

Intitulé de la certification
Science des données et apprentissage profond au service de l'intelligence artificielle
Description du métier, de l'activité ou de la situation professionnelle à partir duquel le dispositif de formation visant la certification est initié:

Le présent dispositif répond au besoin d'exploitation intelligente des masses de données acquises cette dernière décennie, ayant pour finalité la mise au point de systèmes de reconnaissance et de décision automatiques visant à améliorer les process de production, la surveillance et l'exploitation d'installations, la personnalisation des services (médecine, diffusion de l'information, transport, etc.). Ce dispositif a pour but d'apporter aux ingénieurs les compétences nécessaires à la mise en place de logiciels informatiques automatisés issus de l'intelligence artificielle.

Référentiels

Référentiel de compétences	Référentiel de certification	
	Modalité(s)	Critères
C1. Recueillir des données publiques ou privées exploitables pour l'entraînement de modèles d'apprentissage automatique	MODALITES D'EVALUATION E1. Mise en situation professionnelle À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit sélectionner les outils mathématiques et les méthodes d'apprentissage adaptés et justifier de leur exactitude grâce à un raisonnement logique. E2. Mise en situation professionnelle Contenu : À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit traiter les données, entraîner un modèle d'apprentissage automatique ou d'apprentissage profond et calibrer ses paramètres de sorte à maximiser ses performances.	(C1) Capacité à : <ul style="list-style-type: none"> • Déterminer l'architecture de stockage des données ; • Construire une séquence de requêtes pour rapatrier les données ; • Exploiter les bibliothèques du langage Python pour remettre en forme les données.
C2. Analyser les caractéristiques et la qualité des données afin d'évaluer leur potentiel pour des applications d'intelligence artificielle		(C2) Capacité à : <ul style="list-style-type: none"> • Choisir les techniques de statistiques descriptives adaptées aux données ; • Exploiter les bibliothèques du langage Python pour calculer ces statistiques ; • Interpréter les valeurs obtenues au regard de l'ensemble de la chaîne de traitement visée.
C3. Exploiter les architectures big data du cloud afin de manipuler de grandes masses de données et d'en extraire une information pertinente		(C3) Capacité à : <ul style="list-style-type: none"> • Mettre en place et exploiter un cluster HDFS pour le stockage distribué et parallèle large échelle ; • Décrire les étapes d'extraction d'information selon le modèle MapReduce ;

	<p>Rendu attendu : Rapport écrit comprenant une synthèse du code informatique développé pour mener à bien l'élaboration du jeu de données et le modèle d'apprentissage, une analyse et une critique constructive des résultats numériques permettant d'améliorer l'ensemble.</p>	<ul style="list-style-type: none"> • Utiliser l'infrastructure logicielle Spark pour manipuler les données et mettre en place les schémas MapReduce.
<p>C4. Analyser des besoins métiers et déterminer comment ils peuvent être solutionnés par une succession de tâches d'intelligence artificielle</p>	<p>E3. Projet professionnel Contenu : À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit identifier les problématiques et les enjeux rencontrés par l'entreprise. Il/elle doit traduire ces derniers en objectifs réalisables et mettre en place une chaîne complète de traitement des données et d'apprentissage (incluant la sélection et le calibrage du modèle approprié) permettant de répondre à ces objectifs. Enfin, le/la candidat(e) doit communiquer son protocole expérimental et ses conclusions, et fournir un regard critique sur les limites de son approche.</p>	<p>(C4) Capacité à :</p> <ul style="list-style-type: none"> • Restreindre un périmètre d'action dans lequel l'intelligence artificielle peut avoir un impact ; • Établir un lien entre les besoins et l'état de l'art technologique en intelligence artificielle ; • Définir la nature et la quantité des données à exploiter ; proposer une architecture d'acquisition et de stockage des données à exploiter
<p>C5. Développer et mettre en œuvre une architecture de traitement de données répondant à une problématique industrielle ciblée en utilisant les bibliothèques logicielles de référence</p>	<p>À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit identifier les problématiques et les enjeux rencontrés par l'entreprise. Il/elle doit traduire ces derniers en objectifs réalisables et mettre en place une chaîne complète de traitement des données et d'apprentissage (incluant la sélection et le calibrage du modèle approprié) permettant de répondre à ces objectifs. Enfin, le/la candidat(e) doit communiquer son protocole expérimental et ses conclusions, et fournir un regard critique sur les limites de son approche.</p>	<p>(C5) Capacité à :</p> <ul style="list-style-type: none"> • Formaliser une problématique concrète en termes mathématiques ; • Extraire et mettre en forme les données à exploiter ; • Choisir plusieurs modèles d'apprentissage automatique adaptés à la formalisation mathématique et aux données.
<p>C6. Construire une architecture d'apprentissage profond permettant d'analyser automatiquement des données visuelles et textuelles</p>	<p>À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit identifier les problématiques et les enjeux rencontrés par l'entreprise. Il/elle doit traduire ces derniers en objectifs réalisables et mettre en place une chaîne complète de traitement des données et d'apprentissage (incluant la sélection et le calibrage du modèle approprié) permettant de répondre à ces objectifs. Enfin, le/la candidat(e) doit communiquer son protocole expérimental et ses conclusions, et fournir un regard critique sur les limites de son approche.</p>	<p>(C6) Capacité à :</p> <ul style="list-style-type: none"> • Choisir le type d'architecture neuronale adaptée aux données visuelles ou textuelles et à la tâche à accomplir ; • entraîner un réseau de neurones profond
<p>C7. Adapter une architecture d'apprentissage profonde pré-entraînée aux spécificités des problématiques applicatives.</p>	<p>Rendu attendu : Une soutenance composée d'une présentation de 20 minutes faisant état du contexte, des problématiques et des enjeux de l'entreprise, des éléments techniques de la chaîne de traitement mise en place et des conclusions de l'analyse critique de l'approche et de ses limites. Cette présentation est suivie d'un échange de 10 minutes avec le jury permettant d'évaluer en détail les compétences acquises.</p>	<p>(C7) Capacité à :</p> <ul style="list-style-type: none"> • Choisir et charger une architecture neuronale profonde pré-entraînée en fonction de la tâche à accomplir ; • ajuster une architecture pré-entraînée sur des données liées à une application spécifique
<p>C8. Comparer et évaluer les performances de différents modèles d'apprentissage automatique</p>	<p>À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit identifier les problématiques et les enjeux rencontrés par l'entreprise. Il/elle doit traduire ces derniers en objectifs réalisables et mettre en place une chaîne complète de traitement des données et d'apprentissage (incluant la sélection et le calibrage du modèle approprié) permettant de répondre à ces objectifs. Enfin, le/la candidat(e) doit communiquer son protocole expérimental et ses conclusions, et fournir un regard critique sur les limites de son approche.</p>	<p>(C8) Capacité à :</p> <ul style="list-style-type: none"> • Entraîner des modèles d'apprentissage automatique de natures variées sur un même jeu de données ; • Calculer des indicateurs de performance adaptés.
<p>C9. Communiquer autour des données à destination des décideurs, des experts métier, du grand public</p>	<p>À partir d'un cas d'entreprise réel ou fictif, le/la candidat(e) doit identifier les problématiques et les enjeux rencontrés par l'entreprise. Il/elle doit traduire ces derniers en objectifs réalisables et mettre en place une chaîne complète de traitement des données et d'apprentissage (incluant la sélection et le calibrage du modèle approprié) permettant de répondre à ces objectifs. Enfin, le/la candidat(e) doit communiquer son protocole expérimental et ses conclusions, et fournir un regard critique sur les limites de son approche.</p>	<p>(C9) Capacité à :</p> <ul style="list-style-type: none"> • Décrire et expliquer l'environnement des données, le périmètre d'action de l'intelligence artificielle, ses bienfaits et ses limites ; • Expliquer une démarche pratique et décrire les besoins matériels nécessaires au traitement d'un cas concret ; • Vulgariser les méthodes mises en œuvre dans le traitement d'un cas concret ; • Communiquer les performances et les limitations de différents modèles d'apprentissage automatique ;

		<ul style="list-style-type: none"> • Informer relativement à l'éthique, la sécurité et au droit d'utilisation des données.
<p>C10. Dialoguer avec les acteurs de l'entreprise pour aller d'un prototype à la mise en production</p>		<p>(C10) Capacité à :</p> <ul style="list-style-type: none"> • rédiger une documentation informatique ; • communiquer les bonnes pratiques d'utilisation des méthodes d'apprentissage automatique ; • restituer les évolutions techniques des bibliothèques logicielles de référence ; • vulgariser des concepts mathématiques liés à l'intelligence artificielle.
<p>C11. Respecter la réglementation et les bonnes pratiques d'utilisation des données, concernant le respect des données personnelles, de sécurité et d'éthique</p>		<p>(C11) Capacité à :</p> <ul style="list-style-type: none"> • Mettre en pratique la réglementation concernant la sécurité, l'éthique et le droit des données ; • Se tenir informé des évolutions réglementaires concernant l'utilisation des données.