



FFCCERTT 12 500 € 26 jour(s)
ERXBD01



[Formation certifiante] Data Architect - MLOPS, GENAI, DATA SCIENCE

OBJECTIFS

- Concevoir une architecture de données et de traitements robuste et évolutive, y compris la modélisation, la génération des données et le stockage des données
- Intégrer les solutions d'intelligence artificielle dans des architectures de données existantes ou dans des architectures en conception (byDesign) avec une approche frugale
- Produire des pipelines de données efficaces pour l'entraînement, le déploiement et la gestion des modèles de science de données et d'intelligence artificielle
- Gérer le cycle de vie des modèles, la supervision des performances et l'automatisation des processus de déploiement
- Limiter les risques dans la conception, le déploiement et l'exploitation de l'architecture de données

PROGRAMME

Introduction

L'analyse exploratoire des données

- Explorer les données
- Traitement préalable des données : Assurer La Complétude, Gérer Les Outliners, Nettoyage, Normalisation, Atténuation, Discrétisation
- Les données en tant qu'actifs de l'entreprise

Architecture d'une application IA

- Besoins, objectifs et contraintes
- Approche monolithique vs microservices
- Intégration de la sécurité dans le cycle de vie du développement applicatif

Enjeux de l'IA



DATES ET LIEUX

Du 07/09/2026 au 12/05/2027 à Paris

PUBLIC / PREREQUIS

Ingénieur ou chef de projet souhaitant développer des compétences dans le domaine de conception et déploiement de systèmes de données massives pour évoluer vers un poste tel que celui d'architecte des données (architecte data), d'ingénieur MLOps ou de responsable projet data.

COORDINATEURS

Pietro GORI

Enseignant-chercheur au département Image, Données, Signal de Télécom Paris et au laboratoire LTCl. Ses recherches portent principalement sur l'anatomie computationnelle, l'analyse des formes, l'apprentissage statistique et l'imagerie médicale.

Fabian SUCHANEK

Enseignant-chercheur à Télécom Paris. Il a fait ses recherches à l'Institut Max Planck en Allemagne, chez Microsoft Research Cambridge/UK, chez Microsoft Research Silicon Valley/USA, et à l'INRIA Saclay. Il est l'auteur principal de YAGO, une des plus grandes bases de

- Risques juridiques et éthiques
- Risques produit des applications de science de données et IA
- Accompagnement du changement pour une transformation numérique avec l'IA
- Définir une stratégie de qualité de données (data quality)

Systemes de traitement de données massives

- Applications cloud-natives
- Cadrer un projet de données intensives pour la science de données et l'IA
- Intégration de la sécurité dans le cycle de vie du développement logiciel
- Mapreduce
- Au-delà de mapreduce : spark, stratosphere
- Message passing interface
- Calculs distribués sur des graphes : graphlab, pregel, giraph
- Infrastructure et analyse continu
- Optimisation et déploiement continu de modèles

Apprentissage machine: de la théorie aux design patterns

- Produire un modèle d'apprentissage automatique
- Apprentissage supervisé, deeplearning et IA générative : éléments de la théorie de vapnik chervonenkis, support vector machines et arbres de décision
- Apprentissage en ligne
- Apprentissage statistique distribué
- Techniques d'échantillonnage
- Variables latentes et clustering
- Chaînes de markov cachés
- Réseaux de neurones, apprentissage par renforcement et traitement d'images
- Design patterns applicatifs et modèles

Traitement du langage naturel et Transformers

- Tokenisation
- Marquage de partie de discours
- Représentation de document
- Word embeddings
- Transformers, llms et infrastructures de calcul

Génération de données

- Retrieval augmented generation
- Fine tuning à partir de modèles pré-entraînés
- Génération des données visuelles
- Déployer un modèle d'IA générative

connaissances publiques dans le monde.

Stockage de données massives

- Sgbd relationnels distribués classiques
- Systèmes de fichiers distribués hdfs/gfs
- Stockage à grande échelle
- Stockage clés-valeurs par table de hachage distribuée (dynamo)
- Stockage par arbre distribué (bigtable, hbase)
- Systèmes newsq (google spanner, sgbd en mémoire, mysql cluster)

Apprentissage distribué : fouille de Graphes

- Distribution d'algorithmes d'indexation, d'apprentissage et de fouille
- Index Inversé
- Factorisation de matrice
- Échantillonnage
- Pagerank

Synthèse et Conclusion

Appelez le 01 75 31 95 90
International : +33 (0)1 75 31 95 90

contact.exed@telecom-paris.fr / executive-education.telecom-paris.fr