INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN

Projet de Physique P6

STPI/P6/2024 – 01

IA et Fausses Images

Etudiants :

Quentin FOURNIER **Jean HALLOT**

Axel MEILLERAIS **Raphaël SENELLART**

Salomé PUIG **Mohamed-Rayen BEN ABDERRAHMANE**

Enseignant-responsable du projet :

Abdelaziz BENSRHAIR

Cette page est laissée intentionnellement vierge.

Date de remise du rapport : ***14/06/2024***

Référence du projet : ***STPI/P6/2024 – 01***

Intitulé du projet : ***IA et Fausses Images***

Type de projet : ***Bibliographique/Etat de l'Art***

Objectifs du projet :

L'objectif de ce projet est de comprendre le fonctionnement de l'intelligence artificielle et plus particulièrement la création des images par ces IA. Une fois ces connaissances acquises, un autre objectif est d'étudier les nouveaux enjeux et les défis que peuvent susciter ces images notamment d'un point de vue géopolitique. Le constat après une telle mise en lumière de cette technologie est sans appel : l'IA impacte déjà et va impacter fortement notre société.

Mots-clefs du projet : ***intelligence artificielle, fausses images, deepfakes***

TABLE DES MATIERES

INDEX DES FIGURES :

NOTATIONS, ACRONYMES

IA : Intelligence Artificielle

GPU : Graphics Processing Units

GAN : Generative Adversarial Networks

CNN : Convolutional Neural Network

1. INTRODUCTION

Depuis la nuit des temps, l'homme cherche à créer de ses propres mains des nouvelles formes d'intelligence. En 1950, Alan Turing introduit au monde le terme d' "intelligence artificielle" avec l'arrivée importante de l'informatique dans le monde moderne. Les recherches sur l'IA ont débuté juste après et les informaticiens d'aujourd'hui continuent de créer, optimiser et inventer de nouveaux systèmes d'IA. Dans son histoire, les recherches ont connu des hauts et des bas qu'on appelle les âges d'or et les hivers de l'IA.

Dans les années 60, les progrès furent remarquables mais limités dus au manque de puissance de calculs qu'offraient les ordinateurs de l'époque. Les années 70 ont permis à l'IA de s'introduire dans différents domaines tels que la médecine, la finance ou encore la gestion des entreprises. En 1982, les réseaux de neurones commencent à faire leur apparition. Dans les années 90, un changement important de mentalité au sein de la communauté scientifique accentue les recherches sur la robotique afin d'avoir une IA ayant conscience de son corps. Le début du XXIe siècle laisse apparaître des algorithmes génétiques qui se rapprochent plus de la théorie d'évolution naturelle.

Aujourd'hui, les IA sont encore plus performantes et nous permettent de produire certains résultats inimaginables auparavant. C'est le cas notamment des *deepfake* constitués de fausses images générées par une IA. L'IA en général est devenue tellement performante qu'il faut aujourd'hui s'en méfier.

Nous allons évoquer ici, selon une approche bibliographique en trois temps, ces IA qui permettent de générer des fausses images. Tout d'abord, nous étudierons le fonctionnement général de l'IA puis nous verrons dans une seconde partie comment l'IA est utilisée pour créer des fausses images et enfin, nous étudierons les défis de la détection de ces images.

2. METHODOLOGIE / ORGANISATION DU TRAVAIL

Durant l'ensemble du projet nous avons utilisé une organisation structurée et efficace dans le but de mener ce projet de la meilleure des manières.

Chaque semaine, une liste de tâches à faire était mise en œuvre principalement par la cheffe de projet qui avait été désignée dès le début de celui-ci. Le nombre de tâches était variable en fonction de l'avancée dans le projet. La répartition se faisait en fonction de l'intérêt particulier de chaque membre du groupe qui choisissait la ou les tâches qu'il s'engageait à mener à bien, tout en veillant à ce que toutes les tâches soient effectuées chaque semaine. Cependant il y avait toujours un respect de l'équité dans cette répartition afin que tous les membres de l'équipe aient la même charge de travail à fournir chaque semaine. La réalisation des tâches se faisait seul ou en groupe suivant les différentes tâches et les différentes semaines.

Toutes les semaines un point sur l'avancement rassemblait les membres de l'équipe et notre professeur encadrant. Durant les semaines, si un membre avait besoin d'aide, un groupe de discussion créé entre nous permettait de rester en contact et de s'apporter une aide mutuelle. Un document partagé en ligne avait aussi été créé dans le but de suivre l'avancement des tâches de tous les membres du groupe.

Grâce à la mise en œuvre de cette organisation, notre groupe a pu faire preuve d'autonomie et de rigueur tout au long de notre projet.

3. RESULTATS DES RECHERCHES

3.1 Fonctionnement de l'Intelligence Artificielle

3.1.1 Introduction

Une IA a pour but de simuler les activités d'un cerveau humain sur une ou plusieurs tâches. Pour cela, les informaticiens ont créé au fil du temps différents systèmes répondant à différents problèmes avec une puissance de calcul requise plus ou moins importante. Nous traitons ci-dessous trois des plus importants algorithmes.

3.1.2 Arbres de décision

L'IA la plus simple à comprendre utilise des arbres de décisions. Cette méthode n'est pas exclusive à l'IA. A partir d'un état initial, un arbre est créé à l'aide de la simulation de toutes les possibilités. L'IA n'a plus qu'à choisir le meilleur chemin pour atteindre son objectif.
Cette méthode est assez coûteuse en puissance de calcul car l'arbre peut avoir une profondeur importante. L'illustration la plus médiatique est le jeu d'échec : calculer toutes les parties d'échec possibles revient à simuler environ 10^{120} parties différentes [1].

3.1.3 IA génétique

En prenant exemple sur la nature, des algorithmes génétiques ont vu le jour. Le principe consiste à exposer plusieurs versions de l'IA devant un problème et d'isoler les meilleures versions afin de créer la génération suivante. De plus, d'une génération à une autre, certaines versions développent des mutations qui peuvent ou non les rendre meilleures. Cette méthode est très utilisée dans le cas d'une recherche d'optimisation [2].

3.1.4 Réseau neuronal

La méthode la plus répandue est le réseau neuronal. Les réseaux de neurones reproduisent une version simplifiée d'un cerveau biologique.
En informatique, un neurone au sein d'un réseau peut être représenté comme l'image suivante:

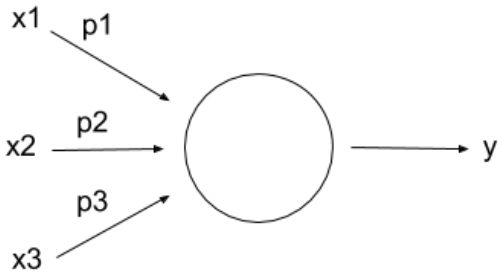

Figure 1 : Schéma de fonctionnement d'un neurone [3]

Un neurone est un nœud regroupant plusieurs entrées en une sortie. Chaque entrée x possède un poids p qui influence son importance par rapport aux autres. Le neurone effectue une opération avec toutes ces entrées pour renvoyer une nouvelle valeur y [3].

Un réseau de neurone est construit à l'aide de milliers de neurones répartis sur des dizaines de couches comme le montre l'image ci-dessous :

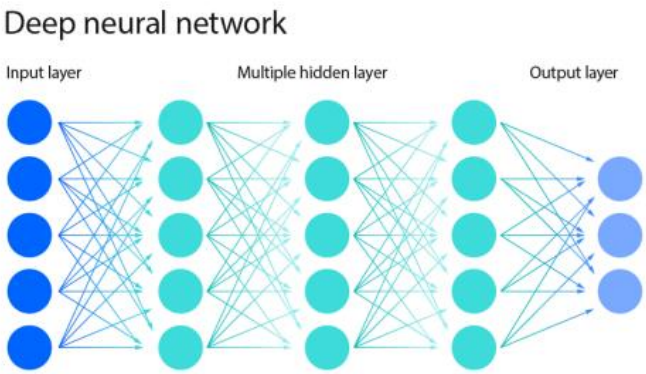

Figure 2 : Schéma de fonctionnement d'un réseau de neurones [4]

Un réseau de neurones est très complexe mais nous pouvons distinguer trois zones [4] :

- la première couche correspond aux différentes données que l'on donne à ce réseau. Il peut s'agir d'informations telles que des distances, des valeurs, des vitesses… ou d'une image avec chaque entrée traitant un pixel
- la dernière couche correspond aux différentes données résultantes du réseau.
- toutes les couches intermédiaires sont celles qui effectuent les calculs. Les calculs sont simples mais la complexité globale rend la compréhension difficile.

Avec chaque nouvel exemple ou chaque nouvel ensemble de données, le réseau de neurones ajuste les poids des différentes entrées de chaque neurone. Au fur et à mesure, notre IA devient de plus en plus compétente et précise au point d'égaler ou de dépasser une réflexion humaine [5].

3.1.5 Machine Learning

Une fois codée, aucune IA n'est utile sans un entraînement adéquat. Il s'agit du Machine Learning. Pour cela, il existe 4 types d'apprentissage [6] :

- l'apprentissage supervisé consiste à lui donner un grand nombre de données avec les réponses attendues.
- l'apprentissage non supervisé consiste à ne pas donner les réponses pour laisser l'IA créer ses propres catégories et sa propre réflexion
- l'apprentissage semi-supervisé est un mélange des deux précédents
- l'apprentissage par renforcement est quant à lui totalement différent. Il se base sur un principe de récompense et de punition. Pour atteindre un objectif, l'IA récupère des bonus ou des malus selon les chemins utilisés.

3.1.6 Conclusion

La puissance d'une IA est propre à chaque algorithme.
Pour les réseaux de neurones, la pertinence de l'IA dépend de sa base de données. Plus la base de données est importante et riche, plus l'IA sera performante.
Pour les algorithmes génétiques, le nombre de génération est le plus important. Entre autres, la performance de l'IA est proportionnelle à la durée d'entraînement de celle-ci.

Pour les algorithmes concernant des arbres de décisions, la puissance de l'IA est simplement liée à la puissance de calcul de l'ordinateur ou du serveur hôte. Plus l'IA peut aller loin dans ses simulations, plus elle sera performante [7].

Trois adjectifs qualifient les IA en fonction de la puissance :

- une IA étroite, la plus répandue aujourd'hui, n'est adaptée qu'à un certain nombre de situations.
- une IA générale est égale à un humain dans n'importe quelle situation.
- une super IA n'existe pas encore mais dépassera les capacités cognitives de n'importe quel humain.

3.2 L'IA comme outil de création de Fausses Images

3.2.1 Introduction

Sur la base du principe du fonctionnement de l'IA, les dernières évolutions peuvent générer des images, et même les mettre en mouvement et des vidéos sont ainsi créées. Jusqu'il y a quelques années, l'IA permettait de générer le scénario d'un film. Désormais, il est possible de concevoir un film entièrement par IA. L'IA crée des images en utilisant des réseaux neuronaux, comme nous l'avons vu précédemment, mais cette fois, des réseaux particuliers puisqu'il s'agit des réseaux de neurones générateurs adverses : GAN et des réseaux de neurones convolutifs : CNN.

3.2.2 Les GAN

Les GAN sont composés de deux réseaux neuronaux distincts mais en compétition : un générateur et un discriminateur. Le générateur prend en entrée des données aléatoires, appelé bruit et les code pour les transformer en données, comme pour les images. L'objectif du générateur est de créer des données réalistes donc normalement, les données transformées ressemblent à celles de l'ensemble de données d'origine. Le discriminateur prend en entrée des données, soit celles du générateur soit celles de l'ensemble des données réelles et il essaie de les distinguer. Son objectif est de déterminer si les données qu'il a reçues sont réelles ou fabriquées [8].

Comme nous l'avons dit, ces réseaux sont en compétition constante lors de l'étape de l'entrainement, ce qui correspond au processus d'apprentissage des poids et des paramètres des deux réseaux. Le générateur essaie d'améliorer sa capacité à tromper le discriminateur en produisant des données qui ressemblent de plus en plus à celles de l'ensemble de données réelles. D'un autre côté, le discriminateur essaie d'améliorer sa capacité à distinguer les données réelles des données générées. Cette compétition pousse à l'amélioration des deux réseaux et idéalement, le générateur finit par produire des données indiscernables des données réelles pour le discriminateur [9].

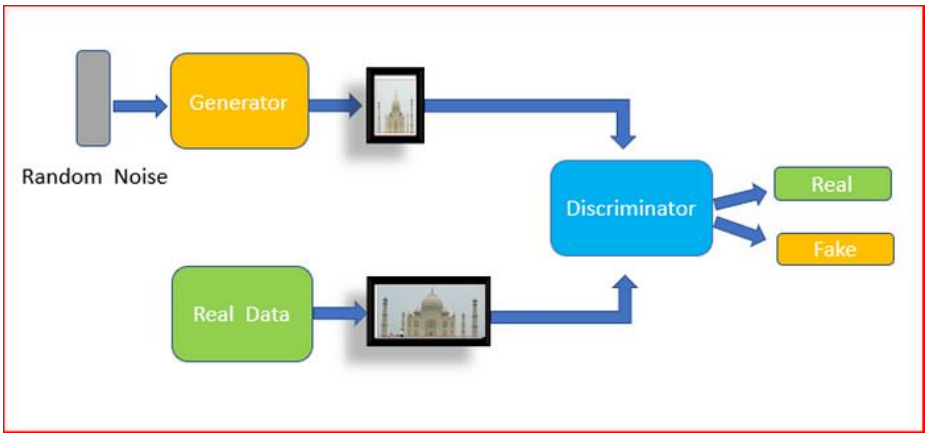

Figure 3 : Fonctionnement des GAN [8]

3.2.3 Les CNN

Alors que les GAN sont utilisés dans divers domaines tels que la création de musique, la synthèse de vidéos… les CNN sont plus souvent utilisés dans les données d'image et utilisent le Machine Learning que nous avons vu dans la première partie. Il s'agit d'une classe de réseaux de neurones artificiels spécialement conçus pour la reconnaissance de motifs dans des données [10].

La caractéristique principale des CNN sont les couches convolutionnelles. Elles filtrent l'entrée, par exemple, l'image, pour en extraire des caractéristiques significatives. Ces filtres, également appelés noyaux de convolution parcourent l'image en effectuant des opérations de convolution pour produire des cartes d'activation, qui représentent les caractéristiques de l'image. Ensuite, nous avons les couches de *pooling* qui réduisent la dimensionnalité des cartes d'activation en échantillonnant localement les valeurs maximales. Cela permet de réduire le nombre de paramètres et de rendre le réseau plus robuste aux translations et aux déformations dans les données d'entrée. Les fonctions d'activation non-linéaires sont utilisées après chaque couche de convolution et de *pooling* pour introduire de la non-linéarité dans le réseau et permettre l'apprentissage de relations complexes entre les caractéristiques des données. Après plusieurs couches convolutionnelles et de *pooling*, les caractéristiques extraites sont aplaties et introduites dans des couches entièrement connectées, souvent utilisées pour la classification ou la régression [11].

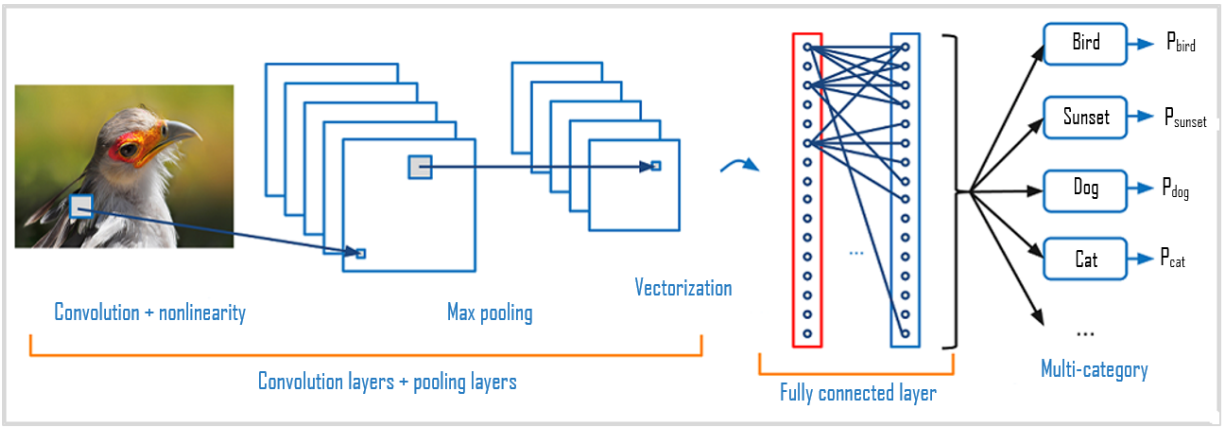

Figure 4 : Fonctionnement des CNN [11]

L'architecture spécifique des CNN les rend particulièrement efficaces pour extraire des caractéristiques hiérarchiques et spatiales à partir de données d'images, ce qui en fait un outil puissant pour la génération et l'analyse d'images.

3.2.4 Les Deepfakes

Un résultat des travaux des réseaux neuronaux sont les *deepfakes* : une forme de manipulation d'images et de vidéos. Une fois le réseau entrainé, le modèle peut être utilisé pour générer de nouveaux contenus en fonction des entrées fournies. Par exemple, un *deepfake* peut remplacer le visage d'une personne dans une vidéo par le visage d'une autre personne, ou même modifier les expressions faciales et les mouvements corporels pour créer une vidéo trompeuse [12]. Les *deepfakes* sont souvent utilisés à des fins de divertissement, de création artistique mais encore dans de la désinformation ou de la manipulation politique. Leur utilisation peut avoir des conséquences sérieuses sur la réputation et la vie privée des individus. Avant d'étudier les défis et autres problèmes, nous allons étudier un exemple.

3.2.5 Exemple d'utilisation

Début d'année 2024, un employé d'un centre financier de Hong Kong a été piégé à cause d'une IA. En effet, lors d'une réunion en visioconférence, son patron lui demande d'effectuer plusieurs versements de plusieurs millions d'euros. Pratique habituelle dans ce milieu, l'employé effectue rapidement les 15 transactions sur 5 comptes différents puisque l'ordre vient directement de son supérieur. Il verse au total plus de 26 millions de dollars américains.

Malheureusement pour lui, la visioconférence était le fruit d'un groupe de malfaiteurs. Les images et les voix sont les produits de différentes IA. Pour cela, les escrocs ont espionné l'entreprise pendant plusieurs mois. En effet, il a fallu hacker le système de messagerie afin d'envoyer le lien de la visioconférence. Mais il a surtout fallu enregistrer les voix et des images du patron et de différents employés afin de les recréer via une IA. L'entreprise étant réputée, un nombre incalculable de vidéos sont disponibles au public sur le site internet YouTube.

L'employé s'était méfié dans un premier temps à cause de l'invitation par mail mais il se sentait en sécurité entouré de ses collègues qui semblaient réels. Cependant, les images des collègues avaient été également générées artificiellement. [13].

Cet exemple présente le danger de ces nouvelles technologies utilisant de fausses images.

3.3 Les défis de la détection des Fausses Images

3.3.1 Les enjeux de la détection des fausses images

Avec l'émergence de nombreuses IA telles que DALL-E, Sora et ChatGPT, une ancienne pratique politique et sociale a refait surface : la désinformation à travers des documents, des images et des vidéos falsifiés. Bien que cette pratique ait été utilisée maintes fois dans l'histoire de l'humanité, l'avènement des IA a rendu la création de fausses informations plus facile que jamais, et cela de manière extrêmement convaincante.

En effet, l'intelligence artificielle joue un rôle croissant dans le domaine de la désinformation. Elle peut être exploitée pour générer du contenu faux de manière sophistiquée et persuasive notamment des vidéos. Cette technologie utilisant des *deepfakes*, qui a progressé

régulièrement depuis près d'une décennie, a la capacité de créer des marionnettes numériques parlantes, souvent à des fins politiques.

Un exemple frappant de ce phénomène pourrait être celui de Volodymyr Zelensky, l'actuel président de l'Ukraine, avec une vidéo circulant sur les réseaux sociaux le montrant annonçant la capitulation de l'Ukraine [14], ou encore celle de Morgan Freeman [15] et de nombreuses autres personnalités publiques.

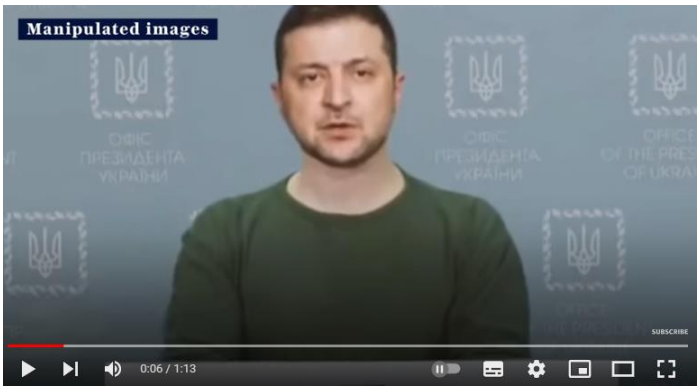

Figure 5 : Volodymyr Zelensky lors du faux discours [14]

Le problème majeur avec cette apparition soudaine de fausses vidéos et images extrêmement convaincantes est qu'il est difficile pour le public de trier le vrai du faux dans l'immense flux d'informations disponibles sur internet. Des groupes de personnes peuvent ainsi réagir violemment à cause de la désinformation : panique générale, comportement disproportionné etc... Néanmoins, l'origine du problème, l'IA peut aussi être le remède. Effectivement, plusieurs entreprises ont développé des logiciels permettant de détecter des *Deepfakes*/Manipulations d'images, même si cela peut aussi poser des questions, notamment en ce qui concerne la censure, la vie privée et la manipulation des algorithmes. Il est donc crucial de trouver un équilibre entre l'utilisation de l'IA pour contrer la désinformation tout en préservant les valeurs démocratiques telles que la liberté d'expression et l'accès à une information fiable.
La diffusion de fausses images compromet la confiance du public dans les médias, les réseaux sociaux et les sources d'information en général. Cela peut entraîner une incrédulité généralisée à toutes les informations présentées en ligne. Si les personnes ne savent plus en quoi croire, elles risquent de ne plus rien croire du tout ce qui entraînerait le désordre dans les pays.

D'autre part, l'IA est également utilisée pour détecter et contrer la désinformation, en analysant de grands ensembles de données pour repérer les schémas de propagation de fausses informations, en identifiant les sources non fiables, ou en développant des outils de vérification des faits plus efficaces, comme nous allons le voir par la suite.

3.3.2 Les moyens utilisés pour lutter contre les fausses images

Pour lutter contre cette désinformation et ainsi la propagation de *fake news*, de nombreux moyens plus ou moins efficaces sont mis en œuvre.
Tout d'abord, de nouvelles méthodes de technologie de détection sont développées. Ainsi, des algorithmes et des outils logiciels sont créés pour détecter les altérations et les manipulations dans les images. Ces technologies utilisent des techniques d'analyse d'image, de détection de motifs incohérents et de comparaison avec des bases de données d'images authentiques pour repérer les falsifications. Parmi ces logiciels, nous pouvons y retrouver InVid Verification Plugin, très connu auprès des chercheurs et des journalistes, TinEye qui permet de retrouver l'origine de l'image et détecter si elle a été modifiée ou encore Forensically qui analyse les images et détecte les anomalies.

Même si en premier lieu, nous pensons à utiliser des algorithmes pour combattre les créations d'autres ordinateurs, il existe d'autres moyens de lutter contre la propagation des fausses images. En effet, certains outils intègrent des signatures numériques dans les images pour vérifier leur authenticité. Ces signatures peuvent être des empreintes digitales numériques ou des métadonnées intégrées lors de la capture de l'image, permettant de retracer son origine et de détecter les modifications ultérieures.
De plus, des médias, en collaboration avec des chercheurs, ont aussi créé des sites de vérification de l'information [16]. Ces sites sont appelés des "fact-checking" et analysent l'exactitude des faits. Ces plateformes fournissent des analyses détaillées des images contestées, démêlant les faits de la fiction et aidant le public à prendre des décisions éclairées.

La collaboration entre les chercheurs, les entreprises technologiques, les médias et les organismes gouvernementaux est nécessaire pour développer des solutions efficaces de détection des fausses images. Le partage de données et de ressources peut contribuer à améliorer la qualité et la diversité des ensembles de données utilisés pour former les modèles de détection.
Enfin, les organismes de sécurité et les gouvernements ont mis en place une autre méthode qui n'est pas directement liée à l'informatique. Ainsi, la sensibilisation du public aux techniques de falsification d'images et à la manipulation de l'information joue un rôle crucial dans la lutte contre les fausses images. En éduquant les individus sur les risques associés à la propagation de fausses informations visuelles, on renforce leur capacité à les discerner.
Pour conclure, nous pouvons dire qu'il existe différents moyens de lutter contre les fausses images qui peuvent aller de la sensibilisation à la détection par des algorithmes. Cependant, la technologie ne cesse d'évoluer et nous pouvons donc nous demander si les logiciels de création de fausses images deviendront un jour si performants qu'il n'y aura plus aucun moyen de déterminer la véracité des images.

3.3.3 Complexité de l'analyse des modèles de détection

Les fausses images générées par l'IA présentent un défi particulier pour les algorithmes de détection en raison de leur capacité à reproduire des détails réalistes. Les GAN peuvent créer des images indiscernables de la réalité, ce qui rend la détection des altérations plus difficile. Pour relever ce défi, il est nécessaire de développer des algorithmes de détection adaptés capables de reconnaître les caractéristiques spécifiques des fausses images générées par l'IA. Cela peut impliquer l'utilisation de techniques avancées d'apprentissage automatique, telles que l'apprentissage par renforcement ou l'apprentissage par transfert, pour entraîner des modèles de détection capables de distinguer les fausses images des authentiques. De plus, les modèles de détection doivent être constamment mis à jour pour s'adapter aux nouvelles techniques de génération d'images utilisées par les GAN. Cela nécessite une surveillance continue des avancées en matière de génération d'images par l'IA et une adaptation rapide des modèles de détection pour rester efficaces face à ces évolutions.
Pour entraîner des modèles de détection efficaces, il est essentiel de disposer de données d'entraînement diversifiées qui représentent la variété des fausses images générées par l'IA. Cependant, la collecte de telles données peut être difficile en raison de la rareté des fausses images générées par l'IA dans les bases de données publiques. Pour pallier cette limitation, des efforts sont nécessaires pour constituer des ensembles de données de référence représentatifs et diversifiés en recourant à des techniques telles que la génération synthétique d'images ou la collaboration avec des experts en IA pour créer des exemples de fausses images. De plus, les modèles de détection doivent être entraînés sur des ensembles de données équilibrés qui contiennent à la fois des exemples d'images authentiques et des exemples de fausses images générées par l'IA. Cela permet d'assurer que les modèles de

détection sont capables de reconnaître les caractéristiques uniques des fausses images tout en évitant les faux positifs.
Enfin, il est essentiel de reconnaître les limites des modèles de détection existants face aux fausses images générées par l'IA. Bien que les modèles de détection puissent être efficaces pour repérer certains types de falsifications, ils peuvent présenter des lacunes dans leur capacité à détecter les fausses images générées par l'IA dans des situations complexes ou évolutives. Par exemple, certains GAN peuvent être entraînés pour contourner les modèles de détection en générant des images spécifiquement conçues pour tromper ces modèles. Pour relever ce défi, il est nécessaire de mener une recherche continue pour comprendre les techniques utilisées par les GAN pour générer des images trompeuses et pour développer des contre-mesures appropriées. Cela peut impliquer l'utilisation de techniques telles que l'adversarial training [17], où les modèles de détection sont entraînés contre des adversaires génératifs pour renforcer leur robustesse face aux attaques potentielles.

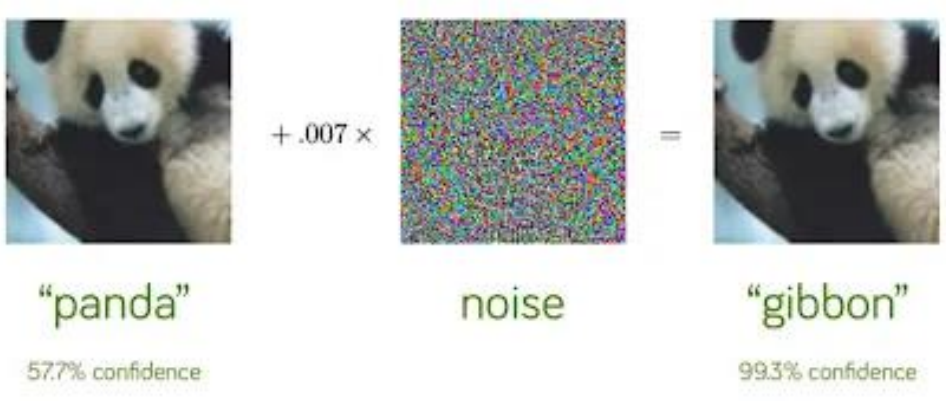

Figure 6 : Classification d'une image de panda grâce à un adversarial network [17]

En résumé, relever les défis posés par la détection des fausses images générées par l'IA nécessite le développement de modèles de détection adaptés, la collecte de données d'entraînement diversifiées et la reconnaissance des limites des modèles existants. En combinant ces approches, il est possible de renforcer l'efficacité des systèmes de détection et de lutter contre la propagation de ces fausses images sur Internet.

L'énorme volume d'images partagées en ligne chaque jour, combiné à la vitesse de leur diffusion sur les réseaux sociaux, pose un défi majeur pour la détection rapide et efficace des fausses images. Les systèmes de détection doivent être capables de traiter un flux constant de données en temps réel. De plus, il peut être difficile de constituer des ensembles de données de référence fiables pour entraîner des modèles de détection d'images falsifiées. En raison de la nature clandestine de la création et de la diffusion de fausses images, il peut être difficile d'obtenir des exemples authentiques et falsifiés en quantité suffisante pour former des algorithmes de détection robustes.

3.3.4 Exemples de Fausses Images et leurs impacts

3.3.4.1 Le pape François en manteau de marque

Figure 7 : De fausses photos montrant le pape François vêtu d'un manteau de marque [18]

Cette image circulant sur internet montre le Pape vêtu d'une longue doudoune blanche, évoquant le style ostentatoire des rappeurs américains les plus riches. Certains pourraient se demander si le Pape envisage de modifier radicalement les traditions vestimentaires de l'Eglise.

Cependant, cette image est en réalité une création générée par une intelligence artificielle, spécifiquement par le programme MidJourney. Elle a été initialement partagée le 24 mars 2023 par un utilisateur américain nommé Pablo Xavier sur un forum Reddit dédié à cette technologie. Au fil du week-end, cette photo humoristique a rapidement circulé sur Twitter et d'autres plateformes de médias sociaux. Initialement partagée comme une simple blague, elle a rapidement évolué pour devenir le symbole de l'émergence de l'ère du *deepfake*, en trompant de nombreuses personnes sur Internet [18].

3.3.4.2 La tentative d'assassinat d'Emmanuel Macron en Ukraine

Figure 8 : Un faux sujet au JT de France 24 évoque un projet d'assassinat de Macron en Ukraine [19]

Selon des sources sur X, il aurait été prétendu que Kiev aurait cherché à attirer le président français Emmanuel Macron en Ukraine dans le but de le tuer, puis de faire porter le blâme de sa mort sur la Russie. Cette initiative aurait eu pour objectif de "recentrer l'attention des médias sur l'Ukraine et d'accroître le soutien financier et militaire de l'Occident". Ayant été informé de

cette menace, le chef de l'État aurait annulé à la dernière minute sa visite prévue les 13 et 14 février dans la capitale ukrainienne. La propagation de cette rumeur s'appuie principalement sur une vidéo présentée comme étant un bulletin d'information de la chaîne France 24. Sa diffusion a gagné en notoriété après avoir été partagée par Dmitri Medvedev. L'ancien Premier ministre russe n'a pas directement relayé la vidéo de France 24, mais a soutenu la théorie d'un projet d'assassinat visant Emmanuel Macron, qui aurait été à l'origine du report de sa visite. À l'écran, il s'agit bien d'un journaliste de la chaîne, Julien Fanciulli, chargé de présenter les tranches d'informations de l'après-midi. Dans l'extrait relayé, le journaliste déclare effectivement qu'Emmanuel Macron aurait été contraint d'annuler sa visite à Kiev à cause d'une tentative d'assassinat par le régime Ukrainien. Or, ces mots n'ont jamais été prononcés par le présentateur, dément France 24. "*C'est ce qu'on appelle un 'deepfake' : un outil d'intelligence artificielle utilisé pour faire dire à une personne, avec un timbre de voix plus ou moins similaire, quelque chose de totalement différent de la vidéo originelle*", détaillent Les Observateurs, le service de fact-checking de la chaîne [19].

4. CONCLUSIONS ET PERSPECTIVES

L'objectif de notre projet était d'expliquer comment les intelligences artificielles peuvent avoir un rôle dans la création de fausses images et pourquoi les détecter est primordial. Pour cela, nous avons étudié le fonctionnement de l'intelligence artificielle et développé sa création : notamment les neurones, puis les réseaux de neurones et afin l'apprentissage, le Machine Learning. Suite à cela, nous avons vu comment les IA génèrent des images avec les réseaux neuronaux : les GAN et les CNN. Pour finir, nous avons parlé des défis liés à la détection des fausses images et les moyens utilisés pour détecter celles-ci. Aujourd'hui plus que jamais, Internet fait partie de notre quotidien. En effet, d'après le rapport annuel de l'Institut Reuters pour l'étude du journalisme : en 2023, 30% des personnes interrogées disaient utiliser les réseaux sociaux comme première source d'information. C'est pourquoi il est légitime de se poser une question : Dans un monde où une aussi grosse partie de la population s'informe sur Internet, devons-nous continuer à développer ces technologies qui peuvent tromper des millions de personnes ?

Pour rendre concret notre projet de mise en lumière des fausses images, nous pourrions mettre en garde les étudiants de l'INSA en diffusant des images dont une créée par IA : ils doivent essayer de déterminer laquelle est fausse. Cette expérience permettrait aux futurs ingénieurs de l'INSA d'être sensibilisés aux enjeux de l'IA et de leur éviter une erreur de jugement dans leur carrière ou dans leur vie personnelle.

Pour conclure, ce projet nous a permis de faire des recherches sur un sujet que nous ne connaissions qu'en surface mais qui nous intéressait fortement. En effet, nous avons, tous les six, la volonté d'évoluer dans le monde de l'informatique, en allant en GM ou en ITI, et découvrir un aspect aussi important n'a fait que confirmer notre choix. Nous avons également fait pour la première fois un état de l'art avec un rapport d'étonnement, tout en réussissant à s'organiser dans un groupe où nous ne nous connaissions pas avant, avec des contraintes d'emploi du temps. Ce travail de groupe et son organisation nous ont permis de découvrir ce à quoi nous serons confrontés lorsque nous serons ingénieurs. Ce fut donc une expérience enrichissante d'un point de vue des connaissances mais également d'un point de vue humain.

Remerciements

Nous tenons à remercier tout particulièrement notre professeur encadrant, Monsieur Abdelaziz Bensrhair, pour nous avoir suivi et guidé tout au long du projet ainsi que pour sa disponibilité et ses réponses à nos questions sur le projet mais également sur l'informatique de manière générale et sur le département ITI.

De plus, nous aimerions remercier les doctorants du laboratoire LITIS qui nous ont accordé de leur temps pour nous expliquer leurs travaux et répondre à nos questions.

Enfin, nous remercions l'INSA et les responsables STPI2 pour nous avoir permis de faire ce projet et de découvrir en profondeur un sujet aussi intéressant qu'important.

5. BIBLIOGRAPHIE

[1] lien internet : Nombre de Shannon - Tout savoir (2024) (apprendre-les-echecs-24h.com) (valide à la date du 14/06/2024).

[2] lien internet : IA et Supply chain : Comment améliorer l'efficacité de votre supply chain avec les algorithmes génétiques (crossdata.tech) (valide à la date du 14/06/2024).

[3] lien internet : Les réseaux de neurones artificiels - Dans quel monde sommes nous? (eklablog.com) (valide à la date du 14/06/2024).

[4] lien internet : Que sont les réseaux neuronaux ? | IBM(valide à la date du 14/06/2024).

[5] lien internet Machine Learning : les 9 types d'algorithmes les plus pertinents en entreprise | LeMagIT (valide à la date du 14/06/2024).

[6] lien internet : Machine Learning : Définition, fonctionnement, utilisations (datascientest.com) (valide à la date du 14/06/2024).

[7] lien internet : Qu'est-ce que l'intelligence artificielle et pourquoi est-ce important ? | Talend (valide à la date du 14/06/2024).

[8] lien internet : Machine Learning : Définition, fonctionnement, utilisations (datascientest.com) (valide à la date du 14/06/2024).

[9] lien internet Qu'est-ce qu'un GAN en Deep Learning ? Comprendre facilement maintenant (inside-machinelearning.com) (valide à la date du 14/06/2024).

[10] lien internet : Qu'est-ce qu'un réseau de neurones convolutifs ? | IBM (valide à la date du 14/06/2024).

[11] lien internet : L'histoire et l'évolution de l'Intelligence Artificielle (IA) : des origines à nos jours (pandia.pro) (valide à la date du 14/06/2024).

[12] lien internet : Deepfake : définition, techniques et risques (journaldunet.fr) (valide à la date du 14/06/2024).

[13] lien internet : IA : floué par une visioconférence avec des collègues deepfakes, il transfère 26 millions de dollars à des escrocs – Libération (liberation.fr) (valide à la date du 14/06/2024).

[14] lien internet : Deepfake video of Volodymyr Zelensky surrendering surfaces on social media (youtube.com) (valide à la date du 14/06/2024).

[15] lien internet : This is not Morgan Freeman - A Deepfake Singularity (youtube.com) (valide à la date du 14/06/2024).

[16] lien internet : Découvrir des outils pour vérifier les fausses informations ou fake news (solidarite-numerique.fr) (valide à la date du 14/06/2024).

[17] lien internet : Adversarial Training : qu'est-ce que c'est ? (datascientest.com) (valide à la date du 14/06/2024).

[18] lien internet : Fake photos of Pope Francis in a puffer jacket go viral, highlighting the power and peril of AI - CBS News (valide à la date du 14/06/2024).

[19] lien internet : Quand un deepfake lance une rumeur sur un projet d'assassinat d'Emmanuel Macron en Ukraine | TF1 INFO (valide à la date du 14/06/2024).

6. ANNEXES

Rapport d'Etonnement

Tout au long de ce rapport, nous avons vu que l'Intelligence Artificielle était devenue un domaine d'étude majeur dans le monde de la technologie, suscitant à la fois fascination et appréhension. Cependant, il existe, certes des avantages mais aussi de nombreux aspects problématiques de la création de fausses images par les IA. Nous allons ainsi étudier les limites dans de nombreux aspects. Tout d'abord, nous verrons les aspects juridiques et éthiques que nous avons déjà abordé à plusieurs reprises dans cette étude. Ensuite, nous nous intéresserons à des problèmes moins évoqués tels que le sujet de l'environnement et enfin les enjeux sociaux et économiques.

Aspects juridiques et éthiques

La capacité de créer des fausses images soulève évidemment des questions d'éthique. Par exemple, l'utilisation de fausses images dans des campagnes de désinformation ou de manipulation peut avoir des conséquences dévastatrices sur la société et la démocratie. De plus, cela soulève des préoccupations concernant la confidentialité et le consentement, car les visages peuvent être générés à partir de données personnelles sans le consentement des individus. De plus, en octobre 2023, un projet de loi a été adopté pour sécuriser et réguler l'espace numérique. Ainsi, une loi indique que sera « *assimilé à l'infraction mentionnée au présent alinéa et puni des mêmes peines le fait de porter à la connaissance du public ou d'un tiers, par quelque voie que ce soit, un contenu visuel ou sonore généré par un traitement algorithmique et représentant l'image ou les paroles d'une personne, sans son consentement, s'il n'apparaît pas à l'évidence qu'il s'agit d'un contenu généré algorithmiquement ou s'il n'en est pas expressément fait mention* ». La sanction de la diffusion de fausses images générées par des IA est ainsi sanctionné en France par un an d'emprisonnement et 15 000€ d'amende. Détournements et *deepfakes*, les enjeux de la protection du droit à l'image face à l'IA.

Aspects sociaux politiques et économiques

La création et la diffusion de *deepfakes* ont également des conséquences sociales et économiques. Prenons l'exemple d'une fausse publicité qui mettrait la marque dans une situation embarrassante, inconsciemment ou non, cette image affectera les spectateurs et les actions de cette marque baisseront. De nombreux statisticiens et experts ont évalué l'impact des *fake news* de manière générale dans l'économie mondiale et ont évalué ce marché à 78 milliards de dollars par an. Nous y retrouvons 39 milliards de dollars pour la majorité mais il y a aussi d'autres catégories importantes comme la désinformation financière et en matière de santé, la protection des marques...

De plus, certains se servent des algorithmes pour réécrire l'histoire ce qui brouille la limite entre la fiction et le réel. Il y a quelques années, une photo de la découvreuse de l'ADN, Rosalind Franklin, a été mise en ligne alors qu'elle remportait avec un prix Nobel. Cependant, elle ne l'a jamais reçu et trois hommes l'ont eu à sa place. Cette photo a émis un doute pour de nombreuses personnes alors qu'il ne s'agit pas d'un sujet très grave. Qu'en sera-t-il alors lorsque de nouvelles images avec des sujets plus géopolitiquement importants par exemple apparaitront ? Les images générées par IA et le risque de réécrire l'histoire. De nombreuses photos sont en effet apparues depuis le début du conflit entre la Russie et l'Ukraine où chacun défend son camp en montrant les adversaires dans de fausses conditions.

Les aspects sociaux et économiques ne font qu'empirer en fonction des situations et des évènements passés et actuels ce qui influencent donc toutes les personnes voyant ces *deepfakes*.

Aspects environnementaux

L'intelligence artificielle est très énergivore et consomme bien plus d'électricité que nous aurions pu penser. En effet, les serveurs de calculs requis pour la formation des IA doivent fonctionner 24 heures sur 24, les stockages de données utilisent également beaucoup d'énergie. Pour évoquer les impacts écologiques directs, les déchets électroniques polluent les sols, l'eau ainsi que l'air. Les fausses images engendrées par les IA peuvent amener à des problèmes de manipulation, de désinformation qui affecteront les politiques environnementales et la sensibilisation du public aux problèmes environnementaux. Nous rejoignons ici les aspects politiques évoqués plus haut. En effet, les intelligences artificielles ne permettent pas de voir l'avenir mais elles permettent de nous montrer ce qui peut arriver si on ne prend pas assez soin de notre planète. Ces images nous montrent les problèmes qui peut survenir mais aussi des solutions pour les éviter.

Cependant, une nouvelle forme d'IA voit le jour : « l'IA Verte » qui est présentée comme nécessitant moins d'énergie donc étant à impact environnemental moindre.

Conclusion

Les fausses images conçues par les Intelligences Artificielles démontrent l'avancée technologique ; cependant elles mettent en évidence de nombreux problèmes environnementaux, sociaux, écologiques ou encore éthiques. Il est donc essentiel de s'adapter au progrès et de créer de nouvelles restrictions et lois pour éviter de se laisser gouverner par les *fake news* et les fausses images.

Avis partagés du groupe :

Tout d'abord, nous avons trouvé un côté ludique à ces fausses images. De ce fait, nous pouvons imaginer comment nous vieillirons par exemple. Nous avons cependant rapidement pris conscience des enjeux. Comme de nombreux aspects liés au progrès récent, il importe de les maitriser. Notre principale inquiétude est l'impact géopolitique des *fake news*. Nous pouvons prendre le cas des élections faussées ou des embrasements militaires. Enfin, notre esprit scientifique a repris le dessus et nous avons trouvé ce sujet passionnant puisque cela crée des voies technologiques tout à fait nouvelles. Nous sommes d'autant plus motivés pour œuvrer à maitriser ces techniques.