

Mis à jour le 07/08/2025

S'inscrire

# Formation ONNX Edge AI Deploy

2 jours (14 heures)

## Présentation

ONNX Edge AI Deploy est un ensemble d'outils et de pratiques permettant de convertir, optimiser et exécuter des modèles IA dans le format ONNX sur des dispositifs à faible puissance tels que des cartes embarquées.

ONNX, pour Open Neural Network Exchange, est un standard ouvert favorisant l'interopérabilité entre frameworks comme PyTorch et TensorFlow. Ce format facilite le déploiement rapide, stable et optimisé de modèles IA sur une large gamme de plateformes Edge.

Notre formation ONNX Edge AI Deploy vous permettra de maîtriser le cycle complet de la préparation à l'exécution de modèles ONNX dans un environnement contraint. Vous apprendrez à convertir vos modèles vers le format ONNX, à les optimiser pour un fonctionnement fluide sur du matériel embarqué, et à les déployer sur des plateformes comme Jetson Nano, Raspberry Pi ou Coral.

Une place importante est donnée aux aspects pratiques du debug, de la quantization et de la mesure de performance pour garantir un déploiement industriel de qualité.

À l'issue de cette formation, vous serez en mesure de construire une pipeline complète de traitement, de la data jusqu'au déploiement, tout en assurant une interopérabilité maximale avec les outils du marché.

Cette formation vous présentera la dernière version stable d'[ONNX Runtime v1.22.1](#) et les bonnes pratiques associées.

## Objectifs

- Comprendre l'architecture et le rôle du format ONNX
- Savoir convertir et optimiser des modèles pour des cibles Edge
- Déployer des modèles sur des plateformes embarquées

- Exploiter ONNX Runtime et ses accélérateurs matériels
- Diagnostiquer les erreurs et surveiller les performances

## Public visé

- Ingénieurs Data / IA
- Développeurs en systèmes embarqués
- Architectes Edge Computing
- Chercheurs et makers souhaitant industrialiser un POC IA

## Pré-requis

- Bonne maîtrise du Python et de la manipulation de modèles ML
- Connaissances de base en déploiement applicatif (Linux embarqué)
- Expérience préalable avec PyTorch ou TensorFlow recommandée

## Programme de formation ONNX Edge AI Deploy

### Comprendre ONNX et le cadre de l'Edge AI

- Introduction à ONNX : objectifs, architecture, interopérabilité
- Panorama des frameworks compatibles : PyTorch, TensorFlow, etc.
- Concepts clés de l'Edge AI : contraintes matérielles et optimisation
- Avantages du format ONNX pour le déploiement cross-plateforme
- Études de cas : usage de modèles ONNX en production
- Atelier : Conversion d'un modèle PyTorch vers le format ONNX

### Explorer le format et les outils ONNX

- Structure d'un fichier modèle ONNX
- Utilisation d'ONNX Checker pour valider les modèles
- Visualisation graphique avec Netron
- Debug et compatibilité inter-frameworks
- Limites actuelles du format ONNX

### Optimiser les modèles pour les environnements Edge

- Notions de quantization dynamique et statique
- Réduction des dimensions : pruning et compression
- Exploitation de ONNX Runtime pour le Edge
- Accélération matérielle : CPU, GPU, NPU
- Atelier : Optimisation d'un modèle ONNX avec quantization et test de performance

### Déployer avec ONNX Runtime sur plateforme Edge

- Présentation de ONNX Runtime : versions, options d'exécution
- Configuration sur Jetson Nano, Raspberry Pi, etc.
- Installation et intégration dans un projet Python
- Benchmarks de latence et consommation mémoire
- Gestion des erreurs d'exécution

## Créer une pipeline Edge IA complète

- Chaînage : prétraitement ? modèle ONNX ? post-traitement
- Utilisation de capteurs et caméra en temps réel
- Exécution embarquée autonome ou connectée
- Déclenchement de traitements conditionnels
- Atelier : Déploiement d'une application Edge IA complète avec ONNX Runtime

## Surveiller, mettre à jour et industrialiser

- Surveillance des performances sur dispositif Edge
- Déploiement via conteneurs Docker ou scripts automatisés
- Stratégies de mise à jour de modèles embarqués
- Considérations de sécurité et robustesse
- Checklist de déploiement Edge IA en production

## Sociétés concernées

Cette formation s'adresse à la fois aux particuliers ainsi qu'aux entreprises, petites ou grandes, souhaitant former ses équipes à une nouvelle technologie informatique avancée ou bien à acquérir des connaissances métiers spécifiques ou des méthodes modernes.

## Positionnement à l'entrée en formation

Le positionnement à l'entrée en formation respecte les critères qualité Qualiopi. Dès son inscription définitive, l'apprenant reçoit un questionnaire d'auto-évaluation nous permettant d'apprécier son niveau estimé sur différents types de technologies, ses attentes et objectifs personnels quant à la formation à venir, dans les limites imposées par le format sélectionné. Ce questionnaire nous permet également d'anticiper certaines difficultés de connexion ou de sécurité interne en entreprise (intraentreprise ou classe virtuelle) qui pourraient être problématiques pour le suivi et le bon déroulement de la session de formation.

## Méthodes pédagogiques

Stage Pratique : 60% Pratique, 40% Théorie. Support de la formation distribué au format numérique à tous les participants.

## Organisation

Le cours alterne les apports théoriques du formateur soutenus par des exemples et des séances de réflexions, et de travail en groupe.

## Validation

À la fin de la session, un questionnaire à choix multiples permet de vérifier l'acquisition correcte des compétences.

## Sanction

Une attestation sera remise à chaque stagiaire qui aura suivi la totalité de la formation.