



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

COLLÈGE DES HUMANITÉS

SCIENCES HUMAINES ET SOCIALES

**Transformation de la société par l'émergence
d'une Intelligence Artificielle générale**

Auteurs :

Johnny ESPIN (MFE)

Ruben FISZEL (IN)

Dirigé par :

Romain BIONDA

Véronique ROHRBACH

Projet SHS de 1^{ère} année master

Hommes/Machines

27 avril 2016

Table des matières

Introduction	3
1 Possibilités et plausibilité de l'évolution de l'IA	6
1.1 L'IA et le <i>machine learning</i>	6
1.2 Les réseaux de neurones	7
1.3 Le <i>Reinforcement Learning</i>	10
2 Fondamentaux économiques	12
2.1 Objectifs de l'économie	12
2.2 Fondamentaux des économies de marché	13
2.2.1 Micro-économie et <i>homo economicus</i>	13
2.2.2 Le système capitaliste	15
3 Le bouleversement apporté par l'IA	17
3.1 Un changement inéluctable	17
3.2 Acceptation sociale de l'IA	19
3.3 Quelles solutions ?	22
Conclusion	22

Introduction

La littérature de science-fiction regorge de mondes où l'intelligence de la machine rivaliserait avec celle de l'homme. Ce fantasme qui fait couler beaucoup d'encre pourrait ne plus relever seulement de la science-fiction. En effet, l'augmentation exponentielle de la puissance de calcul informatique ces dernières décennies, ainsi que la découverte de nouveaux algorithmes, laissent présager un avenir optimiste techniquement dans le domaine de l'intelligence artificielle. De plus, le concept de machines intelligentes ne semble plus si surprenant. Nous vivons après tout dans une ère de *smart-devices*, où beaucoup sont équipés de smart-phones, et travaillent majoritairement sur des ordinateurs. Déjà en 1996, Deep Blue battait le champion du monde des échecs, Garry Kasparov et plus récemment *AlphaGo* a battu le champion du monde de Go, Lee Sedol, jeu pourtant réputé à l'abri des IAs par ses combinaisons infinies de positions. Cependant, cela correspond à des applications spécialisées de l'intelligence artificielle ; celle qui nous intéresse se veut beaucoup plus générale que cela, et serait capable de raisonner sur son environnement à l'instar de l'homme.

Il est évident qu'un débat social et philosophique autour de l'émergence d'une telle intelligence artificielle existe, au vu du nombre d'articles journalistiques et opinions diverses qui apparaissent régulièrement. Dès lors, il est intéressant de réfléchir aux implications de cette IA sur la manière dont l'homme interagit avec son environnement, mais avant cela, il semble nécessaire de nous remettre dans le contexte contemporain et d'analyser la situation à laquelle nous nous trouvons confrontés. Pour certains scientifiques et journalistes, nous nous trouvons à l'aube d'une nouvelle époque : *The Economist* utilise d'ailleurs ce terme pour intituler une de ses lettres ¹ ; l'inventeur et futuriste Ray Kurzweil parle de "singularité" ² ; James Barrat, l'auteur de *Our Final Invention : Artificial Intelligence and the End of the Human Era* essaie d'expliquer pourquoi le monde scientifique est inquiet au sujet du développement rapide de l'IA ³.

¹"The Dawn of Artificial Intelligence", *The Economist*, 9 mai 2015, <http://www.economist.com/news/leaders/21650543-powerful-computers-will-reshape-humanitys-future-how-ensure-promise-outweighs>, visité le 9 mars 2016.

²Ray Kurzweil, *The Singularity is Near*, New-York, Viking Press, 2005.

³James Barrat, "Why Stephen Hawking and Bill Gates Are Terrified of Artificial Intelligence", *The Huffington Post*, 4 septembre 2015, <http://www.huffingtonpost.com/james-barrat/hawking-gates->

Supposons donc que nous nous trouvons bien dans une ère de transition dont l'aboutissement sera une société où cohabiteront êtres humains et machines douées d'une intelligence semblable à celle de l'homme. L'idée principale de ce travail est de s'intéresser aux implications directes et indirectes que ce changement aurait sur notre façon de vivre. Plus précisément, nous nous intéresserons à l'évolution, que nous postulons nécessaire, du système économique afin d'intégrer optimalement une nouvelle "espèce".

Néanmoins, avant de nous lancer dans cet exercice spéculatif, il convient de considérer l'étape antérieure à cette transition. Malgré la conviction de beaucoup de personnes influentes citées ci-dessus, est-il réellement plausible d'espérer (ou de craindre) l'avènement de l'IA ? Afin d'aborder cette problématique, il est important de bien définir ce que nous entendons par IA, et dans quel cadre celle-ci peut espérer se développer.

Tout d'abord, qu'entendons nous par machine douée d'intelligence artificielle ? De prime abord, il est nécessaire de s'accorder sur une définition de l'intelligence. Nous utiliserons cette définition informelle proposé par Legg et Hutter en 2007⁴, se basant sur de nombreuses définitions de psychologues et chercheurs en intelligence artificielles : "L'intelligence mesure l'abilité d'un agent à réaliser des objectifs dans une gamme très diverse d'environnements". Cette définition est intéressante car elle comprend de nombreuses propriétés associées à l'intelligence. Les recherches en neurosciences s'accordent sur le fait que chez l'homme, l'intelligence se base sur des *briques* élémentaires appelées communément des neurones. Cette brique est reproductible de manière simplifiée par un ordinateur ou un composant électronique. C'est l'agencement en couches *layers* et le nombre de neurones présents qui permet l'émergence d'une intelligence complexe. Nous nous baserons sur ces observations pour montrer la plausibilité technique d'une intelligence artificielle se rapprochant de la nôtre, basée sur des réseaux de neurones. Le point critique du développement des IAs se situe lorsque nous serions capable de créer une intelligence artificielle supérieure à la nôtre, aussi petite soit cette différence, car elle-même sera en théorie capable de créer une intelligence supérieure à elle. Ce moment est appelé la "singularité" technologique. Le champ des possibles pourrait ainsi devenir infini.

La possibilité de transition vers un monde où l'intelligence artificielle sera présente dans notre société dépendra fortement de la capacité de cette dernière à accepter un changement drastique dans sa manière de fonctionner. Une forte présence due à une intelligence autre que celle de l'homme pose déjà aujourd'hui de grands problèmes moraux. Pouvons-

artificial-intelligence_b_7008706.html, visité le 9 mars 2016.

⁴Legg, Shane and Veness, Joel, "An Approximation of the Universal Intelligence Measure", eprint <http://arxiv.org/abs/1109.5951>, visité le 9 mars 2016.

nous accepter le fait de déléguer nos responsabilités les plus humaines à une machine ? Serons nous capables d'accepter de ne plus être ce qui est de plus intelligent sur terre ? Pourrions-nous accepter de mettre notre vie entre les mains d'une IA ?

Ensuite, nous nous intéresserons aux conséquences économiques d'un tel avènement. Nous revisiterons certains fondamentaux économiques pour mieux comprendre les bouleversements qui s'annoncent. En effet, nous vivons depuis longtemps dans une économie de marché. Dans une telle économie, les agents productifs sont en compétition pour servir la demande. Ils ont à leur disposition des salariés, des matières premières et des machines-outils. Le capital d'une entreprise permet alors d'investir dans de telles machines, d'acheter les matières premières et de verser des salaires. Il existe d'une part le revenu du capital reversé aux agents ayant fourni le capital de l'entreprise et d'autre part le revenu du travail reversé aux salariés sous forme de salaire. Nous faisons l'hypothèse que l'avènement de l'IA rendra le travail⁵ salarial caduc et donc *in fine* supprimera le revenu du travail. Comment l'économie et la société s'adapteront-elles à cette révolution ?

Enfin, nous proposerons une réflexion sur un monde futuriste où vivront harmonieusement des machines intelligentes et des hommes *libérés* de la contrainte du travail en tant que source de rémunération.

⁵Nous considérons seulement le travail en tant qu'activité lucrative et excluons le bénévolat et toute autre activité dont l'intérêt premier n'est pas la rémunération. Ce détail est important car nous ne supposons pas que l'avènement de l'IA empêcherait les hommes d'exercer une activité à titre personnel. Par exemple, exercer sa passion pour la politique ou l'enseignement.

Chapitre 1

Possibilités et plausibilité de l'évolution de l'IA

1.1 L'IA et le *machine learning*

Le terme d'intelligence artificielle en informatique regroupe tout un ensemble d'algorithmes et de méthodes pour permettre la réalisation de tâches sophistiquées. Celles qui nous intéressent particulièrement sont celles relevant du *machine learning*. En effet jusqu'à présent, l'IA se concentrait sur la résolution de problèmes dont les données lui sont présentées de manière complètement structurée. Dans presque tous les domaines, les performances de ces intelligences spécialisées dépassent de loin les capacités de l'homme, atteignant des *super-human performances*. Peut-t-on malgré tout qualifier ces outils "d'intelligents" ? Dans une certaine mesure, ces méthodes se résument à appliquer une recette de cuisine écrite par un développeur, lui, humain. Ainsi, personne ne s'offusque de la supériorité des calculatrices à réaliser des opérations bien au-delà de nos capacités mentales en une fraction de seconde.

C'est pourquoi, si l'intelligence développée reste pour l'instant considérée le propre de l'homme, c'est grâce à sa capacité d'extraire des structures, appelée *pattern matching*, dans le caractère versatile de l'information auquel il fait face, non seulement pour ses travaux les plus complexes mais aussi pour son quotidien le plus banal. Ainsi, alors que conduire une voiture ne représente pas un challenge intellectuel insurmontable, cela reste un véritable défi même pour les ordinateurs les plus puissants. En effet, le conducteur est exposé par ses sens à une quantité immense d'information qu'il traite très naturellement. Chaque instant, notre vision nous fournit une image très détaillée du monde qui nous entoure et pourtant nous ne nous laissons pas submerger. Nous simplifions l'information en un ensemble de formes et réalisons tout un ensemble d'opérations afin de simplifier celle-ci. Cette capacité d'abstraction est une des composantes essentielles de ce

qui forme notre intelligence. En science, ce traitement de l'information correspond à une réduction de dimension. La dimension en mathématique correspond très grossièrement au nombre d'éléments requis pour décrire complètement une entité. Par exemple, notre position sur la terre est de dimension 2, la latitude et la longitude. L'information non traitée est considérée comme de très grande dimension. Ainsi, une image est d'aussi grande dimension que le nombre de pixels qui la compose. Catégoriser une image la réduit de cette espace à grande dimension vers un espace à une dimension, la catégorie de l'image. Abstraire c'est réduire la dimension de l'information.

En effet, la capacité de conduire est-elle innée ? Non, nous l'acquérons par l'apprentissage : nous nous adaptons à notre environnement. L'apprentissage et l'abstraction sont en réalité des concepts très liés. Nous apprenons à simplifier l'information par un processus complexe, mais qui consiste à observer et à changer notre interprétation de notre environnement selon son évolution. Ainsi, nous apprenons à associer des sons à des mots, et des mots à des concepts. Cependant, nous n'apprenons pas tout, et nous sommes biologiquement préparés à pouvoir extraire des *patterns* sans que cela puisse être explicable seulement par de l'apprentissage ¹. Par exemple, le cerveau peut distinguer une voix parmi d'autres et cela, dès la naissance.

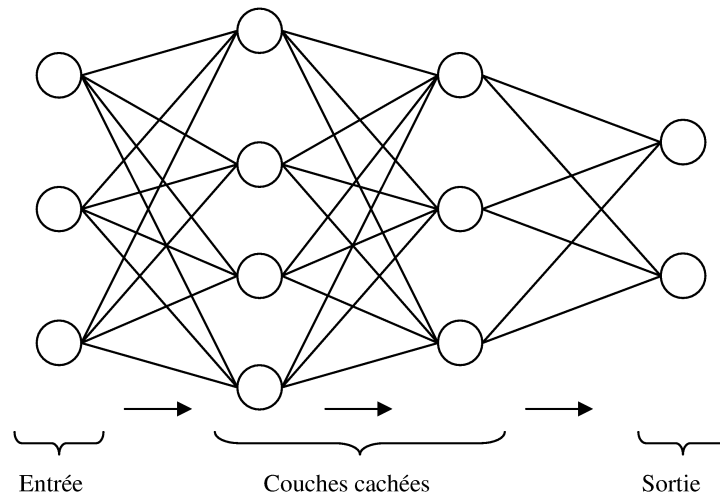
1.2 Les réseaux de neurones

Les réseaux de neurones sont une construction informatique qui se veut inspirée de la construction biologique du cerveau et tentent de reproduire des opérations effectuées en son sein, par les neurones. Ce sont des approximateurs de fonctions universelles, c'est à dire qu'ils ont pour but de se comporter le plus possible comme la fonction qu'ils apprennent sans jamais l'atteindre complètement. Une fonction peut-être par exemple de classer une image en une catégorie, ou décider de la prochaine direction à prendre. En réalité, toute transformation de l'information est assimilable à une fonction. Contrôler notre corps peut aussi se voir comme une fonction qui prend en entrée nos sens, notre mémoire, notre pensée immédiate et notre personnalité et décide du signal électrique à transmettre par nos nerfs à nos muscles. Le but de l'apprentissage pour un réseau de neurone est donc de lui faire se rapprocher le plus possible de la fonction qu'il approxime.

Un réseau de neurones est structuré en couche de neurones. Chaque couche est connectée complètement et uniquement à la couche qui la précède et qui la suit. Le *Deep Learning* qualifie l'usage de réseaux de neurones avec plus d'une couche entre la couche d'entrée et la couche de sortie. De manière générale, plus un réseau est composé de couches, plus

¹Jean-Pierre Changeux, "L'Inné et l'Acquis dans la Structure du Cerveau", *La Recherche*, mai 2000, <http://www.larecherche.fr/actualite/aussi/inne-acquis-structure-du-cerveau-texte-integral-01-05-2000-71955>, visité le 27 avril 2016.

FIGURE 1.1 – La structure d'un réseau de neurone



il est capable de *profondeur*² dans sa capacité d'abstraction. Au passage de chaque couche, l'information est décomposée en autant de neurones qui composent la couche et est mécaniquement transformée selon les paramètres de cette couche. Ces paramètres sont les poids de chaque connection entre les neurones qui définissent l'importance de chaque bout d'information en la mettant à une échelle a ($Y = aX$) puis un biais b lui est additionné ($Y = aX + b$). Un réseau de neurones est constitué de milliers voire de millions de ces paramètres. Un réseau de neurones est donc une fonction paramétrique. Ces paramètres sont ajustés lors de l'apprentissage et restent constants par la suite. C'est pourquoi il est nécessaire de l'entraîner.

Parmi les nombreux succès récents les plus impressionnants des réseaux de neurones et du *Deep Learning*, on peut compter la victoire d'*AlphaGo* sur le champion du monde de Go, Lee Sedol,³ le développement d'une plate-forme de voitures autonomes par NVIDIA⁴, une adoption unanime par les équipes concourants pour les compétitions de reconnaissance d'images⁵ avec d'excellents résultats, proches de la capacité de l'homme à classer des images, et enfin son usage pour apprendre à jouer à des jeux atari 2600 à partir de simples pixels d'images et sans instructions pour atteindre sur certains titres des performances qualifiées de super-humaines⁶. *AlphaGo* est particulièrement intéressant car le

²Christopher Olah, "Deep Learning, NLP, and Representations", blog, 7 juillet 2014, <http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/>, visité le 27 avril 2016.

³https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol, visité le 27 avril 2016.

⁴<http://www.nvidia.com/object/drive-px.html>, visité le 27 avril 2016.

⁵<http://image-net.org/challenges/LSVRC/2015/results>, visité le 27 avril 2016.

⁶Saxena Shalini, "AI Masters 49 Atari 2600 Games Without Instructions", *ars technica*, 27 février 2015, <http://arstechnica.com/science/2015/02/ai-masters-49-atari-2600-games-without-instructions/>, visité le 27 avril 2016.

Go possède infiniment trop d'états de jeu pour pouvoir tous les explorer. Contrairement à *Deep Blue* pour les échecs, la suprématie de la machine ne repose pas sur sa capacité à explorer toutes les positions atteignables à partir d'une position, mais à réfléchir à la qualité d'une position selon des critères de représentation appris lors de son apprentissage, à l'instar de l'homme. Si les entreprises d'IA utilisent des jeux pour montrer la capacité de leur recherche, c'est parce que les jeux sont modélisables par des processus de décision markovien ⁷. Un processus de décision markovien est un modèle général très flexible pour décrire un grand nombre de problèmes qui demandent de prendre des décisions dans un environnement constitué d'aléatoire, ce qui permet d'intégrer la complexité du monde qui nous entoure. Les avancées dans la résolution de ces jeux est donc révélatrice car l'intelligence requise pour les résoudre est similaire à celle requise pour bon nombre d'applications bien plus concrètes.

L'apprentissage pour un réseau de neurones consiste à minimiser une fonction, appelée fonction d'erreur. Cette fonction d'erreur mesure simplement la distance entre le résultat prédit par le réseau de neurones et le résultat attendu. Intuitivement cela veut dire que le réseau de neurones s'adapte lors de l'apprentissage pour que son erreur soit minimale sur les exemples qu'on lui fournit. L'apprentissage n'est alors qu'un problème d'optimisation de la minimisation d'une fonction d'erreur. L'adaptation se déroule par l'ajustement des nombreux poids ou coefficients internes, aussi appelés paramètres du réseau de neurones, de telle façon à ce qu'ils se dirigent dans une direction minimisant l'erreur. Cette direction est connue grâce aux gradients respectifs (une sorte de quantité de changement) des paramètres de la fonction d'erreur. En effet, la technique de la descente de gradient se base sur l'observation qu'en ajustant les paramètres dans la direction perpendiculaire au gradient, il est possible de réduire l'erreur petit à petit, à mesure qu'on le nourrit d'exemples dont on a la solution, un corpus labelé. Les labels ou solutions sont écrits manuellement par des humains, ce qui peut prendre un temps conséquent. On appelle cela le *supervised learning* car la machine est "assistée" par les solutions dans son apprentissage. Les paramètres d'un réseau de neurones sont ajustés lors de l'apprentissage grâce aux exemples qu'on lui fournit. Outre la puissance de calcul, la qualité du corpus d'apprentissage est très importante. La qualité d'un corpus se mesure principalement à la diversité des exemples qui permettent au réseau de neurone d'essayer de généraliser plutôt que d'apprendre par cœur les solutions.

Le *Big Data*, c'est-à-dire le recueillement d'un grand nombre de données des utilisateurs fourni de plein grès par et sur eux, permet de construire des corpus labelés gigantesques contre peu de frais. Ainsi, les habitudes de consommation des magasins en ligne ou les

⁷https://fr.wikipedia.org/wiki/Processus_de_d%C3%A9cision_markovien, visité le 27 avril 2016.

publications sur les réseaux sociaux sont autant de mines d'or d'information pour les entreprises.

1.3 Le *Reinforcement Learning*

L'apprentissage humain ne requiert pas de corpus labelé, et comment espérer que la machine dépasse un jour les capacités de l'homme si c'est lui qui doit *in fine* lui fournir les solutions et donc les connaître. L'apprentissage sans corpus labelé s'appelle l'*unsupervised learning*. Pour réaliser cet apprentissage sans corpus, l'agent contrôlé par l'IA interagit avec son environnement, observe les conséquences de ses actions et en extrait des informations utiles. Avec suffisamment d'interaction, l'agent connaît de mieux en mieux son environnement et est capable de l'explorer plus efficacement et d'optimiser son interaction. On peut ainsi penser aux robots aspirateurs, qui découvrent l'espace qui les entoure grâce à leurs capteurs. Après suffisamment d'explorations, ils ont acquis suffisamment d'information pour trouver le parcours optimal qui maximisera l'espace d'aspiration en un temps donné. Une des méthodes principales d'apprentissage est le *Reinforcement Learning* qui permet d'apprendre sur la base de récompenses, positives et négatives, à chaque interaction avec l'environnement. Ce modèle se veut basé sur le modèle biologique. En effet, les connections entre les neurones se renforcent ou décroissent lors de l'apprentissage. Ces méthodes permettent à des IAs d'explorer seules toutes sortes de modèles et ainsi découvrir des structures qui nous échappent à nous leur concepteur. Ainsi, AlphaGo, l'IA qui a récemment battu le champion du monde de Go, a créé la surprise aux experts en jouant des coups très originaux mais non moins très efficaces. En effet, AlphaGo a exploré en profondeur le Go en jouant des centaines de millions de parties contre lui-même et ne souffre pas des biais sur le style de jeu acquis au fil de l'histoire du Go. Ainsi, l'apprentissage machine, et les réseaux de neurones en particulier, possèdent deux grands atouts. Ils sont capables d'apprendre à généraliser et à extraire des structures très fines à partir de données considérables en un temps fonction seulement de la puissance de calcul et ils sont capables d'explorer profondément et sans assistance de très nombreux modèles. Mais malgré ces avantages, le *machine learning* n'est pas pour autant miraculeux et il reste encore beaucoup de progrès à espérer.

Il est un peu réducteur de croire que les réseaux de neurones simulent parfaitement les neurones biologique. En effet, les recherches en neurosciences montrent que ceux-ci font preuve de mécaniques extrêmement complexes, telles que la synchronisation temporelle, qui ne sont pas du tout exploitées par les réseaux de neurones. De plus, le cerveau humain est capable de généraliser et d'apprendre à partir de bien moins d'exemples qu'un réseau de neurones qui doit souvent ingérer des corpus de millions d'exemples avant d'être utile. De plus, nous sommes capables de transmettre notre compréhension de notre environ-

nement de manière efficace alors qu'il n'y a que peu de solutions comparables pour un réseau de neurones. Enfin, l'apprentissage en *machine learning*, comme beaucoup de problèmes d'optimisation, se heurte aux minima locaux, qui sont des ensembles de paramètres sub-optimaux, qui empêchent d'atteindre les minima globaux. Cela montre qu'il reste encore beaucoup à faire avant de pouvoir apprendre aussi efficacement que le modèle biologique.

Les résultats actuels des réseaux de neurones sont très impressionnants. De plus, on est loin d'avoir exploité tout le potentiel du *machine learning* qui va profiter de l'amélioration de la puissance de calcul mais aussi des nombreuses améliorations de la recherche dans l'exploration et l'apprentissage. Il n'y a, à priori, aucun problème qui ne semble être fondamentalement impossible à résoudre par l'IA. Dans tout les domaines, la suprématie intellectuelle de l'homme semble comptée.

Chapitre 2

Fondamentaux économiques

Dans ce chapitre, nous discutons en premier lieu des objectifs de l'économie, puis des principes fondamentaux qui régissent une économie de marché. Nous précisons ici que le type de système économique (par exemple capitaliste) se définit par la part d'influence que l'Etat a sur l'économie au sens défini ci-dessous. Il nous semble important de discuter de ces fondamentaux parce qu'il s'agit à notre avis d'un domaine où l'avènement de l'IA aura une grande influence.

2.1 Objectifs de l'économie

Nous nous essayons à une définition de l'économie en citant quelques-uns de ses objectifs. L'objectif de l'économie est l'accroissement du produit intérieur brut (PIB, valeur produite par les entreprises ainsi que par le gouvernement, plus consommation et imports-exports) et l'abaissement de l'inflation (augmentation des prix des biens consommables et services). En parallèle, elle cherche à augmenter le revenu du travail *per capita* (voir section 2.2.2) et à créer des conditions favorables afin d'attirer des investisseurs. Dès lors, une économie saine correspond à une économie où les investisseurs et producteurs prospèrent grâce à un environnement profitable, mais aussi où les individus jouissent d'un certain bien-être : haute rémunération en rapport au prix des biens. Notons ici que la rémunération des employés se doit de suivre l'inflation des prix afin que les individus continuent de jouir d'un confort minimal. Il existe plusieurs solutions au problème d'optimisation économique. Nous appelons ces solutions des *systèmes économiques*, et actuellement, le système qui prévaut est le capitalisme. Ce dernier suppose intrinsèquement que l'optimisation du bien-être collectif dépend de la recherche individuelle de profit : les individus chercheront à optimiser leur bien-être, et les entreprises voudront maximiser leurs profits (voir 2.2.1) ; en conséquence, l'augmentation de production et de capacité de développement permettra une prospérité globale.

Dans la suite de ce chapitre, nous allons voir comment le concept d'individu intervient

dans les modèles économiques, puis nous allons discuter de certaines interactions plus ou moins directes entre le système capitaliste et l'individu, et enfin nous allons voir comment ces interactions pourront être fondamentalement bouleversées par l'avènement de l'IA.

2.2 Fondamentaux des économies de marché

2.2.1 Micro-économie et *homo economicus*

Afin de pouvoir réaliser l'influence que pourrait avoir l'avènement d'une IA sur une société capitaliste, il est important de comprendre les principes de base qui régissent cette dernière. En particulier, nous nous intéressons ici à certains axiomes de la théorie micro-économique. La micro-économie "analyse le comportement des individus et des entreprises sur le marché afin de comprendre leurs choix liés à la consommation, l'offre et la demande, et la détermination des prix de divers produits (financiers ou de détail)"¹. Un ensemble de solutions au problème micro-économique a été développé à la fin du 19^e siècle et est regroupé dans ce que l'on nomme aujourd'hui la théorie néo-classique de l'économie. Non seulement, cette théorie est celle qui est en place, mais elle jouit également de l'acceptation de la part de la population académique et industrielle, malgré de nombreuses critiques à son sujet et le fait que des extensions sont toujours étudiées. Les trois hypothèses fondamentales sont² :

1. Les individus ont des préférences rationnelles en ce qui concerne les états de la nature qui sont identifiables et qui peuvent être associés à des valeurs.
2. Les individus maximisent leur utilité de consommation et les entreprises maximisent leurs profits.
3. Les acteurs agissent indépendamment les uns des autres et leurs décisions sont basées sur toute l'information existante et pertinente.

Le premier point sera au centre de notre analyse, mais avant d'en discuter nous éclaircissons les deux points suivants. L'optimisation des profits de la part des entreprises semble raisonnable ; celles-ci sont en quelque sorte des entités dont le but est de créer de la valeur ajoutée, autrement dit du profit pour le ou les propriétaires de l'entreprise. Quant aux individus, ceux-ci cherchent à maximiser leur "utilité de consommation"³. L'utilité est une notion abstraite dont le but est de modéliser les préférences des individus. En effet, il semble plausible qu'un individu préfère plus de richesse à moins, ou encore que l'apport d'utilité d'un kilogramme de fruits soit plus important pour une personne pauvre que

¹<http://www.investopedia.com/terms/m/microeconomics.asp>, visité le 1 mars 2016.

²E. Roy Weintraub, <http://www.econlib.org/library/Enc1/NeoclassicalEconomics.html>, visité le 1 mars 2016.

³Andriy Blokhin, <http://www.investopedia.com/ask/answers/072915/what-utility-function-and-how-it-calculated.asp>, visité le 26 avril 2016.

pour une personne riche ⁴. Ce sont ces deux comportements relationnels qui définissent l'utilité d'un individu. Une notion importante dérivant de celle-ci est l'aversion au risque. Selon la classe d'utilité, l'individu peut redouter le risque, être neutre vis-à-vis de celui-ci, ou encore preneur de risque. Le concept de risque est quelque chose sur quoi nous ne nous étendrons pas plus, car pour notre travail il suffit de comprendre que les individus cherchent à maximiser leur utilité.

Passons désormais à la troisième hypothèse. Celle-ci implique que les décisions que prennent les individus et les entreprises ne sont aucunement influencées par ce qui se passe sur le marché au même instant. Autrement dit, à un moment donné, lorsque les décisions doivent être prises (et en supposant qu'elles ont lieu au même moment pour tous les acteurs), personne ne peut échanger de l'information avec un autre acteur du marché. De plus, l'hypothèse implique que tout le monde a accès aux mêmes informations⁵, c'est-à-dire à toute information rendue publique par les entités sur le marché.

Finalement, la première partie de la première hypothèse est celle dont l'analyse nous concernera le plus. La deuxième partie de celle-ci fait allusion à l'existence de quantités (matérielles ou abstraites) permettant l'identification des états de la nature. Par exemple, un assureur pourrait être concerné par l'événement "client A se fait cambrioler", événement qui se traduit par des pertes pour sa compagnie s'il se produit, et des gains s'il ne se produit pas. Cet événement est associé à une probabilité d'occurrence, et son résultat à une quantité d'argent positive ou négative pour l'assurance. Autant la probabilité que l'argent sont des entités identifiables et quantifiables. La première partie définit ce que les scientifiques appellent communément *homo economicus* : "un individu rationnel théorétisé par les économistes afin de pouvoir établir des prédictions concrètes dans leurs modèles, et ainsi pouvoir les tester et les vérifier"⁶. L'hypothèse qui peut sembler la plus plausible est en fait celle qui pose le plus de problèmes. Entre autres, certains économistes comportementaux et neuroscientifiques critiquent ouvertement le concept d'individu rationnel⁷. Sans cette hypothèse, une grande majorité des modèles micro-économiques ne peuvent plus être appliqués, néanmoins, comme nous allons le montrer ci-dessous, l'avènement de l'IA pourrait remédier à ce problème en apportant aux économistes un nouveau type d'agent : *machina economicus*⁸.

⁴L'importance est ici définie comme l'apport de valeur d'un bien relatif à la richesse existante.

⁵Rappelons qu'il est illégal d'utiliser de l'information privée (*insider information*) sur un marché d'échanges.

⁶<http://www.investopedia.com/terms/h/homoeconomicus.asp>, visité le 1 mars 2016.

⁷Stefan Schneider, "Homo economicus - or more like Homer Simpson?", *Deutsche Bank Research*, 29 juin 2010, https://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD000000000259291.pdf, visité le 27 avril 2016.

⁸David C. Parkes et Michael P. Wellman, "Economic reasoning and artificial intelligence", *Science*, 17 juillet 2015, pp. 267-272.

2.2.2 Le système capitaliste

Un des piliers de l'économie des pays développés est le système capitaliste dans lequel elle est intégrée. En effet, celui-ci s'appuie sur la propriété privée des moyens de production et du capital. La production des biens et services est basée sur l'offre et la demande du marché au lieu d'être planifiée par une unité centrale⁹. Un des objectifs principaux dérivant du capitalisme est le profit du producteur. Comme nous l'avons expliqué précédemment, une des hypothèses de la micro-économie est que les entreprises maximisent leurs profits. Cette hypothèse n'est d'autre qu'une conséquence du modèle capitaliste.

Le système capitaliste est encore largement critiqué aujourd'hui, malgré les échecs successifs d'autres systèmes économiques pour des raisons variables qui ne sont pas le centre d'intérêt de ce travail. Néanmoins, ses auteurs sont d'avis que la raison principale de ces échecs a été la mauvaise gestion économique et sociale de l'unité centrale (Etat). De plus, un des défauts du système capitaliste est justement un manque de contrôle qui, dans les pires cas, a provoqué des crises, au début financières et locales, mais s'étendant par la suite globalement et socialement. Ci-dessous nous analyserons les apports que pourrait avoir une IA dans ce contexte économique global, mais avant cela nous discutons ici des implications micro-économiques et sociales observées¹⁰ du capitalisme.

D'un point de vue social, l'avènement de crises économiques aggrave en général le fossé entre les classes sociales (inégalités). Nous ne pensons pas que les crises économiques sont une conséquence du capitalisme, mais que ce dernier ne permet pas une régulation appropriée des marchés, par définition. Depuis l'éclatement de la bulle immobilière en 2008, les principaux gouvernements mondiaux ont tenté l'instauration de nouvelles régulations, mais l'animosité des acteurs sur le marché (principalement dans les pays anglo-saxons) a par exemple repoussé jusqu'en 2019 l'introduction des nouveaux accords sensés limiter le risque pris par les institutions financières. De plus, lorsque le gouvernement agit sur les marchés, par exemple lors des programmes d'allègement quantitatifs (*Quantitative Easing*), ceci crée une dévaluation de la monnaie qui, en conséquence, réduit le pouvoir d'achat des habitants de la zone monétaire. Cette réduction de pouvoir d'achat se ressent le plus fortement dans les classes moyennes et basses¹¹. Ces inégalités ne s'élargissent pas seulement le long de la hiérarchie de classes, mais aussi au travers des tranches d'âge de la population. En effet, la toujours plus grande nécessité d'accroissement des profits, et l'augmentation de la demande d'emploi implique une inégalité dans les revenus entre

⁹<http://www.investopedia.com/terms/c/capitalism.asp?layout=infini&v=4C&adtest=4C>, visité le 7 mars 2016.

¹⁰Nous nous situons dans un contexte empirique contemporain, et non dans un modèle théorique du capitalisme.

¹¹Myret Zaki, *La fin du dollar*, pp. 153-157, Éditions Favre, Lausanne, 2011

une personne jeune avec peu d'expérience et une personne âgée qui apporte plus à l'entreprise. Cet écart se creuse d'avantage à chaque génération et crée une situation de précarité chez les jeunes¹².

Revenons maintenant à quelques fondamentaux de l'économie et leur lien avec le système capitaliste. Dans le capitalisme à l'américaine, l'objectif principal des entreprises est celui de la maximisation des profits. Dans ce cadre-là, deux concepts font émergence : celui de concurrence entrepreneuriale et celui de revenu du travail (à l'opposé du revenu du capital). Sans concurrence, les profits pourraient en principe être incommensurables. En effet, une entreprise qui arrive à développer un produit dont le marché ne peut plus se passer, peut en principe choisir le prix de ce produit à sa propre convenance. En revanche, dès qu'il y a concurrence avec une autre entreprise développant le même produit et cherchant à maximiser ses profits, chaque compagnie se doit d'ajuster ses prix de telle sorte à maintenir une marge la plus grande possible. Dès lors, l'optimisation des coûts de production devient très importante, et ceux-ci prennent en compte le salaire des employés. Il y a donc d'une part, le revenu du travail qui correspond au prix que les entreprises sont prêtes à payer pour les services d'un employé dont les capacités permettent d'optimiser les profits de celles-ci, et d'autre part, le revenu du capital, concept fondamentalement lié au capitalisme et qui dépend directement de la dynamique des marchés économiques. Le premier est un concept qui a existé parallèlement au développement de la société depuis des millénaires, prenant selon la période une représentation physique différente : il correspond au concept de rémunération, nécessaire à la survie et à l'épanouissement de l'individu, correspondant à un apport mesurable de celui-ci à l'entité le rémunérant. Au contraire, le deuxième n'est qu'un artefact du capitalisme : son existence dépend de la possibilité d'investir dans un marché où les prix des avoirs est corrélé au bien-être de l'économie. Les entreprises n'ont que peu d'influence directe sur ce dernier, par contre, ce sont elles qui déterminent le revenu du travail correspondant à leur optimisation des profits. Ainsi, dans un système capitaliste, il est raisonnable de supposer que ceux-ci vont être minimisés.

En résumant, l'avènement de l'IA nous amène à considérer deux solutions possibles au problème économique. D'une part, une optimisation accrue d'un système capitaliste par une réduction des coûts des entreprises et une augmentation de l'efficacité des marchés grâce à son caractère de *machina economicus* ; et d'autre part, une évolution du système économique vers un système où le gouvernement aura une position plus importante et où la gestion se verra confiée à une IA afin d'éviter les échecs du passé.

¹²Barr Caelainn et Malik Shiv, "Revealed : the 30-Year Economic Betrayal Dragging Down Generation Y's Income", The Guardian, 7 mars 2016, <http://www.theguardian.com/world/2016/mar/07/revealed-30-year-economic-betrayal-dragging-down-generation-y-income>, visité le 26 avril 2016.

Chapitre 3

Le bouleversement apporté par l'IA

3.1 Un changement inéluctable

Les progrès technologiques ont toujours profondément changé nos sociétés à travers les ères. Le travail s'est transformé grâce aux nombreux bénéfices de l'automatisation. Nous sommes passés d'une société majoritairement agricole, vers une société à majorité ouvrière, à une société de cols blancs. Il serait tentant de croire que l'IA s'inscrit simplement dans cette liste, qu'elle est une rupture technologique, certes, mais pas la première. Les auteurs ont la conviction que jamais dans l'histoire l'homme n'a traversé une transition technologique de cette ampleur. Nous ne sommes pas préparés, comment le pourrions-nous ?

Les scientifiques eux-mêmes se surprennent toujours du rythme technologique. Nombreux sont les exemples d'éminents scientifiques qui ont sous-estimés son rythme. La puissance de calcul grandit toujours de manière exponentielle, suivant la prédiction de la Loi de Moore. L'exponentialité est un concept très difficile à appréhender tant il est contre-intuitif.

Les avancées récentes ont été permises par les réseaux de neurones profonds. Les réseaux de neurones profonds permettant le *Deep Learning* ont ceci de magique et fascinant que même leurs chercheurs ont une compréhension très incomplète de son efficacité surprenante ¹. Et nous n'en sommes qu'à ses balbutiements.

Le progrès technologique et l'automatisation n'ont eu de cesse de prouver leurs nombreux bienfaits. Jamais auparavant, l'homme n'a pu bénéficier d'une espérance de vie si grande,

¹Andrej Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks", blog, 21 mai 2015, <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>, visité le 7 mars 2016.

de conditions de vie si bonnes et d'un accès à la connaissance si facile. La pauvreté mondiale ne cesse de reculer et ce même alors que la population croît toujours à un rythme soutenu. La technologie a battu Malthus.

Le vivant et l'homme sont des merveilles biologiques incroyablement versatiles, capables de réaliser un grand nombre d'opérations. Malgré cela, nous ne pouvons plus rivaliser avec la machine et ce depuis bien longtemps. Les muscles mécaniques travaillent sans relâche, avec une puissance immense, et leur forme s'adapte à leur fonction pour les rendre toujours plus spécialisés et donc, toujours plus performants.

Les réseaux de neurones et plus généralement l'IA sont un autre type d'automatisation. Elles remplacent notre travail intellectuel. Au fur et à mesure que les travaux manuels ont été laissés à la machine, les sociétés se sont attelées à des tâches de plus en plus intellectuelles. Ces tâches se reposaient sur notre pouvoir créatif, analytique ou sur notre capacité à superviser la machine dans son effort. Mais l'IA permet maintenant de nous remplacer dans le rôle de cavalier. Il est très illusoire de croire que nous pourrions trouver d'autres emplois permettant de compenser la masse salariale remplacée juste parce qu'il le faut. Il paraît difficile de concevoir une société où nous serions tous chercheurs, poètes ou programmeurs d'intelligence artificielle. Il serait bien provocateur d'appeler l'homme obsolète, mais il serait bien déraisonnable de croire en une loi économique immuable et invisible de création ou de déplacement d'emplois. Elle s'est simplement avérée vraie empiriquement par le passé car la machine n'avait pu concurrencer l'homme que dans ses capacités physiques. Si l'IA n'envisage pas de remplacer l'homme complètement tous les métiers, elle promet néanmoins de le soulager des métiers les plus répétitifs et les plus chronophages. Il n'existe très peu, voire aucun métier qui puisse se targuer d'être indéfiniment résistant à l'automatisation ². Même la création artistique n'est pas tout à fait à l'abri. Il existe des robots compositeurs de musique qui produisent des résultats indifférenciables de ceux d'un compositeur de chair et de sang ³.

Le remplacement des emplois se fera de manière graduelle, des emplois les moins qualifiés vers les plus qualifiés. Il existe déjà une partie de la population remplissant un emploi qui pourrait dès à présent être en partie ou complètement remplacée. Ainsi, les voitures autonomes pourront bientôt remplacer une large partie des emplois du transport. Des camions pourraient ainsi sillonner les continents sans pause, de jour comme de nuit, et sans demander de salaire. Les secrétaires se voient remplacés par des assistants personnels sur *smartphone* et dans le *cloud*. L'IA est capable de chercher et de trouver des

²Nassos Stylianou, Tom Nurse, *et al.*, "Will a Robot Take Your Job?", *BBC*, 11 septembre 2015, <http://www.bbc.com/news/technology-34066941>, visité le 27 avril 2016.

³William Hochberg, "When robots write songs", *The Atlantic*, 7 août 2014, <http://www.theatlantic.com/entertainment/archive/2014/08/computers-that-compose/374916/>, visité le 27 avril.

similitudes ainsi que de comprendre des millions de documents à la seconde, plus vite et plus précisément que les para-juristes. De plus, l'IA n'a pas besoin d'être absolument parfaite, il lui suffit simplement d'être plus efficace que les hommes qu'elle remplace. Par exemple, chez les médecins, le nombre de diagnostics erronés entraînant des effets dramatiques est considérable : En 1991, une étude d'Harvard montrait que environ 15% d'effets indésirables provenaient de diagnostics erronés ⁴. Watson, le robot-médecin d'IBM, permet, selon l'historique du patient et ses symptômes, de faire un diagnostic parmi les millions de diagnostics potentiels et de prescrire des médicaments en tenant compte du nombre immense de contre-indications. Il fait peu de doute que nous serions plus facilement prêts à remplacer notre médecin par Watson si cela nous permettait d'obtenir des diagnostics plus précis, plus personnalisés, et plus rapides. Ces exemples sont les plus frappants car ce sont des domaines qui vont devoir affronter beaucoup de réticences. Le contact humain du médecin peut sembler primordial. Il est vrai qu'à compétences égales, il est plus réconfortant de ne pas laisser son destin reposer sur les décisions d'une machine. Mais à l'instar des caissières automatiques remplaçant peu à peu leurs homologues humains, la transition se fera petit à petit, à mesure que l'écart de compétence se creusera. Ce sont des exemples parmi tant d'autres, et aucun emploi ne semble à l'abri, mis de côté les problèmes éthiques liés à certaines responsabilités. Les taux de chômage vont croître rapidement et largement.

3.2 Acceptation sociale de l'IA

Si l'on admet l'avènement graduel de l'IA, nous devons nous intéresser à la question de la possible acceptation ou du refus social. En effet, en ce qui concerne les nouvelles technologies qui peuvent avoir des répercussions importantes au niveau global, il se peut que la population s'oppose à leur établissement, ou du moins qu'elle n'y soit pas favorable. Ce refus peut avoir plusieurs origines, néanmoins nous nous intéresserons ici plus particulièrement à la peur et à la méfiance associée aux nouvelles technologies.

Concernant l'IA, depuis longtemps scientifiques et autres philosophes et personnalités ont fait part de leur inquiétude quant au comportement que celle-ci pourrait avoir. Nous citons un article publié dans *The Huffington Post* :

“Elles (les IAs) se concentreront sur leur auto-défense et sur la recherche de ressources leur permettant d'atteindre leur but le plus facilement possible. Elles se battront contre nous afin de survivre, et elle ne voudront pas être éteintes.”⁵

⁴L.L. Leape, T.A. Brennan, N. Laird, *et al.*, “The Nature of Adverse Events in Hospitalized Patients. Results of the Harvard Medical Practice Study II”, *N Engl J Med*, février 1991, pp. 377-384.

⁵James Barrat, “Why Stephen Hawking and Bill Gates Are Terrified of Artificial Intelligence”, *The Huffington Post*, 4 septembre 2015, visité le 5 avril 2016.

L'article dont cette citation est tirée cite une étude publiée par un scientifique spécialisé dans la recherche sur l'IA. Il existe une multitude d'études de ce genre qui font part de leur inquiétude concernant le comportement que pourrait avoir une IA. A en croire les médias et bon nombre d'oeuvres de science-fiction⁶, en première instance, que le développement d'une IA ne soit pas une si bonne idée que cela, en tenant compte du fait que nous aimerions garder un certain contrôle sur celle-ci. Notre but est ici de préciser que ces personnes ne discutent que des cas extrêmes où l'IA est créée et laissée à elle-même, et non pas d'autres possibilités telles que l'implémentation d'une IA ayant certains *firewalls* l'empêchant d'agir à sa volonté. En effet, il n'y a, à priori, aucune raison de donner accès à une IA aux codes de lancement nucléaire d'un gouvernement ; des mesures peuvent être prises pour qu'elle ne puisse pas y accéder même si elle le "désirait". Dès lors, il nous semble que cette peur de l'IA tombe dans la catégorie de la paranoïa, comme le furent les "machines auto-mobiles" au cours du 18^e siècle. En effet, le développement des machines automobiles provoqua un sentiment de peur au sein d'une grande partie de la population : des objets sensés être inertes, non seulement commencèrent à se mouvoir d'eux-mêmes, mais aussi devinrent bruyants. Le parallèle peut être fait avec le développement de l'IA : celle-ci serait une "machina sapiens", elle serait douée d'une certaine intelligence et pourrait prendre des décisions de par elle-même. La science-fiction a réussi à inculquer une méfiance prononcée envers ces machines pensantes, et dès lors nous pouvons nous attendre à une certaine crainte lors de la possible implémentation de l'IA dans la société. Nous pensons que malgré quelques voix rugissant leur inquiétudes, lorsque l'avènement de l'IA aura lieu, celle-ci prendra gentiment place au sein de notre société. La question qui se pose désormais est de savoir dans quels domaines les gens seront prêts à l'accepter.

Afin d'illustrer cette tension, nous allons prendre deux exemples simples. En premier lieu, imaginons un travail de type industrie à la chaîne qui, aujourd'hui, nécessite qu'un employé s'occupe constamment de la machine à laquelle il est assigné. Mis à part les motifs syndicaux ou personnels⁷, si ce salarié est remplacé par une IA qui arrive à gérer la production de la machine de manière optimale, il n'y aurait pas, à notre avis, d'opposition sociale à ce remplacement. D'autre part, et comme deuxième exemple, prenons le cas d'un chirurgien cardiaque, qui doit prendre des décisions, parfois morales, souvent très délicates, lors de l'opération de l'un des ses patients. Certes, beaucoup de décès sont causés par une erreur dite "humaine" (manque d'expérience ou de connaissances, stress, manque de sommeil, etc), mais il semblerait (du moins en tenant compte de nos connaissances actuelles) que dans certains cas la décision du médecin, qui sem-

⁶Un exemple typique serait celui de *The Terminator*, 1985. Plus récemment, le film *Ex Machina*, 2015, reprend la problématique de la perte de contrôle sur l'IA.

⁷Il est logique que l'employé puisse ne pas vouloir perdre son travail, mais nous ne traitons pas ce problème ici

blait illogique, de continuer une opération, ait sauvé la vie de plus d'un patient. Dès lors, comment mesurons-nous l'apport qu'aurait une IA sur une table d'opération, et malgré tout, si nous acceptons le fait qu'une IA aura au final de bien plus grandes capacités chirurgicales que le médecin le mieux qualifié, les gens seront-ils prêts à mettre leur vie entre les mains d'un programme informatique ?

Ces deux exemples résument le dilemme qu'affrontera la société lors de l'avènement de l'IA. Il faudra peut-être utiliser cette nouvelle technologie de manière spécifique et réglementée afin, d'une part de garder le contrôle sur une "entité" dont le comportement peut être impossible à prédire, et d'autre part, obtenir une acceptation sociale afin que son utilisation rentre dans les mœurs. Nous supposons que ce n'est que de cette manière que l'IA pourra s'intégrer à la société humaine de manière optimale.

Dans le contexte économique actuel, serons-nous capable de supporter un tel chômage présagé ? Les crises économiques profondes apparaissent dès que les taux de chômage dépassent 25% et il fait nul doute que ce pourcentage sera franchi dans un futur proche. Les inégalités vont fortement se creuser entre les groupes accompagnant ce progrès, et les populations le subissant. Dans la première catégorie, nous avons les ingénieurs et scientifiques, les managers, et les détenteurs de capital, et dans la seconde le reste de la population. En effet, l'IA relevant de la propriété intellectuelle, le capital sera toujours à même de créer un revenu. Par opposition, les emplois disparaissant, les revenus du salaires seront moindre car réservés à une plus petite partie de la population.

Dès lors, il paraît nécessaire de repenser les fondamentaux économiques de notre civilisation. Après tout, la conception du travail comme une valeur en soi n'est que assez récente. En effet, pendant longtemps en Europe, le teint blanc était très prisé car, il signifiait la non-appartenance à la classe laborieuse. De plus, les citoyens grecs réservaient le travail aux esclaves et s'adonnaient à l'artisanat, la politique ou l'art de la guerre. Certains se permettent d'ailleurs de tracer des parallèles entre les travailleurs les moins éduqués et les plus fragiles, condamnés aux petits boulots et les Hilotes, les esclaves grecs non-marchandise des spartiates ; et de dénoncer une forme d'esclavage moderne. Ne devrions-nous pas nous réjouir de la libération offerte par l'automation. Le travail est-il une fin ? Sans nier le sentiment de satisfaction que peut procurer un projet, ou un emploi, il ne faut pas oublier que l'emploi est source de souffrance pour une partie non négligeable de la population. Après tout, la racine étymologique de travail est *tripalium*, un instrument de torture. Ainsi, on peut envisager qu'un certain nombre d'emplois que l'on pourrait qualifier d'enrichissants pour les individus qui les effectuent, ne disparaissent pas mais se mutent simplement en bénévolats et travaux associatifs au service de la population. Ainsi la motivation première de tout individu serait sa réalisation à

travers ses activités et non pas le degré de confort qu’il peut espérer de sa rémunération. De plus, être un politicien ne pourrait ne plus être considéré comme un travail réservé à des professionnels et nous deviendrions tous plus impliqués à des degrés divers dans la vie politique et la “vie de la cité”.

3.3 Quelles solutions ?

Le revenu de base universel est une des solutions envisagées pour préparer les sociétés à faire face à ces changements radicaux. Elles promet un revenu supplémentaire à tout individu, sans discrimination aucune, ni même de fortune ou d’emploi. Le montant doit être suffisant pour permettre à tous de vivre dignement. Ce revenu ferait disparaître la nécessité d’emplois alimentaires, éradiquer la pauvreté et permettre à chacun de poursuivre son enrichissement personnel si tel est son souhait. Dans une société d’abondance, ou tout les produits de première nécessité existent en grande quantité, il paraît raisonnable de pouvoir en allouer une partie divisée universellement. Ce revenu de base universel pourrait être rendu possible par un *Quantitative Easing* direct qui consisterait à créditer directement les comptes des citoyens par le biais de la création monétaire et participer au renforcement de l’unité centrale (Etat). Le revenu de base universel amène de nombreuses questions, mais le rejeter sans proposer d’alternative est dangereux. La transformation de la société par l’IA se fera rapidement, dans tout les secteurs et dans un futur plus proche que l’imaginaire collectif ne le conçoit. Le rythme des technologies s’accélère, celui de l’IA a fait des progrès inespérés ces dernières années. Pour beaucoup, nous étions encore bien loin de pouvoir espérer surpasser l’homme au Go. C’est maintenant fait, et demain, quelle tâche résistera indéfiniment à la puissance de l’IA ?

Conclusion

En premier lieu, il nous paraît très probable qu’une révolution technique soit en marche. Dans une société où le rythme des technologies s’accélère toujours plus, l’intelligence artificielle peut espérer inéluctablement approcher, égaler puis dépasser les capacités intellectuelles de l’homme dans un avenir en somme plus si lointain.

En second lieu, nous discutons des fondamentaux économiques qui régissent notre société. Il est important de comprendre comment le système fonctionne afin de pouvoir conclure sur les apports de l’IA. Nous avons discuté du fait que la théorie moderne de l’économie se base sur l’hypothèse d’existence de l’*homo economicus*, mais que cette dernière est supposée erronée. De plus, un système capitaliste se base sur une optimisation de la richesse qui passe par un accroissement des revenus et une diminution des coûts. L’avènement de *machina economicus* permettra une application plus satisfaisante de ces théories ainsi qu’une optimisation accrue du modèle économique capitaliste, tout en permettant un contrôle plus poussé de l’économie de marché. Néanmoins, nous pensons que son influence pourra être plus grande ; nous avons, en effet, mentionné les limitations du système capitaliste quant à la gestion des crises financières et sociales. Au lieu d’une économie capitaliste où le gouvernement n’a que peu d’influence⁸, nous développons dans le troisième chapitre un point de vue où celui-ci a une place plus importante, et où la gestion de l’IA permettrait un épanouissement de la société.

Enfin, nous examinons les transformations de la société par l’émergence de l’IA. Cette révolution d’origine technique doit s’accompagner d’une révolution économique et sociale, sans quoi nous pourrions risquer une implosion de la société. En tant que solution, nous avons mentionné le revenu universel accompagné d’un système moins capitaliste, mais beaucoup d’autres sont encore à inventer. L’IA nous apparaît comme un outil extrêmement puissant dont il faudra faire un usage sage si l’on veut l’orienter au bénéfice de la société toute entière. Le premier pas pour accompagner ces changements, nous y adapter harmonieusement et embrasser leur potentiel, est de prendre conscience de leur imminence et de leur nécessité.

⁸Rappelons qu’un des objectifs de l’économie est d’accroître l’apport au PIB des entreprises ou du gouvernement.

Bibliographie

- James Barrat, “Why Stephen Hawking and Bill Gates Are Terrified of Artificial Intelligence”, *The Huffington Post*, 4 septembre 2015, http://www.huffingtonpost.com/james-barrat/hawking-gates-artificial-intelligence_b_7008706.html, visité le 9 mars 2016.
- Ben Goertzel et Cassio Pennachin, *Artificial General Intelligence* (PDF), Springer, ISBN 3-540-23733-X, 20 mars 2013, (2006).
- Thomas E. Graedel *et al.*, “Criticality of Metals and Metalloids”, *Proceedings of the National Academy of Sciences of the United States of America* 112.14, 2015.
- Ray Kurzweil, *The Singularity is Near*, New-York, Viking Press, 2015.

Dans la première partie de notre travail, nous nous intéressons entre autres à la plausibilité technique de l’IA. Cette oeuvre du futuriste Kurzweil soutient la thèse que les progrès technologiques sont exponentiels. Elle fait la prédiction d’une puissance de calcul suffisante pour dépasser l’intelligence humaine combinée dès 2045. Bien sûr, la puissance de calcul ne fait pas une intelligence artificielle à elle seule. Mais Kurzweil mise sur le développement des capacités d’exploration du cerveau afin de l’imiter et d’ainsi mettre à profit la puissance de calcul et une mémoire comparable à celle du cerveau. Kurzweil propose ici un agenda concret se basant sur les progrès technologiques passés. Celui-ci justifie les spéculations de la première partie.

- Shane Legg et Joel Veness, “An Approximation of the Universal Intelligence Measure”, *arXiv* eprint 1109.5951, septembre 2011.
- Jean-Pierre Changeux, “L’Inné et l’Acquis dans la Structure du Cerveau”, *La Recherche*, mai 2000, <http://www.larecherche.fr/actualite/aussi/inne-acquis-structure-du-cerveau-texte-integral-01-05-2000-71955>, visité le 27 avril 2016.

- Christopher Olah, “Deep Learning, NLP, and Representations”, blog, 7 juillet 2014, <http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/>, visité le 27 avril 2016.
- Saxena Shalini, “AI Masters 49 Atari 2600 Games Without Instructions”, *ars technica*, 27 février 2015, <http://arstechnica.com/science/2015/02/ai-masters-49-atari-2600-games-without-instructions/>, visité le 27 avril 2016.
- David C. Parkes et Michael P. Wellman, “Economic Reasoning and Artificial Intelligence”, *Science*, 17 juillet 2015, pp 267-272.

Dans la deuxième partie de notre travail, nous discutons des fondements de l'économie de marché ainsi que du bouleversement supposément apporté par l'IA. Cet article publié dans *Science* en juillet 2015 par Parkes (Harvard) et Wellman (Michigan) discute, d'une part, de la validité de l'hypothèse fondamentale en théorie néoclassique de l'économie qui postule l'existence d'agents rationnels participant aux marchés (*homo economicus*) et d'autre part, du développement d'une *machina economicus*, une intelligence artificielle qui, par construction, serait l'agent rationnel décrit dans les susdits modèles. Leur analyse démontre que nous pouvons déjà aujourd'hui observer une certaine influence de l'IA sur les marchés, ne serait-ce que via la masse d'échanges due au trading algorithmique, et que le développement d'une *machina economicus* pourrait même faire évoluer nos modèles économiques afin de mieux décrire une interaction dite “multi-agents”. Nous utilisons cet article de référence afin de justifier la plausibilité économique d'une société gérée par une (ou des) IA(s) et la transition vers des modèles économiques optimaux qui peuvent fonctionner grâce à l'existence de cette dernière.

- Stefan Schneider, “Homo economicus - or more like Homer Simpson?”, *Deutsche Bank Research*, 29 juin 2010, https://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD0000000000259291.pdf, visité le 27 avril 2016.
- Barr Caelainn et Malik Shiv, “Revealed : the 30-Year Economic Betrayal Dragging Down Generation Y's Income”, *The Guardian*, 7 mars 2016, <http://www.theguardian.com/world/2016/mar/07/revealed-30-year-economic-betrayal-dragging-down-generation-y-income>, visité le 26 avril 2016.
- L.L. Leape, T.A. Brennan, N. Laird, *et al.*, “The Nature of Adverse Events in Hospitalized Patients. Results of the Harvard Medical Practice Study II”, *N Engl J Med*, février 1991, pp. 377-384.

- R. Sutton et A. Barto, *Reinforcement Learning : An Introduction*, Cambridge MA, MIT Press, 1998.
- Volodymyr Mnih, *et al.*, *Human-Level Control Through Deep Reinforcement Learning*, *Nature*, 2015, pp. 529-533.
- Volodymyr Mnih, *et al.*, *Playing Atari with Deep Reinforcement Learning.*, *arXiv* eprint 1312.5602, décembre 2013.

Cette référence nous intéresse dans la première partie car elle décrit comment techniquement égaliser l'homme dans la tâche à priori complexe de jouer à des jeux ATARI 2600. Le tour de force est de ne donner en entrée au réseau que les pixels qui composent l'image et aucune autre information. Ils font usage des techniques les plus modernes de *reinforcement learning* et de *deep learning* tout en apportant leurs innovations propres comme l'*experience replay buffer*. Sur certains des jeux de la plateforme, le réseau de neurones atteint des performances super-humaines, c'est-à-dire qu'il dépasse les performances du meilleur joueur humain à ce jour. Sur d'autres titres de jeu où le réseau est moins efficace, les auteurs notent qu'il serait envisageable de modifier le réseau afin de simuler une mémoire au sein de la structure du réseau de neurone afin d'imiter la mémoire à court terme de l'être humain.

- Andrej Karpathy, "The Unreasonable Effectiveness of Recurrent Neural Networks", blog, 21 mai 2015, <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>, visité le 7 mars 2016.
- Jerry Kaplan, *Humans Need Not Apply : A Guide to Wealth and Work in the Age of Artificial Intelligence*, New Haven, Yale University Press, 2015.
- Erik Brynjolfsson, *Race Against The Machine : How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy*, Digital Frontier Press, Kindle Edition, 2011.
- Federico Pistono, *Robots Will Steal Your Job, But That's OK : How to Survive the Economic Collapse and Be Happy*, CreateSpace, Kindle Edition 2014.
- "The Dawn of Artificial Intelligence", *The Economist*, 9 mai 2015, <http://www.economist.com/news/leaders/21650543-powerful-computers-will-reshape-humanitys-future-how-ensure-promise-outweighs>, visité le 9 mars 2016.

- Nassos Stylianou, Tom Nurse, *et al.*, “Will a Robot Take Your Job?”, *BBC*, 11 septembre 2015, <http://www.bbc.com/news/technology-34066941>, visité le 27 avril 2016.
- William Hochberg, “When Robots Write Songs”, *The Atlantic*, 7 août 2014, <http://www.theatlantic.com/entertainment/archive/2014/08/computers-that-compose/374916/>, visité le 27 avril.
- “The Future Availability of Natural Resources”, The World Economic Forum, *World Scenario Series*, novembre 2014, <https://www.weforum.org/reports/future-availability-natural-resources/>, visité le 27 avril 2016.
- Myret Zaki, *La fin du dollar*, Editions Favre, Lausanne, 2011.
- James Cameron, *The Terminator*, 1984.
- Alex Garland, *Ex Machina*, 2015.