

## Mise en Cohérence des Objectifs du TIPE (MCOT)

**Optimisation des performances d'un réseau de neurones artificiels par l'usage d'un réseau de portes logiques**

Le sujet choisi porte sur deux thématiques informatique l'une logicielle autour de l'intelligence artificielle et l'autre logicielle autour des réseau de portes logiques programmables. En effet l'intelligence artificielle est une technologie en pleine explosion et la question du matériel informatique et électronique utilisé est une problématique majeure tant écologiquement, qu'économiquement ou qu'en terme de performances des modèles. Ce sujet m'intéresse depuis très longtemps et ce projet était un bon moyen d'accroître mes connaissances autour du sujet. Cette étude multifactorielle à pour objectif d'avancer des pistes afin de déterminer quelle technologie de puce logique est la plus efficace dans tout les domaines sus-cités pour faire tourner un modèle primitif de reconnaissance faciale en utilisant des composants grands publics.

**Positionnement thématique :**

Informatique (Intelligence Artificielle), Électronique (Général), Mathématiques(Statistiques)

**Mots-clés :**

- Réseau de neurones
- Processeur d'unité centrale
- Réseau de portes programmables in situ
- Reconnaissance Faciale
- régression linéaire multiple

**Bibliographie commentée**

Ces dernières années ont vu l'explosion des performances des modèles d'intelligence artificiel, le principal d'entre eux Générative Trained Transformer (GPT) passant en 2018 d'un peu plus de 100 millions de parametres à plus de 1000 milliards aujourd'hui [1] l'impact écologique de l'intelligence artificielle est aussi en pleine expansion, on estime que chaque entraînement d'un modèle comme ceux que nous avons aujourd'hui émet 300 tonnes de CO<sub>2</sub> [2] soit l'objectif d'émission annuel de 150 personnes tel que décrit dans les accords de paris [3] de plus l'IA a aussi un impact

### Mise en Cohérence des Objectifs du TIPE (MCOT)

économique important, l'installation de nouveaux datacenters pour l'IA devrait coûter en 2030 presque 50 milliard de dollars américains par an. [4] en témoigne aussi le triplement en une année du cours de l'action de NVIDIA [5] principal fabricant de puce pour l'IA qui est devenue l'entreprise la plus valorisée en bourse dans le monde dépassant les 3 500 milliards de dollars de valorisation [6] signe que l'enjeu autour des puces pour l'intelligence artificielle est plus que majeur. C'est pourquoi nous sommes en droit de nous demander si certaines de nos technologies actuelles ne pourraient pas être des pistes sérieuses dans le développement des technologies autour de l'intelligence artificielle. L'une d'entre elles est le Réseau de portes programmables in situ (FPGA en anglais) ceux-ci ont de nombreux avantages et sont déjà utilisés dans de nombreux domaines comme les télécommunications où ils sont des composants essentiels des routeurs en raison de leur très grande vitesse de traitement d'opérations simples comparée à un processeur. De plus leur très faible consommation pourraient en faire des composants plus durables et leur reprogrammabilité leur offre une grande flexibilité[7], c'est une piste que certaines grandes entreprises du domaine comme Intel ou Microsoft, ont commencées à explorer. [7][8] La principale problématique aujourd'hui pour l'utilisation de cette technologie pour l'IA est l'absence quasi totale d'outils permettant la mise en place d'un modèle sur ces composants, problématique qu'Intel a commencé à résoudre avec son logiciel FPGA AI Suite [9], qui reste néanmoins compatible avec très peu de FPGA aujourd'hui.

### Problématique Retenue

Il s'agit d'étudier les performances d'un modèle primitif de reconnaissance faciale sur deux types de cartes. Un FPGA, le Altera Cyclone IV de Intel et un Microprocesseur, le Cortex-A72 équipé sur un micro-ordinateur Raspberry-Pi 4.

### Objectifs du TIPE du candidat

Tout au long de ce projet, je vais :

- Développer et entraîner un modèle d'intelligence artificielle en Python à l'aide du module Keras de TensorFlow de Google
- Compiler le modèle sur la carte FPGA à l'aide des outils limités à ma disposition
- Mesurer et comparer le temps de réponse, la consommation électrique et la précision des réponses du modèle sur les deux cartes
- Modéliser ces mesures pour une échelle internationale et industrielle des Intelligences Artificielles

## Mise en Cohérence des Objectifs du TIPE (MCOT)

**Références Bibliographiques**

- [1] [Makeuseof.com:GPT-1 to GPT-4: Each of OpenAI's GPT Models Explained and Compared](#)
- [2] [University of Massachusetts Amherst: Energy and Policy Considerations for Modern Deep Learning Research](#)
- [3] [Rapport Du GIEC prévoyant un objectif pour 2050 de 20 Gigatonnes de CO<sub>2</sub> en estimant une population de 10 Milliards d'Humains](#)
- [4] [Rapport de la banque de Rotschild sur le coût de l'évolution du numérique](#)
- [5] [Google Finance: NASDAQ NVDA cours de l'action sur 1 an](#)
- [6] [Journaldunet.com : Nvidia dépasse Apple et devient l'entreprise la plus valorisée au monde](#)
- [7] [Intel : FPGA vs GPU pour le Deep Learning](#)
- [8] [FPGAs de l'intelligence artificielle \(IA\) | Altera®](#)
- [9] [Intel: FPGA AI Suite](#)

Retrouvez les lien Hypertextes des Références Bibliographiques en scanant ce qrcode ou sur <https://mcot.elyotverger.com>

