**CONSERVATOIRE NATIONAL DES ARTS ET METIERS**

CENTRE REGIONAL DE TOULOUSE

**MÉMOIRE PROBATOIRE**

présenté en vue d’obtenir

l’UE ENG221 – Information et communication pour l’ingénieur

Spécialité : Informatique

Option : Informatique Modélisation Optimisation

**Par**

**Nathan TENIERES**

**Apport de l'intelligence artificielle**

**pour**

**améliorer la fiabilisation des transformations de données**

Session 2022 – 2023

Jury :

REMERCIEMENTS

SOMMAIRE

**TABLE DES FIGURES & TABLEAU 10**

**INTRODUCTION 11**

1. L’INTELLIGENCE ARTIFICIELLE 12

*Résumé du chapitre : Dans ce chapitre, je vais faire une présentation de l’intelligence artificielle. Présenter et illustrer les trois types d’IA : IA, Machine Learning et Deep Learning avec pour chacun d’eux une présentation sommaire des principaux algorithmes. Je finirai par donner des exemples d’application où l’IA a été utilisé (Alexa, robot, voiture autonome…). Ce chapitre est une entrée en matière sur les bases de l’IA.*

1.1 Présentation historique12

1.2 Les trois familles d’IA12

1.2.1 Intelligence artificielle étroite (ANI)14

1.2.2 Intelligence artificielle générale (AGI)14

1.2.3 Superintelligence artificielle (ASI)14

1.3 Distinction Intelligence Artificielle, Machine Learning et Deep Learning13

1.3.1 Programmation symbolique

1.3.2 Machine Learning

1.3.2.1 Les modes d’apprentissages et leurs algorithmes

*=> supervise, non supervisé, semi-supervisé et par renforcement*

1.3.3 Deep Learning

1.3.3.1 Les réseaux de neurones (artificielle, récurrent et convolutif)

1.4 Domaines d’application13

Transition : rebondir sur un des domaines d’applications, on ne peut pas avoir une IA efficace et pertinente sans avoir une certaine qualité de donnée.

1. FIABILITÉ & QUALITÉ DE LA DONNÉE16

*Résumé du chapitre : Au sein de ce chapitre, j’axerai donc mon étude sur la fiabilité et la qualité de la donnée. Nous ne nous pouvons pas avoir une IA efficace sans une certaine fiabilité et qualité de la donnée. Ainsi, aux prémices de ce chapitre, j’expliquerai le cycle de vie de la donnée au sein des organisations (génération, collecte, stockage, traitement, consommation…).*

*J’étudierai ensuite le processus de fiabilisation de la donnée. Ce processus désigne le mécanisme de contrôle et d’amélioration permanente de la qualité des données d’un système d’informations. La fiabilité des données désigne des données complètes et exactes. Une donnée fiable, qu’est-ce que c’est ? J’exposerai également les méthodes afin de mesurer et d’assurer la cohérence des données.*

*Nous en venons ensuite au paragraphe abordant la qualité de la donnée au sein duquel j’en donnerai une définition, j’exposerai les causes de dégradation de la donnée, les méthodes et les outils pour nous permettre de calculer cette qualité (SLO/SLI/SLA), et je parlerai également de la norme ISO 8000 :61 qui régit la qualité de la donnée.*

2.1 Cycle de vie de la donnée12

2.2 Fiabilité12

2.2.1 Processus de fiabilisation12

2.2.2 Les méthodes de cohérence12

*=> test-retest, inter-juges, parallèle, cohérence interne*

2.3 Validité ???

* *La validité des données mesure si les données mesurent ce qu’elles sont censées mesurer, tandis que la fiabilité mesure la cohérence des mesures. Ces deux concepts sont donc importants afin de déterminer la qualité des données.*

2.4 Qualité12

2.4.1 Définition12

2.4.2 Les causes de dégradation12

2.4.3 Data SLA/SLI/SLO12

2.4.4 Norme ISO 8000-61:201612

Transition : IA pour améliorer la transformation des données pour une meilleure fiabilisation et qualité, plus officient que l’homme

1. L’IA POUR UNE MEILLEURE TRANSFORMATION DES DONNÉES17

*Résumé du chapitre : La transformation des données est le processus de conversion de données d’un format, d’une structure ou d’une représentation à une autre. Dans ce chapitre nous verrons 6 apports de l’IA dans la transformation des données :*

* ***Nettoyage de données****: avec l’apprentissage automatique pour identifier les erreurs de données (données manquantes, valeurs aberrantes, doublons…). Utilisation d’algorithme de clustering pour identifier les doublons, et les réseaux de neurones pour prédire les valeurs manquantes*
* ***Normalisation de données****: algo de normalisation pour transformer les données à une échelle commune. => techniques d’apprentissage automatique telles que la régression linéaire*
* ***Combinaison des données*** *en provenance de différentes sources : algorithme de Matching pour identifier les correspondances entre les données de différentes sources. Les réseaux de neurones peuvent être utilisé pour prédire les valeurs manquantes en utilisant les données disponibles.*
* ***Réduction de données****: techniques d’apprentissage automatique, telles que la réduction de dimensionalité, peuvent être utilisées pour réduire la taille des données en conservant les caractéristiques les plus importants. => Algo de clustering*
* ***Conversion de données****: algo de traitement du langage naturel (PNL) peuvent être utiliser pour transformer les données textuelles en une représentation numérique. => Algo de PNL*
* ***Reconnaissance d’images****: la vision par ordinateur peut être utiliser pour extraire des informations à partir d’images ; détection d’objets ou la segmentation d’images. Réseaux de neurone convolutif, algorithme de reconnaissance faciale pour extraire des informations à partir d’images de visage et les convertir en un format structuré.*

*Pour résumer, nous avons* ***4 techniques d’IA*** *courantes utilisées pour la transformation de données :* ***l’apprentissage automatique, le traitement du langage naturel (PNL), la vision par ordinateur et les réseaux de neurones.***

3.1 Définition et plan d’attaque12

3.2 Les six apports de l’IA pour une transformation des données efficiente12

3.2.1 Nettoyage de données (data cleaning)12

3.2.1.1 Les trois types d’anomalies (ponctuelles, contextuelles, collectives)12

3.2.1.2 Les méthodes de détection12

*=> distinction valeur aberrante/ atypique*

*=> Algo IA pertinent : boxplot, régression linéaire, plus proches voisins, clustering, fôret d’isolation (isolation forest),*

3.2.2 Normalisation de données12

3.2.3 Combinaison de données12

3.2.4 Conversion de données12

3.2.5 Reconnaissance d’images12

1. CONCLUSION40

* *Amélioration de la fiabilisation des transformations de données grâce à l’IA*
* *Meilleure fiabilité == meilleure qualité des données == IA plus pertinente*

**GLOSSAIRE 43**

**BIBLIOGRAPHIE 44**

**ANNEXES 45**

INTRODUCTION

1. L’INTELLIGENCE ARTIFICIELLE

*Ce chapitre a pour objectif d’introduire l’intelligence artificielle (IA). Une présentation historique succincte en sera donnée afin d’en comprendre ses origines et ses évolutions. Dans cette continuité, une explication et une distinction entre les trois familles d’IA feront l’objet des points suivants. Ce paragraphe aboutira avec des exemples d’utilisations où l’IA a été mise en œuvre.*

* 1. Présentation historique

L’intelligence artificielle (IA) est un domaine de l’informatique qui se concentre sur la création de machines capables d’effectuer des tâches qui, autrement, nécessiteraient une intelligence humaine. Cette pratique permet à l’Homme de mettre un système informatique sur la résolution de problématiques complexes intégrant de la logique.

**L’intelligence artificielle avant 2000**

L’histoire de l’IA remonte aux années 1950, lorsque les scientifiques et les ingénieures ont commencé à explorer la possibilité de créer des machines capables de penser et d’apprendre comme des êtres humains. Les premiers travaux ont été menés par des chercheurs tels que Alan Turing, qui a développé le concept de la machine de Turing, considéré comme le premier modèle théorique d’un ordinateur. En découlera le Test de Turing qui mesure la capacité d’une machine à tenir une conversation humaine.

Dans ces mêmes années, un groupe de chercheur américain dirigé par John McCartly et Marvin Minsky, a lancé un le projet de recherche en intelligence artificielle au MIT. Leur objectif était de développer des programmes informatiques capables de simuler des fonctions cognitives telles que la reconnaissance de la parole, la compréhension du langage naturel et la résolution de problèmes complexes.

**Hiver de l’IA**

En 1974 arrive une période appelée « hiver de l’IA ». Elle fait référence à une période de stagnation dans le développement de l’intelligence artificielle, caractérisées par une réduction du financement et de l’intérêt pour la recherche dans ce domaine. Cependant, cette période de stagnation a été suivie par une période de progrès rapide. En effet, dans les années 80, les systèmes experts (ordinateur capable de se comporter comme un humain) connaissent un grand succès et sont à l’origine de la relance des projets de rechercher en IA. A la suite de quoi, les gouvernements soutiennent financièrement les projets académiques. Le marché de l’IA atteint un milliard de dollars.

**Années 1990 et 2000**

Dans les années 1990 et 200, l’IA a connu une renaissance grâce à l’augmentation de la puissance de calcul, l’amélioration de algorithmes d’apprentissage automatique et l’essor d’internet. Les applications de l’IA se sont multipliées dans de nombreux domaines, tels que la reconnaissance d’image et vocale, la traduction automatique, la robotique et les véhicules. En 1997, l’IA « Deep Blue » créé par IBM a battu le champion du monde d’échec Garry Kasparov.

**2000 – 2010 : l’IA représente un enjeu de société**