

## I. Introduction

- Présentation de l'IA et des FPGA
- Importance de l'IA et des FPGA dans le monde technologique actuel

## II. Fondamentaux de l'IA et des FPGA

- Qu'est-ce que l'IA? Vue d'ensemble et applications principales
- Qu'est-ce qu'un FPGA? Fonctionnement et avantages par rapport aux circuits intégrés traditionnels

## III. Pourquoi le FPGA est idéal pour l'IA?

- Flexibilité et reconfigurabilité des FPGA
- Haute performance et efficacité énergétique
- Capacité à accélérer les algorithmes d'IA

## IV. Les différentes entreprises et leurs projets utilisant l'IA avec le FPGA

- Exemples d'entreprises leaders (Xilinx, Intel, etc.)
- Étude de cas : Projets spécifiques et applications de l'IA sur FPGA
- Collaborations entre entreprises technologiques et secteurs industriels

## V. Défis et limitations

- Complexité de la programmation des FPGA
- Coût et accessibilité
- Perspectives sur les solutions aux défis existants

## VI. Les métiers émergents et les compétences requises

- Ingénieur en conception de FPGA
- Développeur algorithmes d'IA spécialisé en FPGA
- Architecte de systèmes embarqués
- Expert en optimisation de performances pour l'IA sur FPGA
- Formation et compétences requises pour ces métiers

# I. Introduction

Aujourd'hui, nous explorons deux révolutions technologiques qui transforment notre façon de concevoir et d'interagir avec le monde numérique : l'Intelligence Artificielle (IA) et les Field-Programmable Gate Arrays (FPGA). L'IA, avec sa capacité à apprendre, analyser et prendre des décisions, s'est imposée comme un moteur d'innovation dans des domaines allant de la santé à l'automobile. Parallèlement, les FPGA offrent une flexibilité sans précédent dans le hardware, permettant aux concepteurs de créer des circuits personnalisés pour accélérer spécifiquement les tâches d'IA.

Cette synergie entre l'IA et les FPGA ouvre des portes vers des performances informatiques inégalées, une efficacité énergétique et une personnalisation à la volée qui étaient inimaginables il y a seulement quelques décennies. Ensemble, elles forment le socle de la prochaine génération de technologies, capables de répondre aux exigences croissantes de notre société connectée.

Dans les pages suivantes, nous allons détailler comment l'IA et les FPGA fonctionnent, pourquoi leur combinaison est si puissante, et explorer les entreprises et projets pionniers qui mènent cette avancée.

## II. Fondamentaux de l'IA et des FPGA

### Qu'est-ce que l'IA? Vue d'ensemble et applications principales

Aujourd'hui on est forcé de constater que nous générons énormément de données et nous pensons très fortement que de ces données nous sommes capables d'en générer de la valeur en prenant des décisions. Le processus de prise de décision consiste tout d'abord à traiter ces données pour en extraire de l'information et ensuite d'agglomérer ses informations pour en créer des connaissances et basé sur ses connaissances nous allons être en mesure de prendre des décisions plus ses décisions vont reposer sur des connaissances qui s'appuie sur des informations couvrant l'ensemble des données à disposition plus cette décision pourra être pertinente et généré de la valeur.



Figure 1 : Illustration de l'utilisation de l'IA

## Qu'est-ce qu'un FPGA? Fonctionnement et avantages par rapport aux circuits intégrés traditionnels

Le changement induit par l'avènement du machine learning (et en particulier de sa branche du deep learning) est que les modèles utilisés ont des besoins qui diffèrent des programmes habituels. Afin d'être entraînés plus efficacement, ils ont besoin d'ordinateurs effectuant de nombreux calculs en parallèle notamment pour les réseaux neuronaux, puisqu'il faut calculer en même temps l'état de nombreux neurones), alors qu'un CPU est un processeur généraliste adepte du calcul séquentiel.

Les FPGA (Field-Programmable Gate Arrays) sont un type de circuit intégré qui peut être reconfiguré après sa production. Les FPGA existent depuis quelques décennies et sont notamment utilisés pour prototyper d'autres processeurs. Mais ils ont aussi un avantage-clé pour les chercheurs en IA. En premier lieu, ils peuvent être personnalisés pour s'adapter à chaque modèle d'IA testé. Ils sont également plus puissants que les GPU.

Les transistors dans les puces électroniques sont organisés en portes logiques qui exécutent des fonctions spécifiques. Les opérations effectuées par les processeurs se déroulent grâce à un flux électrique traversant ces portes logiques, permettant ainsi l'exécution de calculs. Les FPGA, ou Field-Programmable Gate Arrays, comprennent une multitude de ces portes logiques, organisées en une grille programmable et reprogrammable selon les besoins pour réaliser différentes fonctions.

Contrairement aux puces standards telles que les microprocesseurs (CPU) de nos ordinateurs, les processeurs graphiques (GPU), les systèmes sur puce (SoC), et d'autres puces spécialisées (ASIC, etc.), dont la structure et l'organisation sont permanentes, les puces FPGA offrent une flexibilité remarquable.

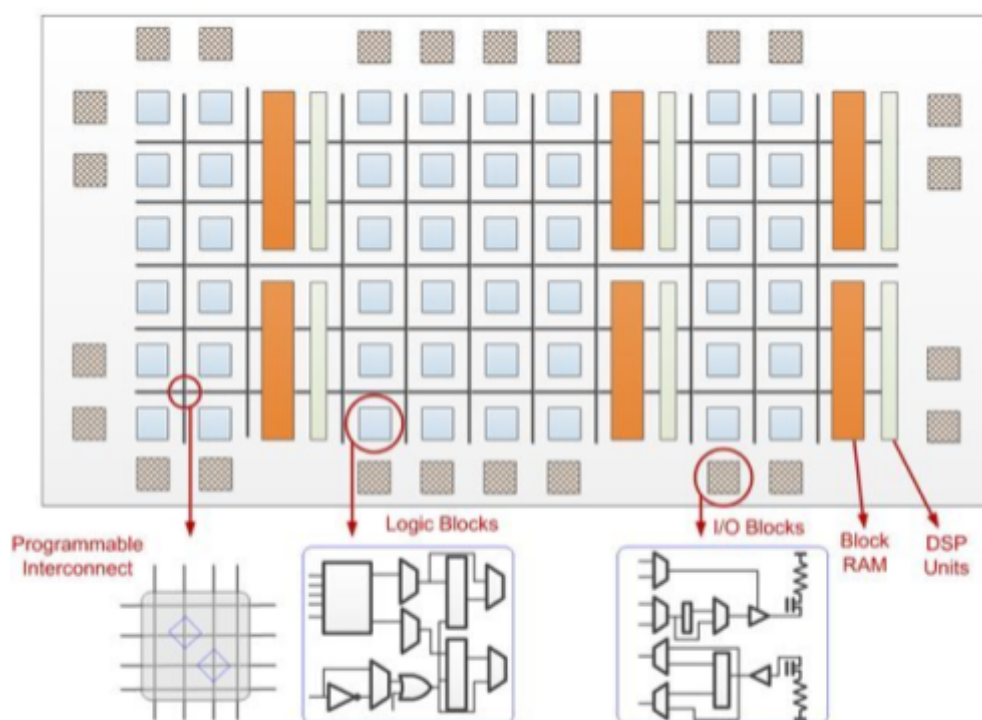


Figure 2 : Illustration de la programmation d'un FPGA

Elles se distinguent par leur capacité à être programmées, offrant ainsi une adaptabilité sans précédent. Les FPGA ne se limitent pas à un nombre fixe de transistors ou d'unités de calcul; elles contiennent des cellules logiques que les développeurs peuvent configurer presque librement, ouvrant la voie à une vaste gamme d'applications et d'usages.

### III. Pourquoi le FPGA est idéal pour l'IA?

**Flexibilité :** Les FPGAs sont hautement configurables, ce qui permet aux ingénieurs de personnaliser le matériel pour les besoins spécifiques d'un algorithme d'IA, offrant une adéquation presque parfaite entre le matériel et l'algorithme.

**Efficacité énergétique :** Les FPGAs sont plus efficaces en termes d'énergie par rapport aux processeurs traditionnels lorsqu'ils exécutent des tâches d'IA spécifiques, ce qui est crucial pour les systèmes embarqués et les applications à faible consommation d'énergie.

**Accélération des performances :** Ils peuvent accélérer les calculs nécessaires pour l'IA, notamment en traitant plusieurs opérations en parallèle, ce qui réduit le temps nécessaire pour l'inférence et l'apprentissage.

**Latence réduite :** Les FPGAs peuvent réduire la latence, un avantage important pour les applications d'IA en temps réel comme la reconnaissance vocale et le traitement d'images.

**Adaptabilité :** L'IA évoluant rapidement, les FPGAs permettent une mise à jour facile du matériel pour s'adapter aux nouveaux algorithmes et modèles sans avoir besoin de remplacer le matériel existant.

Les soc :



Figure 3 : Soc Stratix 10 Intel

Le Stratix 10 d'Intel Altera est une carte FPGA haut de gamme conçue pour des performances élevées, offrant des capacités avancées pour le traitement des données et des applications d'IA, grâce notamment à la prise en charge de la mémoire HBM2. Elle est idéale pour des charges de travail complexes nécessitant une grande vitesse et une efficacité énergétique.



Figure 4 : Soc Virtex Ultrascale+ de AMD Xilinx

La série Virtex Ultrascale+ de AMD Xilinx, quant à elle, est reconnue pour son efficacité énergétique et ses performances exceptionnelles dans le traitement en temps réel et les applications nécessitant de grandes quantités de données, la rendant parfaite pour l'IA et le traitement de données volumineuses.

Les cartes FPGA comme le Stratix 10 d'Intel Altera et la série Virtex UltraScale+ de AMD Xilinx peuvent être utilisées dans diverses applications avancées. Par exemple, elles peuvent accélérer le traitement des données pour l'intelligence artificielle, comme le deep learning et l'analyse prédictive, gérer des charges de travail complexes dans les data centers, permettre des calculs haute performance pour la recherche scientifique, et soutenir les systèmes embarqués dans l'automobile pour la conduite autonome.

Ces 2 constructeurs de FPGA offrent également des solutions diverses dans plusieurs domaines en créant des PCB avec de multiples entrées et sorties que ce soit dans la radio et communication avec les ports SMA ou encore des ports pour la fibre optique.

#### Virtex UltraScale+ Boards, Kits, and Modules



Figure 5 : Exemples de Evaluation Board que crée Xilinx pour ses soc

## IV. Les différentes entreprises et leurs projets utilisant l'IA avec le FPGA

### Exemples d'entreprises leaders (Xilinx, Intel, etc.)

Les deux leaders sur le marché des FPGA sont Xilinx et Altera. Ce dernier a été racheté par Intel en 2015 pour 17 milliards de dollars et Xilinx par AMD en 2022 pour 50 milliard de dollars.

Grâce à cette « fusion », AMD renforce son portfolio de technologies. Déjà numéro 2 des CPU dans les PC et serveurs, numéro 2 des GPU dans les PC et serveurs, fournisseur numéro 1 des puces pour consoles de jeu, AMD devient, grâce à Xilinx, le numéro 1 mondial des FPGA.

## Microsoft

Dans son projet Brainwave, Microsoft utilise des FPGA pour accélérer l'exécution de réseaux de neurones profonds sur sa plateforme cloud Azure. Ces FPGAs sont utilisés pour offrir une performance en temps réel sur des tâches d'IA complexes.

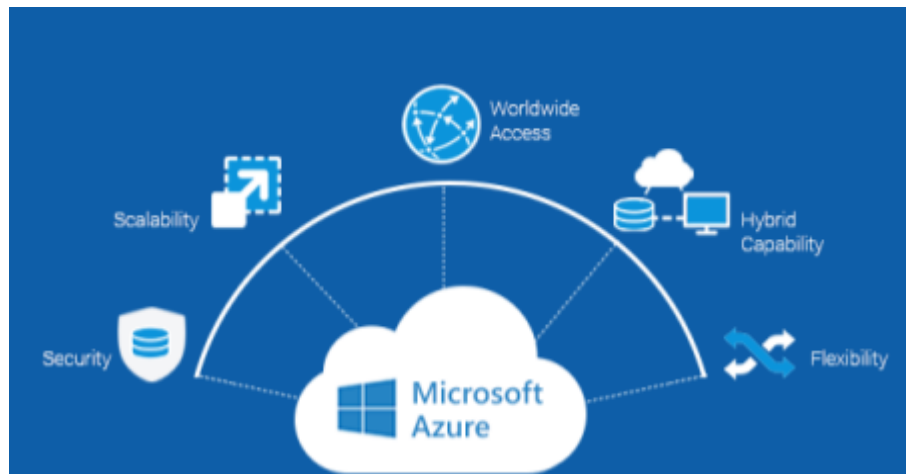


Figure 6 : Principale caractéristique de Microsoft Azure

Azure est la plateforme de cloud computing de Microsoft, offrant une gamme étendue de services pour répondre aux besoins variés des entreprises. Que ce soit pour l'hébergement de sites Web et de bases de données, le déploiement d'applications web, l'intelligence artificielle, l'apprentissage automatique ou l'IoT, Azure propose des solutions adaptées. Avec Azure, les entreprises bénéficient d'une flexibilité accrue grâce à des services pay-as-you-go, ne payant que pour les ressources qu'elles utilisent réellement.

La plateforme cloud Azure c'est plus de 200 produits et services cloud, open source avec la prise en charge de l'ensemble des langages et frameworks. C'est aussi 1 milliards d'investissement par an en terme de sécurité pour protéger leurs clients contre des cybermenaces.



Figure 7 : Services et plateforme que propose Microsoft Azure



Azure est utilisé dans plusieurs domaines tels que la santé, les services financiers, le secteur public, la distribution et l'industrie.

Azure permet un contrôle complet sur l'environnement informatique grâce à la possibilité de créer et de gérer des machines virtuelles dans le cloud. La sécurité est également une priorité, avec des services de stockage cloud sécurisés et des centres de données régionaux permettant une distribution mondiale des applications.

## Xilinx

Xilinx développe des plates-formes et des cartes spécialement conçues pour l'IA, comme la série Alveo, qui sont optimisées pour les charges de travail d'IA et de Machine Learning et utilisées dans les services financiers, l'Internet des objets (IoT), les médias et le divertissement lancé en 2018 dans une conférence de presse.



Figure 8 : Conférence de presse lancement FPGA Alveo



Fast

Highest Performance

Grâce à l'utilisation de la carte accélératrice Alveo, les performances sont jusqu'à 90 fois supérieures à celles des CPU (Exemple : EC2 c4.8xlarge) sur des charges de travail clés, tout en maintenant un coût trois fois inférieur aux serveurs équipés d'Intel Xeon Platinum.

Comparé aux solutions basées sur le GPU, la carte Alveo offre un débit d'inférence 4 fois plus élevé et une latence 3 fois inférieure, ce qui en fait un choix optimal pour des applications nécessitant une grande efficacité de traitement.



### Adaptable

#### Accelerate Any Workload

L'utilisation de l'apprentissage automatique, le traitement vidéo et d'autres charges de travail peuvent être exécutés de manière optimale en utilisant la même carte accélératrice, offrant ainsi une solution polyvalente pour diverses tâches.

En utilisant du matériel reconfigurable, il est possible de s'adapter plus rapidement aux évolutions des algorithmes de charge de travail que les cycles de production des cartes d'accélération à fonction fixe.



### Accessible

#### Cloud ↔ On-Premises Mobility

Possibilité de déployer des solutions dans le cloud ou sur site de manière interchangeable, en fonction des besoins de l'application, la flexibilité est optimisée pour répondre aux exigences changeantes.

Des applications prêtes à l'emploi sont disponibles pour les charges de travail courantes, mais il est également possible de créer des applications personnalisées grâce à l'outil de développement d'applications, offrant ainsi une adaptabilité maximale aux besoins spécifiques.



Figure 9 : Les différentes utilisations de la série Alveo



AMD et ses partenaires offrent une gamme variée de solutions pour différentes charges de travail. Que ce soit dans le stockage computationnel, les bases de données et l'analytique de données, la technologie financière, le calcul haute performance, l'accélération réseau, la vidéo et l'imagerie, l'apprentissage automatique ou encore les outils et services, AMD s'engage à fournir des solutions innovantes pour répondre aux besoins diversifiés de ses clients. Ces solutions sont conçues pour optimiser les performances, améliorer l'efficacité opérationnelle et offrir des capacités de traitement avancées, tout en assurant une qualité de service et un accompagnement personnalisé à chaque étape du processus.

Comme on peut le voir ci-dessous les cartes Alveo rangées dans des racks de datacenter pour permettre une puissante puissance de calcul.



Figure 10 : Installation des cartes Alveo dans des racks de datacenter

## VI. Les métiers émergents et les compétences requises

### **Ingénieur FPGA**

Un ingénieur FPGA, avec un salaire annuel moyen de 42 000 € est généralement détenteur d'un diplôme de niveau bac+5 en électronique numérique, en systèmes électroniques ou en électronique, peut trouver de nombreuses applications de l'IA dans son domaine.

- Diplôme d'ingénieurs en électronique numérique
- Diplôme d'ingénieur du spécialité systèmes électroniques
- Master en électronique

Avec son expertise en électronique et la capacité à concevoir et à programmer des FPGA, un ingénieur peut contribuer à développer des systèmes embarqués qui intègrent des fonctionnalités d'IA. Par exemple, il pourrait participer à la conception de circuits spécialisés pour accélérer des algorithmes d'apprentissage automatique sur des dispositifs embarqués, permettant ainsi des applications telles que la détection d'objets, la reconnaissance vocale ou la surveillance intelligente contrairement au CPU qui fonctionne de façon séquentiel, donc un vrai gain de performance et de temps.

### **Ingénieur en systèmes embarqués**

Un ingénieur en systèmes embarqués, avec un salaire annuel moyen de 39 900 € et titulaire d'un diplôme d'ingénieur spécialisé en électronique, en systèmes électriques et électroniques embarqués, ou d'un master en informatique ou en ingénierie des systèmes complexes, possède un bagage solide pour intégrer l'IA dans ses domaines d'expertise.

- Diplôme d'ingénieur spécialisé en( électronique, systèmes électriques et électroniques embarqués etc ...)
- Master mention informatique
- Master ingénierie des systèmes complexes

Les systèmes embarqués sont omniprésents dans notre monde dans l'automobile, la sécurité avec les capteurs, les performances et la connectivité . L'IA peut être intégrée dans ces systèmes pour améliorer leurs fonctionnalités, leur efficacité et leur capacité à interagir de manière intelligente avec leur environnement.

## Sources :

<https://www.cairn.info/revue-flux-2023-3-page-24.htm>

<https://hypertec.com/blog/fr/pourquoi-il-vous-faut-un-circuit-integre-prediffuse-programmable-fpga-dans-votre-nouveau-serveur/>

<https://www.xilinx.com/products/boards-and-kits/alveo.html>

<https://docs.xilinx.com/v/u/en-US/wp504-accel-dnns>

<https://azure.microsoft.com/fr-fr/resources/cloud-computing-dictionary/what-is-azure#:~:text=Les%20partenaires%20informatiques%20utilisent%20Azure,Azure%20pour%20atteindre%20ses%20objectifs.>

<https://www.elsys-design.com/fr/ingenieur-fpga/#:~:text=Pour%20devenir%20ing%C3%A9nieur%20FPGA%2C%20il,une%20sp%C3%A9cialisation%20en%20%C3%A9lectronique%20num%C3%A9rique.>