrastructure			
Conception de	C.2.3.1. Modéliser une	L'évaluation est réalisée au	CE1 - Pertinence de la définition de

<p>l'infrastructure des données :</p> <ul style="list-style-type: none"> - <i>Définition d'une infrastructure de gestion des données.</i> - <i>Choix des composants (matériels de stockage, matériels de traitement, logiciels, réseaux, etc.).</i> 	<p>infrastructure compatible avec le SI existant, en collaboration avec la DSI, afin de concevoir un support à la gestion des données (stockage, exploitation et partage des données) :</p> <ul style="list-style-type: none"> - <i>Définir, structurer et représenter visuellement une infrastructure de gestion des données à l'aide d'un outil de modélisation.</i> - <i>Identifier et sélectionner les composants disponibles en interne ou sur le marché (matériels, logiciels, réseaux etc.) et nécessaires à la création d'une infrastructure de gestion des données à partir de critères d'évaluation définis (performance, sécurité, compatibilité, et les coûts).</i> - <i>Évaluer la compatibilité des composants avec l'environnement SI de l'organisation, en collaboration avec la DSI.</i> 	<p>moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.</p> <p>À partir d'un contexte donné, le candidat doit concevoir et déployer une infrastructure de gestion des données :</p> <ul style="list-style-type: none"> - Concevoir et mettre en place (installer / tester) une infrastructure de gestion des données. - Automatiser les flux de données (script d'automatisation). - Justifier ses choix techniques. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>l'infrastructure :</p> <ul style="list-style-type: none"> - <i>Les besoins en gestion des données sont identifiés (création, stockage, exploitation, partage).</i> - <i>L'outil de modélisation utilisé fournit une représentation visuelle de l'infrastructure proposée, incluant les composants nécessaires et les flux de données est pertinent.</i> - <i>L'infrastructure modélisée répond efficacement aux besoins métiers en matière de gestion des données.</i> - <i>La représentation visuelle de l'infrastructure est claire et compréhensible par les parties prenantes.</i> - <i>L'infrastructure respecte les normes de sécurité (établies par le RSSI), de gestion des données, les contraintes budgétaires, etc.</i> <p>CE2 - Sélection des composants :</p> <ul style="list-style-type: none"> - <i>La sélection des composants (matériels, logiciels, réseaux) nécessaires pour l'infrastructure est basée sur des critères de performance, de sécurité, et de coûts.</i> - <i>Les composants sélectionnés sont compatibles avec l'environnement SI existant.</i> - <i>La structure des composants garantit la cohérence des flux de données, sans entraîner de redondance.</i> <p>CE3 - Validation de la conception :</p> <ul style="list-style-type: none"> - <i>Les tests effectués (validés avec la DSI) permettent de vérifier la compatibilité de l'infrastructure au SI existant.</i> - <i>La solution proposée répond aux besoins identifiés et aux contraintes techniques et organisationnelles.</i>
--	--	--	---

			<ul style="list-style-type: none"> - <i>Si nécessaire, des mécanismes d'automatisation sont intégrés à l'infrastructure, améliorant l'efficacité opérationnelle et réduisant les interventions manuelles.</i>
<p>Déploiement de l'infrastructure des données</p>	<p>C.2.3.2. Installer et tester l'infrastructure de gestion des données afin d'en garantir l'opérationnalité, la disponibilité et l'interopérabilité avec le SI.</p>		<p>CE1 - Installation et configuration Installation de l'infrastructure :</p> <ul style="list-style-type: none"> - <i>Les spécifications techniques et les besoins du système d'information (SI) sont respectés : les composants nécessaires (serveurs, bases de données, outils de gestion, etc.) sont correctement déployés et configurés.</i> - <i>La configuration de l'infrastructure est réalisée selon les meilleures pratiques (établies par la DSI), en veillant à ce que tous les paramètres soient ajustés pour garantir la performance et l'intégration avec les systèmes.</i> <p>CE2 - Tests et validation de l'infrastructure :</p> <ul style="list-style-type: none"> - <i>Des tests sont effectués sur les fonctionnalités essentielles du cycle de vie de la donnée: la collecte, le traitement, le stockage, et la récupération des données.</i> - <i>Des tests de disponibilité sont réalisés :</i> <ul style="list-style-type: none"> - <i>L'infrastructure est accessible en continu et capable de gérer les charges de travail prévues sans interruptions non planifiées.</i> - <i>L'interopérabilité avec le système d'information existant est validée, notamment en testant l'intégration avec d'autres systèmes et outils.</i>

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION

BLOC 3. Créer et /ou adapter un modèle d'apprentissage

Dans le cadre de la création ou de l'adaptation d'un modèle d'apprentissage, le titulaire de la certification opère dans un environnement de travail complexe nécessitant une expertise approfondie en science des données et en ingénierie des modèles. Il commence par choisir et installer l'infrastructure logicielle et matérielle adaptée au projet, en tenant compte des spécificités des données, des objectifs métiers et des contraintes techniques. Le contexte de cette activité est souvent caractérisé par la nécessité d'intégrer des données provenant de sources hétérogènes et de qualité variable, ce qui implique une préparation minutieuse des données.

Cette phase de préparation comprend l'évaluation de la qualité des données disponibles, la correction des anomalies (données manquantes, doublons, incohérences) et le contrôle systématique de leur qualité, souvent avec un recours à des outils d'automatisation pour garantir une fiabilité et une efficacité maximales. Le titulaire doit également assurer une gestion fine des données pour en tirer des jeux exploitables par le modèle d'apprentissage, via des processus de normalisation et de transformation.

Le choix du modèle d'apprentissage se base sur une évaluation approfondie des besoins du projet, en fonction des objectifs prédéfinis et des contraintes techniques (ressources de calcul, volumétrie des données). Une fois le modèle sélectionné, il procède à son entraînement, en séparant les jeux de données en ensembles d'entraînement, de test et de validation. Cette étape est cruciale pour garantir que le modèle soit capable de générer des résultats fiables et reproductibles.

Enfin, dans un souci d'applicabilité des résultats, le titulaire met en place une API permettant d'interfacer les modèles avec les systèmes utilisateurs, facilitant ainsi l'exploitation des prédictions générées. La visualisation des résultats est souvent un complément clé pour permettre aux parties prenantes de comprendre et d'interpréter les performances du modèle dans un cadre opérationnel. Tout au long de ce processus, le titulaire s'assure que chaque étape respecte les exigences de performance, de sécurité et de qualité définies par le projet.

A.3.1 Caractérisation des features et identification d'un modèle d'apprentissage

Choix, installation et paramétrage de l'environnement de travail	C3.1.1. Configurer l'environnement de travail nécessaire à l'exploitation des données (c'est-à-dire pour réaliser des analyses ou des opérations) dans des délais	L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.	CEI - Pertinence de la configuration de l'environnement pour l'exploitation des données : - <i>La configuration de l'environnement est adaptée aux besoins spécifiques des analyses ou des opérations sur les données (en fonction des types</i>
---	---	--	--

	<p>adaptés (temps de calcul).</p>	<p>À partir d'un contexte donné, le candidat doit caractériser les features et identifier un modèle d'apprentissage adapté :</p> <ul style="list-style-type: none"> - Configurer l'environnement de travail. - Mettre en place un processus de nettoyage automatisé ou non des données. - Proposer et paramétrer un modèle d'apprentissage. - Valider sa robustesse. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p><i>d'analyses ou des opérations requises).</i></p> <ul style="list-style-type: none"> - <i>Le choix de ressources est justifié en fonction des exigences de performance pour les opérations et analyses (temps de calcul, capacité de traitement).</i> - <i>La configuration permet de répondre aux délais de traitement des données, des besoins de performance et des contraintes techniques.</i> <p>CE2 - Conformité de la configuration de l'environnement pour l'exploitation des données :</p> <ul style="list-style-type: none"> - <i>La configuration respecte les normes et les meilleures pratiques.</i> - <i>La configuration garantit la qualité et la sécurité des données pendant les analyses et opérations.</i> - <i>Des mesures de sécurité adéquates sont mises en place, en lien avec la politique de sécurité établie par le RSSI.</i> - <i>Les processus d'analyse et d'opération sont sécurisés.</i> - <i>La configuration respecte les réglementations en vigueur concernant l'exploitation des données (ex. RGPD).</i> - <i>Les mécanismes de contrôle d'accès et de gestion des permissions sont définis et correctement configurés</i> <p>CE3 - Qualité et efficacité de la configuration de l'environnement pour l'exploitation des données :</p> <ul style="list-style-type: none"> - <i>La configuration permet une exploitation efficace des données :</i> <ul style="list-style-type: none"> - <i>Les temps de calcul sont adaptés</i> - <i>La gestion des processus d'analyse et d'opération est fluide.</i> - <i>La configuration garantit une performance</i>
--	-----------------------------------	--	---

			<p><i>optimale des opérations et des analyses (réduction des temps de traitement et des goulots d'étranglement).</i></p> <ul style="list-style-type: none"> - <i>La documentation des configurations est complète et claire, facilitant la gestion et l'ajustement des paramètres.</i>
<p>Préparation des données (ou nettoyage des données) :</p> <ul style="list-style-type: none"> - Évaluation de la qualité des données. - Correction des anomalies : données manquantes, doublons, erreur de saisie, données aberrantes, cohérence et qualité des sources. - Contrôle de la qualité des données. - Si besoin, automatisation des règles de nettoyage. 	<p>C.3.1.2. Mettre en place un processus de nettoyage des données automatisée ou non et le lancer afin d'améliorer la qualité des données :</p> <ul style="list-style-type: none"> - <i>Appliquer des analyses statistiques descriptives et / ou naviguer visuellement au sein des données afin de détecter des anomalies.</i> - <i>Supprimer / corriger les anomalies manuellement et à l'aide d'outils / logiciels de nettoyage des données adaptés.</i> - <i>Vérifier / contrôler la qualité des données - en continu - à l'aide d'outils de contrôle qualité adaptés afin de garantir la précision, la complétude et la cohérence.</i> - <i>Automatiser le processus de nettoyage à l'aide de d'un langage de programmation adapté (ex : Python)</i> 		<p>CE1 - Qualité de l'analyse des données :</p> <ul style="list-style-type: none"> - <i>Le processus de nettoyage est adapté aux besoins spécifiques du projet et aux caractéristiques des données.</i> - <i>Des analyses statistiques descriptives et des visualisations permettent d'identifier les anomalies (valeurs manquantes, incohérences, outliers) de manière exhaustive.</i> - <i>Les outils de visualisation et d'exploration des données permettent de naviguer et de comprendre les jeux de données complexes.</i> <p>CE2 - Efficacité du nettoyage des données :</p> <ul style="list-style-type: none"> - <i>Les anomalies détectées sont corrigées ou supprimées à l'aide d'outils de nettoyage adaptés (ex : Pandas, OpenRefine).</i> - <i>La correction des anomalies garantit la précision et la cohérence des données.</i> - <i>La correction des anomalies respecte les règles de bonnes pratiques.</i> - <i>Le processus de nettoyage est ajusté selon les spécificités des jeux de données (taille, structure, type), que ce soit de manière manuelle ou automatisée.</i> <p>CE3 - Automatisation et contrôle continu :</p> <ul style="list-style-type: none"> - <i>Le processus de nettoyage est automatisé à l'aide de scripts ou d'outils dédiés (ex : Python, R), minimisant les interventions manuelles et</i>

			<p><i>réduisant les risques d'erreurs humaines.</i></p> <ul style="list-style-type: none"> - <i>Le processus de nettoyage automatisé respecte les standards d'intégrité des données, en assurant la traçabilité des modifications, etc.</i> - <i>Des contrôles qualité continus sont mis en place. Ils permettent de garantir la cohérence et la précision des données traitées en temps réel.</i>
<p>Choix / configuration du modèle d'apprentissage</p>	<p>C3.1.3. Identifier un modèle d'apprentissage adapté aux contraintes (notamment techniques) et aux besoins métiers. Le cas échéant, créer un modèle d'apprentissage :</p> <ul style="list-style-type: none"> - <i>Choisir un algorithme adapté aux objectifs visés (analyse prédictive ...).</i> - <i>Programmer un algorithme à l'aide des hyperparamètres afin de contrôler le processus d'entraînement.</i> 		<p>CE1 - Pertinence du modèle sélectionné :</p> <ul style="list-style-type: none"> - <i>Le modèle d'apprentissage est choisi en fonction des besoins métiers et des contraintes techniques, garantissant l'adéquation entre les objectifs visés et le modèle (analyse prédictive, classification, etc.).</i> - <i>L'algorithme tient compte des exigences et contraintes techniques (capacité de traitement, volumes de données, temps de calcul, etc.).</i> - <i>L'algorithme sélectionné (régression, arbres de décision, réseaux de neurones, etc.) répond aux exigences en termes de performances et de précision des résultats.</i> <p>CE2 - Optimisation des hyperparamètres :</p> <ul style="list-style-type: none"> - <i>Les hyperparamètres du modèle sont ajustés et permettent d'optimiser le processus d'entraînement et la performance du modèle (taux d'apprentissage, nombre d'itérations, régularisation, etc.).</i> - <i>Le modèle est testé et la configuration est validée en fonction des résultats.</i> <p>CE3 - Adaptation du modèle aux spécificités métiers :</p> <ul style="list-style-type: none"> - <i>Si nécessaire, un modèle d'apprentissage sur mesure est créé, intégrant des algorithmes</i>

			<p>spécifiques et conformes aux besoins métiers particuliers non couverts par les modèles standards.</p> <ul style="list-style-type: none"> - Le modèle est compatible avec l'infrastructure technique existante et répond aux contraintes de performance.
--	--	--	---

A.3.2 Entraînement d'un modèle d'apprentissage et mise à disposition des résultats

<p>Prétraitement des données : normalisation (scaling) des données, transformation des caractéristiques.</p>	<p>C3.2.1. Préparer et transformer des données (standardisation, harmonisation, encodage, etc.) afin de les adapter au modèle d'apprentissage.</p>	<p>L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.</p> <p>À partir d'un contexte donné, le candidat doit entraîner un modèle d'apprentissage et mettre à disposition les résultats :</p> <ul style="list-style-type: none"> - Entraîner un modèle d'apprentissage (intégrant une analyse exploratoire des données, le choix des traitements à opérer sur les données et la préparation des données). - Réaliser une analyse prédictive à l'aide d'un modèle d'apprentissage entraîné, testé et validé. - Mettre à disposition les données via une API. - Rédiger un rapport complet. 	<p>CE1 - Qualité de la préparation des données :</p> <ul style="list-style-type: none"> - Les données sont nettoyées et les valeurs manquantes sont traitées (imputation, suppression, etc.) - Le nettoyage permet d'éviter les biais et d'optimiser le modèle d'apprentissage - Les données sont filtrées et normalisées selon les besoins spécifiques du modèle. <p>CE2 - Efficacité de la transformation des données :</p> <ul style="list-style-type: none"> - Les techniques de transformation sont appliquées de manière appropriée, telles que la standardisation (mise à l'échelle) et l'harmonisation. - Les données exploitables par le modèle. - Les encodages spécifiques (catégorisation, encodage one-hot, etc.) sont utilisés. - Les variables catégorielles sont transformées. <p>CE3 - Compatibilité des données avec le modèle :</p> <ul style="list-style-type: none"> - Les données transformées sont prêtes pour l'entraînement, avec un format adapté à la structure du modèle (régression, classification, etc.).
---	--	--	---

<p>Entraînement des modèles d'apprentissage :</p> <ul style="list-style-type: none"> - Préparation des jeux de données (séparation train / test). - Entraînement / test / validation du modèle d'apprentissage. 	<p>C3.2.2. Entraîner un modèle d'apprentissage (exemples : supervisé, non supervisé, par renforcement) afin de prédire la valeur d'un KPI, classifier la donnée tabulaire, du texte ou des images dans des catégories pré-définies :</p> <ul style="list-style-type: none"> - Préparer les jeux de données afin de mettre les variables sous une échelle commune et faciliter l'analyse et l'interprétation des résultats de l'entraînement du modèle. - Créer/ choisir et configurer un modèle d'apprentissage à l'aide d'une technique d'entraînement adaptée (ex : régression, classification). - Définir la procédure d'entraînement et entraîner le modèle avec les jeux de données adaptés. 	<p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>CE1 - Qualité de la préparation des données :</p> <ul style="list-style-type: none"> - Les jeux de données sont mis sous une échelle commune (normalisation, standardisation). - Les données sont nettoyées et prétraitées de manière cohérente pour éviter les biais ou les erreurs durant l'entraînement. <p>CE2 - Pertinence du modèle d'apprentissage :</p> <ul style="list-style-type: none"> - Le modèle sélectionné (régression, classification, etc.) correspond aux besoins de la tâche (prédiction d'un KPI, classification de données). - La configuration des hyperparamètres et des algorithmes est adaptée aux caractéristiques des données (taille, variabilité) et aux objectifs. <p>CE3 - Efficacité de la procédure d'entraînement :</p> <ul style="list-style-type: none"> - La procédure d'entraînement est définie en fonction du type de modèle et des jeux de données, incluant des techniques telles que la validation croisée. - L'entraînement du modèle est réalisé de manière efficace, avec un suivi régulier des performances pour ajuster si nécessaire.
	<p>C3.2.3. Créer les processus de test (procédures / outils) et les lancer en vue de confirmer / valider la mise en production du modèle d'apprentissage.</p>		<p>CE1 - Pertinence des processus de test :</p> <ul style="list-style-type: none"> - Les tests couvrent les différents cas d'usage du modèle (prédictions, classifications, etc.). - Les outils et méthodes de test choisis (unitaires, intégration, performance) sont adaptés aux caractéristiques du modèle. <p>CE2 - Efficacité et exhaustivité des tests :</p>

			<ul style="list-style-type: none"> - Les processus de test sont automatisés autant que possible, garantissant des tests reproductibles. - Les résultats des tests sont analysés, permettant de détecter et de corriger les erreurs avant la mise en production. <p>CE3 - Documentation et validation des tests :</p> <ul style="list-style-type: none"> - Les résultats des tests sont documentés avant le passage en production, assurant la conformité du modèle. - Les résultats des tests valident l'opérationnalité et l'efficacité du modèle
<p>Evaluation / Interprétation du modèle d'apprentissage</p>	<p>C3.2.4. Évaluer le modèle d'apprentissage selon les métriques définies afin d'en déterminer la performance, la capacité prédictive et de raisonnement.</p>		<p>CE1 - Pertinence des métriques d'évaluation :</p> <ul style="list-style-type: none"> - Les métriques choisies sont pertinentes et adaptées au type de modèle et à ses objectifs (précision, rappel, F1-score, AUC, etc.). - Les métriques permettent de couvrir tous les aspects clés : performance globale, capacité prédictive, et raisonnement du modèle. - <p>CE2 - Qualité de l'évaluation des performances :</p> <ul style="list-style-type: none"> - L'évaluation est rigoureusement menée sur des jeux de données de validation et de test, garantissant des résultats fiables et représentatifs. <p>CE3 - Capacité prédictive et cohérence du raisonnement :</p> <ul style="list-style-type: none"> - La capacité prédictive du modèle est vérifiée à l'aide de scénarios variés, avec des analyses pour identifier les biais potentiels ou des

			<p>comportements non désirés.</p> <ul style="list-style-type: none"> - Le modèle démontre une cohérence avec des prédictions justifiées par les caractéristiques apprises.
<p>Mise à disposition des résultats pour l'utilisateur (via une API)</p> <ul style="list-style-type: none"> - Création et / ou mise en place d'une API (intermédiaire entre les données ou modèle d'apprentissage et l'utilisateur). - Exploitation / visualisation des résultats. 	<p>C.3.2.5. Identifier et configurer une API compatible et l'intégrer afin de permettre l'accès aux résultats par les utilisateurs finaux. Le cas échéant créer une API :</p> <ul style="list-style-type: none"> - . - Définir les flux de données (entrées / sorties de l'API, et les formater si nécessaire) afin de sélectionner une API adaptée aux échanges entre le modèle et l'utilisateur - Installer / configurer l'API. - Tester les différentes interactions entre le modèle/ l'API / l'utilisateur. 		<p>CE1 - Qualité de la configuration de l'API :</p> <ul style="list-style-type: none"> - L'API choisie est adaptée aux besoins métiers et aux objectifs de l'utilisation des résultats du modèle par les utilisateurs finaux. - La définition des flux de données nécessaires (entrées/sorties) sont correctement définis, garantissant une bonne correspondance entre les échanges de données du modèle et les besoins de l'utilisateur. - Le choix de l'API existante est compatible avec l'environnement technique. <p>CE2 - Qualité des tests de l'API :</p> <ul style="list-style-type: none"> - Les différentes interactions entre le modèle et l'API sont testées. - Le résultat des tests est concluant et les flux de données sont correctement échangés. - Les résultats obtenus via l'API répondent aux attentes des utilisateurs finaux. - Le fonctionnement et le paramétrage de l'API respectent les normes et bonnes pratiques en vigueur. - La documentation de l'API est claire et complète, facilitant l'intégration, l'utilisation et la maintenance par les équipes techniques.
	<p>C.3.2.6. Exposer les résultats aux directions / services métiers (via une API) en vue de leur exploitation :</p> <ul style="list-style-type: none"> - Identifier et choisir un 	<p>L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.</p>	<p>CE1 - Qualité de la présentation des résultats :</p> <ul style="list-style-type: none"> - Le support de présentation des résultats est sélectionné en fonction des besoins métiers, des résultats à présenter, et des utilisateurs finaux.

	<p>système d'exposition des résultats (ex : PowerBI, Looker, Streamlit).</p> <ul style="list-style-type: none"> - Construire des tableaux de bord interactifs (ex : Tableau, Shiny, etc.) et / ou des rapports automatisés (ex : Jupyter Notebooks) afin de visualiser les données / les modèles et les insights générés pour l'utilisateur final. - Tenir compte des personnes en situation de handicap dans la construction des tableaux de bord. 	<p>À partir d'un contexte donné, le candidat doit :</p> <ul style="list-style-type: none"> - Choisir un système d'exposition des résultats (ex : PowerBI, Looker, Streamlit). - Construire des tableaux de bord interactifs et / ou des rapports automatisés afin de visualiser les données / les modèles pour l'utilisateur final. - Tenir compte des personnes en situation de handicap dans la construction des tableaux de bord. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés</p>	<ul style="list-style-type: none"> - Le support d'exposition des résultats est clair, complet et accessible, notamment aux personnes en situation de handicap. <p>CE2 - Pertinence de la conception des tableaux de bord ou des visualisations :</p> <ul style="list-style-type: none"> - Les tableaux de bord ou visualisations respectent les bonnes pratiques de la profession <ul style="list-style-type: none"> - interactivité (par exemple avec Tableau, Shiny) ou des rapports automatisés (par exemple avec Jupyter Notebooks) pour visualiser les données, les modèles et les insights générés. - Accessibilité : (descriptions alternatives, contrastes suffisants et navigation clavier-friendly). - Le tableau de bord et les rapports d'analyse sont conformes aux normes de présentation des données et d'accessibilité (ex. RGPD, pour les données personnelles, au RGAA¹). - Les visualisations sont optimisées, rendant ainsi les informations complexes facilement compréhensibles pour les différents utilisateurs. <p>CE3 - Accessibilité des tableaux de bord ou visualisations :</p> <ul style="list-style-type: none"> - Les principes d'accessibilité sont pris en compte, notamment pour les personnes en situation de handicap (respect des standards d'accessibilité WCAG ou au RGAA) : <ul style="list-style-type: none"> - Critère de succès 1.1.1 Contenu non textuel - Critère de succès 1.4.1 Utilisation de la couleur - Critère de succès 1.4.3 Contraste (minimum) - Critère de succès 1.4.4 Redimensionnement du
--	---	---	---

¹ [Référentiel général de l'amélioration de l'accessibilité](#)

			<i>texte</i> - <i>Critère de succès 2.4.2 Titre de page</i>
--	--	--	--

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION

BLOC DE SPÉCIALISATION : OPTION DATA ENGINEERING

BLOC 4. Développer, déployer et optimiser les pipelines de données

Dans le cadre de la conception, du déploiement et de l'optimisation des pipelines de données, le titulaire de la certification est responsable de la conception et de l'implémentation des pipelines de données. Cela inclut la création des circuits d'extraction, de traitement et de chargement des données, en intégrant des solutions adaptées à la nature des flux de données traitées et de leur volume. Le titulaire met en place un système d'ordonnancement permettant de gérer et d'automatiser le fonctionnement des pipelines, en assurant leur bon déroulement à chaque étape du cycle de vie des données.

Pour s'assurer du bon fonctionnement et de l'interopérabilité des pipelines, il conçoit et exécute les tests afin de s'assurer de la conformité des pipelines aux exigences techniques et métiers. Ces tests permettent de vérifier et d'apprécier la fiabilité et la performance des processus de traitement, notamment en termes de rapidité, de gestion des erreurs et de stabilité sur le long terme.

L'optimisation des pipelines fait partie intégrante de l'activité. Le titulaire identifie les points d'amélioration dans les performances des processus et applique les ajustements pour garantir la fluidité du système, de minimiser les coûts et délais de traitement, en cohérence avec les objectifs de l'organisation.

<p>Création des pipelines de données</p> <ul style="list-style-type: none"> - <i>Création des circuits d'extraction, de traitement et de chargement.</i> - <i>Mise en place d'un système d'ordonnancement et gestion des pipelines</i> 	<p>C4.1.1. Mettre en place un pipeline de données automatisé et adapté au besoin :</p> <ul style="list-style-type: none"> - <i>Extraire des données structurées / non structurées issues de toutes sources confondues afin de les traiter ou bien de les déplacer dans un emplacement adapté.</i> - <i>Transformer (standardisation,</i> 	<p>L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée.</p> <p>À partir d'un contexte donné, le candidat doit développer et déployer un pipeline de données optimisé :</p> <ul style="list-style-type: none"> - Concevoir, mettre en 	<p>CE1 - Qualité de la mise en place du pipeline :</p> <ul style="list-style-type: none"> - <i>L'extraction des données structurées et non structurées est réalisée à partir de diverses sources (bases de données, fichiers, APIs, etc.). Elle permet de garantir la précision et la complétude des données extraites.</i> - <i>Le pipeline inclut des processus automatisés pour la transformation des données (standardisation, déduplication, tri, vérification), assurant ainsi que les données sont correctement préparées pour leur utilisation</i>
---	--	--	---

	<p>déduplication, tri, vérification, etc.) des données afin de les adapter à leur utilisation finale.</p> <ul style="list-style-type: none"> - Charger des données afin de les stocker dans un emplacement adapté (magasin de données type data lake, entrepôt de données etc.). 	<p>place (installer / tester/ surveiller) un pipeline de données</p> <ul style="list-style-type: none"> - Évaluer et analyser les performances des flux. - Proposer et mettre en place des améliorations. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p>	<p>finale.</p> <ul style="list-style-type: none"> - Le chargement des données transformées est fait dans un emplacement de stockage adapté (data lake, entrepôt de données, etc.). <p>CE2 - Automatisation et validation du pipeline :</p> <ul style="list-style-type: none"> - Le pipeline est automatisé pour gérer les processus d'extraction, de transformation et de chargement des données sans intervention manuelle, en utilisant des outils et technologies adaptés. - Des tests sont réalisés. ils permettent : <ul style="list-style-type: none"> - de valider le bon fonctionnement du pipeline, - de vérifier que les données sont correctement extraites, transformées et chargées, - de détecter et de corriger les anomalies éventuelles. - Un système de surveillance est mis en place. Il permet : <ul style="list-style-type: none"> - d'assurer le bon fonctionnement du pipeline, - de fournir des alertes en cas de défaillance ou d'anomalie dans le traitement des données.
	<p>C4.1.2. Mettre en place un système d'ordonnancement des flux de données afin de programmer leur déclenchement.</p>		<p>CE1 - Qualité de la conception du système d'ordonnancement :</p> <ul style="list-style-type: none"> - Un système d'ordonnancement est mis en place. Il permet : <ul style="list-style-type: none"> - de répondre aux besoins spécifiques des flux de données, - d'assurer une planification précise des tâches selon les exigences opérationnelles. - Les outils d'ordonnancement choisis sont

			<p><i>compatibles avec l'infrastructure existante et adaptés aux types de flux de données à gérer.</i></p> <ul style="list-style-type: none"> - <i>La configuration du système inclut une documentation claire des horaires, des dépendances et des priorités des flux de données.</i> <p>CE2 - Cohérence de la mise en œuvre et validation du système d'ordonnancement :</p> <ul style="list-style-type: none"> - <i>Le système d'ordonnancement est mis en place et testé.</i> - <i>Le système d'ordonnancement garantit prévoit que les flux de données sont déclenchés conformément au planning défini.</i> - <i>Les processus d'ordonnancement sont surveillés pour détecter et résoudre les éventuels problèmes de déclenchement ou de performance.</i> - <i>Les ajustements nécessaires sont effectués. Ils permettent :</i> <ul style="list-style-type: none"> - <i>d'optimiser l'efficacité du système d'ordonnancement,</i> - <i>de garantir une exécution fluide des flux de données.</i>
<p>Test des pipelines de données</p>	<p>C4.1.3. Créer les processus de test et les lancer afin de valider la mise en production des pipelines de données.</p>		<p>CE1 - Qualité de la conception des processus de test :</p> <ul style="list-style-type: none"> - <i>Les processus de test conçus couvrent toutes les étapes critiques des pipelines de données, incluant la collecte, le traitement et le stockage.</i> - <i>Les processus sont documentés avec des scénarios de test clairs, incluant des cas de test fonctionnels, de performance, et de sécurité.</i> - <i>Les outils et méthodologies de test sont choisis en fonction des besoins spécifiques des pipelines et de l'infrastructure utilisée.</i>

			<p>CE2 - Exécution des tests :</p> <ul style="list-style-type: none"> - Les processus de test sont lancés dans un environnement de pré-production qui reflète fidèlement l'environnement de production. - Les tests sont effectués systématiquement pour valider le bon fonctionnement des pipelines, en détectant et en corrigeant les anomalies ou les erreurs avant la mise en production. - Les résultats des tests sont analysés pour garantir que tous les critères de performance, de qualité et de sécurité sont satisfaits.
<p>Optimisation des pipelines de données.</p>	<p>C4.1.4. Concevoir et mettre en place un système de contrôle / de suivi automatisé des flux de données afin de valider en continu l'intégration des données :</p> <ul style="list-style-type: none"> - Surveiller les flux à l'aide des outils de gestion de la qualité des données adaptés (ex : GreatExpectation, SODA) afin de détecter les anomalies. - Suivre les flux à l'aide d'un outil d'observation adapté (ex : Openlineage) afin de les tracer. 		<p>CE1 - Qualité de la conception et de la mise en place :</p> <ul style="list-style-type: none"> - Le système de contrôle / suivi est conçu pour surveiller et tracer en continu les flux de données, avec des outils adaptés à la gestion de la qualité (ex : GreatExpectations, SODA). - Les outils choisis sont compatibles avec l'infrastructure existante et permettent une détection rapide des anomalies dans les flux de données. - La configuration du système inclut une documentation claire sur les processus de surveillance et de suivi des flux, avec des alertes en cas d'anomalies. <p>CE2 - Efficacité de la surveillance et de la traçabilité :</p> <ul style="list-style-type: none"> - Le système de contrôle permet une détection automatisée des anomalies via des outils d'observation des flux (ex : Openlineage), garantissant une traçabilité complète. - Des tests sont effectués pour assurer que les anomalies et incohérences dans les flux de données sont rapidement détectées et résolues.

			<ul style="list-style-type: none"> - Les flux sont suivis et tracés en temps réel, garantissant l'intégrité et la qualité continue des données intégrées. <p>CE3 Efficacité du système de suivi des flux :</p> <ul style="list-style-type: none"> - Le pourcentage d'anomalies détectées et corrigées automatiquement correspond aux besoins. - Le taux de traçabilité des flux de données (par rapport aux flux traités) correspond aux besoins.
	<p>C4.1.5. Ajuster les requêtes et les paramétrages des pipelines en prenant appui sur différentes approches / méthodes (ex : méthodes de parallélisation) afin d'améliorer les performances du système (ex : réduction des temps d'exécution).</p>		<p>CE1 Qualité des ajustements apportés :</p> <ul style="list-style-type: none"> - Les requêtes et les paramétrages des pipelines sont ajustés en utilisant des approches appropriées (ex : parallélisation, optimisation des requêtes) pour améliorer les performances du système. - Les modifications sont basées sur une analyse approfondie des performances actuelles et sont documentées. - Les modifications assurent une compréhension des ajustements réalisés. - Les ajustements sont testés dans un environnement contrôlé avant le déploiement. - Les ajustements améliorent réellement les performances sans introduire de nouvelles erreurs. - <p>CE2 - Efficacité des améliorations de performances :</p> <ul style="list-style-type: none"> - Les ajustements permettent une réduction mesurable des temps d'exécution des pipelines, avec des comparaisons avant et après pour évaluer l'impact. - Des outils de reporting sont utilisés pour évaluer

			<i>l'efficacité des modifications en termes de performances (ex : réduction des temps de traitement, amélioration de la vitesse de réponse).</i>
--	--	--	--

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION

BLOC DE SPÉCIALISATION : OPTION DATA SCIENCE

BLOC 5. Optimiser des modèles d'apprentissage

Dans le cadre de l'optimisation des modèles d'apprentissage, le titulaire de la certification est responsable du suivi et du contrôle du cycle de vie des modèles déployés. Cela inclut la supervision des performances des modèles, depuis leur phase d'entraînement initiale jusqu'à leur mise en production, en passant par les étapes d'amélioration continue et d'itération.

Le titulaire veille à optimiser l'efficacité des modèles en ajustant les paramètres et en réévaluant les données utilisées pour l'entraînement, tout en s'assurant que l'infrastructure supporte les charges de travail associées. Il s'agit de veiller à ce que les modèles opèrent avec consommation minimale des ressources tout en garantissant une rapidité et une précision maximales dans leurs prédictions.

En plus de ces tâches d'optimisation, le titulaire met en œuvre des processus d'automatisation pour le déploiement des modèles d'apprentissage. En utilisant des méthodologies d'intégration continue (CI) et de déploiement continu (CD), il est responsable de l'intégration fluide des nouvelles versions des modèles ou de nouveaux modèles dans les environnements de production. Cela permet de garantir une mise à jour rapide et sans interruption des services, tout en minimisant les risques d'erreur et en maximisant la stabilité des systèmes.

Suivi / contrôle du cycle de vie des modèles d'apprentissage	C5.1.1. Concevoir et mettre en place un système de suivi du cycle de vie du modèle d'apprentissage (et de ses fonctionnalités) afin de détecter des anomalies (dérives de	L'évaluation est réalisée au moyen d'un projet professionnalisant, sous la forme d'une mise en situation professionnelle reconstituée et d'une étude de cas.	CE1 - Pertinence et qualité du système de suivi du cycle de vie du modèle : - <i>Le système de suivi permet de surveiller les différentes étapes du cycle de vie du modèle (entraînement, validation, déploiement,</i>
---	---	--	--

	<p>données ou bien dérives conceptuelles), de surveiller la qualité des features ou bien de déterminer le moment et la méthode de ré-entraînement du modèle.</p>	<p>Mise en situation reconstituée Le candidat doit suivre et optimiser le cycle de vie d'un modèle :</p> <ul style="list-style-type: none"> - concevoir et mettre en place un système de suivi performant du cycle de vie de ses modèles d'apprentissage. - évaluer les performances de l'infrastructure sous-jacente au modèle d'apprentissage. - ajuster les paramètres d'entraînement du modèle - optimiser sa performance - automatiser le déploiement du modèle. - tester et valider les changements de code. - Envoyer le code en production. <p>L'ensemble du travail est présenté à un évaluateur (expert technique) lors d'une soutenance distancielle dont la durée et le déroulement sont calibrés.</p> <p>Etude de cas</p>	<p>ré-entraînement).</p> <ul style="list-style-type: none"> - Le système de suivi permet de surveiller les dérives des données et les dérives conceptuelles du modèle, en intégrant des indicateurs appropriés. - Les outils choisis pour le suivi (monitoring, alertes) sont adaptés aux besoins spécifiques du modèle et compatibles avec l'infrastructure existante. - La configuration du système intègre des mécanismes de détection précoce des anomalies, avec documentation des processus et alertes. - Les indicateurs de performance choisis permettent d'évaluer la qualité des features, les erreurs de prédiction et la stabilité du modèle sur le long terme. <p>CE2 - Efficacité de la surveillance et de la détection :</p> <ul style="list-style-type: none"> - Le système permet une détection automatique des dérives de données (data drift) et conceptuelles (concept drift) en temps réel, garantissant une réactivité rapide. - Des procédures sont mises en place pour la surveillance continue des features, assurant leur qualité et leur pertinence pour le modèle d'apprentissage. - Le système détermine efficacement le moment optimal pour le ré-entraînement du modèle, en fonction des indicateurs de performance. <ul style="list-style-type: none"> - Par exemple : pourcentage de dérives détectées par rapport au total des anomalies observées. - Réduction du nombre d'erreurs ou d'imprécisions du modèle grâce à la surveillance proactive des dérives. - Les mécanismes de ré-entraînement
--	--	--	--

		<p>Dans le cadre d'une étude de cas, le candidat doit évaluer et ajuster les performances d'un modèle :</p> <ul style="list-style-type: none"> - Analyser un système de surveillance déjà mis en place, proposer des améliorations pour : <ul style="list-style-type: none"> - renforcer la détection des anomalies - garantir une surveillance optimale de la qualité des données d'entrée. - Examiner les performances de l'infrastructure actuelle, en identifier les points faibles et proposer des solutions concrètes. - Ajuster les étapes d'entraînement du modèle et optimiser sa performance via des modifications ciblées. 	<p><i>automatique ou semi-automatique respectent les bonnes pratiques de gestion des modèles et les exigences métiers.</i></p> <ul style="list-style-type: none"> - <i>Le système de suivi respecte les normes et réglementations en vigueur, notamment la transparence des modèles, la gestion des biais, etc.</i>
<p>Optimisation des performances des modèles d'apprentissage :</p> <ul style="list-style-type: none"> - <i>Via l'entraînement des modèles.</i> - <i>Via l'infrastructure de gestion des modèles d'apprentissage.</i> 	<p>C5.1.2. Évaluer les performances de l'infrastructure sous-jacente au modèle d'apprentissage selon des critères définis afin de réduire les coûts et / ou les temps de calcul.</p>		<p>CE1 - Pertinence et qualité de l'évaluation des performances de l'infrastructure :</p> <ul style="list-style-type: none"> - <i>Les critères d'évaluation des performances (coûts, temps de calcul, ressources utilisées) sont définis en accord avec les objectifs du projet.</i> - <i>L'évaluation des performances est réalisée en identifiant précisément les points de faiblesse de l'infrastructure.</i> - <i>Les outils d'analyse utilisés sont appropriés pour mesurer les performances de manière fiable et précise.</i> <p>CE2 - Efficacité des recommandations d'optimisation :</p> <ul style="list-style-type: none"> - <i>Des recommandations concrètes sont proposées pour réduire les coûts ou améliorer les temps de calcul, avec une évaluation de leur impact.</i> - <i>Les ajustements recommandés sur l'infrastructure prennent en compte l'équilibre entre performance, coût, et disponibilité.</i>
	<p>C5.1.3. Ajuster les paramètres des procédures d'entraînement / de test / validation, à l'aide de techniques adaptées (hyperparamètres, choix des caractéristiques, ajustement</p>		<p>CE1 - Pertinence des ajustements des hyperparamètres :</p> <ul style="list-style-type: none"> - <i>Les hyperparamètres sont ajustés de manière méthodique, en tenant compte des caractéristiques des données et des objectifs du</i>

	<p>d'architecture de neurone, etc.) afin d'optimiser la qualité et les performances des données.</p>		<p>modèle.</p> <ul style="list-style-type: none"> - Les caractéristiques choisies permettent d'améliorer les performances prédictives du modèle. - Des techniques de recherche d'hyperparamètres (grid search, random search, etc.) sont mises en place. - L'optimisation des hyperparamètres améliore de manière mesurable les performances du modèle par exemple : en pourcentage de précision, rappel, etc.). - Le temps d'entraînement ou de validation est réduit après ajustement des paramètres (en pourcentage ou temps absolu).
<ul style="list-style-type: none"> - Automatisation du déploiement du modèle d'apprentissage (approche CI / CD) (nouveau modèle ou nouvelle version du modèle existant) 	<p>C5.1.4. Automatiser le déploiement afin d'intégrer et de livrer en continu les évolutions apportées au modèle d'apprentissage (modifications de code) :</p> <ul style="list-style-type: none"> - Déposer le code modifié dans un outil de contrôle de version afin de partager / de fusionner le code. - Créer des tests CI / CD afin de contrôler / valider les changements de code. - Envoyer le code en production et l'exécuter de manière automatique. 		<p>CE1 - Qualité du dépôt de code et de l'intégration :</p> <ul style="list-style-type: none"> - Le pipeline CI/CD permet de déployer les modifications du code sans dégradation des performances du modèle en production. - Le code modifié est déposé et partagé dans un outil de contrôle de version - Le processus de fusion et de validation du code garantit une intégration fluide. - L'outil de version utilisé est compatible avec les besoins du projet et l'infrastructure existante. - La documentation des modifications et des processus d'intégration continue est claire et conforme aux bonnes pratiques. <p>CE2 - Efficacité de l'automatisation des tests et du déploiement :</p> <ul style="list-style-type: none"> - Les tests sont créés et exécutés pour valider les changements de code avant leur mise en production. - Les tests sont alignés avec les objectifs de qualité et de performance du modèle.

			<ul style="list-style-type: none">- <i>Le déploiement en production est réalisé de manière automatisée et fiable, assurant l'intégration des nouvelles évolutions.</i>- <i>Des tests sont effectués pour garantir la stabilité du code en production et l'efficacité des changements implémentés.</i>- <i>Les temps de déploiement sont optimisés grâce à l'automatisation.</i> <p>CE3 Conformité aux normes et exigences :</p> <ul style="list-style-type: none">- <i>Le processus d'intégration et de déploiement automatisés sont sécurisés.</i>- <i>Le processus de déploiement automatisé respecte les réglementations en vigueur (ex. RGPD) en matière de gestion des données et de traçabilité des changements.</i>
--	--	--	--