ARTIFICIAL INTELLIGENCE ETHICS: FROM EMERGING STANDARDS TO PERSPECTIVES OF IMPLEMENTATION

Proceedings
International Colloquium
Rabat, July 22nd & 23rd, 2022



المجلس الوطني لحقوق الإنسان
ⴰⵙⵇⵇⵉⵎ ⴰⵏⴰⵎⵓⵔ ⵏ ⵉⵣⵔⴼⴰⵏ ⵏ ⵓⴼⴳⴰⵏ
Conseil national des droits de l'Homme



ARTIFICIAL INTELLIGENCE ETHICS: FROM EMERGING STANDARDS TO PERSPECTIVES OF IMPLEMENTATION

Proceedings
International Colloquium
Rabat, July 22nd & 23rd, 2022

TABLE OF CONTENTS

Concept note 4
Allocution de Madame Amina Bouayach 9
Allocution de Monsieur Karim Hendili 11
Mot d'ouverture de Monsieur Abdellatif Miraoui 16
Mot d'ouverture de Monsieur Amine Mounir Alaoui 20
Mot d'ouverture de Monsieur Omar Seghrouchni 25
Mot d'ouverture de Mme Sarah Lamrani 30

AI ethics, global, regional and national initiatives 32
Amal El Fallah Seghrouchni
Lecture dans les standards émergents en matière de l'éthique de l'IA 33
Céline Caira
OECD Principles on Artificial Intelligence 40
Diane Galbois-LeHalle
Approche européenne pour une régulation éthique et juridique de l'IA : une approche fondée sur les risques 42
Mark Gaffley
Code and codification: the possibility of rights-respecting AI 48
John C. Havens
IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems 52

AI at the service of Human Right 54
John Schwartz
AI in lifelong learning 55
Workineh Kelbessa
Artificial Intelligence and the Environment 57
Raja Bensaoud
L'intelligence artificielle sous l'angle du droit 63
Abdelmajid Elouadi
Industrie 4.0 et applications de l'IA 68

Ethics of AI, state of initiatives in Morocco 72
Zakaria Rachid
Approche nationale en matière d'organisation du digital et de l'IA 73
Saida Belouali
L'éthique de l'IA dans l'action de la maison de l'IA à Oujda 76
Abdessalam Jaldi
Le débat sur l'intelligence artificielle au Maroc : le besoin de concilier entre le progrès

et la protection des libertés et droits fondamentaux 79
Salah Baïna
Université – Entreprise : quelle transition digitale ? 87

AI Ethical Codes, Discussion and Implementation Prospects 94

Ravit Dotan
AI Ethics for Investors 95
Jean Pierre Llored
De la bioéthique aux codes éthiques de l'IA : limites et perspectives 97
Ayman Boughanmi
The ethical dilemma of AI: between irresistible evolutions and improbable regulations 114
Nicolás Duque Buitrago
Intelligence artificielle et droits de l'Homme : une réflexion éthique sur l'universalisme et le Modus vivendi 120
Yves Demazeau
Conception éthique de systèmes d'IA de très grande taille 134
Elizabeth D. Gibbons
Critique of AI codes of ethics from a human rights perspective 139
Malik Boumediene
Existe-t-il un Droit International de l'Intelligence Artificielle ? 141
Saida Belouali
Toumi Bouchentouf
Tension entre automatisation et responsabilisation en éthique appliquée à l'Intelligence Artificielle 155

CONCEPT NOTE

The National Human Rights Council (CNDH) has the deep-seated conviction that human rights are pertinent to the field of Artificial Intelligence. It also encourages the ongoing comprehensive reflection and analysis within the national, regional and global context.

The various initiatives being developed by international, regional, and national bodies in an extensive array of sectors prove that Artificial Intelligence is undeniably one of the mechanisms that facilitate access and the enjoyment of fundamental rights and freedoms.

Withal, the use of Artificial Intelligence is not devoid of risks. Its impacts on human beings' lives have been a concern for the Council, where digitization is a lever for societies' growth and evolution nowadays. Indeed, the use of AI already affects human rights and raises preoccupations about private life, education, employment, healthcare, social care, health, mobility, law enforcement, and even the maintenance of law and order. Therefore, it is critical for institutions concerned with the prevention of human rights violations, protection of victims, and promotion of values and principles of human rights to assist and monitor the development and implementation of Artificial Intelligence.

This has driven the CNDH to engage in an international reflection on AI and human rights, starting from the four following fundamental determinants:

- Promote technological development and use of AI in Morocco, in line with a constructive approach to human rights and the values of a democratic society;
- Study and adequately assess Artificial Intelligence development, use, and its impacts on human rights;
- Hold stakeholders accountable for the different uses of AI;
- Enable human beings to benefit from the possibilities, opportunities, and advantages of using AI in full respect of human rights.

Since the end of 2019, the Council Council has initiated wide-ranging consultations among stakeholders including experts, researchers, civil society, companies, government institutions, etc.

This process aims to follow the development of AI in Morocco within the context of achieving a balance between technological progress, on the one hand, and the protection and promotion of human rights on the other hand, as stated by Amina Bouayach, the CNDH Chairperson.

In May 2020, an interactive platform on the right to freedom of opinion and expression in the digital space TaabiratRaqmya.ma went live. This platform aimed to interact with experts and the public opinion about this right and its restrictions, through the international principles establishing the freedom of opinion without harassment and the freedom to seek, receive and share news, information and ideas with others by all means and regardless of borders.

In April 2021, the first seminar on "Artificial Intelligence and Digital Citizenship" was held as part of the CNDH national consultations on artificial intelligence interactions with human rights. This seminar primarily focused on the possibilities given to algorithms and AI to make decisions. It also paid particular attention to the potential impacts of certain uses of AI and violations of fundamental rights and freedoms that may occur.

The Council organized, beginning of December 2021, the international seminar "Artificial Intelligence and Human Rights – Organization, Ethics, and Guiding Principles: an international benchmark" to open up the debate to global and regional initiatives and exchange experiences, as well as standards, guidelines, regulations and governing principles in the field.

Moroccan and foreign stakeholders representing institutions, universities, schools, centers, and enterprises involved in scientific research related to new technologies and the use of AI took part in this event.

On this basis, national and international experts adopted the Declaration of Rabat on "Artificial Intelligence and Digital Citizenship - For an Artificial Intelligence respecting Human Rights" and also declared the constitution of the scientific committee to provide consultation, support, and guidance on the implementation of the recommendations announced in the Declaration of Rabat.

In this Declaration, the National Human Rights Council committed to organizing an international colloquium in July 2022.

Hence, the CNDH is organizing along with its academic partners the international colloquium entitled "Artificial Intelligence ethics: from emerging standards to perspectives of implementation" which will take place in Rabat from July 22nd to 23rd, 2022 at the Council's headquarters.

This colloquium will gather international speakers from Africa, South America, North America, and Europe to further discuss available standards and tools to better assist in evaluating the development of AI that respects and protects fundamental rights and freedoms.

It will also be an opportunity to deepen the debate, exchange further expertise and share best practices about the key challenges for implementing ethical Artificial Intelligence such as upgrading the legal and organizational frameworks, the regulation of the education, research, business and development sectors, the composition and prerogatives of control and monitoring bodies.

NOTE CONCEPTUELLE

Le Conseil national des droits de l'Homme (CNDH) est profondément convaincu de l'importance qui échoit aux droits de l'Homme dans le domaine de l'Intelligence Artificielle. Il encourage également une réflexion et une analyse approfondies continues dans le contexte national, régional et global.

Les diverses initiatives développées par les instances internationales, régionales et nationales dans un large éventail de secteurs prouvent que l'Intelligence Artificielle est indéniablement l'un des mécanismes à même de faciliter l'accès et la jouissance des droits et libertés fondamentaux.

Cependant, l'utilisation de l'intelligence artificielle n'est pas dénuée de risques. Ses impacts sur la vie des êtres humains sont une préoccupation pour le Conseil, tandis que la digitalisation est aujourd'hui un levier de croissance et d'évolution des sociétés. En effet, l'utilisation de l'IA affecte déjà les droits de l'Homme et soulève des inquiétudes concernant la vie privée, l'éducation, l'emploi, l'environnement, les soins de santé, la protection sociale, la santé, la mobilité, l'application de la loi et même le maintien de l'ordre. Par conséquent, il est essentiel que les institutions concernées par la prévention des violations des droits de l'Homme, la protection des victimes et la promotion des valeurs et des principes des droits de l'Homme, accompagnent et surveillent le développement et la mise en œuvre de l'Intelligence Artificielle.

Ceci a amené la CNDH à s'engager dans une réflexion internationale sur l'IA et les droits humains, à partir des quatre déterminants fondamentaux suivants :

- Promouvoir le développement technologique et l'utilisation de l'IA au Maroc, en accord avec une approche intégrée des droits de l'Homme et les principes de la démocratie;
- Étudier et évaluer de manière adéquate le développement et l'utilisation de

l'Intelligence Artificielle, ainsi que ses impacts sur les droits de l'Homme ;

- Tenir les parties prenantes responsables des différentes utilisations de l'IA;
- Permettre aux êtres humains de bénéficier des possibilités, opportunités et avantages de l'utilisation de l'IA dans le plein respect des droits de l'Homme.

Depuis la fin de l'année 2019, le Conseil a initié de larges consultations auprès des parties prenantes, notamment des experts, des chercheurs, la société civile, des entreprises, des institutions gouvernementales, etc.

Ce processus vise à suivre le développement de l'IA au Maroc dans le cadre de la réalisation d'un équilibre entre le progrès technologique, d'une part, et la protection et la promotion des droits de l'Homme, d'autre part, tel que déclaré par Amina Bouayach, la Présidente du CNDH.

En mai 2020, une plateforme interactive sur le droit à la liberté d'opinion et d'expression dans l'espace numérique « TaabiratRaqmya.ma », a été mise en ligne. Cette plateforme visait à interagir avec des experts et l'opinion publique sur ce droit et ses restrictions, à travers les principes internationaux établissant la liberté d'opinion sans harcèlement et la liberté de chercher, recevoir et partager des informations et des idées avec d'autres personnes par tous les moyens et indépendamment des frontières.

En avril 2021, le premier séminaire placé sous la thématique: "Intelligence artificielle et citoyenneté numérique" s'est tenu dans le cadre des consultations nationales du CNDH sur les interactions entre l'intelligence artificielle et les droits humains. Ce séminaire a principalement porté sur les possibilités données aux algorithmes et à l'IA de prendre des décisions. Il a également porté une attention particulière aux impacts potentiels de certains usages de l'IA et aux violations des droits et libertés fondamentales qui peuvent en découler.

Le Conseil a ensuite organisé, début décembre 2021, le séminaire international "Intelligence artificielle et droits Humains – Organisation, éthique et principes directeurs : un benchmarking international" pour ouvrir le débat aux initiatives mondiales et régionales et échanger les expériences, ainsi que les normes, les lignes directrices, les règlements et les principes directeurs dans le domaine.

Des intervenants marocains et étrangers représentant des institutions, des universités, des écoles, des centres de recherche et des entreprises impliqués dans la recherche scientifique liée aux nouvelles technologies et à l'utilisation de l'IA ont pris part à cet événement.

Sur cette base, les experts nationaux et internationaux ont adopté la déclaration de Rabat sur "l'Intelligence Artificielle et la citoyenneté numérique – Pour une intelligence artificielle respectueuse des droits de l'Homme" et ont également déclaré la constitution du comité scientifique chargé d'accompagner, orienter et soutenir la mise en œuvre des recommandations annoncées dans la déclaration de Rabat.

Dans cette déclaration, le Conseil national des droits de l'Homme s'est engagé à coordonner l'organisation d'un colloque international en juillet 2022.

Ainsi, le CNDH organise avec ses partenaires académiques le colloque international intitulé "Ethique de l'Intelligence Artificielle : des standards émergents aux perspectives de mise en œuvre" qui aura lieu à Rabat du 22 au 23 juillet 2022 au siège du Conseil.

Ce colloque réunira des intervenants internationaux de l'Afrique, de l'Amérique du Sud, de l'Amérique du Nord et de l'Europe afin de discuter plus en détail des normes et outils disponibles pour mieux évaluer le développement d'une IA qui respecte et protège les droits et libertés fondamentaux. Il sera également l'occasion d'approfondir le débat, d'échanger de nouvelles expertises et de partager les bonnes pratiques autour

des principaux défis à relever pour mettre en œuvre une intelligence artificielle éthique, tels que la mise à niveau des cadres juridiques et organisationnels, la réglementation des secteurs de l'éducation, de la recherche, des entreprises et du développement, la composition et les prérogatives des organismes de contrôle et de surveillance.

Allocution de Madame Amina Bouayach

Présidente du Conseil national des droits de l'Homme

Mesdames et Messieurs, chers amis,

Je tiens à rappeler, qu'à l'issue du séminaire international de décembre 2021, les participants ont rendu publique « la Déclaration de Rabat » qui, entre autres, a prévu cette rencontre internationale et a désigné un comité scientifique composé de plusieurs experts et acteurs de différentes disciplines et de plusieurs pays.

Je tiens aujourd'hui, encore une fois, à les remercier pour leur engagement et disponibilité en vue de mutualiser notre plaidoyer à cet effet.

C'est ainsi qu'en se substituant à l'humain, Mesdames et Messieurs, l'utilisation de l'intelligence artificielle affecte certains droits et soulève des préoccupations et des interrogations relatives à la vie privée, à l'éducation, au travail, aux soins sociaux, à la santé, à la mobilité, à l'application de la loi et au maintien de l'ordre public.

Ainsi, le CNDH a-t-il décidé d'accompagner le développement de l'IA dans notre pays, avec pour mission la consolidation de la protection des droits de l'Homme et de leur promotion.

Notre souci relatif aux impacts négatifs de l'utilisation de l'IA sur les droits de l'Homme est fondé sur un certain nombre de faits et observations, qui sont les résultats de notre monitoring sur la situation des droits de l'Homme, de plusieurs consultations et la tenue d'un séminaire national et un autre international, autour des questions des droits humains dans la sphère digitale et les systèmes de l'intelligence artificielle.

Depuis le début de ce mandat ; nous étions très attentifs aux enjeux de l'espace numérique, d'abord, comme étant un espace de protection et de promotion des droits de l'Homme mais également, et malheureusement, un espace de chambres d'écho et de diffusion de discours d'incitation à la haine, à la discrimination, à la violence, autrement dit, un espace, aussi, de diffusion, d'une culture, fondamentalement, opposée et contraire aux droits de l'Homme.

Notre rapport annuel de 2019 avait défini, en analysant la situation des droits de l'Homme au Maroc, l'émergence d'un modèle naissant des libertés publiques, faisant le constat de l'élargissement de l'espace public au monde digital, et depuis, le CNDH assure le monitoring de l'espace civique numérique qui constitue une des thématiques de sa stratégie d'action.

Aujourd'hui, nous élargissons cette consultation, au volet de l'éthique de l'IA et au partage des initiatives mondiales, régionales et nationales, à cet effet, et les perspectives possibles de leur mise en œuvre.

Mesdames et Messieurs ;

Nous sommes convaincus de la contribution du digital et des systèmes de l'IA pour l'accès aux droits, mais nous sommes conscients des impacts sur la vie privée, les discriminations envers les femmes, les races et leurs couleurs, les personnes en situation de handicap,

Nos préoccupations sont énormes à cet effet, comment mettre en place des règles déontologiques, en vue de réduire l'impact et les retombées des systèmes de l'IA sur les droits de l'Homme en général et surtout sur ces catégories en particulier.

Peut-on être assuré que tout système d'IA, n'utiliserait guère lors de son élaboration, par exemple, les données relatives à la couleur de la peau, ou de genre, qui pourraient être un ou des éléments d'élimination pour un recrutement, une promotion ou d'accès à une information.

Le défi, donc, de mettre en place une éthique est majeur, certes. Nous avons pris note de plus de 300 documents des Organismes des Nations Unies, des organisations intergouvernementales, d'INDH, d'universités, d'entreprises, … qui proposent des codes éthiques mais l'implémentation nécessiterait une persévérance, monitoring et des actions multidimensionnelles de sensibilisation.

Aussi, le Secrétaire général de l'ONU, par exemple, avait déclaré, en février dernier, que : « Nous avons besoin de plus de femmes qui développent une intelligence artificielle qui soit véritablement au service de tous et toutes et qui œuvre en faveur de l'égalité des genres »[1]. La Haute Commissaire des Nations Unies aux Droits de l'Homme avait, de sa part, déclaré que « Les technologies d'intelligence artificielle peuvent avoir des effets négatifs, voire catastrophiques si elles sont utilisées sans prendre suffisamment en compte la manière dont elles affectent les droits humains ».

Certaines Institutions Nationales des Droits de l'Homme, Mesdames et Messieurs, comme le CNDH du Maroc, les INDH d'Australie, de France, d'Irlande et du Danemark ont publié des recommandations dans ce sens et sont actives dans le monitoring de l'impact de l'IA sur les droits de l'Homme, mais, également, même au sein de la communauté des INDHs, nous sommes loin de l'implication de la majorité à un champ, qui, de plus en plus, influence le courant de la garantie des droits dans notre quotidien.

Nous avons veillé, lors de ce colloque, en s'interrogeant sur les perspectives de mise en œuvre de l'éthique, à inviter des collègues d'autres INDHs, des experts et des acteurs de 11 pays (Maroc, États-Unis, Canada, France, Afrique du Sud, Ethiopie, Colombie, Mexique, Tunisie, Emirats Arabes Unis et Australie) en vue d'échanger autour de la menace que l'utilisation généralisée par les États et les entreprises de l'intelligence artificielle, y compris le profilage, la prise de décision automatisée et les technologies d'apprentissage automatique, qui affecte la jouissance du droit à la vie privée et des droits associés[2] et la violation des droits fondamentaux tels que les libertés individuelles et politiques (liberté d'association, d'expression…etc.).

Quelle éthique mettre en place pour protéger les droits de l'Homme dans un environnement d'utilisation accélérée de l'IA?

Au CNDH, nous serons très attentifs au débat, aux propositions de différents codes d'éthique et surtout aux difficultés et aux possibilités de leur mise en œuvre.

Merci de votre participation.

1- https://www.un.org/africarenewal/fr/derni%C3%A8re-heure/nous-avons-besoin-de-plus-de-femmes-qui-d%C3%A9veloppent-une-intelligence-artificielle-au

2- Link : https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27469&LangID=E

Allocution de Monsieur Karim Hendili

Directeur PI du Bureau de l'UNESCO région Maghreb

Madame la Présidente du Conseil national des droits de l'Homme,

Monsieur le Ministre de l'Enseignement Supérieur, de la Recherche Scientifique et de l'Innovation,

Monsieur le Président de la Commission Nationale du Contrôle de la Protection des Données à Caractère Personnel,

Madame la Représentante de la Ministre Déléguée auprès du Chef du Gouvernement, Chargée de la Transition Numérique et de la Réforme de l'Administration,

Monsieur le Représentant du Conseil Economique, Social et Environnemental,

Mesdames et Messieurs,

Je voudrais commencer par dire que c'est un réel plaisir et un honneur d'être parmi vous aujourd'hui.

Madame la Présidente, merci de votre accueil.

Comme il a été dit, il y a eu la signature d'un accord extrêmement important entre le Conseil national des droits de l'Homme et l'UNESCO et qui finalement marque et renforce notre engagement auprès du Royaume du Maroc sur les différentes questions qui touchent aux droits de l'Homme. Nous sommes très heureux que cette coopération qui s'articule autour de tous les domaines de compétence de notre organisation par rapport à sa mission, amène à des actions communes en lien avec des thèmes et des enjeux marquants, auxquels a été confronté le monde depuis le milieu du 20ème siècle, qui demeurent source de défis majeurs, tels que l'égalité des chances et l'accès aux services fondamentaux auxquels chacune et chacun est en droit d'inspirer. L'éthique de l'intelligence artificielle est l'un de ces thèmes.

Je tiens justement à réitérer notre gratitude à notre partenaire le CNDH, à tous les institutionnels et aussi au monde universitaire. Je remercie Monsieur le Ministre des éléments qu'il a partagé avec nous, c'est extrêmement inspirant et important pour la suite. Puis, bien sûr, sans oublier le rôle fondamental de la société civile par rapport à cet engagement qui est important et au cœur des priorités de l'UNESCO.

Je vais très modestement partager avec vous quelques éléments de réflexion qui ont mené au travail qu'a effectué l'UNESCO, et qui ont aussi abouti à cette fameuse recommandation adoptée en Novembre dernier. Permettez-moi de revenir à la source, à savoir cet extrait du préambule de l'acte constitutif de l'UNESCO et qui est devenu le slogan de notre organisation et qui dit la chose suivante : « les guerres prenant naissance dans l'esprit des hommes et des femmes. C'est dans l'esprit des hommes et des femmes qu'il faut diriger les défenses de la paix ». Cette phrase inspirée par les ravages de la seconde guerre mondiale est devenue le symbole de l'action de l'UNESCO pour la paix dans le monde et elle traduit aussi la reconnaissance claire, pleine et entière du fait que les défis auxquels doit faire face l'humanité sont en majeure partie issus de ce que l'esprit humain est capable de conceptualiser et de créer.

En d'autres termes, ce que nous créons est un reflet de ce que nous sommes ou pourrions devenir pour le meilleur ou pour le pire. En effet, l'Histoire nous enseigne que maintes inventions ont pu, au gré des époques et des contextes, constituer une avancée majeure pour le développement humain, ou au contraire mettre en pratique les pires desseins et encore les détourner de leur but originel pour devenir des outils de discrimination conscients d'asservissement et de destruction à grande échelle menant souvent aux pires tragédies.

Nous ne nions pas que l'humanité a

parsemé des exemples qui illustrent cette réalité. Le développement de l'intelligence artificielle n'échappe pas à cette lecture nourrie de 75 ans de réflexions et d'analyses que l'UNESCO et tous ses Etats membres tentent de porter. La question qui se pose à tout individu aujourd'hui n'est pas de savoir s'il a, doit et peut faire partie du processus d'évolution, mais plutôt de savoir comment cela doit s'opérer étant donné que l'IA est déjà partie intégrante de ce processus. A l'heure, où l'intelligence artificielle générale semble être une inéluctable étape, qu'on décrit à tort ou à raison, décisive pour l'avenir de l'humanité.

Tant que l'accès à une IA consciente, sur le plus ou moins long terme, suscite de l'inquiétude, il convient de faire preuve de la plus grande lucidité et tempérance possibles pour accompagner cette évolution, qui, en quelques sortes, changera de façon irréversible la vie de tous les êtres humains d'aujourd'hui et de demain.

L'UNESCO a un vrai rôle à jouer à l'heure de ces mutations, d'abord parce que les applications de l'IA touchent directement ses domaines de compétences. Monsieur le Ministre a rappelé le domaine de l'éducation, on pourrait parler aussi du domaine de la culture, des sciences et du travail sur la biodiversité. L'UNESCO a un programme hydrologique international, un système d'alerte précoce aux tsunamis, tout ceci est rattaché quelque part à l'évolution de l'IA.

L'éducation sera totalement transformée, Monsieur le Ministre a évoqué toutes les possibilités jusque là envisagées que peut offrir l'IA par rapport à l'accès à ce droit fondamental, qu'est l'éducation et qui pourraient décliner cette description sur tous les domaines de compétences. Puis, la question des compétences, justement, est développée pour évoluer dans un monde toujours plus automatisé et va devenir de plus en plus centrale, c'est-à-dire qu'il faudra acquérir les compétences pour pouvoir utiliser l'IA.

L'UNESCO a aussi un rôle très important, Madame la Présidente l'a évoqué dans la question de la réduction des inégalités qui persistent encore dans plusieurs régions du monde et qui seront probablement renforcées par l'IA, si on n'y prête pas attention. L'élimination du cloisonnement entre les pays, entre les sexes, mais aussi en termes de ressources et de connaissances pourrait permettre davantage au citoyen de contribuer à la transformation numérique qui est en cours.

Le passage à mission, à sa dimension internationale, l'UNESCO qui réunit des chercheurs, des philosophes, des programmeurs, des responsables politiques, ainsi que des représentants du secteur privé et de la société civile a jugé idéal de mener des débats sur les questions éthiques que posent le développement exponentiel de l'IA. Car, en effet, la recherche dans le domaine de l'IA avance à très grande vitesse, tandis que les environnements juridiques, sociaux et éthiques qui devraient l'encadrer évoluent plus lentement.

Plusieurs questions se posent : Jusqu'où peut aller l'autonomie d'une machine ? Qui a la responsabilité en cas d'incident ? Quelles sont les valeurs qui peuvent être inculquées en machine ? Quel niveau d'apprentissage a-t-on le droit d'attendre ou auquel on peut aspirer ? Toutes ces questions, parmi bien d'autres évidemment, restent aujourd'hui soit sans réponses, soit elles ont des réponses qui ne sont pas totalement complètes.

On a constaté que des algorithmes formés en langage humain ordinaire avaient acquis des préjugés fondés sur des stéréotypes, à partir des données textuelles présentes dans notre culture tous les jours. Donc, comment ne pas s'inquiéter des dangers qui peuvent apparaître des machines de comportements ségrégationnistes, en faisant la différence entre les personnes, voire racistes, voire hostiles.

Evidemment, on peut imaginer les pires scénarios. Il y a effectivement des raisons de s'inquiéter. Puis, la question de la protection

de la vie privée, la question de la surveillance, évoquée par Monsieur le Ministre, qui a également évoqué l'exemple de la Chine, où il y a même des systèmes de classement des citoyens en fonction du comportement. Vous avez une IA capable, quand bien même, vous essayez d'utiliser les subterfuges pour vous déguiser et changer votre démarche,, grâce aux algorithmes, de reconnaître la personne en question. C'est assez exceptionnel. Il y a eu beaucoup d'écrits là-dessus bien évidemment.

Les risques sont bien réels, la conscience de ces risques sera large et partagée plus la capacité de les gérer sera grande et efficace. Il y a quelques risques déjà évoqués, ceux des biais qui fonctionnent via des algorithmes, la question évidemment de la discrimination évoquée par Madame la Présidente. On se rend compte aujourd'hui dans les entreprises spécialisées dans l'innovation de l'IA que les femmes sont ultra minoritaires. Elles représentent seulement 22% des chiffres des professionnels de la technologie. En conséquence, un important risque de biais lié au genre, accentué par des bases de données, refléterait une société qu'on pourrait appeler « patriarcale ». On peut dire que ce risque nous amène aussi à d'autres risques, à savoir l'usage mal contrôlé ou malveillant de l'IA, l'atteinte à la vie privée, etc.

Puis, la mauvaise maîtrise de l'IA, conçue pour un but dont les conséquences sont absolument dramatiques, voire irréversibles, est plus préoccupante que l'on se penche sur la question de sa capacité d'apprentissage automatique. J'ai évoqué l'IA générale et aussi cette question fondamentale de l'accès à la conscience qui fait évidemment débat au sein de la communauté scientifique. Pour certains, c'est quelque chose qui n'est pas d'actualité pour le moment, pour d'autres, on n'y est déjà, sauf que ce n'est pas encore suffisamment développé. Le débat est ouvert.

En tous cas, nos collègues de la section biotique et éthique dans le secteur des sciences humaines et sociales disent qu'il faut toujours avoir la possibilité de savoir finalement qui a été à l'origine de l'algorithme. Pour l'instant, c'est une mesure qui doit être mise en place pour faire un peu le rôle de pare-feu si on veut dire. La responsabilité de chaque décision doit revenir à l'être humain et non pas à la machine.

Enfin, il est aussi important de se pencher sur le risque que fait peser la technologie du numérique sur l'environnement et ça paraît être une relation indirecte. Mais, si l'intelligence artificielle peut aider mieux à appréhender les problèmes climatiques, puisqu'elle a été utilisée évidemment et pour cause, son utilisation peut soit apporter quelque chose de positif en ce sens, soit être dangereuse pour l'environnement.

En effet, selon les données collectées par l'UNESCO, l'extraction des données concerne près de 10% de l'énergie mondiale et les centres de données pourraient consommer jusqu'à 20% de l'électricité mondiale d'ici 2025. Ce sont des chiffres extrêmement élevés et préoccupants qui montrent bien justement ce rapport permanent entre bénéfices et points négatifs, que l'on peut avoir par rapport au développement de toute technologie. Monsieur le Ministre a rappelé l'exemple du nucléaire et c'est une chose qui se trouve vraiment au cœur de l'actualité concernant l'IA.

Il est donc important de retenir le fait que même si la recherche fondamentale en matière d'IA est dans l'ensemble motivée par le développement durable et le bien-être, des biais existent et des dérives sont toujours possibles. C'est la raison pour laquelle l'UNESCO qui est un forum universel, unique, qui dispose d'une expérience de plus de 20 ans dans l'élaboration d'instruments normatifs nationaux relatifs à la bioéthique et à l'éthique des sciences et des technologies, s'est engagée dans un large processus de concertation impliquant toutes les parties prenantes et qui a pu mener à l'adoption en novembre 2021 de la Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle, qui est le premier instrument international dans cette question.

Placer l'humain au centre de toute politique de développement est en fin de compte l'approche qui guide les Nations Unies dans leur ensemble, par rapport au Programme de développement durable à l'horizon de 2030 et les 17 objectifs qui ont été adoptés à cet effet. L'autre pilier de la genèse de la Recommandation sur l'IA est le large processus participatif et multipartite. Je tiens à insister là-dessus, parce que c'est ce qui a vraiment permis d'élaborer le texte de la recommandation et de cadrer aussi son action et ensuite permettre son adoption par les Etats membres.

D'ailleurs, je tiens à saluer le fait que le Royaume du Maroc fait partie des premiers pays sur les 193 Etats membres de l'UNESCO qui ont lancé la mise en œuvre de la recommandation sur l'éthique de l'intelligence artificielle adoptée en novembre dernier, lancée en mars dernier à l'occasion de la visite de la Directrice Générale de l'UNESCO au Maroc, par le Ministre de l'éducation nationale, du préscolaire et du sport. Évidemment, nous ne pouvons que féliciter le Royaume du Maroc, c'est un rôle précurseur, puisqu'il fait partie des premiers pays qui sont au nombre de deux ou trois au monde à être relancés de façon très claire sur cette question. Au niveau du continent africain, le Maroc est pionnier en la matière et c'est extrêmement important.

Je tiens à cet effet, évidemment en tant que bureau de l'UNESCO du Maghreb basé à Rabat, à réitérer la disponibilité de toutes les institutions qui sont engagées dans cette mission, puisqu'il y a eu la mise en place d'une commission dédiée à l'IA au service du bien commun. Il y a eu beaucoup d'actions qui ont été engagées aussi et pas mal d'activités qui ont préparé cette adoption par le Maroc de l'engagement vis-à-vis de l'application de la recommandation sur l'IA.

Mais, on peut considérer qu'effectivement l'avènement de l'IA est une chance pour l'humanité, mais aussi une source de questionnement et d'appréhension qui sont tout à fait légitimes. L'IA est présente partout et sa place dans la vie des personnes ne cesse de s'affirmer, à tel point que de temps à autre on peut se demander si un point de retour n'a pas été franchi, c'est tout le débat qu'il y a sur l'éthique. C'est une nouvelle frontière de l'humanité selon certains et une fois qu'elle sera franchie, peut-être qu'une nouvelle forme de sinisation verra le jour.

Pour conclure, il faut parler de la révolution numérique qui est indissociable du développement de l'IA qui a bouleversé les codes culturels, sociaux et économiques. Le numérique a tendance à être présent dans différents domaines, notamment dans le domaine de la culture et du patrimoine, où le numérique devient une fin plutôt qu'un outil. D'ailleurs, on fait de la modélisation 3D et de la numérisation, parce qu'il faut le faire non pas parce que ça peut être utile. On le fait parce que quelque part c'est dans l'ère du temps et ça fait partie des dérives qu'il faut prendre en compte par rapport à la gestion de la question du numérique et de l'IA.

Au-delà des perspectives exceptionnelles qu'a ouvert le développement de l'IA, on peut parler du domaine du nucléaire qui peut aussi être détourné à des fins militaires, le domaine médical aussi qui a grandement gagné du développement de l'IA, mais aussi avec un encadrement éthique et moral très important. Il y a certaines dimensions philosophiques, voire métaphysiques qui sont apparues et qui sont désormais indissociables des considérations scientifiques et techniques. Cela démontre combien il est nécessaire que les avancées de la recherche en matière de l'IA intègrent à part entière ces autres considérations afin d'assurer l'équilibre, certes fragile, mais nécessaire, entre innovation et développement technologique d'une part, et préservation des droits de l'Homme et des libertés fondamentales et du libre arbitre d'une autre part.

Les défis et les enjeux qui se posent sont à la hauteur de la mutation profonde. Je pense qu'il y a un consensus là-dessus provoqué par l'avènement de l'IA. Elle est en train d'accélérer la transformation de nos vies et les problématiques liées à l'éthique sont

considérables. Notre réponse à ces défis tracera les prémisses d'un nouveau monde que nous transmettrons aux nouvelles générations suivantes. Un nouveau monde que nous voulons plus juste, plus pacifique et plus prospère. Nous sommes toujours en appui à nos Etats membres. Je réitère l'appui de l'UNESCO au Royaume du Maroc dans cette démarche.

Je souhaite conclure ce propos en remerciant à nouveau le CNDH de l'invitation et de l'opportunité d'être parmi vous aujourd'hui. Encore une fois c'est un grand honneur. Je souhaite plein succès à vos travaux. Merci beaucoup.

Mot d'ouverture de Monsieur Abdellatif Miraoui

Ministre de l'Enseignement Supérieur, de la Recherche Scientifique et de l'Innovation

Madame la Présidente,

Messieurs les directeurs,

Monsieur le Président du Conseil Économique, Social et Environnemental,

Madame la représentante de Madame la Ministre de la Transition Numérique,

Mesdames, Messieurs, chers ami.e.s,

C'est d'abord un honneur pour moi de participer à ce colloque, j'ai tout de suite répondu favorablement parce que je trouve que le sujet est plus que d'actualité.

D'abord c'est un sujet qui aujourd'hui est devenu très important sur le plan socio-économique et qui va encore le devenir de plus en plus durant les prochaines années, mais qui est aussi un sujet qui interpelle tout le monde.

Je m'excuse souvent de le comparer à la découverte du nucléaire, dont on a vu beaucoup d'avantages mais aussi les limites de ce qui pourrait en être fait.

C'est un sujet très pertinent et qui traite d'une question à double tranchant, qui met en quelque sorte en relief les opportunités sur le plan humain puisqu'il va aider l'être humain à se développer mais qui aussi interpelle sur les grandes questions que vous allez probablement débattre parce que ce sont des questions qui touchent aux droits de l'Homme.

Aujourd'hui l'ancrage de ces moyens, de ces techniques de cette technologie de l'Intelligence Artificielle n'est plus discuté mais une nécessité impérieuse car on ne peut pas s'en priver, pour les raisons citées auparavant, parce qu'elles ont beaucoup de bienfaits mais il va falloir qu'on arrive à les essayer de les réglementer et de les utiliser dans un cadre bien défini.

Les budgets importants de la recherche et développement sont conséquents, les start-ups spécialisées représentent pratiquement 12% du capital investissement mondial qui d'ici 2030 atteindra les 15 trillions de dollars, il s'agit de sommes énormes et qui vont transformer notre façon de vivre. Les projections qui sont en cours aujourd'hui prédisent que le nombre d'assistants vocaux alimentés par l'Intelligence Artificielle devrait atteindre 8 milliards d'ici 2023. Aujourd'hui même, dans les grandes entreprises mondiales, vous n'avez plus d'être humain à l'écoute mais plutôt des robots munis d'IA.

Vous avez vu ce qui se passe sur le plan éducation et formation, nous sommes également dans ces réflexions d'introduire l'Intelligence Artificielle pour essayer d'optimiser nos pédagogies.

Vous avez vu maintenant qu'il y a dans un certain nombre de pays, notamment la Chine, où on a introduit les caméras avec IA autour des salles de classe qui permettent de détecter à quel moment l'enfant a en quelque sorte perdu le contact avec la pédagogie ou avec son enseignant.

Vous voyez également toutes les caméras qu'on peut trouver dans les amphithéâtres aujourd'hui pour essayer de voir quelles sont les réactions des individus, on fait ça de plus en plus dans les recrutements sur le plan international. Alors, c'est discutable.

Le débat est toujours ouvert mais j'allais dire que l'avancée de cette technologie est telle qu' on ne peut que suivre et en même temps de réglementer ; d'essayer de trouver un cadre qui soit le plus humain possible.

32% des emplois actuels seraient profondément transformés par ces progrès de l'Intelligence Artificielle d'ici 2030-2040.

Bien évidemment, c'est un enjeu économique mais également un enjeu de droits humains qu'il faut absolument qu'on arrive à débattre sur la scène internationale,

et la présence des Nations Unies parmi nous est un point essentiel parce qu'il faut absolument qu'on arrive à trouver le juste milieu.

Je citerais la Haut-Commissaire des Nations Unies aux droits de l'Homme, Michelle Bachelet, qui avait déclaré à ce sujet « les technologies de l'IA peuvent avoir des effets négatifs, voire catastrophiques, si elles ne tiennent pas suffisamment compte de la manière dont elles affectent les droits de l'Homme»[3], et d'autres organismes et organisations ont adopté une position similaire tels que la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST), le Conseil de l'Europe qui a créé en décembre 2020 le Comité ad hoc sur l'Intelligence Artificielle (CAHAI).

De telles positions trouvent également leurs racines, entre autres, dans certains facteurs de risque. Le risque d'empiètement des algorithmes sur les droits humains, la vie privée et la confidentialité ; on le voit de plus en plus d'ailleurs même dans nos activités régulières. Le risque de creusement de l'écart de développement au vu des capacités inégales d'appropriation de ces technologies.

J'ai parlé tout à l'heure de la Chine qui a beaucoup investi depuis des années mais vous avez également d'autres pays en particulier les États-Unis qui ont investi des milliards pour ne pas dire des trilliards depuis un certain temps.

La robotique aujourd'hui va remplacer de plus en plus l'être humain, ne serait-ce que dans les administrations publiques, de plus en plus, je pense que dans les 5 à dix années à venir nous devons absolument trouver des solutions pour nos humains ; qu'est-ce qu'ils feront vu les qualifications de travail que les robots vont prendre et le Maroc bien évidemment n'est pas en reste dans tous ces développements. Aujourd'hui nous avons déjà pas mal d'entreprises et nous sommes fiers de ces compétences qui sont en train d'être développées dans notre pays, que cela soit pour des applications nationales ou pour des besoins d'entreprises internationales, mais c'est un levier également pour la concrétisation de plusieurs stratégies sectorielles.

Dans le rapport sur le nouveau modèle de développement, nous avons abordé ce sujet de manière consistante de telle sorte que l'on puisse se l'approprier et pouvoir prendre une place que ce soit du côté des droits de l'Homme mais aussi du côté socio-économique.

Si je dois donner quelques chiffres concernant l'enseignement supérieur, c'est d'abord 1500 thèses doctorales qui sont soutenues.

Quand on parle d'Intelligence Artificielle, souvent, le commun des mortels, ne regarde que l'actualité de l'IA, alors que moi-même lorsque j'étais doctorant fin des années 1980 j'avais fait des travaux usant de l'IA, les réseaux de neurones que l'on a utilisé pour plusieurs applications sauf qu'elles ont pris des formes complètement différentes avec ce progrès qu'on a fait dans le numérique et qui va encore prendre plus de place parce qu'aujourd'hui on est en train de parler d'IA avec les moyens dont on dispose sur le plan informatique et avec l'informatique quantique qui se profile. Probablement dans les 5 à 10 ans, une rupture technologique personne ne la devine mais qui va complètement transformer ce domaine même de l'IA j'allais dire, qui restera dans la partie scientifique et dans la partie théorique la même thématique mais elle va prendre un élan considérable avec cette transformation.

Plus de 2000 thèses sont en cours de préparation dans notre pays et dans tous les domaines, toutes les applications que vous pouvez imaginer de la mécanique à l'électricité à des applications dans

3- Haut-Commissariat des Nations Unies aux droits de l'homme, communiqué de presse « Michelle Bachelet : les risques d'atteinte à la vie privée liés à l'intelligence artificielle nécessitent une action urgente » lien du communiqué

l'économie, dans la sociologie… sont des travaux qui aujourd'hui sont menées dans nos laboratoires et dans ce nouvel élan de renouvellement de l'enseignement supérieur marocain nous avons mis au cœur le digital et bien évidemment l'IA et la data science font partie intégrante de tout ce développement.

Nous allons commencer par la première qu'on a appelé code 212, code pour coding et 212 pour l'indicatif téléphonique du Maroc que nous sommes en train d'installer dans toutes les universités marocaines qui sont des espaces complètement ouverts et qui permettent de donner une double compétence à nos jeunes. Nous savons que les diplômes sont nécessaires mais ils ne sont pas suffisants et donc nous souhaiterions donner des doubles compétences particulièrement dans le digital et nous visons plus de 100.000 certificats d'ici 2026 dans les domaines de l'IA, de la data science mais du coding en général… et que bien évidemment que l'alignement de l'offre de formation du capital humain aux besoins actuels et futures du secteur de l'économie parce que dans ce secteur il y a un développement considérable.

Nous avons la chance dans notre pays d'avoir une caractéristique de jeunesse qui est une dérivée positive et que cette jeunesse est de plus en plus, allez je le dis, « geek » et qu'il faut absolument profiter de cette opportunité pour entrer dans l'ère de l'IA.

Votre sujet d'aujourd'hui m'intéresse grandement parce que l'on devrait l'enseigner de plus en plus dans le cadre de toutes les formations de manière transversale, parce que les droits de l'Homme ça devrait être comme l'air que l'on respire et il faut absolument qu'on arrive à ce que nos jeunes lauréats soient bien imprégnés, des inventions, des conceptions de nouveaux produits et des innovations pour intégrer ce paramètre et il faut absolument qu'on arrive Madame la Présidente à partager et à avoir des synthèses pour qu'on puisse les intégrer, particulièrement dans les formations technologiques.

Les orientations de Sa Majesté que Dieu l'assiste dans ces éléments de la digitalisation mais aussi du respect de l'être humain nous conduisent à progresser sur le plan intellectuel et sur le plan socio-économique tout en intégrant tous ces éléments sur l'aspect de la formation en particulier.

Bien évidemment les droits de l'Homme c'est d'offrir une vie digne à l'être humain mais c'est également de le protéger des méfaits qu'on peut avoir suite à des inventions ou suite à des technologies.

Je m'excuse toujours si je reviens au nucléaire qui était une découverte extraordinaire et qui n'est pas finie, avec les nouvelles versions du nucléaire qui sont en train de se profiler où les déchets seront réutilisés et donc ce côté négatif pourrait peut-être diminuer mais il reste encore la possibilité de l'utiliser dans des guerres ou de l'utiliser dans des moments, de folie.

Et j'ose comparer l'IA avec cela puisqu'elle peut être aussi nucléaire dans son utilisation, nucléaire sur tous les volets ; le respect de la dignité humaine mais aussi de priver l'individu d'un certain nombre de droits qu'aujourd'hui il a défendu depuis des siècles et que petit à petit va de plus en plus diminuer si on n'y prend pas garde mais qu'il faut absolument gérer de manière intelligente et ça ne se gère pas qu'à l'échelle d'une nation mais se gère à l'échelle de l'humanité et qu'il n' y a plus de frontières c'est comme pour le développement durable ou le climat. Vous pouvez économiser ce que vous voulez, vous pouvez gérer comme vous le souhaitez mais si vous êtes tout seul à le faire et les frontières ne sont pas étanches surtout sur ces aspects donc il faut absolument avoir une intelligence collective à l'échelle internationale pour pouvoir la gérer.

Mais nous sommes décidés dans notre ministère à ne pas laisser notre place parmi les nations les plus développées dans ce

secteur, en tout cas notre continent africain a une opportunité extraordinaire dans ces sujets et nous devons mener ces avancées de pair avec le respect des droits humains les plus élémentaires.

Je reprends pour finir, vous imaginez ce que l'Intelligence Artificielle pourrait nous faire gagner pour l'éducation. L'éducation aujourd'hui, les technologies ont beaucoup évolué et on est en capacité aujourd'hui d'éduquer sur mesure avec ces technologies là et il faut qu'on arrive à en profiter pour les enfants qui habitent les endroits les plus reculés grâce aux technologies en particulier celles de l'IA, qui pourraient servir aussi à leur amener leurs droits les plus humains ; le droit au travail, le droit à la dignité qui ne viendra que lorsque l'on est bien élevé et bien éduqué.

Bien évidemment la technologie peut se servir d'outils comme l'IA pour essayer de progresser.

Merci et bravo pour l'organisation.

Mot d'ouverture de Monsieur Amine Mounir Alaoui

Président de la commission de la société du savoir et de l'information
Conseil Économique, Social et Environnemental

J'ai le difficile rôle d'être la dernière barrière avant un peu de repos, de fraîcheur et éventuellement du déjeuner.

Je voudrais vous remercier Madame la Présidente, chers collègues du CNDH, d'avoir choisi cette thématique qui est en effet une thématique extrêmement importante aujourd'hui.

Mais avant de revenir sur le rapport de la commission, que j'ai l'honneur de présider au niveau du Conseil Économique, Social et Environnemental sur la transition numérique, je voudrais rappeler quelques éléments que l'auditoire ne connaît peut-être pas :

Le premier cours d'Intelligence Artificielle au Maroc a eu lieu en 1989, c'est-à-dire il y a à peu près 35 ans, donc ce n'est pas quelque chose de totalement nouveau. C'est quelque chose qui a été mise en place il y a très longtemps et qui n'a pas éventuellement connu l'évolution industrielle ou l'évolution scientifique qu'elle aurait pu avoir.

Deuxième date, -je fais le vieux qui est dans le secteur depuis 40 ans ici mais c'est important parce que tout à l'heure Monsieur le Ministre a parlé du fait que c'est vraiment le moment, moi je voudrais être un peu plus critique et dire que c'était le moment il y a déjà longtemps et que l'on a pris énormément de retard-. Même si le Maroc a connu une trajectoire extrêmement innovante dans les années 1990 en particulier puisque Internet est arrivé il y a exactement 29 ans et 3 semaines au Maroc ; cela veut donc dire que les jeunes qui ont plus de 29 ans sont nés au Maroc alors que Internet existait déjà dans ce pays. Nous ne sommes donc pas en train de parler de quelque chose de nouveau mais de quelque chose de très ancien, puisque si l'on prend la population qui a moins de 30 ans au Maroc il s'agit d'un bon tiers de la population.

Donc le premier élément sur lequel je voudrais insister est lorsque l'on parle de toutes ces problématiques de ne pas parler au futur mais de parler plutôt au passé, nous sommes dans une réalité aujourd'hui : le numérique est une réalité, l'Intelligence Artificielle est une réalité, qu'on le veuille ou pas, que cela soit bien ou mauvais. Maintenant comme cela a été dit par mes prédécesseurs il serait bien de favoriser les aspects positifs et d'essayer de contrôler les aspects négatifs, mais de ne pas dire que « ça va être » ou « ça sera » ; c'est vraiment quelque chose qu'il faut bannir de notre réflexion.

Je vous renvoie de toutes les façons à l'excellent rapport que fait tous les ans l'ANRT (Agence Nationale de Réglementation des Télécommunications) sur l'état des technologies de l'information au Maroc, dans lequel on voit très bien l'évolution. Aujourd'hui on est à des chiffres tout simplement incroyables avec une vraie rupture au niveau du numérique à l'âge de 60 ans, donc si vous voyez les courbes en dessous de 60 ans le minimum c'est 85% d'usage et au-dessus de 60 ans on passe brutalement à 44%. On a donc un vrai gap générationnel qui fait que malheureusement, -et c'est un point qui me tient particulièrement à cœur-, les politiques qui sont faites par ces gens qui sont au-dessus de 60 ans pour les gens qui sont en dessous de 60 ans ne correspondent pas à la réalité, pour la simple raison qu'ils ne vivent pas dans le même monde. Lorsque l'on parle aujourd'hui d'introduire les technologies de l'information à l'école c'est comme dire d'introduire le réfrigérateur à l'école car c'est quelque chose que les enfants ont à la maison et qu'ils utilisent tous les jours donc il n'y a pas besoin de l'introduire à l'école.

Que faut-il introduire à l'école ? C'est le bon usage de ces technologies, et dans un cadre où justement on respecte les droits des uns et des autres. Je reviens et j'insiste, je suis

tout à fait d'accord avec Monsieur le Ministre tout à l'heure et je suis sûr que le Ministre de l'Éducation nationale du Préscolaire et des Sports est également dans le même cadre, c'est que la partie droits de l'Homme et respect des autres et leurs valeurs est fondamentale à l'école parce que c'est ce qui nous permet de nous protéger par la suite ; et de faire en sorte que l'usage que l'on va faire de l'Intelligence Artificielle ou d'autres technologies est un usage en correspondance avec nos valeurs.

Cela me donne une bonne liaison par rapport aux travaux du Conseil Économique, Social et Environnemental. Cela fait 12 ans que l'on existe et cela fait 12 ans que l'on sort pratiquement tous les ans un rapport spécifique sur les technologies de l'information parce que c'est quelque chose qui nous semble être fondamental au niveau économique, mais surtout au niveau social car comme je le disais tout à l'heure il y a un gap entre les générations qui est très important.

Le premier, -et là je vais faire la liaison avec mon ami Mr Seghrouchni-, rapport que nous avons fait de manière spécifique sur les technologies de l'information était sur l'open data c'est-à-dire que la donnée est quelque chose de fondamental pour l'avenir et quelque chose d'absolument structurel pour tout ce qui est Intelligence Artificielle.

Malheureusement au Maroc -et je pense que le CNDH a eu des interventions aussi par rapport à l'accès aux données, etc.- c'est un vrai problème mais au-delà des données personnelles, aujourd'hui nous avons un problème hygrométrique. Je suis chercheur à l'université, -il y a des collègues chercheurs dans la salle-, et je veux faire une étude sur l'évolution de la pluviométrie au Maroc... eh bien il est extrêmement difficile d'obtenir ces données pour ne pas dire impossible.

Il s'agit là d'un point clé au niveau de toute réflexion sur l'Intelligence Artificielle et son usage, et la relation avec les droits de l'Homme est très simple : si je ne peux pas faire un travail sur l'hygrométrie et l'évolution de l'eau, le droit de chacun de nos concitoyens d'avoir accès à de l'eau potable, qui nous semble aujourd'hui tellement naturel, peut être nié tout simplement parce que je ne peux pas faire de travail universitaire et de recherche dessus. Donc cet aspect de l'open data est fondamental.

Il y a beaucoup d'autres rapports mais je ne vais pas y revenir parce que cela serait trop long, mais pour résumer : au niveau de la transition numérique au Maroc il y a eu beaucoup, beaucoup, beaucoup d'initiatives. Malheureusement ces initiatives ont souvent été des initiatives ponctuelles, il n'y a jamais eu de vision globale qui intègre l'ensemble des parties prenantes et quand je dis parties prenantes c'est les 36 ou 37 millions de Marocains. On est tous partie prenante dans le numérique que cela soit au niveau de l'accès, que cela soit au niveau de l'usage, aujourd'hui que cela soit en tant que marchandise -nous sommes tous des marchandises-, la CNDP (Commission Nationale de Contrôle de la Protection des Données à Caractère Personnel) est bien placée pour le savoir. Pour Google, pour Facebook, pour Instagram etc. qui utilise notre ego -on va dire cela comme ça- pour absorber toutes les données qu'ils peuvent pour pouvoir mettre en place des modèles et piétiner nos droits -comment dire- avec notre assentiment inconscient et ça c'est un véritable problème.

Ce qui est assez intéressant c'est qu'effectivement nous avons certaines administrations qui sont extrêmement -comment dire- à cheval sur les règles, et qui vont avant de vous demander des informations dire que « on correspond à la délibération numéro tant de la CNDP » etc. et le citoyen va dire « ah ouais mais moi je n'ai pas envie de donner cette information à telle administration je ne sais pas ce qu'ils vont en faire ». Le même citoyen va mettre ses photos, les photos de ses enfants, de ses parents, de ses grands-parents et dire où est-ce qu'il habite sur Facebook, etc. de manière totalement volontaire et gratuite. Donc c'est là l'aspect social qui pose problème, c'est-

à-dire que l'on vit dans une société qui est complètement déconnectée par rapport à la réalité ; la réalité c'est que l'on donne les informations. Moi je ne vois pas pourquoi je ne donnerais pas mes informations pour passer mon permis de conduire ou aller titiller le Ministère du Transport parce qu'ils ont demandé telle information et que je ne vois pas l'intérêt de la donner, et d'un autre côté donner ces informations de manière gratuite – mais ces informations sont le moteur effectivement de l'Intelligence Artificielle aujourd'hui.

J'aimerais parler des principales recommandations -et c'est là la clé au niveau de la transition numérique- en accord complet avec le ministère en charge du secteur et avec d'autres ministères, c'est d'abord leadership et gouvernance. Comme je le disais, il s'agit de quelque chose qui nous concerne tous et donc aujourd'hui ce n'est pas l'affaire d'un ministère mais l'affaire de tous les ministères, l'affaire de toutes les institutions, l'affaire de toutes les entreprises, et l'affaire de tout le monde. Mais ce « tout le monde » est trop nombreux, chacun a ses intérêts ; il faut absolument coordonner ce travail. Et pour coordonner ce travail il faut une institution qui soit forte et qui ait la légitimité de le faire et qui, pas qu'accessoirement, de manière très importante ait les moyens de le faire. C'est bien de créer des institutions mais si on ne leur donne pas derrière les moyens de remplir leurs missions cela ne sert strictement à rien, c'est une case que l'on coche mais pas un travail que l'on fait.

Deuxième élément, qui a aussi été révélé par la crise de la Covid, c'est l'accessibilité. Il faudrait que tous les Marocains et toutes les Marocaines puissent avoir accès à internet, aujourd'hui c'est quasiment le cas mais "le quasiment" ne veut pas dire 100%. Or aujourd'hui nous avons les moyens d'être à 100%, pourquoi ? Parce que nous avons un système au Maroc qui est très bien et qui fonctionne très bien qui s'appelle le fonds de service universel. Ce fonds de service universel selon nos calculs a suffisamment d'argent pour couvrir 100% des Marocains, attention cela ne veut pas dire couvrir 100% du territoire mais 100% des citoyens et citoyennes.

Troisième point, l'État digital. Aujourd'hui nous n'avons pas une seule procédure administrative qui soit à 100% numérique, on en a 0. On parle de digitalisation de l'administration, on parle de beaucoup de choses, c'est vrai que des efforts ont été fait mais aujourd'hui on est à 0 procédure à 100%. On a des procédures qui sont à 99% mais que l'on soit à 99% ou 0% c'est la même chose, pourquoi ? Parce que le 1% où il faut se déplacer à l'administration plutôt que de faire une partie en ligne et une partie à l'administration ben je vais à l'administration je fais tout et au moins j'ai quelqu'un en face de moi et je suis tranquille. On a donc aujourd'hui un vrai problème au niveau de l'état digital qui doit se développer.

L'élément droits de l'Homme, à ce niveau-là c'est qu'aujourd'hui d'un point de vue pratique, il ne s'agit pas de règles, il ne s'agit pas de lois, je pars d'un point de vue pragmatique, selon les auditions que nous avons fait c'est que malheureusement le numérique est toujours à charge jamais à décharge. Quand vous allez à une administration et que l'on vous dise « ah mais sur l'ordinateur vous n'avez pas payé vos impôts », « ah si si j'ai payé mes impôts », « ah vous amenez des papiers », « très bien » j'amène les papiers qui prouvent que j'ai payé mes impôts mais il faut que cela soit des papiers avec des cachets je ne peux simplement pas amener un lien avec mon compte bancaire qui montre que j'ai été débité par les impôts, non je ne peux pas ; ça doit être un papier avec un cachet et donc on revient à ce problème-là. Et là on est donc à charge, par contre à décharge ça ne marche pas c'est-à-dire que si vous allez chez votre percepteur ou chez le bureau d'impôts -je reprends cet exemple qui nous concerne tous- « mais sur votre ordinateur moi je suis à jour, est-ce que vous pourrez le dire à quelqu'un qui me dit que je n'ai pas payé il y a 3 jours » « ah non non non

même si sur l'ordinateur c'est bon, il faut que vous ameniez le papier » donc le papier est toujours là et le numérique jamais à décharge ; c'est-à-dire le fait que vous ayez une information sur un ordinateur au niveau d'une administration n'est pas équivalent à avoir le papier correspondant.

Il y a donc l'aspect IA et droits de l'Homme mais je pense qu'il y a aussi une grande réflexion à mener sur la citoyenneté numérique ; quelle est la relation entre le citoyen et son double, son avatar sur internet que ce soit au niveau du Ministère de la Justice, que cela soit au niveau de la police, que cela soit au niveau des impôts et les droits qu'il a en tant que citoyen numérique. Alors il y a une partie des droits qui concerne les données personnelles qui sont protégées par la CNDP mais il y a toute une partie interaction pour laquelle il n'y a pas de règles particulières.

Après l'État digital, il y a la stratégie industrielle. C'était très beau d'avoir les premiers cours en Intelligence Artificielle en 1989 mais en 2022 -donc 33 ans plus tard- nous n'avons aucune entreprise qui a été créée dans l'IA à ce moment-là et qui ait évolué. Cela veut donc dire que l'on avait la connaissance, on avait l'information, mais la partie industrielle et développement industriel n'a pas eu lieu pour des raisons diverses et variées. Je peux vous donner des exemples qui étaient extrêmement intéressants. Vous savez que le CHU Ibn Sina, le dispatching des urgences était fait par un système intelligent en 1992-1993, alors ce qu'il s'est passé par la suite c'est que malheureusement l'hôpital n'avait pas de budget pour acheter un ordinateur assez puissant parce qu'à l'époque cela coûtait très cher et ils ont donc laissé tomber le projet, cela avait fonctionné avec l'université un certain temps puis ça s'était arrêté. Je vous parle de 1992-1993, il y a 30 ans, donc nous avons un vrai problème de développement depuis les idées, depuis notre intelligence collective, -et Dieu seul sait que les jeunes Marocains sont extrêmement dynamiques dans ce secteur-, et la réalisation par la suite au niveau industriel.

Au niveau industriel pour revenir à l'Intelligence Artificielle, une feuille de route pour l'IA qui soit claire et établie est fondamentale.

Je salue tous les efforts qui sont faits mais on est très très très loin du compte, aujourd'hui les projets d'Intelligence Artificielle qui sont financés au niveau universitaire etc. on est en train de parler de quelques millions de dirhams alors que les projets minimalistes dès que l'on traverse la Méditerranée sont financés à plusieurs dizaines de millions d'euros ; donc on ne peut pas concurrencer, on ne peut pas travailler, on ne joue pas dans la même cour et on a un vrai problème qui est lié à l'industrialisation.

Autre élément fondamental qui fait partie -et là je suis heureux que l'UNESCO soit présente- des recommandations majeures c'est tout ce qui est culture et média. Aujourd'hui au niveau du numérique, on consomme la culture à travers le numérique. Vous savez quel est le premier bien culturel vendu en Europe ? Ça n'est ni le cinéma, ni les livres, ni la musique… c'est les jeux vidéo, un produit 100% numérique est le premier bien culturel vendu en Europe.

La relation entre culture et numérique est extrêmement importante et nous avons un problème fondamental au Maroc, c'est que nous produisons malheureusement très peu de produits culturels au niveau numérique. Il y a une production qui a évolué au niveau musical au niveau cinématographique par exemple, même au niveau livre -même si là-dessus nous avons un rapport sur la lecture au Maroc qui est d'ailleurs intéressant dans ce sens-là- mais la production qui va dans le numérique est relativement limitée.

Pour terminer, comme c'est IA et droits de l'Homme je ne pouvais pas parler sans ressortir, à mon sens et avis en tant qu'ancien du secteur, ces deux niveaux qui posent vraiment problème. On a un problème au niveau national de l'IA et droits de l'Homme dans l'utilisation des données

mais également la mise à la disposition des données pour pouvoir vraiment évoluer dans le secteur et faire évoluer le secteur de l'IA.

Mais le principal problème de droits que je vois, et de droits de l'Homme de manière plus particulière, c'est au niveau international. Aujourd'hui on est en train de parler d'IA et de droits de l'Homme mais celui qui viole le plus les droits de l'Homme au niveau de l'IA s'appelle Alphabet c'est la maison-mère de Google. Ils ne vous demandent pas votre avis, ils font ce qu'ils veulent de vos informations et s'ils ne sont pas d'accord avec vous de manière automatique ou de manière humaine on ne sait pas parce que nous n'avons aucune traçabilité. Ils font ce qu'ils veulent. Alors j'ai parlé d'Alphabet mais c'est tout autant vrai pour Facebook, pour Instagram, pour WhatsApp donc en fait on a un vrai problème. C'est qu'aujourd'hui le droit tel qu'il a été conçu jusqu'à il y a 10 ans était un droit des états aujourd'hui nous avons un droit des entreprises et ce droit des entreprises qui est imposé par les entreprises quel que soit le lieu où vous êtes est un droit qui prend sa source dans le droit américain et donc quelque chose qui est autorisé aux États-Unis pourrait être autorisé au Maroc sur Facebook, par contre quelque chose qui est autorisée au Maroc peut être banni de Facebook de manière complètement arbitraire tout simplement parce que Facebook est une entreprise de droit américain et qu'ils sont obligés de respecter un certain nombre de normes et de conditions qui ne sont pas les nôtres et qui ne sont pas forcément nos valeurs.

L'exemple le plus typique sur lequel je voudrais terminer est celui de la violence.

La gestion de la violence chez les américains n'est pas la gestion de la violence chez les marocains ni chez les Européens d'ailleurs, et nous sommes beaucoup plus proches des européens dans ce cadre et des pays arabes voire africains que des américains. Les américains ont une gestion de la violence qui est une gestion au quotidien et qui est normale ce qui, donc, fait que l'on se retrouve face à des choses sur les réseaux qui sont des choses inacceptables au Maroc. À l'inverse, il y a un certain nombre de choses qui sont tout à fait normales chez nous qui correspondent à nos valeurs qui vont être bannies.

Il y a grande une réflexion à mener autour de tous ces algorithmes, cette façon de fonctionner et surtout qui les maîtrisent et qui a le droit d'arrêter ces algorithmes et les modifier.

Dernier point scientifique, adressé aux universitaires ici, mon problème aussi en tant qu'ex professeur universitaire est qu'il n'y a pas d'algorithmes à proprement parler dans beaucoup d'Intelligence Artificielle c'est de l'apprentissage automatique. Cet apprentissage automatique, pour ceux qui ne connaissent pas beaucoup les modèles mathématiques, est fait aujourd'hui sans que l'on sache comment l'ordinateur procède. Autrement dit, en bout de course, vous n'avez pas un algorithme à l'ancienne où on sait comment cela va fonctionner et on peut justifier, aujourd'hui beaucoup de décisions prises par des systèmes dits intelligents sont prises dans un cadre que l'on ne saurait pas expliquer tout simplement par la technologie qui est utilisée, une technologie qui pour la première fois dans l'histoire de l'humanité échappe partiellement à ceux qui l'ont créée. Mais comme il y a beaucoup de business autour, les gens qui ont l'argent et qui pourraient développer la recherche dans ce sens se disent "oh là là on n'y va pas parce qu'il faut d'abord que l'on gagne beaucoup d'argent et après quand on aura un problème éventuellement on reviendra là-dessus".

Je vous remercie et je vous souhaite très bon courage pour vos travaux.

Mot d'ouverture de Monsieur Omar Seghrouchni

Président de la Commission Nationale de Contrôle de la Protection des Données à Caractère Personnel

Je remercie le CNDH de nous donner cette opportunité d'intervenir.

Madame la Présidente,

Monsieur le Directeur représentant de l'UNESCO,

Monsieur le représentant du Conseil Économique, Social et Environnemental,

Madame la représentante de la Ministre de la Transition Numérique et de la Réforme de l'Administration,

Je suis heureux d'intervenir avec ce titre, d'autant plus que ce titre n'est pas le mien. En fait, il n'y a pas de modèle marocain de la protection des données. La protection des données d'ailleurs comme les droits humains est un concept universel et dès lors que l'on commence à se mettre dans l'optique de gérer une spécificité ou des spécificités qui peuvent être exagérées par moment on se retrouve assez éloigné de l'universel.

Je parlerais donc de la protection des données et de l'Intelligence Artificielle, et je parlerais en particulier des pratiques que nous développons pour protéger la vie privée ; puisqu'on protège les données pour protéger la vie privée des citoyens à travers ou dans le contexte de cette Intelligence Artificielle.

La loi 09-08 a été adoptée en 2009 tandis que la Commission Nationale de Contrôle de la Protection des Données à Caractère Personnel a été établie en 2010. Cette loi est une loi pour la protection des personnes physiques à l'égard des traitements de données à caractère personnel. Qu'est-ce que cela veut dire ? On protège les personnes physiques, à l'égard des traitements de données à caractère personnel. On va donc commencer par trois couches, à la fin de mon intervention on aura d'autres couches mais pour l'instant il y a trois couches.

Il y a donc les données à caractère personnel et il y a les traitements que l'on opère sur ces données, et tout cela pour protéger les personnes physiques. Les traitements que l'on peut effectuer sur ces données sont d'origines diverses, il peut y avoir la gestion de vos données d'état civil pour produire un extrait d'acte de naissance, il peut y avoir la gestion de vos données de santé pour produire un pass vaccinal, comme il peut y avoir la gestion de vos notes dans un concours pour produire une attestation d'acceptation ou une attestation de refus d'intégration dans une institution… il peut y avoir différentes décisions -donc traitements- qui sont prises sur la base de ces données à caractère personnel. Lorsque certains veulent entrer dans une boîte de nuit et que selon le faciès le videur leur dit oui ou non, c'est un traitement de données à caractère personnel.

Avec le temps, ce traitement là avec la technologie s'élabore de plus en plus et nous sommes en train de converger vers des logiques de décisions automatiques pour différentes raisons. Monsieur le Directeur de l'UNESCO l'a évoqué, en fait il y a une balance à réaliser entre l'efficacité, la rapidité et la valeur ajoutée supposée ; ensuite ce que cela peut générer en termes de respect de l'être humain, de vie privée et des valeurs que l'on souhaite avoir au sein de nos sociétés.

Ainsi, la gestion des traitements de données à caractère personnel et le seuil que l'on va accepter pour certains éléments va dépendre foncièrement des valeurs que l'on souhaite avoir au sein de nos sociétés. On se plaît parfois à donner un exemple sur la Chine, l'Australie, l'Amérique Latine etc, mais souvent on s'enferme dans une analyse technologique techniciste en omettant l'endroit où l'on souhaite faire la greffe et est-ce qu'on a envie de le faire.

Qu'est-ce qu'on observe alors ? On observe que nous sommes tous passés dans une

société de traçabilité, de traçabilité avancée. Nous sommes tous devenus producteurs de données qu'on le veuille ou non par nos smartphones, par la réponse à des enquêtes, par plein de choses... donc on est dans une société de traçabilité. Et ce qu'il faut que l'on arrive à réglementer en notre sein -et j'utilise volontairement « réglementer », j'évoquerai tout à l'heure l'agencement avec les notions d'éthique- le niveau, le seuil de traçabilité que l'on souhaite avoir et si, dans une société démocratique, on a envie de tracer tous nos faits et gestes tels qu'ils soient et s'ils sont tracés est-ce qu'on a envie qu'ils soient accessibles par les uns et les autres ou est-ce qu'on a envie de positionner un seuil de traçabilité en dessous duquel on réglemente comment on accède à la donnée.

Alors à la CNDP comme chez toutes nos consœurs à l'étranger,

On va considérer que lorsque l'on est en procédure judiciaire ou lorsqu'il y a des éléments liés à la sécurité intérieure ou extérieure de l'Etat, il y a des règles qui ne sont pas forcément les règles que l'on va appliquer pour tout un chacun de façon commerciale ou politique. Pourquoi commerciale ou politique ? Parce que l'exploitation de ces données dans le passé, il y a quelques décennies, elle pouvait infléchir ou influencer des politiques commerciales ; cela s'appelait le ciblage. C'était la façon avec laquelle on va vous convaincre d'acheter quelque chose, mais ces dernières années on voit bien que cela peut impacter les orientations politiques. Mais elle peut également impacter les orientations de gouvernance c'est-à-dire qu'un gouvernement demain va avoir besoin et a déjà besoin et manipule déjà des données à caractère personnel pour élaborer des orientations de gouvernement, de gestion, et de planification des politiques publiques.

Il faut savoir aussi pourquoi je réfute la notion d'un modèle marocain, même si on l'applique en prenant en compte le contexte marocain ; c'est une marocanité dans la façon de faire mais la valeur reste universelle. On ne peut pas tailler la valeur à notre mesure on est tous des êtres humains, et je dirais que bientôt on va tous être des données donc il faut que l'on reste dans cette dimension universelle.

Par rapport à ces traitements-là, on travaille et on a travaillé avec nos consœurs à l'étranger parce que le droit de la protection des données est un droit qui 1) se construit au fur et à mesure, il faut être en alerte. On ne peut pas se permettre d'être dans une logique où on légifère dix ans après le phénomène, on doit être capable de suivre, d'être en alerte, d'être agile et de trouver des mécanismes d'encadrement qui puissent préserver justement la vie privée de l'être humain et l'être humain. On en reparlera.

Donc, le droit de la protection des données se construit au fur et à mesure et il se construit à l'international ; et nous avons choisi de façon définitive de travailler à l'international pour anticiper et comprendre.

Il existe par exemple au niveau africain, un réseau des autorités africaines de la protection des données et la CNDP du Maroc en assure le secrétariat permanent. Il existe également à l'échelle mondiale ce qu'on appelle la GPA-Global Privacy Assembly qui est ce qu'on appelle chez nous l'ONU de la propriété des données et le Maroc, en la CNDP, a eu le plaisir et l'honneur d'être élu en octobre dernier membre du comité exécutif de huit membres. On estime qu'on est plein dedans.

Pourquoi je cite cela ? Ce n'est pas pour faire montrer les acquis de la CNDP mais parce que ce que je vais évoquer là a une contenance internationale, parce que nous allons direct à l'objectif et nous nous renseignons sur ce qui se passe dans chacun des pays pour ne pas perdre de temps à trop réfléchir et à ne pas avancer.

Donc avec cette vision-là, on parle souvent de protection des données mais il faut savoir que c'est un héritage qui est en cours de changement et d'évolution dans le monde. Dans les années 1980, on parlait de protection des données parce que ce

qui allait devenir le digital par la suite et qui était l'informatique c'était les fichiers plats et les bases de données ; donc protéger les données à caractère personnel c'était la manière de l'époque de dire qu'on protégeait l'être humain, le citoyen, du contexte informatique naissant.

Aujourd'hui, nous allons tous -et ce n'est pas la volonté de la seule petite CNDP- toutes les autorités de protection dans le monde sont en train d'aller dans ce qu'on appelle la protection du citoyen au sein de l'écosystème numérique. La protection du citoyen au sein de l'écosystème numérique n'est pas qu'une problématique de données ou une problématique technique c'est également des problématiques -ceux qui m'ont déjà entendu parler vous pensez que je radote, mais en fait je ne radote pas je suis juste constant pour vous dire la même chose-.

La manière avec laquelle on peut manipuler des phrases dans des tweets, des SMS, etc. fait qu'on est en train de former des générations de jeunes à ne pas rédiger. Ceux qui ont quelques cheveux blancs se rappellent des rédactions, des dissertations, de la synthèse, du résumé, de l'analyse etc. c'est cela qui fait qu'un être humain est citoyen parce qu'il est capable d'appréhender une situation, de l'analyser, de prendre position, de défendre et de faire un plaidoyer.

Alors on n'est pas en train de dire qu'il faut confisquer les réseaux sociaux aux jeunes -mes filles ne me le pardonneraient pas-, je suis juste en train de dire qu'il faut gérer et faire en sorte que ces usages-là ne provoquent pas des défaillances par ailleurs (incapacité de rédiger, incapacité de maîtriser l'orthographe, incapacité de maîtriser une grammaire quel que soit la langue le problème n'est même pas du choix de la langue, incapacité d'analyser, de faire un plaidoyer etc.). Il faut que l'on veille tous ensemble à nous battre pour former non pas des consommateurs du numérique mais des citoyens du numérique, et c'est là où la problématique devient compliquée.

Pourquoi compliquée ? Parce que nous avons un challenge, nous tous, toutes les institutions qui peuvent être amenées à travailler sur ces sujets-là, qui est d'y aller en cohérence et pas de façon exclusive, je vais expliquer pourquoi. En fait, si nous avons un verre devant nous, il est impossible de dire qu'il appartient à une seule personne : il y a celui qui l'a fabriqué, celui qui va le remplir, celui qui va boire dedans, celui qui va le laver, celui qui va reboire dedans… donc le même objet est manipulé par différentes personnes selon différentes facettes.

Aujourd'hui le numérique et l'Intelligence Artificielle c'est un concept qui est « mono-stakeholder » mono-partie prenante -et on doit tous arriver à gérer le verre correctement pour éviter de le casser ou de le vider, on doit en faire le meilleur usage-. Pourquoi je vous dis cela ? Parce que l'Intelligence Artificielle qui est devenue ces dernières années et ces derniers mois on va dire, moi quand je reçois des CV ces derniers-temps chez tous il y a écrit « Intelligence Artificielle », « data scientist » et quelques mots clés et quand vous interviewez les gens en tout cas la partie artificielle de l'intelligence elle est à revoir.

Pour nous, l'Intelligence Artificielle est une chose très très sérieuse et qui doit être appréhendée avec beaucoup de sérieux. Il y a 2 grandes familles au moins -après je ne m'étalerai pas puisqu'il y a des spécialistes dans la salle- :

La famille de « j'ai plein de données, je ne suis pas capable de l'analyser tout seul, je me dote d'une capacité de calcul et de super calculateurs qui vont me permettre de faire des calculs à grande échelle très vite et arriver à une décision », donc il y a le côté artificiel par rapport à un cerveau normal qui est d'accélérer et d'étendre ma capacité à prendre une décision par rapport à un ensemble de données. »

Il y a une autre famille qui est moins connue, qui est un changement de paradigme c'est-à-dire que je ne décide pas comme par le passé en faisant 1 puis 2 puis

3 puis 4 et comme j'ai un super calculateur j'arrive très vite avec des données de masse c'est une autre façon d'analyser, une autre façon de gérer on parlerait même peut-être plus d'algorithme ou bien il faudrait étendre le concept parce que si on demande à qui appartient l'algorithme je vous dirais à Al Khawarizmi mais au-delà de la boutade je dirais qu'il y a cette nouvelle façon de faire qui nécessite beaucoup d'investissements en recherche, en formation et en attention.

Que l'on prenne la famille 1 ou la famille 2 que je viens d'évoquer, tout cela aboutit à des traitements et des décisions automatiques et donc quand on parle de protection des personnes physiques à l'égard des traitements des données à caractère personnel. Pour nous, l'Intelligence Artificielle est un traitement parmi d'autres, est un paradigme parmi d'autres à considérer dans les traitements. Une fois qu'on a dit cela... je ne suis pas en train de vous dire poussez-vous c'est à nous ce n'est pas le sens de mon propos, le sens de mon propos c'est que c'est un sujet multi-parties prenantes et on doit veiller à préserver la synergie entre les différentes institutions à cet effet.

La communauté de la protection des données travaille depuis des années sur cette problématique, elle a abouti notamment avec un des cercles (il y en a plusieurs) qui est le Conseil de l'Europe et sa Convention 108. Je rappelle que le Maroc a ratifié la Convention 108 en avril 2019, cette convention et son comité a sorti un document après plusieurs mois de travail qui date du 25 janvier 2019 où sont relatés un certain nombre de recommandations, un ensemble d'orientations, de choses qui sont très naturelles.

Aujourd'hui il faut que le citoyen puisse avoir un recours en cas de décision automatique qui ne lui convient pas, donc cela veut dire qu'il faut imposer que les faiseurs de décisions automatiques respectent certaines normes et soient auditables. Il faut que les algorithmes soient déchiffrables, décodables, explicables, -donc tout cela dans une intervention de panel inaugural je ne vais pas m'attarder dans le détail ce ces choses-là- mais nous travaillons dessus et nous espérons pouvoir continuer à travailler dessus avec le CNDH et d'autres institutions.

La capacité tout simplement, l'exemple le plus patent c'est dans les films américains, ce n'est pas un drone qui de façon automatique va tuer un Afghan ; il faut qu'il y ait une décision humaine derrière un tableau à quelques milliers de kilomètres qui va dire ok ou pas ok. C'est donc ces mécanismes-là qui doivent transpirer dans la société et pour transpirer dans la société on a besoin à ce que toutes les institutions qui protègent la société et son citoyen que ce soit le CNDH, le Conseil de la Concurrence, le Conseil Économique Social et Environnemental ou la CNDP chacune de ces institutions a son rôle à porter par rapport à cette chose-là

Je souhaiterais passer à une autre partie de mon intervention en insistant sur l'importance des usages, la différence entre la gestion et l'augmentation des usages, et la potentielle réglementation de la technologie qui est impossible. On ne peut pas réglementer la technologie, on cherche à réglementer les usages. Lorsque vous avez une voiture ce n'est pas parce qu'il est marqué 240km/h sur votre compteur que vous allez rouler à 240 km/h, il y a des règles, un code de la route et un permis de conduire qu'il faut avoir et puis il y a une signalétique sur tous les chemins ; on s'arrête à un feu (théoriquement), on ne va pas sur le passage piéton etc.

Souvent, et on s'est retrouvé durant cette pandémie dans des situations très opérationnelles où on nous dit voilà on a un truc là qu'on va utiliser on dit ouais mais est-ce que à bord de votre voiture vous appuyez sur le champignon jusqu'à 200km/h, non vous roulez en ville à 60km/h... Eh bien avec votre caméra à reconnaissance faciale peut-être qu'il y a des règles à respecter.

Je parle de la reconnaissance faciale parce

que c'est évidemment une technologie de progrès mais ce n'est pas une biométrie comme les autres, quand vous posez votre empreinte digitale vous êtes là on ne vous coupe pas le doigt pour venir le poser à votre place, quand on regarde votre iris vous êtes là que vous soyez conscient ou non, que vous soyez sachant ou pas, décalé numériquement -pour ne pas dire analphabète- ou pas… c'est autre chose mais vous êtes là.

Avec la reconnaissance faciale, on peut vous identifier alors que vous êtes dans une foule ou que vous marchez dans la rue. La question donc n'est pas cela mais la question c'est quelles décisions automatiques vont être prises après coup, donc il s'agit d'une technologie sans consentement. Dans certains cas de maintien de l'ordre public on n'a pas vraiment de problème, dans d'autres cas notamment les cas commerciaux on se pose la question et on voudrait réfléchir.La CNDP avait lancé un moratoire pour stopper les projets ou plutôt on les a mis sous version expérimentale, c'est-à-dire on a dit aux gens attention ce n'est pas aussi simple, ce n'est pas parce que des Chinois vont vouloir vous vendre des caméras pas chères qu'il faut en équiper tout le territoire.

Nous avons réfléchi et nous sommes arrivés en 2020 à une délibération qui a été très bien négocié avec la Banque Centrale pour le secteur bancaire qui était de dire « ben on ne va pas autoriser chaque banque de créer sa base biométrique et chaque provider de services, chaque assurance, chaque fournisseur de logistique, chaque livreur de pizza à construire sa propre base -parce que cela peut être sympathique qu'il vous mette le téléphone sur le visage et on est sûr de l'identité de celui qui a commandé la pizza-».

On a donc travaillé pour la mise en place d'un titre de confiance nationale au même titre que vous avez une carte bancaire Visa ou Mastercard qui permet d'acheter quelque chose, de payer un repas dans un restaurant, la carte bancaire cautionne que vous pouvez acheter quelque chose mais ne sait pas si vous avez pris de la salade ou un coca cola, de la même manière qu'elle ne connaît pas la vie de votre compte bancaire, c'est juste que votre banque cautionne pour 140 dhs… Qu'est-ce que cela veut dire ? Cela veut dire qu'on se doit dans ces problématiques de séparer les données d'usage des données d'authentification et ça c'est fondamental d'un point de vue droits de l'Homme et d'un point de vue protection de la vie privée. On doit séparer les données d'usage des données d'authentification. Quand on dit ça on se retrouve avec le défi d'architecture c'est-à-dire que l'on considère que la protection des données à caractère personnel et notamment la gestion des traitements -si on reste sur des traitements granulaires on risque de faire des erreurs graves on risque d'annuler un traitement alors qu'on est en train de perturber tout un écosystème- donc on a délibéré en décembre 2020 qu'on va aller vers une approche écosystème.

C'est l'écosystème qu'on doit reconsidérer parce qu'en termes de protection de la vie privée le risque zéro n'existe pas -ce n'est pas la peine de courir derrière le zéro il n'existe pas- donc il faut selon les contextes voir si on accepte 2%, 3%, 5%, 10% de risque et ce sont des choses à considérer au niveau de l'Intelligence Artificielle, on ne peut pas définir l'intégrité d'un traitement d'IA si l'on n'a pas considéré toutes ces choses-là.

Donc je reviens à cette orientation conclue par le comité de la Convention 108 du Conseil de l'Europe le 25 juin 2019, nous travaillons là-dedans et nous continuons à travailler au niveau international là-dedans donc c'est clair que nous sommes attentifs à ce qui se passe par ailleurs parce qu'on est convaincu que tôt ou tard on se retrouvera autour des mêmes tables parce qu'on gère la même chose, on gère les mêmes êtres humains, on gère leurs données et on gère les usages que les uns et les autres souhaitent avoir par rapport à ces données.

Mot d'ouverture de Mme Sarah Lamrani

Secrétaire Général du Ministère de la transition numérique et de la réforme de l'administration

Monsieur le Ministre de l'Enseignement Supérieur de la Recherche et de l'Innovation

Monsieur le Président du Conseil Économique, Social et Environnemental

Monsieur le Président de la Commission Nationale de Contrôle de la Protection des Données

Mesdames et Messieurs,

J'ai le plaisir de participer aujourd'hui à ce colloque, et de vous transmettre les excuses de Madame la Ministre qui n'a pas pu être parmi nous aujourd'hui. Elle a néanmoins tenu à se faire représenter, compte tenu de l'importance majeure de la thématique que nous abordons aujourd'hui, et souhaite plein succès aux travaux de ce colloque.

Le choix d'aborder le thème de l'intelligence artificielle et des droits de l'Homme est un choix crucial, qui vient appuyer les nombreuses réflexions entamées au niveau national et international autour de l'éthique et de l'intelligence artificielle, et de son utilisation dans le respect des droits et des libertés fondamentales.

L'intelligence artificielle s'impose aujourd'hui comme une technologie à fort impact sur le développement socio-économique du Maroc, de l'Afrique et dans le monde dans sa globalité, il est donc essentiel pour nous tous de déployer les efforts nécessaires afin de soutenir les travaux et les initiatives visant à développer cette technologie au Maroc. Ceci vient en droite ligne de la Vision Royale de Sa Majesté le Roi Mohammed VI, que Dieu L'assiste, qui a souligné dans son discours adressé le 05 juillet dernier à la réunion du groupe africain des Ministres des Finances et des Gouverneurs, des banques centrales , des états Africains membres de la banque mondiale et du FMI, je cite « Bien que cette technologie implique des enjeux complexes auxquels il est difficile de faire face, il y'a lieu d'accélérer le rythme de digitalisation de nos sociétés africaines afin de combler le fossé qui nous sépare des pays avancés et pouvoir, dès lors, récolter les fruits de la mutation numérique à l'échelle planétaire ». Fin de citation

En effet l'intelligence artificielle, qui est une des technologies du numérique, est un outil puissant pouvant non seulement répondre à des problématiques du quotidien mais aussi influencer les interactions, les choix et les prises de décisions avec un fort impact sur des secteurs majeurs tels que l'éducation, les sciences sociales et humaines, la culture, la communication et l'information.

Dans ce sens, le 24 Novembre 2021, lors de sa 41ème session, la conférence générale de l'UNESCO a adopté la recommandation sur l'éthique de l'intelligence artificielle, premier instrument normatif mondial, dont le Maroc est fier d'être l'un des premiers pays à participer à sa mise en œuvre.

A travers l'application de cette recommandation, le Ministère de la Transition Numérique et de la Réforme de l'Administration prépare la mise en place d'un Comité National de l'éthique de l'intelligence artificielle, qui aura pour objectifs de déployer un outil d'évaluation de l'impact de l'éthique, et de veiller au respect et suivi des mesures appropriées, notamment législatives, pour intégrer les principes et normes qui y sont énoncés.

Le Ministère se veut ainsi une véritable force motrice pour accompagner ce chantier et pour répondre aux recommandations du Nouveau Modèle de développement concernant le digital et à la Vision Royale portée par Sa Majesté Le Roi Mohammed VI, Que Dieu L'Assiste, qui a appelé à réaliser un développement global et intégré permettant au Maroc d'occuper la place qui lui revient dans un monde transformé par la révolution numérique.

Compte tenu de l'enjeu de cette

thématique, et de l'accélération de l'utilisation des nouvelles technologies, il est aujourd'hui crucial d'accompagner cette évolution et d'appuyer tous les travaux qui ont pour objectif de mettre les systèmes d'Intelligence Artificielle au service de l'humanité, des individus, des sociétés, de l'environnement et des écosystèmes.

Mesdames et Messieurs,

Je tiens à remercier et à féliciter, au nom de Madame la Ministre, le Conseil National des Droits de l'Homme pour l'organisation de ce colloque, réunissant des intervenants et experts nationaux et internationaux et visant à appuyer l'échange de nouvelles expertises et le partage des bonnes pratiques autour des principaux défis à relever pour mettre en œuvre une intelligence artificielle éthique.

Notre mission à tous ici présents est d'œuvrer pour la protection des droits humains dans le domaine des technologies et de l'intelligence artificielle et de contribuer au développement de notre pays pour construire le Maroc de demain.

Merci à tous.

AI ETHICS, GLOBAL, REGIONAL AND NATIONAL INITIATIVES

Amal El Fallah Seghrouchni

Professeure universitaire et Directrice de Ai Movement, UM6P, Maroc, membre du COMEST / UNESCO et Professeure à l'Université Paris La Sorbonne, France

Lecture dans les standards émergents en matière de l'éthique de l'IA

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

Bonjour,

Merci pour l'invitation.

En préliminaire, permettez-moi de vous présenter en quelques mots « AI Mouvement » qui est un centre d'excellence national, à vocation internationale, de recherche et de développement en intelligence artificielle, qui a pour objectif de favoriser l'émergence d'une expertise marocaine en Intelligence Artificielle et en Sciences des Données.

Le AI Mouvement repose sur 7 piliers qui structurent et définissent ses missions :

- Favoriser l'émergence d'un écosystème attractif
- Développer une approche orientée vers le terrain
- Promouvoir la formation inclusive
- Combiner la recherche théorique et appliquée, le développement technologique, l'innovation, et le transfert industriel
- Assurer et maintenir une vision stratégique et géopolitique concernant les transformations induites par l'IA.
- Créer et renforcer des partenariats internationaux et nationaux de qualité.

Il a été créé à la suite d'une série de travaux menés au sein de l'Université Mohammed VI Polytechnique, qui a organisé en 2018 la première conférence sur l'intelligence artificielle pour l'Afrique en collaboration avec l'UNESCO. Cette conférence a abouti à la déclaration de Benguerir, dont l'objectif était de promouvoir et d'adopter l'intelligence artificielle sur le continent africain.

Nous avons établi plusieurs collaborations au niveau national et international. Nous travaillons en étroite collaboration avec l'UNESCO sur divers sujets visant à renforcer les capacités en intelligence artificielle en Afrique.

Le 7 juillet 2022, l'UNESCO et le Centre International d'Intelligence Artificielle Ai Movement de l'Université Mohammed VI Polytechnique (UM6P) au Maroc ont signé un accord de partenariat stratégique et financier de trois ans. Cet accord vise le renforcement des capacités en intelligence artificielle (IA) en Afrique.

Je cite deux programmes phares :

- Campus Africa : Renforcer l'enseignement supérieur en Afrique.
- Exploiter les nouvelles technologies émergentes pour le développement durable en Afrique, notamment par la mise en œuvre de recommandations sur l'éthique de l'intelligence artificielle.

Alors, de quelle intelligence artificielle parle-t-on ?

On se réfère ici à une intelligence artificielle (IA) capable de réaliser des interactions sophistiquées et de reproduire

des processus cognitifs similaires à ceux des humains. Cette IA se caractérise par une autonomie avancée, associée à des capacités telles que la perception détaillée, l'activité sensori-motrice, la coordination fluide et la navigation précise.

Grâce à des techniques d'apprentissage artificiel, elle excelle dans l'acquisition et le traitement des données, ainsi que dans la représentation et la gestion des connaissances. Ses compétences en interaction et cognition sont également remarquables, incluant la communication à travers des langages verbaux et gestuels, ainsi que l'interaction sociale et multimodale.

L'IA dispose également de compétences en raisonnement et planification, lui permettant de faire des prédictions, de concevoir des politiques et de gérer l'exécution des tâches avec rigueur. La supervision est un aspect essentiel, garantissant une surveillance et un contrôle efficaces des processus. Ces capacités permettent à l'IA de simuler des comportements humains complexes, en s'adaptant et en réagissant de manière sophistiquée à son environnement.

L'intelligence artificielle transcende et enrichit le monde numérique en passant de l'automatisation à une véritable autonomisation des systèmes. Elle joue un rôle essentiel dans la protection du domaine digital et dans la création d'interfaces intelligentes qui facilitent les interactions entre les univers réels et virtuels.

En tant que domaine scientifique le plus proche de la cognition humaine, l'IA permet une interaction multimodale, intégrant la voix, les gestes, les émotions, et même le métavers. Elle est capable de prédire les comportements des individus et des groupes, d'expliquer les tendances émergentes et d'offrir une assistance dans divers processus d'apprentissage.

Cependant, l'IA présente également des aspects problématiques. En tant que technologie à double tranchant, elle peut être intrusive, influencer les comportements, engendrer des addictions et créer des « nudges » qui manipulent les décisions des utilisateurs. De plus, les systèmes d'IA qualifiés de « boîtes noires » posent des défis majeurs en matière de transparence et de fiabilité, étant souvent difficiles à expliquer et à comprendre.

Il est donc crucial de mettre en place un cadre éthique et réglementaire rigoureux, couvrant l'ensemble du cycle de vie des systèmes d'IA, depuis leur conception jusqu'à leur déploiement et leur démantèlement, pour garantir leur utilisation responsable et sécurisée.

Blue IA se positionne comme un levier essentiel pour le développement économique. On estime, que d'ici 2025, 45 entreprises prévoient d'employer plus de machines intelligentes que de personnes, ainsi les appareils électroniques qui pourraient bénéficier de l'IA atteindraient le nombre de 6 milliards. A l'horizon de 2045, l'intelligence artificielle pourrait être capable de surpasser toutes les intelligences humaines, transformant ainsi profondément le paysage économique et technologique.

Blue IA se positionne comme un levier essentiel pour le développement sociétal. Elle vise à réduire les inégalités culturelles, économiques, de genre, sociales, cognitives et technologiques. De plus, elle s'engage à aider à lutter contre les violences et les discriminations, contribuant ainsi à la création d'une société plus juste et équitable.

Les impacts de l'intelligence artificielle (IA) se manifestent dans de nombreux domaines de la société. Sur le plan sociétal, l'IA influence le marché du travail, l'égalité, la privacité, les droits de l'Homme et la dignité. Elle a également un impact significatif sur la psychologie humaine, notamment sur les relations interpersonnelles et les addictions.

L'IA influence la justice, la transparence, la responsabilité et le contrôle. Elle modifie aussi le système juridique, notamment par le biais de la legal tech et impacte les domaines de la loi criminelle et le droit de la responsabilité délictuelle.

Sur le plan financier, l'IA joue un rôle crucial dans la gestion des risques, la détection de fraude et l'automatisation. Enfin, l'IA a des répercussions importantes sur l'environnement et la planète, en particulier en matière de durabilité et d'empreinte carbone.

L'intelligence artificielle pose plusieurs défis cruciaux pour les droits de l'Homme et la dignité humaine. La subordination algorithmique peut créer des hiérarchies injustes basées sur des critères algorithmiques, compromettant l'égalité et le respect de chacun.

Les techniques de manipulation et les nudges, utilisés pour influencer les choix, risquent d'affaiblir l'autonomie des individus en les orientant vers des décisions qui ne sont pas nécessairement dans leur meilleur intérêt. La protection de la vie privée et des données personnelles est également en jeu, avec la collecte et l'analyse d'informations sensibles souvent sans consentement pleinement éclairé. Enfin, les biais et discriminations, qu'ils soient intégrés dans les algorithmes ou exogènes, peuvent exacerber les inégalités et engendrer des injustices.

1. SUBORDINATION ALGORITHMIQUE

Nous observons une transition d'un système où les décisions sont prises exclusivement par des humains à un modèle où les décisions sont faites par des machines, parfois pour nous. Cette subordination algorithmique entraîne un partage de l'autorité en faveur des systèmes automatisés. Par exemple, des algorithmes comme ceux de Parcoursup déterminent l'affectation des bacheliers, tandis que des drones peuvent décider de lancer des missiles, et des voitures autonomes peuvent prendre des décisions potentiellement fatales, comme foncer sur un piéton. Les systèmes de recommandation attribuent des crédits et les prix des billets d'avion se modifient automatiquement en ligne. De plus, certaines voitures refusent de démarrer si la ceinture de sécurité n'est pas bouclée.

Avec moins de 0,05 % de la population mondiale capable de coder ou de créer des algorithmes, une question cruciale se pose : à qui confions-nous la gouvernance de ces technologies ?

2. MANIPULATION ET NUDGES - LE PRINCIPE

Les nudges, issus des sciences du comportement et inspirés du design industriel, reposent sur l'idée que des suggestions indirectes peuvent, sans forcer, influencer les motivations, les incitations et la prise de décision des groupes et des individus, au moins de manière aussi efficace sinon plus efficacement que l'instruction directe, la législation ou l'exécution.

Des hommes politiques (comme David Cameron et Barack Obama) ont eu recours aux nudges pour pousser les citoyens à prendre les « bonnes décisions » dans divers domaines.

Les robots empathiques, en particulier ceux qui ressemblent à des enfants, ont un pouvoir d'attraction notable. Une expérience allemande a démontré cela avec un robot qui disait : « Non, non, non, ne partez pas ! J'ai peur de ne jamais revoir la lumière du jour. » Suite à cette interaction, 13 des 43 participants à l'expérience ont été si touchés par la supplication du robot qu'ils ont éprouvé des difficultés à éteindre l'appareil. Cet exemple illustre comment les robots, lorsqu'ils manifestent des émotions, peuvent manipuler les comportements humains et influencer les décisions de manière significative.

3. QUESTION DE VIE PRIVÉE ET PROTECTION DES DONNÉES

La protection de la vie privée et des données personnelles a évolué au fil des années avec plusieurs régulations clés. En 1981, la Convention 108 du Conseil de l'Europe a été adoptée pour encadrer la protection des données personnelles. En 1995, la Directive sur la protection des données a été mise en

place pour réguler la collecte, le traitement et le transfert de données au sein de l'UE. Le Parlement britannique a également légiféré en 1998 avec la Data Protection Act, visant à protéger les données personnelles stockées sur les ordinateurs ou dans des systèmes de classement papier organisés.

Le Règlement général sur la protection des données (GDPR) de l'UE, entré en vigueur en 2018, impose des normes beaucoup plus strictes pour la collecte, le stockage et l'utilisation des données personnelles. Parallèlement, les préoccupations éthiques concernant l'IA ont émergé, notamment avec le débat sur l'IA au service de l'humanité. Il est crucial de garantir que les algorithmes et les corrélations utilisés par l'IA respectent les principes d'éthique et de responsabilité.

4.1 Biais cognitifs

Les biais cognitifs influencent la manière dont les algorithmes traitent les informations, en intégrant des préjugés souvent issus des opinions humaines. Par exemple, un algorithme peut systématiquement associer le mot « femme » à des termes tels que « femme de ménage », « nourrice », « réceptionniste » ou « styliste », tandis que « homme » est associé à des professions comme « chef », « philosophe », « capitaine » ou « financier ». Ces associations reflètent des mécanismes de pensée qui altèrent le jugement et perpétuent des stéréotypes.

4.2 Biais économiques

Les biais économiques dans les algorithmes peuvent être intentionnels ou non. Par exemple, les algorithmes peuvent cibler certains segments de la population de manière biaisée, comme dans le cas d'offres d'emploi qui privilégient un genre par rapport à l'autre. Une autre illustration est la manipulation des résultats de recherche, où des entreprises comme Google ont été condamnées par la Commission européenne en 2017 à une amende de 2,4 milliards d'euros pour avoir favorisé ses propres produits dans Google Shopping, au détriment de ses concurrents.

4.3 Biais statistiques ou de données

Les biais statistiques ou de données incluent le biais d'endogénéité et la variable omise. Omettre des caractéristiques sensibles comme l'origine ethnique peut renforcer les discriminations et nuire à l'équité des algorithmes. Le RGPD pose un dilemme en ce qui concerne la lutte contre la discrimination des femmes si les données sur le genre ne sont pas collectées. Le biais de sélection survient lorsque les caractéristiques de la population étudiée diffèrent de celles de la population générale. James Heckman, lauréat du Prix Nobel d'économie, a montré que ce biais est similaire au biais de variable omise. Des cas pratiques incluent le calcul des scores de crédit (credit scoring) et la sélection des talents, comme l'illustre l'exemple d'Amazon, où un déséquilibre de genre dans les entrées de données a conduit à des biais dans les processus de recrutement.

Vers une prise de conscience globale des droits et de l'éthique de l'IA :

En octobre 2020, le Chili est devenu le premier pays au monde à intégrer les droits cognitifs dans sa constitution. Le Congrès national chilien a adopté un amendement définissant l'identité mentale comme un droit à protéger, interdisant toute forme de manipulation.

Parallèlement, de nombreuses initiatives internationales visent à encadrer l'éthique de l'IA. L'UNESCO a lancé une consultation mondiale sur son projet d'éthique de l'IA durant l'été 2020, et l'OCDE a constitué un groupe d'experts pour élaborer des recommandations éthiques similaires. Le Parlement Européen a adopté trois nouvelles résolutions le 20 octobre 2020, portant sur la responsabilité de l'IA, la propriété intellectuelle liée à l'IA, et l'encadrement éthique des développements en intelligence artificielle, incluant le droit à l'information.

Des centres de recherche à travers le monde, tels que ceux à Montréal et Melbourne, se concentrent également sur ces questions, contribuant à une prise de

conscience globale et à la définition de normes éthiques pour l'IA.

Blue and Red AI :

- **L'IA pour une cybersécurité avancée:** L'intelligence artificielle peut améliorer la cybersécurité en offrant des contrôles plus efficaces tels que des antivirus optimisés, la détection et la prévention des intrusions, ainsi que l'analyse automatisée des courriels. Elle permet aussi la découverte automatique des vulnérabilités, l'analyse des logiciels malveillants, et le développement de stratégies avancées de cyberdéfense.
- Robustesse et vulnérabilité des algorithmes d'IA : Les algorithmes d'IA doivent faire face à des défis en termes de robustesse et de vulnérabilité. Par exemple, l'apprentissage automatique contradictoire exploite les faiblesses des algorithmes pour modifier leur comportement, tandis que les systèmes cyber-physiques alimentés par l'IA peuvent être ciblés par des attaques sophistiquées.
- Mauvaise utilisation de l'IA : L'IA peut également être utilisée de manière abusive pour créer des vidéos truquées, développer des logiciels malveillants, ou mener des attaques d'ingénierie sociale de plus en plus sophistiquées.
- L'IA dans la lutte contre la cybercriminalité : Par ailleurs, l'IA joue un rôle crucial dans la lutte contre la cybercriminalité en améliorant les technologies biométriques, en analysant les fraudes, en renforçant les méthodes de cryptage, et en combattant les fake news.

Plusieurs organisations intergouvernementales jouent un rôle clé dans l'encadrement et la régulation de l'intelligence artificielle. Le Conseil de l'Europe, par l'intermédiaire de son Comité ad hoc sur l'intelligence artificielle (CAHAI), travaille sur les questions liées à l'IA. L'Union européenne mène également des efforts importants dans ce domaine, avec notamment le Groupe d'experts de haut niveau de la Commission européenne, qui se concentre sur le développement d'une IA digne de confiance.

Au niveau international, la première initiative d'experts de l'Organisation de coopération et de développement économiques (OCDE), le AIGO, a été suivie par le réseau d'experts de l'OCDE sur l'IA (ONE AI). L'OCDE a publié une recommandation sur l'IA, et l'Observatoire des politiques de l'IA de l'OCDE (OECD.AI) surveille les politiques dans ce domaine.

Les principes sur l'IA du G20, inspirés par les travaux de l'OCDE, ont été intégrés dans la Déclaration ministérielle du G20 sur le commerce et l'économie numérique. De plus, le G7 a adopté une vision commune de Charlevoix pour l'avenir de l'IA.

En Afrique, le Groupe de travail de l'Union africaine sur l'IA et le Groupe de travail de la Ligue des États arabes sur l'IA se penchent également sur ces enjeux. Enfin, l'UNESCO a émis une recommandation en deux temps pour orienter les politiques et les pratiques en matière d'intelligence artificielle.

Initiatives du secteur privé, organisations professionnelles et ONG :

Diverses initiatives issues du secteur privé, des organisations professionnelles et des ONG contribuent à encadrer l'éthique de l'intelligence artificielle. L'Institut des Ingénieurs Électriciens et Électroniciens (IEEE) a lancé une initiative mondiale pour l'éthique des systèmes autonomes et intelligents, axée sur la conception éthique de ces technologies.

Le Forum économique mondial a publié un livre blanc sur la gouvernance multipartite mondiale des technologies, proposant des approches pour une régulation collaborative. De son côté, l'UNI Global Union a établi les « 10 grands principes pour une intelligence artificielle éthique » pour guider les pratiques dans le secteur.

La Déclaration de Montréal appelle à un développement responsable de l'IA, tandis que la Déclaration de Toronto se concentre sur la protection du droit à l'égalité et à la non-discrimination dans les systèmes basés sur l'apprentissage automatique. Enfin, le Partenariat sur l'IA a formulé des principes pour une IA harmonieuse, soulignant l'importance de principes éthiques communs dans le développement et l'application de l'intelligence artificielle.

Recommandations de l'UNESCO

Après un long processus, l'UNESCO a adopté une recommandation sur l'éthique de l'intelligence artificielle (IA). Les 193 États membres de l'UNESCO ont approuvé ce nouveau texte, qui vise à fournir des "orientations éthiques" pour les technologies d'IA. Bien que cet instrument juridique mondial sur l'IA ne soit pas contraignant pour les signataires, il les incite à mettre en place leurs propres cadres pour lutter contre certains abus, tels que les biais et les violations de la vie privée. L'objectif est de "fournir une direction responsable pour les technologies d'IA" en assurant un "accès équitable" tout en prenant en compte les "dommages et impacts".

Concernant la gestion des données, étroitement liées à l'IA, l'UNESCO souligne la nécessité d'une politique appropriée. Le rapport de l'UNESCO met en avant l'importance d'évaluer "la qualité des données de formation", en considérant notamment "l'adéquation des processus de collecte et de sélection des données". Il est également essentiel de trouver un équilibre entre la collecte de données et la protection de la vie privée.

En revanche, l'UNESCO ne traite pas directement du sujet de la reconnaissance faciale, qui reste un thème central en Europe.

Les questions éthiques touchent l'ensemble du cycle de vie d'un système d'intelligence artificielle. Cela comprend plusieurs étapes essentielles : la recherche, la conception, le développement, le déploiement, l'utilisation.

Les valeurs universelles fondamentales incluent le respect, la protection et la promotion des droits de l'Homme, ainsi que des libertés fondamentales et de la dignité humaine. Elles engagent également à favoriser l'épanouissement de l'environnement et des écosystèmes, à garantir la diversité et l'inclusivité, et à vivre dans des sociétés pacifiques, justes et interconnectées. Pour soutenir ces valeurs, plusieurs principes directeurs ont été établis :

- Proportionnalité et ne pas nuire
- Sûreté et sécurité
- Équité et non-discrimination
- Le développement durable
- Droit à la vie privée et protection des données
- Contrôle humain et détermination
- Transparence et explicabilité
- Responsabilité et obligation de rendre compte
- Sensibilisation et alphabétisation
- Gouvernance et collaboration multipartite et adaptative.

Le lien entre ces valeurs et principes dans le déploiement des recommandations sur l'IA peut se traduire par l'utilisation de deux outils principaux :

- **The Ethical impact Assessment (ElAs)** : un outil conçu de manière agile et universelle, permettant aux développeurs des systèmes d'IA d'évaluer les effets potentiels des ensembles de données, des cadres conceptuels et des algorithmes sur la société. Cet outil vise à aider à ajuster les algorithmes afin de traiter ou atténuer d'éventuels impacts négatifs.
- **The readiness assessment (RA)** : un outil qui soutient les États membres dans l'identification de leur niveau de conformité par rapport aux recommandations. Reconnaissant

que chaque État membre se trouve à des niveaux de préparation variés en matière de développement, d'adoption et d'utilisation des technologies de l'IA, cet outil a pour but d'aider les gouvernements à affiner leurs mécanismes réglementaires.

En intégrant ces outils, nous pouvons garantir que les valeurs et principes directeurs sont effectivement appliqués, favorisant ainsi un avenir où la dignité humaine, la diversité et le développement durable sont au cœur de toutes les initiatives.

Concernant le déploiement des systèmes d'intelligence artificielle (IA) et la mise en œuvre des recommandations éthiques, il est crucial de se demander comment garantir le respect et l'application de ces recommandations ?

Un des sujets qui va arriver très vite sur la table, c'est que si on n'est pas conforme avec les réglementations européennes y aura plus de business avec l'Europe à partir de l'Afrique. Aujourd'hui, il est très important de bien comprendre ce qui se passe aussi au niveau des initiatives européennes. Thierry Breton, le Commissaire Européen, avait dit « Nous voulons que tous les pays qui travaillent avec nous respectent nos normes, nos valeurs, et que les systèmes soient compatibles et consistants avec nos valeurs et nos règles ».

L'Europe a classifiée les niveaux critiques de l'algorithme en quatre catégories :

-Les algorithmes classés comme présentant un risque inacceptable seront interdits.

-Les algorithmes classés comme présentant un risque élevé devront respecter des obligations strictes pour pouvoir être mis sur le marché.

-Les algorithmes classés comme présentant un risque limité seront soumis à des exigences de transparence spécifiques.

-Les algorithmes à risque minimal seront librement autorisés (jeux vidéo, anti-spam, filtres...).

D'ailleurs, c'est une classification qui fait un peu rire parce que vous avez des risques qui ne sont pas acceptables, les systèmes qui présentent des risques très hauts, des systèmes qui présentent des risques limités et dans les minimales risques on a tout ce qui est sérieuse, gamme, etcetera.

Donc, ça me paraît quand même assez compliqué de classer les jeux vidéo dans des systèmes qui ne présentent aucun risque pour la population et on voit qu'il y a les addictions et des suicides dans les hôpitaux psychiatriques.

La question qui va se poser très vite, c'est comment classifier un système d'IA dans l'une ou l'autre des catégories ? Là, c'est ce qu'on appelle en informatique, l'aspect de très haut niveau. Maintenant implémenté, ça va être une autre paire de manches.

En conclusion, je dirai est ce qu'on va vers l'audit des systèmes d'intelligence artificielle ? Il y a dix ans, il n'y avait pas de data scientistes, il n'y avait pas de data protection officer. D'accord, c'est des métiers qui sont apparus avec le RGPD, la protection des données.

Si on nous dit il faut maintenant classer les systèmes, savoir ce qu'ils font, et bien tout un boulevard pour l'audit, pour auditer tous les systèmes, ceux qui sont risqués, seront prohibés. Cela fera du travail pour beaucoup de cabinets de conseil.

Céline Caira

Economist and policy analyst in Artificial Intelligence at the OECD, Canada

OECD Principles on Artificial Intelligence

Today, I will be presenting something that you're probably very familiar with: the OECD AI Principles. These principles were adopted by member countries and partners in 2019. The theme of this colloquium is how we move from principles to practice. As practitioners working in AI, I hope to leave you with some tools, statistics, and data points that you can use in your work.

So, what are the AI principles? In 2019, countries adhered to the recommendation on AI with the objective of promoting trustworthy AI. I won't go into too much detail on this, as you are likely all familiar with it. However, we do have Principle 1.2, which focuses on human-centered values and fairness.

These principles were adopted as the first intergovernmental standard on AI in May 2019 by 46 countries and formed the basis for the G20 principles. From the African continent, South Africa and Egypt have adhered to the OECD AI principles.

So, before we dive deeper, it always helps to start with the definition of what AI is. There are varying definitions, but I'll give you the OECD one. Essentially, an AI system, as we define it, is a machine-based system that can influence the environment by providing a certain input and producing a certain output. It can make independent predictions for a given set of objectives.

These systems operate with varying levels of autonomy, which is at the heart of the risk question. For example, a recommender system for choosing a restaurant for dinner carries different risks compared to an AI system scanning X-rays to detect cancer.

I would also like to echo some of what we've heard today: oftentimes, when people talk to me about AI, they're actually referring to automation and digitalization. Therefore, it's very important to distinguish between what AI is, what automation is, and what digitalization is, as well as the interactions among these concepts.

The OECD AI Principles cover ten principles, divided into two areas. The first are values-based principles, and the one I will highlight today is Principle 1.2, which focuses on human-centered values and fairness, an aspect that relates to democracy and human rights.

We also emphasize that AI should be transparent, explainable, robust, and safe. When things go wrong, it should ultimately be accountable. There are five recommendations for governments, which are actions that can be taken to promote an ecosystem for trustworthy AI. These are particularly important when considering the challenges in preparing for the labor market transition in today's context, where there is often a great deal of anxiety.

People are concerned that their jobs will be replaced by automation and artificial intelligence. Therefore, it is essential for governments to take actions to train, upskill, and reskill their citizens, ultimately ensuring that those citizens trust the AI being deployed in their countries. I'd like to highlight Principle 1.2, which focuses on human-centered values and fairness: "AI actors should respect the rule of law, human rights, and democratic values." This principle applies to various areas, from privacy to data. We have discussed many of these issues this morning, including the Justice Act and access to justice, and the need for safeguards.

When we talk about human oversight, algorithmic impact assessments, and having a human in the loop, we are addressing the core of these principles.

While we have these high-level principles, the challenge is how to move from principles

to practice. OECD countries have adhered to these principles, but the implementation depends on the national context. We support countries in this effort through several pillars, including the OECD AI Policy Observatory, which provides data, trends, and a database of national AI policies, along with a blog.

Internally at the OECD, we have a Working Party on Artificial Intelligence, which includes several working groups. One group focuses on risk classification, another on tools and accountability, and a third on underlying infrastructure.

I would also like to note that in 2019, we published a report on "The State of Implementation of the OECD AI Principles" This report provides a comprehensive synthesis of what countries are doing to move from principles to practice.

The last topic I'll discuss in this section is AI country reviews. At the OECD, we are also helping to measure the implementation of the principles. AI country reviews are essentially tailored assessments of a country using specific indicators to evaluate how it is implementing these principles from a concrete perspective.

We examine areas such as AI governance, trustworthy AI, and the ecosystem for AI research and development. For instance, we assess whether the country has sufficient digital infrastructure, including supercomputers, computing capacity, and access to skills and job programs, as well as international cooperation.

One example of an AI country review, which I'm very pleased to mention, is the first-ever OECD AI country review in the MENA region. We are currently conducting a review of Egypt's implementation of the OECD AI principles. This review focuses on AI in government, startups, and foreign direct investment (FDI), with a specific emphasis on development in health care and agriculture.

For the final part of my presentation, the exciting part, I hope to provide you with tools you can use, along with some data and trends from the OECD AI Policy Observatory. I've tailored this presentation to the region and, where possible, incorporated data from Morocco. We have a database of national AI policies, which is a live repository of over 700 policy initiatives from more than 60 countries. This is quite impressive, considering that from 2015 to 2017, only a few countries had AI strategies.

According to this repository, we have over 700 documents from more than 60 countries and territories in the EU. This is a valuable resource for you as legal practitioners, as it allows us to track the implementation of AI principles and see what other countries are doing, enabling us to compare and draw best practices.

This database is set up and updated frequently based on surveys with national contact points, so we strive to provide the most current data possible.

We have a wealth of information on skills and jobs, research and development (R&D), and venture capital investments. There has been an increase in the number of AI research publications from the MENA region, which is particularly interesting as it began around the turn of the century, coinciding with the rise in popularity of deep learning and machine learning. It is not surprising that the number of AI research publications is steadily increasing.

Looking at domestic and inter
research collaboration for 2021,
data from Morocco, revealing a significant flow of research collaborations with the EU as well as domestically within Morocco.

I should note that there is a trend of growing international collaboration, particularly in research and development in artificial intelligence. When it comes to hiring, it all boils down to people. Hiring is crucial, and we can observe the AI growth rate over time. While we don't have data for Morocco in this instance, we do have snapshots from select countries.

What's particularly interesting is that we can see a downturn in hiring trends. The data clearly indicates fluctuations, an increase followed by a decrease. We speculate that this is likely due to events in 2020, specifically COVID-19. The data shows a hiring freeze and a slowdown in the labor market for companies seeking individuals with AI skills.

The last thing I'll show you today is skills migration between countries. At the bottom, you have countries that have an AI skills deficit, meaning that more AI-skilled workers are leaving the country than coming in. At the top, you see countries with a surplus of AI skills migration, where more workers are arriving.

An interesting point is that Canada was initially ranked lower in terms of skills migration, but after introducing a policy called the AI Skills Strategy and the Global Skills Strategy, we see a notable trend: a few years after the policy was implemented, Canada became one of the top countries in this area. This helps to explain the shift.

Essentially, the takeaway is that policies matter and can help ensure that there is an adequate supply of skilled labor to meet the needs of your AI and innovation economy.

Thank you.

Diane Galbois-LeHalle

Maître de conférences en droit privé, directrice du Master droit de l'intelligence artificielle à l'Institut Catholique de Paris (Faculté de sciences sociales, d'économie et de droit), membre de l'Unité de Recherche EA7403

Approche européenne pour une régulation éthique et juridique de l'IA : une approche fondée sur les risques

Depuis quelques années, plusieurs initiatives des différentes instances de l'Union européenne (UE) pour réguler l'intelligence artificielle (IA) ont été présentées. La présidente de la Commission européenne, Ursula Van der Leyen a fait de la régulation du numérique, des données et de l'IA une des priorités de son mandat. L'ensemble de ces initiatives n'a pas encore pu aboutir en raison des différentes étapes de la procédure législative par lesquelles un texte doit passer avant d'être adopté puis d'entrer en vigueur. Certaines sont d'ailleurs de simples instruments de soft law, et ont vocation à le rester.

Le point de départ de la réflexion au niveau européen peut être daté à 2017 – ce qui, si on y songe, est relativement tardif par rapport aux autres initiatives nationales ou régionales : le gouvernement coréen, par exemple, a travaillé dès 2007 sur un projet de charte éthique des robots. Si la Commission européenne a en principe le monopole du pouvoir d'initiative des textes législatifs, le Parlement européen a le pouvoir de demander à la Commission de soumettre une proposition sur un sujet donné ; pouvoir dont il a largement usé en matière d'IA. Dès début 2017, le Parlement européen a publié une Résolution contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique[4]. L'objectif qui transparaît est double : d'une part, l'UE entend assurer la

4- Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (20152103/(INL).

sécurité de ses citoyens et le respect de leurs droits fondamentaux mais, d'autre part, elle est soucieuse de ne pas freiner l'innovation, afin d'encourager l'investissement des entreprises européennes dans ce secteur, pour les rendre compétitives face aux puissants géants américains ou chinois (les GAFAM[5] et les BATX[6]), et de faciliter le déploiement des petites et moyennes entreprises dans ce secteur.

En réponse à cette sollicitation du Parlement européen, la Commission européenne a publié le 25 avril 2018 une communication sur l'intelligence artificielle pour l'Europe[7]. La garantie d'un cadre éthique et juridique adéquat y figure parmi les principaux objectifs, aux côtés du renforcement de la capacité technologique et industrielle et du recours à l'IA dans l'ensemble de l'économie.

Pour atteindre ces objectifs, la Commission européenne a constitué un groupe d'experts « de haut niveau » – composé d'universitaires, de représentants d'institutions et de sociétés – chargé d'y réfléchir, qui a publié en avril 2019 des « lignes directrices en matière d'éthique pour une IA digne de confiance »[8]. Ce texte de soft law, qui n'a pas une valeur normative, a ensuite servi de base de réflexion au législateur européen pour proposer des textes de loi.

Le souci du respect des droits fondamentaux est le fondement de toutes ces réflexions. Ce guide est toutefois trop flou, en particulier en cas de conflit entre deux droits fondamentaux, de sorte que les instances législatives réfléchissent à une réglementation plus précise. Les principes éthiques et droits fondamentaux promus au sein de l'UE pour une IA éthique (1), précèdent donc l'édiction d'une législation européenne sur l'IA (2).

Les principes éthiques et droits fondamentaux pour une IA éthique

Dans les « lignes directrices en matière d'éthique pour une IA digne de confiance » du groupe d'experts de haut niveau mandaté par la Commission européenne, le caractère éthique de l'IA est l'un des trois aspects d'une IA « digne de confiance », aux côtés d'une IA « robuste » (sur le plan tant technique que social, pour éviter tout préjudice), et d'une IA « licite » (c'est-à-dire conformes aux législations et réglementations applicables, si elles existent). C'est précisément à défaut de législation, ou en raison de leur retard sur les évolutions technologiques, que le respect des exigences éthiques est fondamental.

Le respect des droits fondamentaux se trouve au cœur de l'approche éthique de l'IA promue par l'UE on trouve. Selon l'article 2 du Traité sur l'Union européenne « L'Union est fondée sur les valeurs de respect de la dignité humaine, de liberté, de démocratie, d'égalité, de l'État de droit, ainsi que de respect des droits de l'Homme, y compris des droits des personnes appartenant à des minorités ». Les droits de l'Homme sont ensuite juridiquement déclinés dans la Charte des droits fondamentaux de l'UE et dans la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales du Conseil de l'Europe, prévue par le traité de Lisbonne.

À partir de ces catalogues génériques des droits de l'homme, quatre principes fondamentaux ont été dégagés pour garantir une IA éthique. Leur respect s'impose tant au stade de la mise au point que du déploiement et de l'utilisation des systèmes d'IA. Le premier principe est celui de l'autonomie humaine, autrement dit du droit de chaque individu à l'autodétermination de chacun. Le Règlement général pour la protection

5- Sigle désignant les entreprises Google, Amazon, Facebook, Apple et Microsoft.

6- Sigle désignant les entreprises Baidu, Alibaba, Tencent et Xiaomi.

7- Communication de la Commission au Parlement européen, au Conseil européen, au Conseil, au Comité économique et social européen et au Comité des régions, « L'intelligence artificielle pour l'Europe », COM(2018) 237 final.

8- Commission européenne (Groupe d'experts indépendants de haut niveau sur l'intelligence artificielle), *Lignes directrices en matière d'éthique pour une IA digne de confiance*, Bruxelles, 2019.

des données (RGPD)[9] décline déjà ce droit en prévoyant le droit de ne pas faire l'objet de décision exclusivement automatisées et l'encadrement du profilage[10]. La prévention de toute atteinte ou nuisance aux êtres humains figure également parmi ces principes clés, ce qui renvoie à la protection de la dignité humaine ainsi que de l'intégrité mentale et physique, en particulier pour les personnes vulnérables. Le principe d'équité est quant à lui décliné selon un volet matériel – veiller à l'absence de biais injustes, de discrimination et de stigmatisation – et un volet procédural – donner la possibilité de contester les décisions prises par des systèmes d'IA. Enfin, le quatrième principe requis pour une IA éthique est l'explicabilité des outils utilisés en fonction du contexte et de la gravité des conséquences.

Assurer le respect de ces principes et résoudre les éventuels conflits entre eux passe par le respect d'un certain nombre d'exigences plus précises : action humaine et contrôle humain, robustesse technique et sécurité, respect de la vie privée et gouvernance des données, transparence, diversité, non-discrimination et équité, bien-être sociétal et environnemental, et responsabilité. Certaines de ces exigences sont déjà prises en compte par la réglementation européenne.

Ainsi, le droit fondamental au respect de la vie privée et à la protection des données est déjà largement assuré par le RGPD, qui pose, notamment, les principes de limitation des finalités des traitements, de minimisation des données, de limitation de la conservation, et de confidentialité. Les questions de gouvernance des données sont quant à elles prises en compte avec le tout récent Règlement sur la gouvernance des données (Data Governance Act), publié le 30 mai dernier et qui entrera en vigueur au 24 septembre 2023[11]. Ce texte met notamment en œuvre le « data altruisme », c'est-à-dire la mise en place des mesures pour permettre aux personnes de mettre volontairement à disposition leurs données sans contrepartie pour des objectifs d'intérêt général, tels que les soins de santé, la lutte contre le changement climatique, la recherche scientifique, etc. Les questions de transparence sont également au cœur de la législation européenne, en particulier de la proposition de règlement sur les services numériques (Data Services Act ou « DSA »)[12], et qui visera à réguler les plateformes, grandes pourvoyeuses d'algorithmes.

De manière concrète, le groupe d'experts européen sur l'IA préconise de garantir le respect de ces exigences au cas par cas en programmant dans le système d'IA des comportements ou états que le système devrait toujours suivre, et inversement des restrictions relatives aux comportements ou états que le système ne devrait jamais transgresser, ainsi que des combinaisons des deux ou de garanties démontrables plus complexes concernant le comportement du système. Il recommande également le recours à une étude d'impact en amont dès lors que les droits fondamentaux sont concernés, et une réflexion approfondie sur l'équilibre entre eux lorsque plusieurs sont en conflit (par exemple, le principe de la prévention de toute atteinte, qui pourrait tendre vers la mise en place d'un système de vidéos de surveillance intelligentes, entre en conflit avec le droit au respect de la vie privée).

La Commission européenne a salué les propositions du groupe d'experts, tout en insistant sur les contextes spécifiques, parfois très différents. À titre d'illustration, une application de l'IA suggérant à mauvais escient la lecture d'un livre est beaucoup moins dangereuse qu'une application

9- Règlement (UE) 2016679/ du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 9546//CE (règlement général sur la protection des données).

10- Article 22 du RGPD.

11- Règlement du Parlement européen et du Conseil sur la gouvernance européenne des données (acte sur la gouvernance des données) COM (2020) 767.

12- Proposée le 15 décembre 2020 par la Commission européenne (COM/2020825/ final), sa négociation est sur le point d'aboutir.

posant un mauvais diagnostic de cancer. En conséquence, la Commission européenne entend proposer une réglementation proportionnée, en adoptant une approche fondée sur les risques ; dans cet exemple, l'algorithme de recommandation de livre pourrait faire l'objet d'une surveillance moins stricte que celui de détection des cancers[13].

La régulation de l'IA en fonction du risque

Deux projets de textes législatifs tendant à l'encadrement éthique de l'IA ont été présentés au sein de l'Union européenne. La Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes[14], et la Proposition de la Commission européenne de Règlement sur l'IA (AI act) du 21 avril 2021[15]. À l'heure où ces lignes sont écrites, aucun de ces deux textes n'est entré en vigueur ; ils sont tous deux au stade de projet.

Afin de promouvoir l'adoption de l'IA par les entreprises européennes tout en tenant compte des risques associés à certaines utilisations de cette technologie[16], plusieurs options présentant différents degrés d'intervention réglementaire ont été envisagées : un système de label non obligatoire ; une approche ad hoc secteur par secteur ; des exigences obligatoires pour tous les systèmes d'IA, indépendamment du niveau de risque qu'ils présentent ; une approche transversale et proportionnée fondée sur les risques, éventuellement accompagnée de codes de conduite pour les systèmes d'IA qui ne sont pas à haut risque. C'est cette dernière option qui a été retenue. Elle doit permettre de promouvoir l'IA tout en tenant compte des risques, avec efficacité sans être excessivement normatif, afin de ne pas créer une charge disproportionnée, en particulier pour les petites et moyennes entreprises.

Le Parlement européen et la Commission s'accordent sur le principe d'une réglementation en fonction du risque, mais diffèrent sur les détails de cette réglementation. Alors que le Parlement européen propose de n'encadrer que les systèmes à haut risque, la Commission propose une approche plus graduée, avec une classification selon trois niveaux de risques : (i) les risques inacceptables sont tout bonnement interdits ; cela renvoie au crédit social utilisé par les autorités publiques et à la reconnaissance faciale dans les lieux publics en temps réel (sauf exceptions liées à la recherche d'auteurs ou de victimes d'infractions ou de menaces pour l'intérêt général, notamment terroristes) ; les IA à hauts risques, encadrées par plusieurs types d'exigence ; les IA à faibles risques, autorisées sans restriction. Le projet d'AI act prévoit en outre un niveau spécifique imposant des obligations de transparence quand le système d'IA interagit avec les personnes, quand il détecte, sur la base de données biométriques, des émotions ou l'appartenance à une catégorie sociale, ou génère ou manipule des contenus.

Le contenu des obligations réglementaires pour les IA à haut risque n'est pas strictement identique dans les deux projets. Si tous deux imposent un contrôle humain – pouvant aller jusqu'à la désactivation du système d'IA par un bouton d'arrêt – et un niveau de robustesse et de résilience, le Parlement européen impose en outre la prise en compte de l'empreinte environnementale

13- Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, « Renforcer la confiance dans l'intelligence artificielle axée sur le facteur humain », 8 avril 2019, COM/2019168/ final.

14- Résolution du Parlement européen du 20 octobre 2020 contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes (20202012/(INL)).

15- Proposition de Règlement établissant des règles harmonisées concernant l'intelligence artificielle (Législation sur l'intelligence artificielle) COM(2021) 206 final.

16- Commission européenne est développée dans son Livre blanc du 19 février 2020 « Intelligence artificielle. Une approche européenne axée sur l'excellence et la confiance »

et prévoit expressément un droit à réparation en cas de préjudice découlant du manquement à une des obligations prévues, alors que la commission ne le prévoit pas. Il propose également de mettre en place une certification de conformité éthique. Inversement, le projet d'AI Act va plus loin sur d'autres points, par exemple en imposant un enregistrement automatique des événements du système d'IA. Il est également plus précis, en distinguant les obligations qui incombent aux fournisseurs, aux distributeurs ou aux utilisateurs professionnels.

La question centrale est, bien entendu, celle de la définition des systèmes d'IA à haut risque, soumis à ces obligations spécifiques. Le Parlement européen propose un double critère, lié au secteur d'une part, et aux usages ou finalités d'autre part[17]. Pour qu'un système d'IA soit qualifié à haut risque, il doit donc à la fois relever d'un secteur qualifié comme tel, et son usage ou sa finalité doit également être considérée à haut risque. Le Parlement européen propose une double liste, exhaustive et cumulative, au sein de laquelle sont énumérés d'une part les secteurs à haut risque (emploi, éducation, soins de santé, transports, énergie, asile, migration, contrôles aux frontières, système judiciaire et services de sécurité sociale, défense et sécurité, finance, banque et assurance), et d'autre part les usages ou finalités à haut risque (recrutement, notation et évaluation des étudiants, affectation de fonds publics, octroi de prêts, commerce, courtage, fiscalité, etc., traitements et procédures médicaux, processus électoraux et campagnes politiques, décisions du secteur public ayant une incidence importante et directe sur les droits et obligations des personnes physiques ou morales, conduite automatisée, gestion du trafic, systèmes militaires autonomes, production et distribution d'énergie, gestion des déchets, contrôle des émissions.

Pour la Commission européenne, la qualification de système d'IA à haut risque doit tenir, de même, du secteur et de l'utilisation du système d'IA. Grosso modo, les secteurs considérés sont sensiblement identiques ; en revanche, les types d'utilisation à haut risque sont listés secteur par secteur, et non pas de manière générique. Par exemple, dans le secteur de l'Éducation et de la formation professionnelle, deux finalités sont considérées comme à haut risque : les systèmes d'IA destinés à être utilisés pour déterminer l'accès ou l'affectation de personnes physiques aux établissements d'enseignement et de formation professionnelle ; et ceux destinés à être utilisés pour évaluer les étudiants ou les candidats à l'accès aux établissements d'enseignement. Autre exemple, encore plus détaillé : dans le secteur des autorités répressives : les systèmes d'IA destinés à être utilisés pour mener des évaluations individuelles des risques visant à déterminer la probabilité qu'une personne physique commette une infraction ou récidive, ou le risque encouru par les victimes potentielles d'infractions pénales ; les détecteurs de mensonges ou d'émotions ; les systèmes type deepfake ; les systèmes d'IA destinés à être utilisés pour évaluer la fiabilité des preuves au cours d'enquêtes ou de poursuites pénales ; les systèmes d'IA destinés à être utilisés pour prédire la survenance ou la réitération d'une infraction pénale réelle ou potentielle sur la base du profilage de personnes physiques ; les systèmes d'IA de profilage de personnes ; les systèmes d'IA destinés à être utilisés pour l'analyse de la criminalité des personnes physiques.

17- Notons que la définition des systèmes d'IA à haut risque proposée dans le cadre de la Résolution contenant des recommandations à la Commission concernant un cadre pour les aspects éthiques de l'intelligence artificielle diffère de celle proposée par le même Parlement européen, le même jour, dans sa Résolution contenant des recommandations à la Commission sur un régime de responsabilité civile pour l'intelligence artificielle (20202014/(INL)) : le «haut risque» y est défini comme « un risque important, dans un système d'IA opérant de manière autonome, de causer un préjudice ou un dommage à une ou plusieurs personnes d'une manière aléatoire et qui va au-delà de ce à quoi l'on peut raisonnablement s'attendre; l'importance de ce risque dépend de l'interaction entre la gravité de l'éventuel préjudice ou dommage, le degré d'autonomie de décision, la probabilité que le risque se concrétise, la manière dont le système d'IA est utilisé et le contexte de l'utilisation ».

Ce système très détaillé a l'avantage de la sécurité juridique… mais présente l'inconvénient d'être difficilement exhaustif, et rapidement obsolète. Certes, la Commission propose une mise à jour annuelle de cette liste. Le système semble néanmoins beaucoup trop rigide, et présente le risque d'une régulation toujours en retard sur les évolutions technologiques.

En conclusion, si l'Union européenne va globalement dans la bonne direction en tâchant de trouver une troisième voie, entre celle du capitalisme américain et celle de la surveillance étatique chinoise. Elle entend promouvoir de nouveaux standards, à l'instar de ce qu'elle a réussi à imposer avec le RGPD, en promouvant une IA éthique, respectueuse des droits fondamentaux. Toutefois, elle tâtonne encore pour trouver la manière de réglementer, de manière précise et exhaustive, mais également souple et évolutive, les systèmes d'IA.

Mark Gaffley

Director of African Observatory on Responsible AI and Research Fellow at Research ICT Africa, South Africa

Code and codification: the possibility of rights-respecting AI

Good afternoon fellow panellists and speakers. I would first like to extend my gratitude to the National Council for Human Rights of the Kingdom of Morocco and its academic partners for inviting me here to speak today. My name is Mark Gaffley and I am here in my capacity as project manager for the African Observatory on Responsible Artificial Intelligence and as a Research Fellow with Research ICT Africa, a pro-poor policy think tank based out of South Africa. I am also reading towards my PhD in Jurisprudence at the University of Cape Town, South Africa, where my studies contrast Kantian ethics with the theories underpinning key components of algorithms used in artificial intelligence (AI). It is on this comparison of the ethical with algorithmic that I wish to address you all today.

What do I mean by Code and codification: the possibility of rights respecting artificial intelligence? In choosing this title, I was reminded of a keynote address at a colloquium given by French philosopher Jacques Derrida where he was asked to speak to the topic of deconstruction and the possibility of justice. His address, titled Force of Law: The Mystical Foundation of Authority, amongst other things, noted the intimate relationship between law and the enforceability of law. Authority grants one the ability to apply law, if necessary, by force. As Derrida's discussion unfolds, he unpacks this relationship and moves to assessing what might be considered legitimate (and illegitimate) enforcement of law as well as what might be considered just and unjust actions. As a technology, AI is unique in that it carries with it a set of algorithmic norms, represented as code, which I call the 'law of the algorithm'. This law is increasingly applied to govern social situations and settings, and brings with it its own enforceability.

Derrida's discussion is relevant today, and to global debates on AI ethics because this law of the algorithm is more frequently taking on these social functions in the way it interacts with and influences our thoughts, habits and feelings, and is coming up against the human rights and ethical standards we already have codified as law in the form of statues, legislation and regulation. Indeed, the global AI debate is facing similar predicaments around legitimacy and justness in considering how best the seemingly unstoppable force of AI can interact with the immovable object that is the respect, protection, promotion and fulfilment of fundamental human rights and values. We could say this quandary is the very reason we are all here today.

With this in mind, I think the common comparison of AI to a double-edged sword insofar as it can be both helpful and harmful may not be entirely accurate. The more I come to understand what the law of the algorithm is, how it is operating and the authority it carries under its own laws, the more I begin to ask myself whether it is not, in fact, more accurate to describe AI as inconsistent with a rights respecting state of existence. I think this position reveals itself more clearly when we attempt to unpack algorithmic first principles.

I will not attempt to give a comprehensive single definition of what AI is - increasingly, the word AI has become much like ethical concepts of truth and dignity where it can mean different things to different people depending on the contexts in which its meaning is explored or interrogated. For the purposes of today, at a very high level, I think it safe to broadly frame AI as the method or application that exists between input data provided by its designers, developers and, indeed, most of us sitting here today (as data subjects), and output data, being the results, learnings, predictions and expectations derived from the algorithmic method. This

really is the true magic of the technology under enquiry.

It is this algorithmic method that I believe we as ethicists and human rights practitioners need to dedicate more resources to fully understanding - indeed, it is this 'code' which has, for the most part, already been codified into the digital world that we now seek to apply ethical guidelines and principles to, in the form of policy and legislation. As time progresses, I think this task of codifying the code becomes increasingly more difficult as more and more documents are being generated that seek to describe what ethical principles of AI should mean. As an example, a landscape study in 2019 revealed this number to be 84 different documents. Most of these documents find common ground with variations of requirements of fairness, accountability, integrity, responsibility and transparency in the design, development, deployment and governance of AI. These then represent the ideals that make rights respecting AI possible.

However, what we really need to be asking ourselves more intently is whether, fundamentally, AI is indeed capable of being consistent with these ideals. Here I must caveat that I lack the technical expertise to fully comprehend what the theories underpinning AI actually entail, but, if we consider Bayes' theorem, competition theory or utility theory as some of the core theories of algorithmic design and development, we can note the following:

1- Bayes' theorem looks to determine whether a particular state of the world is true by weighing expected costs and benefits, which are usually weighed using financial considerations – put differently Bayes' theorem is about probability and inferences based on the likelihood of particular events occurring, given available evidence;

2- game theory is based on mechanisms of prices and products, and the gaining and spending of incomes – it is a theory about games of strategy in a social exchange economy that try to obtain the optimum result, being maximum resulting satisfaction; and

3- In the AI space, utility theory pins rationality to the maximisation of expected utility.

From this admittedly rudimentary understanding, we can see a philosophy of competition, prediction and performance is inherent in the algorithmic method. Indeed, I would argue that the first principle of AI is about harnessing the efficiencies its designers can extract, according to their own preferences of economic utility. Here I am not talking about intentionality – knowingly or not, the theories and principles designers rely on to develop their systems are founded on this point of view and it is these that we should be targeting for ethical analysis. Efficiency is the tenet that needs to be weighed against non-derogable human rights. This statement is itself a contradiction because how can we even measure something against the non-derogable? Yet, this is the challenge we face for achieving a rights respecting AI.

Creating room for this weighting against rights to take place is exactly what the law of the algorithm accomplishes – in any social setting, AI can only function by reducing human values, norms and principles to the sum of 1s and 0s and not to principles such as fairness and equality. Indeed, early AI pioneers such as Alan Turing, with the Turing test, and John von Neumann in his book The Computer and the Brain, advanced philosophies that reduce biological functions to a series of 1s and 0s thus allowing them to equate the computational with the biological. This kind of view introduces uncertainty about what life means and can be used to ascribe rights to machines, potentially at our own (human) expense. We should never confuse the point that humans do not work like AI, it is the AI that we are modelling to work like us. In my opinion, this conflation downplays the importance of human rights

respecting AI and we need to be mindful of rhetoric that ultimately seeks to muddy the waters in order to develop the certainty it needs to advance certain developmental AI agendas.

Indeed, it is the enhanced capability to measure and predict that distinguishes AI technologies from those that came before – we are no longer looking to physical processes that harness and extract from nature, but instead we are looking to digital processes that harness and extract from human nature. This means the focus of the technology is the very rights we seek to protect. Even if read with principles for responsible use, if the fundamental theories the designers rely on do not first take into consideration rights respecting principles, a human centric approach, is, in all scenarios, most likely an afterthought and not an imperative. I say this because AI principles and rights respecting principles are rooted in inherently different concepts.

The former with optimisation and efficiency and the latter with how we relate to and treat one another.

This point is relevant even where people are not the focus of AI. In Kantian philosophy, freedom is attained through autonomy of the will – i.e. the ability to independently choose between one decision or another. If this is true, then we can note that with the rise in machine autonomy and decision-making, there is a corresponding decrease in human autonomy and decision-making as we outsource this very function to AI systems. This is more obvious in surveillance technologies, or social media and ecommerce, where design choices deliberately steer us down a path of least resistance and most susceptibility. But even, for example, in a medical diagnostics programme or an AgTech solution, over time there will be less reason to doubt the decisions made by AI and more reliance on the guidance they have given. For the Kantian, this represents loss of freedom. In Homo Deus, Yuval Noah Harari speaks of the emergence of the data religion society and the possibility of a 'useless' class coming into existence. I think we need not go this far – the ethical concern is, broadly speaking, the gradual intellectual deadening of society through the way in which AI potentially limits our own autonomy and critical thinking abilities.

This brings me to my final thought – that cautious views of AI development may be considered to be non-progressive or stand in the way of societal advancement. When this is raised, I constantly find myself referring back to a passage by C. S. Lewis. In describing the belief in Progress, Lewis writes:

"any given generation is always [deemed] in all respects wiser than all previous generations. To those who believe thus, our ancestors are superseded and there seems nothing improbable in the claim that the whole world was wrong until the day before yesterday and now has suddenly become right. With such people I confess I cannot argue, for I do not share their basic assumption. Believers in progress rightly note that in the world of machines the new model supersedes the old; from this they falsely infer a similar kind of supercession in such things as virtue and wisdom."

The development of AI cannot leave behind traditional rights and value systems, which may stand in its way or seem antiquated. Though AI application is advancing and increasingly applicable to more problems, its core principles that I spoke to earlier effectively remain the same. Indeed, Bayes' theorem was first conceived in the 1700s and game theory in the 1950s. The literature speaks to how these basic principles of AI have remained unchanged for almost 70 years. And so we should be mindful that whilst AI models are continually being built upon and superseded, the algorithmic method behind them is remaining the same. We should remember that even the most advanced AI applications (for example, Google's image recognition system with its over 300,000 parameters) started their development journey with a single line of

code and a formative method that may be inconsistent with the rights and freedoms we have worked to codify at global, regional and national levels. If this is not interrogated and the focus of ethical scrutiny, then the possibility of rights respecting AI will remain a pipedream and our attempts to bring fairness and accountability to the digital world will likely remain noise to the progression of AI.

Thank you.

John C. Havens

Executive Director of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Member of the World Economic Forum Global Future Council on Human Rights and Technology, USA

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

Today, I want to touch on the nature of innovation in Artificial Intelligence relating to Human Rights, ethics, risk, and sustainability for the planet. I realize that it is a lot to talk about in 5 minutes, but basically, the first idea I want to start with is the idea of risk.

There is so much good work being done around the world with regard to what risk even is; how do we define risk for Artificial Intelligence and other technologies? I would like to introduce -by the way, it is, of course, not just me- these ideas.

I should say, by the way, that I am speaking for myself. I work on staff at the IEEE Standards Association, I am the lead of the sustainability practice for our IEEE SA, and I also have run IEEE's largest Artificial Intelligence and ethics program called "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems"; so I am basing my ideas on both roles.

Most of the time, defining risk happens because we have to be concerned about the known physical risks that a technology will bring on for a user, which is, of course, incredibly important and will never stop, and it should not: that is the due diligence anyone creating technology has to think about. Also, fair known mental stressors for whatever reason a device may cause stress, so these are critical things to think about in terms of risk.

However, the opportunity for innovation is to now broaden that lens of what risk is and keep in mind that we must take care of the planet now -we should have been all along-. We must all, on an individual and societal level, especially as people creating technology or policy, prioritize even the increase of what I call planetary or ecosystem flourishing. What this means is this idea of sort of planet-positive or climate-positive in which you give more to the planet than you take out.

Risk is often associated with what not to do. So say you build a car, and you take from the emissions, you make sure it does not damage the planet; that is great, we have to do that, but now the extra step and the opportunity for innovation is also to say that when we build this car or when we build any technology: how do we give back more to the planet than we are taking out.

Usually, that has to do with systems thinking global supply chain issues, all the stakeholders involved, shareholders to employees to the people using it to partners, and asking the questions: how would this technology end up, will the hardware end up being recycled and reused or is it going to a landfill that is perhaps in the Amazon, in the Global South, and harming without that intention necessarily -or obviously usually people do not intend to harm, but it is still causing those problems for the environment.

So, in 5 minutes, that is the main message I want to leave for you, and you may be saying: but how does this relate to Artificial Intelligence and data?

The answer is from what the science says: in 50 or 60 years, the planet will be here continuing to warm, but it is the people that we have to take care of. The way to do that with all technology and policy is to put human rights and environmental rights together from the standpoint of what we should be prioritizing, as anyone building technology. The answer is look first and be accountable to matrixes like the ESG matrix and the United Nations Sustainable Development Goals. First and foremost, say, is this technology that I am building or that I am recommending from a policy standpoint; will it increase

planetary flourishing and human well-being? Once you know those answers and the matrix for what will make sure that you can prove that those things are happening. This is both innovations, meaning research and development innovation, and more importantly, it is the way that we make sure that our children, grandchildren, and descendants are here to enjoy those technologies for a long time.

I am John C. Havens with the IEEE Standards Association, and there are a couple of standards, IEEE 7010, in particular, that I would recommend if you want to learn more about what I have discussed.

Thank you again for the opportunity to present.

AI AT THE SERVICE OF HUMAN RIGHT

John Schwartz

Head of Global Business Development edX, USA

AI in lifelong learning

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

I'm here to discuss the intersection between artificial intelligence and the future of learning in a post-pandemic world. I want to focus on what higher education looks like now that we've emerged from this pandemic, which, as everyone in this room knows, was a significant turning point. For years, we've been advocating for the power of digital learning, it's cost-effective, time-efficient, and agile, addressing the skills-based competencies that governments and the private sector are grappling with.

During the pandemic, in just a few weeks, hundreds of millions of people transitioned online and were forced to become comfortable with discomfort. Now, the educational landscape is learner-centric; we must meet learners where they are in their journey.

Adnan shared his program for 14-year-olds, preparing them for the workforce and equipping them with essential skills. Education needs to be blended, stackable, and career-focused, especially since many universities and countries face the challenge of a significant percentage of graduates being unemployable.

So, how do we address this? edX is a unique organization, and many of you in the room are familiar with us. We are a consortium of top universities and enterprises providing content in critical areas that workers and learners need globally. edX was founded by Harvard and MIT in 2012, with a mission to expand access to elite content that was historically available only to those who could afford the high fees of Ivy League schools. Fast forward to 2022, we now offer over 4,000 programs to nearly 50 million learners, including 38 of the top 50 universities worldwide. This is remarkable progress in just a decade, making education accessible to everyone, regardless of socioeconomic background, religion, race, or geographic location. We call it "free to degree," as all of our content is accessible for free.

This approach allows us to meet learners at every stage of their lifelong learning journey. As we know, education doesn't stop after college, or even if one doesn't attend college. The jobs my children will encounter have not even been invented yet, and they are just six years away from entering the workforce. We offer courses that refresh skills, all the way up to Ph.D. level, which can be taken online for free from leading providers in each field.

In the last five years, the world has changed significantly. For instance, with MIT Healthcare, most of my appointments can now be done online. I can consult accomplished doctors from anywhere. During the pandemic, I worked out with my family six days a week, accessing countless fitness instructors. In the evenings, we watched movies together, with access to virtually anything we desired.

While the world has evolved, education is gradually recognizing the need for change. Many professors at prestigious institutions were initially reluctant to embrase the power of digital transformation and its potentiel impact on billions of people.

Those in education must embrace this change and transform, as the skills gap due to technology and AI continues to widen. In

a learner-centric world , something I hope all educators will adopt in their terminology, it's about data-enabled, personalized learning and the growth of modular, stackable education. Not everyone learns the same way at the same time; options must be available.

The significant shift resulting from the pandemic has been towards blended learning. I don't think you can find a provost, dean, or professor who wouldn't agree that hybrid learning meets learners where they are, catering to their comfort as digital natives.

It's crucial to determine which skills-based competencies can be taught in higher education to ensure career relevance. Data-enabled experiences are essential, as corporations and governments invest heavily in big data and AI. The volume of data generated each second is staggering. For example, with 44 million learners, we can analyze hundreds of thousands of them and aggregate data to tailor learning experiences, perhaps suggesting a pop quiz or an AI-based video. This powerful data is vital to ensure that the content we develop aligns with learners' needs.

There has also been a substantial shift in how people view non-degree programs compared to traditional degrees. While degrees are still important, alternative pathways to gaining the necessary skills are becoming more prominent.

For instance, consider a MicroMasters program. If someone needs supply chain training to be job-ready, they can take a MicroMasters program from MIT, a world leader in supply chain management. There are no prerequisites; completing the program costs $1,200, and it prepares them for entry-level roles. What's fascinating is that these programs offer credit towards a full master's degree, meaning that someone like Mark, who is studying for his Ph.D, can apply the $1,200 he spent on the MicroMasters as 50% of the credits required for a master's at MIT.

These developments are game-changers in how we perceive education and envision its future. AI plays a crucial role here. I mentioned earlier that I initially had reservations about attending this conference, but we can't serve 55 million people and growing without AI. It helps keep learners engaged and manage vast workflows, like grading millions of essays. This approach allows us to meet learners where they are in their lifelong learning journey.

The good news for academics is that there has been a significant shift in understanding the efficacy of AI. They see positive learning outcomes and recognize that AI is not a threat to their existence, this is encouraging.

This is our moment to seize change and create lasting impact. It's okay to feel uncomfortable; while the pace of change can be daunting, embracing discomfort will propel us to the next level.

Thank you.

Workineh Kelbessa

Professor of Philosophy at Addis Ababa University, PhD., Ethiopia

Artificial Intelligence and the Environment

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

1. INTRODUCTION

Artificial intelligence (AI) is one of the emerging technologies created by human beings. The term "intelligence" has different meanings, including the capacity for abstraction, the ability to learn and adapt, creativity, self-awareness, etc. According to Mathias Risse,

> "Intelligence" is the ability to make predictions about the future and solve complex tasks. AI is such an ability demonstrated by machines, in smart phones, tablets, laptops, drones, self-operating vehicles, or robots. Such devices might take on tasks ranging from household support and companionship (including sexual companionship), to policing and warfare (2019:2).

The European Commission defines AI as follows: "AI is a collection of technologies that combine data, algorithms, and computing power. Advances in computing and the increasing availability of data are therefore key drivers of the current upsurge of AI" (EC, 2020:2). Although AI is not yet capable of completely mimicking the cognitive capacities of human beings, humans have used it to mimic intelligent human behavior. Thus, machines can behave like human beings.

Artificial intelligence has both positive and negative impacts on humans and the environment. Despite this fact, AI ethics has not given enough attention to environmental issues. Most works on AI ethics have been largely anthropocentric (Baum and Owe, forthcoming). The impact of AI on nonhumans has not been fully studied. Benedetta Brevini (2020) also shows that AI's environmental costs have been systematically excluded from the public discourse on AI. In this paper, I will briefly show the impact of AI on the environment.

2. THE ENVIRONMENTAL IMPACT OF ARTIFICIAL INTELLIGENCE

2.1. Positive Environmental Impacts

AI has many positive impacts on the environment. Amongst other impacts, AI can be used to improve factory efficiency and reduce energy costs. AI can also assist humanity in the struggle against climate change by spotting different trends. A report published by Microsoft in association with PWC states that greenhouse gas (GHG) emissions can be reduced globally by applying AI levers and adopting a "wider technology infrastructure" (Herweijer et al., 2018:8). AI programs have the potential to learn and predict weather patterns, which will help study large-scale global climate change (Snow, 2019) and limit its impact. For example, AI helped create a flexible and autonomous electric grid, integrating more renewable energy in Norway (Cho, 2018). Researchers used AI to identify tropical cyclones, weather fronts, and atmospheric rivers with 89 to 99 percent

accuracy (Cho, 2018). AI can thus accelerate "global efforts to protect the environment and conserve resources" by detecting energy emission reductions, removing CO2, promoting greener transportation networks, monitoring deforestation, and predicting unseasonal weather (Alsamhi et al. 2021:91). Companies can reduce their carbon footprint by benefitting from green data centers that can build environmental sustainability (Wyatt, 2021). Data center management can be optimized through the use of robotics and machine learning. "Energy clouds enable companies to expend energy into data centers by using renewable energy and indirect evaporative cooling to develop greener, more sustainable centers and reduce carbon footprints to a net-zero level" (Wyatt, 2021). Also, environmental researchers can share information and data compiled in an AI system, which can positively impact the environment. Moreover, autonomous vehicles can also contribute to environmental suitability by optimizing travel for energy efficiency. Another study also shows that AI can be used to reduce humans' impact on the environment by managing electric grids better, optimizing traffic and transit, saving energy in the construction and maintenance of buildings, improving the planning of tree felling and reforestation, detecting illegal logging on time, and making agriculture more efficient using fewer pesticides (Maclure and Russell, 2021:121-122).

Earthquakes, floods, storms, sea level fluctuations, and other natural disasters can be detected and monitored through AI-driven predictive analytics, as well as drones, advanced sensor platforms, and similar technologies (Bhattacharjee et al., 2022:76). By combining AI with satellite data, humans can detect land-use changes, vegetation, forest cover, and the consequences of natural disasters (Bhattacharjee et al., 2022:78). The early warning provided by these instruments can be used by emergency responders, government officials, and other interested groups to reduce the consequences of natural disasters.

It should also be mentioned that AI can positively impact economic growth, poverty reduction and health promotion, and the realization of human rights. It can be used to predict the use of environmental resources for farming. Some African countries including South Africa, Rwanda, Nigeria, Ghana, Botswana, Uganda, Tanzania, Zambia, Kenya, Benin, Ivory Coast, Côte d'Ivoire, and Mozambique have used AI in the fields of agriculture, energy, health care, public services, financial services, and wild animal management (Novitske, 2018; Alupo et al., 2022). Burkina Faso has used unmanned/unoccupied aerial vehicles (UAVs), commonly referred to as drones, to survey elephant populations (Vermeulen et al., 2013). South Africa used drones to fight rhino poaching (Mulero-Pázmány et al., 2014). Anti-poaching troops use predictive algorithms to perform their regular patrol routes. Drones can also be used effectively to conduct wildlife research in inaccessible terrain without significant habitat disturbance (Mazumdar, 2022). Farmers in India have increased crop yields of groundnuts by 30% using artificial intelligence-enabled systems (Cho, 2018). Farmers can also use AI-driven weather forecasts to determine the type of seeds to sow (Bhattacharjee et al., 2022:69). Additionally, robots such as the Econobox powered by computer vision can detect and eradicate weeds (Frew and Dozier, 2012). This reduces the need to use pesticides that pollute the environment.

Moreover, AI can collect data from inaccessible remote ocean locations to help protect species and habitats (Bhattacharjee et al., 2022:91-92). Robots powered by AI can monitor the ocean's pollution, temperature, and pH levels (Bhattacharjee et al., 2022:92). The information collected by such robots can be used to reduce ocean pollution.

Artificial intelligence is also instrumental in collecting, sorting, and managing different types of waste. Six major application areas for AI in municipal solid waste management were recently identified. "Detection of levels

in the waste bin, prediction of the waste characteristics, forecasting the process parameters, process output, vehicle routing, and approaches used for solid waste management are the key sectors where AI can be applicable" (Bharti et al., 2022:111). Some countries have adopted advanced digital technologies, such as smart waste bins (see Keymakr, 2021; Borchard et al., 2022; Bharti et al., 2022). Some people in the United States, Europe, China, South Korea, and Japan have used AI to detect and sort hazardous waste before recycling. In the US in particular, computer vision or other sensory methods have assisted garbage sorting robots to manage waste. Some scientists at the Danish Technological Institute (DTI) have tried to use artificial intelligence to train robots to sort hazardous waste (DTI, ND). Artificial Neural Networks (ANNs), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Multi-Layer Perception (MLP) models have been used to predict pollutant concentrations and particulate matter issues (Yetilmezsoy et al., 2011). Therefore, it can be said that AI can contribute to human health by minimizing risks associated with waste and the environment.

In short, AI can be used to advance climate change research; develop autonomous vehicles; forecast weather and changes in ozone layer; measure the levels of atmospheric carbon dioxide and nitrous oxides; grow crops more efficiently and make better use of fertile land (Wyatt, 2021); better and more efficiently move food through the supply chain (Wyatt, 2021); reduce water pollution, air pollution, and ocean pollution; monitor endangered species, the health of farm animals (Bhattacharjee et al., 2022:69), and plant development; identify, categorize, and count insects that destroy crops (Bhattacharjee et al., 2022:71); detect, track, and remove invasive species and other changes in biodiversity (Bhattacharjee et al., 2022:78); minimize contaminants in water; reduce waste and increase productivity; detect and sort hazardous waste; forecast waste generation trends; improve the routing of waste collection vehicles (Hamdy et al., 2021); manage the intermittency of renewable energy storage and improve energy storage (Cho, 2018); pinpoint regions that are affected by environmental changes and need assistance, etc.

Paramita Bhattacharjee and her co-authors have an optimistic view of AI's role. They believe that IoT devices, AI, and Machine Learning (ML)-based technologies can effectively achieve various goals, thereby changing the game "for environmental maintenance and enhancement" (2022:92). They further assert that these technologies will enable humanity to achieve all of its future economic and social goals, as economic progress and environmental progress are inextricably linked. However, these authors remain silent on the negative environmental impacts of these technologies. I will briefly discuss this topic in the next section.

2.2. Negative Environmental Impacts

AI has also harmed the environment. AI can contribute to environmental degradation through excessive consumption of the energy and raw materials needed to manufacture AI technologies, store and transmit data, as well as to train artificial neural networks.

> AI with its focus on efficiency through automation might potentially lead to 'over exploitation' of natural resources (e.g. precision agriculture, precision dairy farming, precision mining) if not carefully guided and managed. AI, especially deep learning, could also lead to increased demand for energy, which could be counter-productive for sustainability goals, unless that energy is renewable (from wind, solar, hydro etc.) and that electricity generation is developed hand-in-hand with application deployment (Herweijer et al., 2018:15).

The reliance on power-intensive Graphics Processing Units (GPUs) for machine learning training can lead to environmental degradation (Lawton, 2019). Training an advanced AI model and deep learning systems require high-energy consumption. Training a single AI model can result in greenhouse gas emissions equivalent to the total lifetime emissions of one or several cars (Strubell et al., 2019). AI should go through an extensive training process to achieve deep learning. For example, image recognition AI can be trained on millions of images to recognize different types of images. Such training can contribute to greenhouse gas emissions and environmental degradation. Also, storing large amounts of data requires a lot of energy. "AI--and in particular, deep learning--methods usually require large quantities of data" (Ligozat et al., 2022). All in all, the high carbon emissions caused by AI will exacerbate the problem of climate change. It is worth noting that digital technologies have contributed 3.7% to global greenhouse gas emissions (Shift Project, 2019, cited in Brevini, 2020:3).

Autonomous vehicles could also lead to the proliferation of private cars at the expense of public transportation, walking, cycling, etc. The "decrease in the value of trip time and user tendencies" are identified as the negative effects of autonomous vehicles (Silva et al., 2022). Excessive use of autonomous vehicles can also contribute to greenhouse gas emissions if not handled properly.

While AI can be used to manage e-waste, discarded communication and computational machines that are important for AI to function are already adding to it. Discarded electrical or electronic devices, such as electronic wastes, which include televisions, computers, mobile phones, refrigerators, washing machines, microwaves, batteries, printing devices, X-ray machines, MRI machines, ultrasound machines, children's toys, other types of home appliances, etc. generate e-waste. If not managed properly, e-waste can negatively impact human health and the environment because it contains numerous toxic chemicals and organic compounds.

Another problem with the application of AI technology is that it has been increasing inequality, and, in effect, power disparities in the world. The poor have no access to AI technology and the benefits of innovation. As Joan Sidney Howland said, the technological revolution of the 20th century had its own unpleasant aspect. "Technology has created a chasm, polarizing the 'technological haves' from the 'technological have-notes" (1998:287). Those countries which own AI have used it to exploit poor countries and increased the emissions of carbon dioxide. They have used drones in combat and destroyed the environment in conflict-prone countries, and protected their interest at the expense of other people. Undemocratic leaders in poor countries have done something similar to suppress opposition. They have used the meager resources of their poor countries to buy drones and consolidate their hold on power. Although networked AI can help humans effectively solve some problems and promote human autonomy and agency, it can still threaten human autonomy, agency, and life on earth. Therefore, the open use of AI technologies without a strong regulatory program will negatively impact people and the environment.

3. THE WAY FORWARD

It should be noted that machines can do the unexpected and challenge Mother Earth. Humans should not leave machines alone to operate in the world, as that can have dangerous consequences for humans and the more-than-human world. Therefore, the development and operation of AI in the coming years should be controlled by humans, and there should be appropriate mechanisms in place to regulate digital systems.

As Seth D. Baum and Andrea Owe (forthcoming) suggest, computer scientists

can benefit from environmental ethics to make AI environmentally friendly and socially sound. The interest of both humans and nonhumans should be taken into account when we use and develop AI. Human beings therefore should try to develop and promote AI in an environmentally-responsible manner.

In addition, the overall carbon footprint of AI needs to be reduced. Alternative energy sources, such as solar, wind power, hydropower, ocean energy, and bioenergy should be used and thermal energy techniques should be incorporated into a building design. We need increasingly accurate computer models of the earth's climate and weather systems if we are to be able to preserve and protect them.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the International Colloquium on Artificial Intelligence Ethics: from Emerging Standards to Perspectives of Implementation, held in Rabat, Morocco, from July 22-23, 2022. I am grateful to those present for their comments on that draft. Finally, I specifically thank the Alexander von Humboldt Foundation for supporting my research.

References

Alsamhi, Saeed et al. 2021. "Machine Learning for Smart Environments in B5GNetworks: Connectivity and QoS." Review: Comput Intell Neurosci DOI: 10.1155/2021/6805151. Available at Machine Learning for Smart Environments in B5G Networks: Connectivity and QoS - PubMed (nih.gov).

Alupo, Carity Delmus, Daniel Omeiza and David Vernon. 2022. "Realizing the Potential of AI in Africa: It All turns on Trust," in Towards Trustworthy Artificial Intelligence Systems, Ferreira, M.I.A. and M. O. Tokhi (eds.). Intelligent Systems, Control and Automation: Science and Engineering. Springer, pp. 179-192.

Baum, Seth D. and Andrea Owe. Forthcoming. "Artificial Intelligence Needs Environmental Ethics." Ethics, Policy & Environment. DOI: 10.1080/21550085.2022.2076538.

Bharti, Sharda, Shourat Fatma and Vinay Kumar. 2022. "AI in Waste Management: The Savage of Environment," in Paul, P. K., Amitava Choudhury, Arindam Biswas and Binod Kumar Singh (eds.). Environmental Informatics: Challenges and Solutions. Springer. Singapore, pp. 97-123.

Bhattacharjee, Paramita, Ajitesh Moy Ghosh, and Pabak Indu. 2022. "A Study on the Social and Economic Impact of Artificial Intelligence-Based Environmental Forecasts," in Paul, P. K., Amitava Choudhury, Arindam Biswas and Binod Kumar Singh (eds.). Environmental Informatics: Challenges and Solutions. Springer. Singapore, pp. 67-95.

Borchard, Rebecca, Roman Zeiss and Jan Recker. 2022. "Digitalization of Waste Management: Insights from German Private and Public Management Firms." Waste Management & Research 40(6):775-792. DOI: 10.1177/0734242X211029173.

Brevini, Benedetta. 2020. "Black Boxes, Not Green: Mythologizing Artificial Intelligence and Omitting the Environment." Big Data & Society July-December: 1-5. DOI: 10.1177/2053951720935141.

Cho, Renee. 2018. "Artificial Intelligence—A Game Changer for Climate Change and the Environment." Available at Artificial Intelligence—A Game Changer for Climate Change and the Environment (columbia.edu) [Accessed 7 July 2022].

Danish Technological Institute. No Date. "Robots with Artificial Intelligence to Sort Hazardous Waste." Available at Robots with Artificial Intelligence to Sort Hazardous Waste - Services - Danish Technological Institute (dti.dk) [Accessed 10 November 2022].

Hamdy, Walid, Ashraf Darwish, and Aboul Ella Hassanien. 2021. "Artificial Intelligence for Sustainable Waste Management and Control During and Post COVID-19 Crisis: Critical Challenges," in Hassanien, Aboul

Ella, Ashraf Darwish, Benji Gyampoh, Alaa Tharwat Abdel-Monaim and Ahmed M. Anter (eds.). The Global Environmental Effects During and Beyond COVID-19: Intelligent Computing Solutions. Cham, Switzerland: Springer, 81-91.

Herweijer, Celine, Benjamin Combes, and Jonathan Gillham. 2018. "How AI can enable a Sustainable Future." Microsoft in Association with PWC. Available at How AI can enable a sustainable future (pwc.co.uk) [Accessed 10 November 2022].

Howland, Joan Sidney. 1998. "The 'Digital Divide': Are We becoming a World of Technological 'Haves' and 'Have-Nots?'" The Electronic Library 16(5):287-289.

Keymakr. (2021). "How Artificial Intelligence is Transforming Waste Management." Available at How Artificial Intelligence Is Transforming Waste Management (keymakr.com) [Accessed 11 May 2022].

Lawton, George. 2019. "AI's Carbon Footprint will be an Issue for Enterprises." TechTarget Available at AI's carbon footprint will be an issue for enterprises (techtarget.com) [Accessed 19 July 2022].

Ligozat, Anne-Lauer, Julien Lefevre, Aurélie Bugeau and Jacques Combaz. 2022. "Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions." Sustainability 14, 5172. https://doi.org/10.3390/su14095172.

Maclure, Jocelyn and Stuart Russell. 2021. "AI for Humanity: The Global Challenges," in Braunschweig, Bertrand and Malik Ghallab (eds.). Reflections on Artificial Intelligence for Humanity. Cham, Switzerland: Springer, pp. 116-126.

Mazumdar, Subhendu. 2022. "Drone Applications in Wildlife Research—A Synoptic Review," in Paul, P. K., Amitava Choudhury, Arindam Biswas and Binod Kumar Singh (eds.). Environmental Informatics: Challenges and Solutions. Springer. Singapore, pp. 237-257.

Mulero-Pázmány, M., R. Stolper, L. van Essen, J. Negro, and T. Sassen. 2014. "Remotely Piloted Aircraft Systems as a Rhinoceros Anti-poaching Tool in Africa," PLOS One, vol. 9, no. 1, Jan. Doi: 10.1371/jopournal.pone0083873.

Novitske, Lexi. 2018. "The AI Invasion is coming to Africa (and It's a Good Thing)." Stanford Social Innovation Review, February. Available at https://ssir.org/articles/entry/the_ai_invasion_is_coming_to_africa_and_its_a_good_thing[Accessed [Accessed 5 July 2021].

Risse, Mathias. 2019. "Human Rights and Artificial Intelligence: An Urgently Needed Agenda." Human Rights Quarterly 41(1):1-16.

Silva, Óscar, Rubén Cordera, Esther González-González, and Soledad Nogués. 2022. "Environmental Impacts of Autonomous vehicles: A Review of the Scientific Literature." Science of the Total Environment. DOI: 10.1016/j.scitotenv.2022.154615.

Snow, Jackie. 2019. "How Artificial Intelligence can tackle Climate Change." National Geographic. July 18.

Strubell, E., A Ganesh, & A. McCallum. 2019. "Energy and Policy Consideration for Deep Learning in NLP," in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645-3650. Florence, Italy.

Vermeulen, C., P. Lejeune, J. Lisein, P. Sawadogo, and P. Bouché. 2013. "Unmanned aerial survey of elephants," PLOS One, vol. 8, no. 2, Feb. Doi 10.1371/journal.pone.0054700.

Wyatt, Kandis Y. 2021. "Using AI to Make More Improvements to Our Environment." Available at Using AI to Make More Improvements to Our Environment (apuedge.com) [Accessed 7 July 2022].

Yetilmezsoy, K., B. Ozkaya and M. Cakmakci. 2011. "Artificial Intelligence-based Prediction Models for Environmental Engineering. Neural Netw. World 11(3):193-218.

Raja Bensaoud

Enseignante chercheure en droit du numérique, Co-fondatrice du Think tank Digital Act, Maroc

L'intelligence artificielle sous l'angle du droit

L'IA, prédite de longue date, n'est plus un programme de recherche dans les laboratoires. Les technologies liées à l'intelligence artificielle (IA) connaissent un essor fulgurant et son ascension semble irrésistible.

C'est une banalité de dire que les applications de l'IA ne cessent de se multiplier et plusieurs des techniques qui en sont issues ont intégré notre quotidien et conquis de nombreuses sphères de la société. L'IA est entrée dans la vie de chacun d'entre nous. Elle est dans nos smartphones, dans nos ordinateurs et dans nos objets connectés.

L'IA représente des avantages indéniables, elle peut favoriser le progrès humain et être une source de prospérité économique.

Elle peut contribuer à l'amélioration du bien-être et des conditions de vie, apporter des solutions novatrices à une foule de problèmes humains complexes et dans des domaines aussi divers que ceux de la santé, de l'énergie, de l'agriculture, du tourisme, de la mobilité ou de l'éducation.

Mais si l'IA a des vertus, elle a aussi des vices. L'IA et son omniprésence suscite des appréhensions à la mesure de sa puissance et de son gigantesque champ d'intervention.

Les usages de l'IA dans les domaines à risque (recrutement, l'octroi de prestations sociales, la gestion de services comme la police ou la justice, ...), les rendent autant sources d'avancées majeures que porteurs de risques pour les droits fondamentaux.

1. EST-CE QUE LE DROIT PEUT APPRÉHENDER L'IA?

1/ Les biais discriminatoires

Les décisions basées sur un système d'IA peuvent aboutir à des discriminations. Elles peuvent reproduire des schémas discriminatoires préexistants voire les accentuer.

Ces biais peuvent être volontaires comme c'est le cas pour certains algorithmes de recrutement. Ils peuvent aussi être involontaires voire accidentels, l'effet discriminatoire n'étant pas recherché.

La question concerne toutes les formes de discrimination : qu'elle soit basée sur le genre, l'origine raciale ou ethnique, l'âge, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale.

L'essentiel des exemples sur le sujet nous vient de la doctrine anglo-saxonne et les exemples pris concernent en général des algorithmes déployés aux Etats Unis, par des sociétés privées ou par des administrations publiques.

- Le système pénal américain recourt à un système algorithmique qui mesure le risque de récidive et qui est utilisé par les juges de nombreux Etats américains. Des études ont démontré que cet algorithme comportait des biais discriminatoires qui conduisaient à surévaluer le risque de récidive chez les afro-américains.
- Des chercheurs de l'Université de Californie à Berkeley ont également découvert que les algorithmes perpétuaient les préjugés raciaux dans les décisions de prêt hypothécaire.
- En Ecosse, il a été démontré qu'un système d'ajustement des notes des étudiants basé sur l'IA était biaisé contre les étudiants appartenant à des milieux défavorisés.
- Enfin, l'algorithme mis en place par AMAZON en 2015 pour recruter des

talents, et qui utilisait des centaines de milliers de CV recueillis 10 ans auparavant, et bien cet algorithme a été abandonné, en 2017, à cause de sa propension à discriminer les femmes.

Comment le droit peut-il appréhender ces dysfonctionnements de l'IA qui impactent l'exercice de droits individuels ou collectifs ?

A partir du moment où le droit protège les personnes contre les décisions discriminatoires, il est normal que cette protection soit étendue aux décisions prises à l'aide d'un système d'IA.

Sauf que, il y a une difficulté du fait que ces biais peuvent être indécelables, cachés, et les dommages qu'ils peuvent provoquer peuvent être difficilement identifiables. Cette difficulté rend hasardeuse toute tentative de les corriger, de les supprimer ou de les sanctionner.

2/ Les risques pour les droits des travailleurs ?

L'IA a métamorphosé le monde du travail. Son adoption à travers le management algorithmique ou les technologies prescriptives peut affecter les droits des travailleurs.

Des algorithmes vont ainsi dicter les modalités du travail, comment doit-il être réalisé, dans quel délai, selon quelles procédures. Le chauffeur Uber doit suivre l'itinéraire décidé par l'algorithme. Son comportement est tracé à partir de données mobiles. Le travailleur devient dépendant de l'interface numérique.

Dans un ouvrage collectif « Travail e(s)t liberté ? » les auteurs affirment : « Le

« micro-tâcheron » est lié à son employeur par toute une série d'applications diverses et variées, chats, réseaux sociaux qui demandent une disponibilité presque totale et une attention renforcée, jusqu'au bracelet électronique devenu tristement célèbre dans les entrepôts d'Amazon. » (1)

Le travailleur est commandé, surveillé par l'algorithme via des accessoires connectés, des capteurs. Il devient un rouage d'un système commandé par des outils d'IA. Et cette aliénation technologique peut entraîner des risques pour la santé du salarié et des risques psycho-sociaux.

Le management algorithmique n'est pas l'apanage des entreprises plateformisées ou des pure players. La société Hitachi, par exemple, propose le «Business Microscope», un collier connecté qui peut tracer les déplacements des salariés, les interactions et les communications entre personnes.

Est-ce que le droit positif apporte des réponses à ces dérives ?

Ces pratiques peuvent parfaitement être cadrées par le droit commun et le droit du travail mais aussi par les lois sur la protection des données personnelles. Ces textes apportent une protection au travailleur et des réponses appropriées aux excès du management algorithmique.

Concernant les conditions de travail qui peuvent entraîner des risques professionnels, des problèmes psychosociaux, le droit positif prévoit, à travers les règles édictées dans le code du travail, la responsabilité de l'employeur qui doit protéger la sécurité physique et mentale du salarié. L'employeur ne peut pas se retrancher derrière l'algorithme.

3/ Les risques pour les droits des utilisateurs-consommateurs.

Dans son livre intitulé « Entreprises hyperpuissantes » Lévêque, nous rappelle que

: « Nos choix de consommateurs sont orientés par la publicité ciblée à travers un mécanisme bien particulier : l'échange d'attention contre un service désirable. Google et Facebook sont des marchands d'attention, ils agissent à l'échelle de la planète ou presque, en employant des méthodes sophistiquées – celles du big data et de ses puissants algorithmes. » (2)

Les acteurs du marché numérique se livrent à une véritable compétition pour capter l'attention des utilisateurs.

Et dans cet univers, le consommateur peut s'exposer à différents risques résultant notamment de techniques de vente en ligne basées sur un micro ciblage très fin rendu possible grâce à l'exploitation de données massives sur le comportement des internautes ; ces techniques influencent la liberté de choix et le libre arbitre des utilisateurs.

La notion de dark patterns (ou interfaces sombres) illustre bien ce phénomène : il s'agit d'interfaces algorithmiques qui peuvent altérer ou restreindre le libre choix du consommateur et limiter son droit d'exprimer ses préférences, en l'amenant à prendre des décisions qu'il n'a pas nécessairement souhaitées.

Il faut rappeler que ces profilages en ligne permettent aussi d'influencer les opinions, les croyances ou les émotions et brider ainsi le libre choix du consommateur ou du citoyen et l'autonomie de sa volonté.

Est-ce que le droit permet de cerner de telles situations ?

Au niveau du droit des contrats, il s'agit d'un vice du consentement sanctionné par la nullité du contrat qui aurait été conclu suite à un dark pattern. On appliquerait ainsi le vieux principe de l'autonomie de la volonté.

Sauf que là, il y a des difficultés pratiques, notamment sur le terrain de la preuve mais aussi en raison du montant généralement faible de la transaction et qui fait que le consommateur victime de ce type de préjudice n'est pas encouragé à faire valoir ses droits devant le juge.

Dans cet exemple, on s'aperçoit que le recours au droit existant présente des limites.

4/ La protection des données personnelles

On le sait, l'IA ne peut fonctionner sans les données. Algorithmes et données forment un tout indissociable et c'est pourquoi certains juristes estiment que le recours au droit des données permet de réglementer l'IA de façon indirecte.

Les principes figurent dans la Convention 108 du Conseil de l'Europe sur la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (révisée en 2018) se retrouvent dans plus d'une centaine de législations nationales sur la protection des données.

Le recours à ce droit peut ainsi être une réponse pour réduire les risques de discriminations injustes et illicites, en exigeant la transparence du processus de traitement des données personnelles.

Mais là aussi, le droit des données personnelles peut se heurter à des situations qu'il n'a pas prévu et certains algorithmes peuvent échapper à son emprise.

A titre d'exemple : ce droit s'applique au traitement de données personnelles mais ne cadre pas les modèles prédictifs, car ces modèles sont impersonnels et ne portent pas sur des personnes identifiables.

Par exemple, un modèle prédictif qui dit que 70 % des habitants d'un certain quartier ont telles ou telles habitudes, ne comporte pas de données personnelles puisqu'il ne se réfère pas à des personnes. Le résultat est que le droit des DP est inopérant dans un cas comme celui que je viens de citer.

Pour toutes ces raisons, et d'autres, le droit ne peut rester en retrait.

De nombreuses instances internationales ont alerté sur l'impact de l'IA par rapport aux droits fondamentaux, notamment sur le respect de la dignité humaine, le respect de la vie privée et la protection des données, l'égalité et la non- discrimination, l'accès à la justice et l'accès aux droits sociaux.

Actuellement, il n'existe pas de cadre

juridique, tant à l'échelon national qu'international, pour prévenir ces risques. D'ailleurs, une telle entreprise, se heurterait à plusieurs écueils. J'en citerai trois en particulier, en dehors du champ sémantique, de cet univers de mots nouveaux que le droit doit apprivoiser et domestiquer, :

1/ le premier écueil est lié au caractère hybride de l'IA expliqué dans un livre « Vers un droit de l'IA », signé par Samir Merabet. En effet, l'auteur explique que je cite

« le système informatique intelligent apparaît alors comme une entité hybride qui trouve son origine dans une chose, mais qui adopte, à certains égards, le comportement d'une personne ». (3)

La réglementation de l'IA se heurte ainsi à la summa divisio du droit qui ne reconnaît que deux catégories : les choses, objets du droit, et les personnes, sujets de droit. 2/ L'opacité de l'IA constitue un écueil sérieux pour déceler et prouver d'éventuelles infractions à la législation, notamment aux dispositions juridiques qui protègent les droits fondamentaux, mais aussi d'imputer la responsabilité et de remplir les conditions requises pour prétendre à une indemnisation.

Dans une note publiée au mois d'août 2020 et consacrée à l'application de l'IA au sein du système pénal américain, The National Institute of Justice (Etats Unis), on peut lire : « La plupart des méthodes d'IA les plus courantes ne sont pas des "boîtes noires". Cependant, ce n'est pas le cas pour certains algorithmes de deep learning. Les expliquer n'est pas toujours possible en raison de leur complexité et cela crée un problème de boîte noire pour de nombreux systèmes d'IA basés sur le deep learning. »

2. QUELLE DÉMARCHE POUR UN ENCADREMENT JURIDIQUE DE L'IA ?

Si la nécessité d'un cadre régulatoire n'est guère contestée, les modalités ne sont pas encore précises ou du moins diffèrent selon les Etats et les groupements d'Etats.

Plusieurs initiatives sont lancées pour bâtir un cadre réglementaire notamment au sein de l'Union Européenne et dans différents Etats américains.

Pour ma part, je me contenterai de suggérer des éléments d'ordre méthodologique pour un travail d'encadrement juridique de l'IA.

Démarche proposée :

Procéder à un inventaire, un diagnostic des règles de droit pouvant encadrer les risques liés à l'IA et d'apprécier si elles peuvent répondre, en l'état, aux questions posées.

Il s'agit ainsi d'évaluer la capacité des lois existantes à contrer les risques issus des usages de l'IA, et d'apprécier si de nouvelles mesures réglementaires sont nécessaires pour garantir la protection des personnes.

Dans certains cas, et les exemples que j'avais cités l'attestent bien, le droit positif peut suffire à cadrer l'IA sans procéder à une réingénierie des schémas juridiques existants. Cela suppose d'interroger les notions, les concepts existants mais aussi de mobiliser toutes les ressources et les techniques juridiques à travers des analogies, des recoupements, des extrapolations, pour identifier les réponses appropriées.

Dans d'autres cas, il s'agira de repenser un cadre juridique existant ou de renforcer l'arsenal de droit commun par des règles nouvelles et des garde-fous juridiques pour l'adapter à certaines situations algorithmiques.

Par exemple, pour assurer la neutralité de l'IA, on peut aller chercher des règles soit dans le droit commun soit dans des droits spécifiques comme le droit pénal, le droit de la concurrence ou le droit de la consommation.

La démarche ne peut être que progressive et graduelle car, ne l'oublions pas, il s'agit de ne pas freiner l'innovation avec des normes inadaptées ou contre productives qui conduiraient uniquement à priver la société des bénéfices et des bienfaits des outils

algorithmiques.

Au niveau du fond, une législation en la matière devrait comporter des règles imposant la transparence et l'explicabilité des algorithmes afin de lever le voile sur leur logique de conception et d'exécution, et ce, nonobstant leurs droits de propriété intellectuelle.

Ces exigences sont particulièrement requises lorsque des SIA sont utilisées par des administrations publiques avec des effets sur la vie de millions de citoyens, comme c'est le cas dans le domaine de la fiscalité, la sélection pour l'enseignement supérieur, ou la reconnaissance faciale.

CONCLUSION

Un cadre juridique est nécessaire pour à la fois exploiter utilement le potentiel technologique de l'IA et en même temps gagner la confiance du public dans son utilisation. La confiance, pilier essentiel dans le monde numérique, ne peut s'obtenir sans un cadre juridique clair et lisible pour tous. La reconnaissance d'un droit dédié à l'IA peut instaurer un sentiment de confiance et dissiper les incertitudes qui entourent cette technologie.

Sinon, ce sera, les géants de la tech qui, à nouveau, rempliront les vides juridiques en matière d'IA et comme l'a si bien dit Valère NDIOR dans une tribune : On sera dans « la perspective de voir se créer un panel de Google, un mécanisme d'inspection de Twitter ou un tribunal paritaire d'Uber, dotés de référentiels propres et d'une compétence mondiale, à l'instar du Conseil de surveillance (de Facebook)

… Un tel développement relèverait essentiellement de stratégies de relations publiques et ne saurait se substituer à une protection institutionnelle des droits fondamentaux des utilisateurs. »

Je vous remercie.

Références

(1) « Travail et plateformes numériques : entre exploitation et opportunités », Massimiliano Nicoli, Licas Paltrinieri, M. Prévot-Carpentier dans « Travail et liberté », Ed Erès, 2022.
(2) François Lévêque, « Entreprises Hyperpuissantes Géants et Titans, la fin du modèle global ?», Ed Odile Jacob,2021,
(3) Samir Merabet, « Vers un droit de l'intelligence artificielle », Dalloz 2020
(4) Valère Ndior, Le Conseil de surveillance de Facebook, « service après-vente » de la liberté d'expression ? Recueil Dalloz, 2020

Abdelmajid Elouadi

Professeur à l'Université Ibn Tofaïl, Maroc

Industrie 4.0 et applications de l'IA

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

Tout d'abord, je tiens à féliciter le Conseil National des Droits de l'Homme d'avoir amorcé cette série de conférences sur l'intelligence artificielle et ses implications.

Mon intervention s'intitule « L'industrie 4.0 et les applications de l'intelligence artificielle », justement pour souligner que l'IA n'est pas isolée. Elle s'inscrit dans un contexte d'évolution industrielle et de recherche et développement. Pour ce qui est du plan adopté : je vais introduire ma présentation avec un projet Erasmus, « Enhance Industry 4.0 », lancé avec la Commission européenne en 2021 à l'université Ibn Tofail que je représente. Ce projet de renforcement des capacités vise à développer des fonctions essentielles en matière d'intelligence artificielle, notamment dans les domaines de la maintenance, de la qualité et de la production. Par la suite, j'aborderai l'industrie 4.0 (I4.0) en rapport avec ses composantes et son évolution historique, puis l'intelligence artificielle à travers ses quatre vagues, avant de discuter des implications éthiques et sociales. Je terminerai par les réflexions des experts.

L'intelligence artificielle, l'industrie 4.0, l'éthique et les droits de l'homme sont des chantiers majeurs de formation et de recherche-développement. Ainsi, si l'on souhaite explorer les implications sur les aspects éthiques et les droits de l'homme, il est crucial de former des équipes multidisciplinaires autour de la recherche et du développement, sans laisser uniquement aux mathématiciens et aux informaticiens le soin de concevoir et de prédire l'avenir. Il s'agit de rechercher des solutions, ce qui repose sur le travail d'équipe, la multidisciplinarité et l'apprentissage tout au long de la carrière.

Le projet Erasmus relatif à l'industrie 4.0 a permis de développer un master de recherche à l'ENSA de Kénitra, intitulé « Industrie intelligente et technologies numériques ». Ce master couvre des thématiques telles que le Big Data et l'intelligence artificielle, intégrant toutes les méthodes modernes, avec un volume horaire de 40 % consacré à l'intelligence artificielle. Un module du master aborde également la philosophie des sciences, en se concentrant sur les aspects juridiques, éthiques et les implications sociales.

Au fur et à mesure, nous avons pu nous contextualiser et nous adapter aux débats, tant au niveau national qu'international, concernant ces implications. En quelques mots, le projet a traversé plusieurs phases. Dans un premier temps, nous avons analysé les écarts des programmes de formation existants, tant pour le cycle ingénieur que pour le master. Ensuite, nous avons évalué la maturité des entreprises en lien avec l'industrie 4.0, en ciblant principalement les secteurs de l'automobile et de l'aéronautique, ainsi que d'autres industries où l'intelligence artificielle joue un rôle clé dans la prise de décision et la résolution de problèmes.

Nous avons également développé un cadre d'apprentissage innovant, un contenu de formation et des méthodologies d'apprentissage centrées sur la pratique. Dans le premier programme, nous avons

introduit des jeux pédagogiques, car la formation par le jeu sur des plateformes adaptatives utilise les dernières recherches en matière d'apprentissage. Actuellement, nous avons créé une plateforme de formation et d'apprentissage continu de deux semaines, ayant formé 20 enseignants sur divers aspects de l'intelligence artificielle, du Big Data et de leur application dans les domaines de la qualité, de la maintenance et de la production.

Ainsi, vers les mois de novembre et décembre, nous allons constituer un centre de compétences regroupant des enseignants, sous la supervision du Digital Innovation Hub (DIH). Ce dernier sera un pôle intégrant entreprises, incubateurs, associations professionnelles et organisations défendant les implications de l'intelligence artificielle sur les droits de l'homme.

Ce projet comprend un ensemble de partenaires, notamment : l'Université de Lyon en France, l'Université de BIBA en Allemagne, l'Université DE NOVA de Lisboa au Portugal, l'International Institute of Technology (IIT) en Tunisie, l'Université de Carthage en Tunisie (UCAR), la Fondation École Centrale de Casablanca et l'Université Ibn Tofail de Kénitra.

Dans un premier temps, nous avons réalisé une analyse des écarts par rapport aux programmes existants et aux niveaux de maturité concernant l'industrie 4.0 et l'intelligence artificielle. Ensuite, nous avons développé des compétences, tant dans la formation professionnelle que universitaire. Nous avons établi un ensemble d'axes de formation touchant plusieurs compétences, marquant ainsi un virage important. En effet, nous nous concentrons sur la recherche de compétences en intelligence artificielle, Big Data, agilité, adaptabilité, etc.

Concernant le DIH, c'est la structure où se regrouperont diverses missions avec plusieurs partenaires, notamment startups, associations professionnelles, industries, investisseurs et institutions gouvernementales. Il sera le premier DIH au Maroc, après la mise en place de nombreux DIH en Europe.

Quant aux missions, leur présentation est essentielle, car c'est ici que nous pourrons introduire les implications éthiques et les droits de l'homme. Les résultats obtenus au niveau du Digital Hub seront testés, et c'est à cette étape que nous pourrons intégrer les aspects relatifs aux droits de l'homme. Un soutien à l'investissement sera conditionné par le respect de ces implications, et toute la formation et le développement des compétences devront prendre en compte les enjeux que nous discutons aujourd'hui.

Rapidement, en ce qui concerne les fondamentaux de l'industrie 4.0, le terme est apparu lors de la Foire de Hanovre en 2011. Ce concept a émergé en 2006 face à la prise de conscience que le modèle industriel allemand, fondé sur le leadership, était menacé par la concurrence asiatique, l'innovation technologique et l'arrivée de l'Internet dans l'industrie.

L'industrie 4.0 correspond à une nouvelle manière d'organiser les moyens de production, avec la mise en place d'usines intelligentes qui interagissent entre produits, machines et réseaux. Nous évoluons donc dans un écosystème numérique, qu'il soit interne ou externe.

En regardant l'ensemble des débats actuels, nous assistons à l'aboutissement de plusieurs révolutions, au nombre de quatre. Dans la première révolution, il s'agissait de mécanisation et de l'utilisation de la machine à vapeur, cherchant à établir des connexions au transport et à l'infrastructure. Cette révolution a connu trois étapes : initialement, elle a été perçue comme ridicule, puis dangereuse, avant d'être acceptée comme évidente, un peu comme la production de masse.

Si l'on se remémore le film « Les Temps modernes » de Charlie Chaplin, ce dernier illustre parfaitement ces trois préoccupations. Dans l'étape de production de masse, la question des conditions de

travail a conduit à repenser le management, l'organisation du travail et l'automatisation.

La même dynamique s'applique à l'électronique, à l'informatique et à l'automatisation, qui caractérisent la troisième révolution. Aujourd'hui, nous nous demandons si nous avons dépassé l'étape où l'intelligence artificielle était perçue comme ridicule. En effet, nous l'avons dépassée, mais elle présente aussi des dangers. Selon moi, la recherche et le développement doivent se concentrer sur le savoir-être et le vivre-ensemble, tant dans la recherche que dans la formation, afin de rendre cette étape évidente.

Il ne s'agit pas seulement de l'industrie 4.0 ; à l'échelle mondiale, nous discutons actuellement de l'industrie 5.0, qui met l'accent sur la rapidité des progrès et la prise en compte des préoccupations sociales. L'objectif de l'industrie 5.0 est d'appliquer efficacement une main-d'œuvre combinée de machines et d'humains. Ainsi, il ne s'agit pas de remplacer, mais de collaborer avec les robots. Certes, une rupture existe, comme dans les révolutions précédentes, mais l'industrie 4.0 se concentre davantage sur l'intégration des technologies numériques et physiques, la connectivité, la puissance de calcul, etc.

L'industrie 5.0 introduit un nouveau modèle de collaboration, axé sur l'idée de combiner la créativité et le savoir-faire humains avec la rapidité, la productivité et la cohérence des robots. Cela implique une adaptation, une résilience, un développement durable et une recherche d'ergonomie neuronale.

En ce qui concerne l'industrie 4.0, il existe quatre vagues de l'intelligence artificielle qui façonneront l'avenir. Tout ce que nous observons aujourd'hui n'est que la partie émergée de l'iceberg, tandis que l'essentiel reste caché. Chercher à légiférer ou à réglementer les outils d'IA est une tâche ardue.

Pour résumer, l'intelligence est une expression du pouvoir humain. L'homme a toujours cherché à être aidé ou remplacé. L'histoire nous montre que la nécessité est la mère de l'invention, que ce soit avec la découverte de l'écriture, la révolution agricole, la révolution cognitive ou la révolution industrielle. C'est ce constat qui a permis les avancées que nous connaissons aujourd'hui. L'humain a toujours cherché à prolonger son existence à travers les artefacts qu'il crée.

L'intelligence artificielle se divise en quatre vagues. La première est « l'Internet », avec les moteurs de recommandation et les systèmes algorithmiques qui apprennent à partir de grandes quantités de données étiquetées, comme les « likes » et les commentaires. Ce qui est connecté aujourd'hui l'est principalement dans un but commercial, cherchant à maximiser la connexion des utilisateurs. Si l'on souhaite mesurer notre niveau de socialisation, il est préférable de se concentrer sur notre capacité à nous déconnecter.

La deuxième vague est celle du « business », où l'intelligence s'appuie davantage sur des données déjà étiquetées, contrairement à la première vague, qui dépendait des utilisateurs pour annoter les publications sur les réseaux sociaux.

La troisième vague concerne « la perception ». Grâce à des capteurs et des composants électroniques, des ingénieurs qualifiés peuvent réaliser des prototypes au sein de milliers d'usines.

Pour avoir une idée de ce qui est développé par la Chine et les États-Unis, on estime qu'en cinq ans, la Chine va surpasser les capacités de l'IA perceptive. Par rapport aux États-Unis, elle atteindra 50 % de l'intelligence autonome, ce qui représente une part de présence.

L'IA évolue vers l'androïde ou le cyborg. L'androïde fait référence à une machine, tandis que le cyborg désigne l'humain équipé de puces électroniques dans le cerveau ou sur les organes. Cette évolution pose la question de la coexistence entre l'humain et

l'IA sur le marché du travail, ce qui dépend du niveau de conception et du degré de socialisation recherché.

Qu'en pensent les experts ?

Concernant l'IA, certains envisagent une forme de conscience artificielle. Laurent Alexandre, par exemple, évoque l'idée d'une conscience artificielle dotée de libre arbitre, ce qui soulève des questions philosophiques. Il met en avant les enjeux liés à la prise de décision autonome et à la dangerosité de l'intelligence artificielle. Elon Musk a également mis en lumière les effets négatifs que l'IA pourrait avoir sur l'avenir de l'humanité.

Il est essentiel de ne pas arrêter la réflexion. Nous devons suivre et évaluer les avancées de l'IA, comme nous le faisons dans le domaine de l'écologie. Avant de lancer un projet, il est crucial de réaliser une étude d'impact et de s'assurer qu'un ensemble de référentiels réglementaires soit en place.

Pour conclure, je propose que toutes les instances autonomes, les instances consultatives et les organisations professionnelles, notamment les syndicats et le Conseil national des droits de l'homme, se réunissent autour de cette problématique. Il est nécessaire d'aller au-delà des conférences et de constituer un groupe de travail thématique et d'impact.

Merci.

ETHICS OF AI, STATE OF INITIATIVES IN MOROCCO

Zakaria Rachid

Chef de service Appui à la Recherche et Innovation de l'Agence de Développement Digital (ADD), Maroc

Approche nationale en matière d'organisation du digital et de l'IA

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

D'abord, je tiens à remercier le Conseil National des Droits de l'Homme pour l'organisation de ce colloque qui porte sur un thème très important, celui de la relation entre l'intelligence artificielle et l'éthique.

Aujourd'hui je vais présenter ce que fais l'Agence du Développement du Digital (ADD), qui est une agence publique, placée sous la tutelle du Ministère délégué auprès du Chef du Gouvernement chargé de la Transition numérique et la Réforme de l'Administration. Les missions de l'ADD sont :

-La mise en œuvre, pour le compte de l'Etat, de la stratégie nationale dans le domaine du développement digital ;

-La proposition des orientations générales au Gouvernement en matière du développement du Digital ;

-La promotion de la diffusion des outils numériques au sein de la population marocaine.

Pour ce qui est des domaines d'intervention de l'ADD :

-Smart Gouvernement : l'ADD s'occupe de la transformation digitale de l'ensemble des services publics du Royaume du Maroc ;

-Ecosystème, Digital et Innovation : Il s'agit de développer l'économie digitale dans une vision de faire du Maroc un Hub Digital et Technologique de référence en Afrique ;

-Inclusion sociale et développement humain : L'amélioration de la qualité de vie des citoyens à travers le Digital, et notamment la facilitation de l'accès à plusieurs services publics, tels que l'accès aux soins, à l'éducation, etc. ;

-Environnement et confiance digitale : L'objectif est de développer tout ce qui permet à mettre en place un écosystème digital (e g. développement des infrastructures, du capital humain et des compétences digitales).

L'Agence du Développement du Digital a été créée en vertu de la loi 61-16 de 2017. La constitution de l'agence a été faite en 2018. C'est à partir de 2019 que les premiers chantiers ont été élaborés, tel que la Digital Factory en charge du développement des applications informatiques pour les administrations au Maroc.

La feuille de route de l'agence est structurée autour de 15 chantiers dont 5 sont validés comme prioritaires lors du dernier conseil d'administration de l'agence.

Le premier chantier est une Plateforme d'échange de données entre les administrations et d'interopérabilité qui a comme objectif de lier à travers un système d'information l'ensemble des administrations marocaines. De cette façon, les citoyens et les entreprises peuvent par une seule demande accéder aux documents nécessaires. Le deuxième chantier concerne la mise en place d'un Portail Unique du Citoyen (PUC), dont l'objectif est de regrouper l'ensemble des services publics dans une seule plateforme web permettant aux citoyens d'accéder facilement aux différents

services proposés par les administrations. Il y a également la Digital Factory qui est une entité interne de l'ADD, son objectif est de développer des applications informatiques pour les autres administrations en mode agile.

En plus, des chantiers de la Digitalisation du Parcours Investisseur et la Digitalisation du Parcours Import/Export. Ces trois chantiers entrent dans le cadre du domaine du Smart Gouvernement.

Pour ce qui est des chantiers qui concernent le domaine Ecosystème, Digital et Innovation, il y a tout d'abord le chantier Digital Park qui vise la mise en œuvre d'un parc technologique dans lequel les entreprises et les startups peuvent intégrer le domaine digital. Il y a également la partie Smart Factory- une usine digitale modèle 4.0, l'objectif de ce modèle industriel est d'offrir aux entreprises un espace de formation sur les concepts fondamentaux de la transition numérique et d'accompagner les startups marocaines. Un chantier Digital PME est également en place. Etant donné que la digitalisation et l'informatisation des entreprises joue un rôle important dans leur compétitivité, ce chantier offre un panel d'outils de promotion digitale pour les PME au Maroc. Aussi, un chantier Digital Auto-Entrepreneur qui a comme objectif de digitaliser l'ensemble des procédures du dépôt du dossier auto-entrepreneur et de proposer une plateforme de mise en relation des auto-entrepreneurs avec les clients potentiels.

Parmi les chantiers prioritaires, il y a la mise en œuvre d'un Ecosystème dédié à l'Intelligence Artificielle et le chantier Infrastructure Digitale qui vise la rationalisation des infrastructures par les administrations marocaines. Le chantier d'Appui à l'évolution du cadre réglementaire qui tend à la mise en place d'une réglementation propice pour le développement du Digital au Maroc. A cela s'ajoute le chantier Génération Digitale qui vise le développement des compétences digitales au Maroc à travers plusieurs actions, notamment une plateforme de formation à distance et un programme digital académique de réorientation pour les jeunes sans emploi. Dernièrement, le chantier Défisnationaux.ma qui consiste à mettre en place une forme de plateforme pour les challenges publics et le projet applications pour les jeunes, notamment la plateforme e-himaya qui vise la protection des enfants des dangers de l'internet.

Pourquoi faut-il adopter l'IA et doter le Maroc d'une stratégie nationale dans le domaine de l'intelligence artificielle ? Ce qu'il faut savoir c'est qu'actuellement environ 5 pays dans le monde possèdent des stratégies nationales dans le domaine de l'intelligence artificielle. Les études montrent que l'IA a un potentiel de contribution dans le PIB mondial de 15,7 milliards de dollars, ce qui représente un peu près 1,2 taux de croissance de PIB mondial. Par conséquent, la non adoption de l'intelligence artificielle peut éventuellement creuser des écarts entre les pays qui adopteront l'IA dans les prochaines années et ceux qui tarderont pour le faire. Il y a également un autre potentiel en termes de marchés lié aux logiciels, aux software, hardware et services d'IA. Selon les estimations d'ici 2030 ils peuvent atteindre 1,5 de dollars.

Il faut également noter que l'intelligence artificielle est une révolution qui vient en parallèle avec plusieurs autres révolutions, telle que la révolution dans la robotique et dans la blockchain. L'intelligence artificielle touche un bon nombre de domaine que ce soit l'industrie, le transport, l'éducation, la santé, etc. Dernièrement, il y a lieu de signaler qu'un grand nombre d'entreprises sont passées de simples techniques de Business Intelligence aux techniques de l'intelligence artificielle qui permettent de gagner énormément.

Compte tenu de ces éléments, l'ADD souhaite mettre en place un écosystème dédié à l'intelligence artificielle, comme l'un des chantiers prioritaires validés par le conseil d'administration. Il consiste au développement des activités de recherche

appliquée et d'innovation dans le domaine de l'intelligence artificielle, le développement des compétences nationales dans le domaine de l'IA, à travers la formation, la facilitation de l'émergence des acteurs marocains dans le domaine de l'IA, à savoir les startups, les laboratoires de recherche et la mise en œuvre des investissements en IA. Finalement, le développement des secteurs prioritaires pour le développement de l'application de l'intelligence artificielle au Maroc.

Il faut savoir que dans ce chantier de mise en œuvre d'un écosystème d'IA, nous sommes conscients de l'importance de l'IA et nous avons déjà fait un Benchmark sur ce qui est fait à l'international dans ce domaine, ce qui a été pris en considération dans l'élaboration de notre feuille de route.

Pour cette feuille de route, on va lancer une étude pour définir une feuille de route pour le Royaume du Maroc en prenant en considération la nature du système, des intervenants et également proposer des écosystèmes en IA pour amorcer le développement des activités d'intelligence artificielle.

Comme déjà mentionné, dans le chantier prioritaire d'intelligence artificielle, nous sommes partenaires dans le Programme AL-KHAWARIZMI de recherche appliquée dans le domaine de l'intelligence artificielle. C'est un programme doté d'un budget de 50 millions de dirhams, qui accompagnent 45 projets retenus pour financement dans divers domaines. L'objectif est de contribuer à l'émergence de l'écosystème de l'intelligence artificielle, à travers plusieurs actions d'accompagnement et de recherche afin de préparer ces projets à la valorisation socio-économique.

Dans cette perspective, nous avons organisés des sessions de formation et de sensibilisation sur la propriété intellectuelle-industrielle pour sensibiliser les équipes de recherche sur ce qui doit être protégé par les droits d'auteur. Les projets retenus portent sur plusieurs thématiques, notamment la santé, l'eau, l'énergie, l'industrie, l'éducation, l'emploi, transport et logistique, le tourisme, la justice, etc. Ces 45 projets relèvent de 15 universités et structures de recherche présentes au Maroc. Pour les prochaines étapes du programme, nous envisageons d'appuyer plus les équipes de recherche à travers des ateliers sectoriels.

Saida Belouali

Professeure en éthique appliquée, Co-responsable de la Maison de l'Intelligence Artificielle, Université Mohammed 1er, Maroc

L'éthique de l'IA dans l'action de la maison de l'IA à Oujda

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

Tout d'abord, je tiens à remercier les organisateurs du Conseil National des droits de l'homme, tout particulièrement Monsieur le Secrétaire Général, pour nous avoir réuni autour de cette question épineuse. Je félicite l'ensemble du staff pour la qualité de l'organisation.

Je suis également contente d'avoir rencontré des personnes avec un haut potentiel intellectuel qui vous enrichissent, vous apprennent davantage et vous poussent toujours à explorer d'autres pistes et confronter vos convictions et vos opinions et vous remettre en question tout simplement.

Je vais vous parler d'une expérience que nous avons mise en place à l'Université Mohammed Premier d'Oujda : « La maison de l'intelligence artificielle (MIA) ». C'est une réflexion d'un ensemble de personnes et d'une équipe pluridisciplinaire. C'est aussi l'histoire d'un ensemble de rencontres, un processus également qui s'est fait en continuum autour de cette intelligence collective et de cet ensemble de rencontres.

Commencée il y a quelques années avec des événements, des flagships, et par la suite la mise en place d'un ensemble de projets multidisciplinaires et une équipe qui vient d'horizons différents, une équipe qui se complète, qui s'accepte et qui affronte les connaissances et les doutes. C'est aussi l'histoire d'un réseau national et international et des partenariats de grande qualité, qui s'écoutent, s'apprécient, s'estiment et se complètent.

La MIA est une franchise de la maison de l'intelligence artificielle des Alpes-Maritimes en France. Pour le moment, c'est l'unique franchise en Afrique. Dans la philosophie de la maison de l'IA on a tenté de respecter les orientations de la maison mère, mais on y ajoutant des briques qui sont locales et qui nous intéressaient en tant qu'équipe pluridisciplinaire. On a introduit une dimension recherche et développement qui n'existe pas dans la maison mère. C'est une franchise qui nous a permis de mettre en place un espace autour d'un ensemble de projets. La maison s'est lancée il y a plus d'une année, on a mis en place le contenu scientifique et tous les parcours qui vont faire en sorte l'animer.

Inaugurée récemment à l'occasion de la semaine de l'intelligence artificielle. Elle est constituée de trois espaces : un espace d'exposition, d'expérimentation et de coworking. Nous avons retenu quatre axes qui sont pour nous prioritaires, parce que la réflexion est aussi locale et régionale. Il s'agit du domaine de la santé, le développement durable, le patrimoine logistique transport et tourisme et bien évidemment l'éthique.

La maison de l'intelligence artificielle c'est aussi des applications, du contenu numérique. Les missions fondamentales de la maison de l'IA c'est de développer des cas d'usage et des cas d'expérimentation, mesurer les impacts sociaux et faire de la veille éthique et c'est aussi la mise en

tension d'un réseau de partenariats, mais c'est surtout un collectif de chercheurs et d'experts agissant pour la recherche multidisciplinaire et multisectorielle.

Le campus est placé sur une surface de 1300 m². La philosophie de la MIA est d'acculturer l'intelligence artificielle et de toujours réfléchir en fonction des impacts sociaux. Nous avons utilisé des expérimentations pour arriver à faire cela. Par exemple l'atelier sur la voiture autonome, dans lequel on va expliquer les dilemmes, les choix présentés et la délégation des choix à la voiture autonome qui va prendre les décisions à notre place. Il y a aussi l'atelier drones, l'expérience de délibération éthique pour pousser les personnes à apprendre à raisonner, délibérer et opérer des choix dans des situations différentes et l'expérience de discussion avec un Chatbot, etc.

Nous avons parmi les structures de la MIA l'observatoire de veille éthique. Comme mentionné antérieurement, la philosophie de la MIA c'est remettre l'Homme au centre. Pour ce faire, il fallait mettre en place cet observatoire de veille éthique, situé à l'Université Mohammed Premier de Oujda. La réflexion autour de l'éthique a été entamée vers 2004 et on a commencé à mettre en place un module en éthique appliquée, ayant trait à l'éthique professionnelle dans le cursus des élèves ingénieurs.

Les missions de l'observatoire c'est d'appréhender ces transformations et proposer une réflexion locale. Nous n'avons pas la prétention de répondre à des questions qui nous dépassent, mais nous avons la prétention de réfléchir aux impacts différents sur notre territoire par rapport à toute la politique numérique. Nous souhaitons contribuer à la formation et la recherche, c'est pourquoi on propose des formations et on l'a fait avec HEP France. On a proposé une formation en éthique et on a contribué à la mise en place d'un module sur l'éthique appliquée. Nous souhaitons produire des rapports scientifiques, mais aussi des rapports à destination de responsables publiques, parce qu'on est sur l'aspect d'acculturation.

Vous avez un ensemble d'applications développées au sein de la Maison de l'Intelligence Artificielle, pour vous expliquer que l'éthique est toujours présente dans ce que nous faisons à la MIA. La première application est CONFLIEARN qui propose un tuteur intelligent pour adresser un peu la problématique qu'on a confronté lors du confinement, on a justement posé le souci de la connexion au Maroc. Ce qu'on a proposé c'est un tuteur intelligent et la possibilité de faire des cours en mode offline avec l'introduction d'une dimension éthique.

Tout ce qui se fait au sein de la MIA a absolument toujours un regard éthique, en effet dans l'ensemble des applications on a essayé à chaque fois de poser les questions éthiques qui pourraient éventuellement être posées et auxquelles on essaie de trouver des solutions. Parfois, on ne peut pas trouver des solutions mais l'objectif dans un espace de recherche c'est de trouver des hypothèses de recherche sur lesquels on fait travailler des doctorants qui sont plus de 15 doctorants encadrés par les membres du comité de gestion au sein de la MIA.

Je souhaiterai également vous parler aussi d'un projet sur lequel on travaille au niveau de la MIA et de l'observatoire de l'éthique. Ce projet pose la problématique de l'impossibilité d'une éthique universelle de l'IA. Il s'agit d'essayer de trouver l'intersection entre l'ensemble des codes, des référents et des standards qui sont aujourd'hui autour de l'intelligence artificielle. Donc, est ce que la question de l'éthique de l'IA est une question qui devrait être universelle, ou y a-t-il des spécificités qui devraient être prises en compte ? Quel est le rapport entre les deux et comment trouver la complémentarité dans les débats et les discussions ? Il n'y a pas forcément de tension entre les deux, mais ce que nous avons essayé de comprendre à travers une étude de concordance sur le plan de terminologie en faisant des rapprochements entre les termes et les syntagmes, c'est de dire comment traiter la pertinence contextuelle. L'observatoire de veille éthique

est un ensemble de partenariats, par exemple nous avons un partenariat avec le CNDH, Global AI, etc. J'attire votre attention sur le fait qu'on a aussi une revue qui traite de ces questions.

Il faut aussi que je vous parle de ce qu'est l'éthique et comment elle est conçue au niveau de l'observatoire de veille éthique ? J'ai retenu un visage celui de Paul RICOEUR qui est un philosophe français. Je préfère parler de l'éthique de l'IA, en retenant avec Paul RICOEUR ces trois éléments : l'éthique est une relation, une contextualisation et une démarche.

L'éthique est une relation, pour lui l'éthique est une recherche du bon pour soi, avec autrui et dans des institutions. Donc, on ne peut pas dire que l'éthique se fait indépendamment de l'autre, c'est pour ça que la spécificité est importante. Quand on dit pour soi, mais avec l'autre, on doit prendre en considération la singularité et la spécificité de l'autre dans la complémentarité et c'est quelque chose d'important. D'une autre part, cette relation nous renvoie aussi à l'environnement, car la relation que nous avons avec l'environnement est aussi importante.

L'éthique est une contextualisation, à la différence de la morale qui est prescriptive, l'éthique est une construction. C'est un cheminement, un recommencement, une élaboration morale et c'est pour ça qu'on parle d'enjeux et de dilemmes éthiques. Donc, forcement vous êtes dans la recherche d'un comportement bon et face à l'enjeu éthique, obligés de prendre en considération la relation, le contexte, les données qui s'offrent à vous et de prendre votre décision. Pour arriver au comportement éthique, il faut faire le lien aussi avec le fait que l'éthique est une délibération et un raisonnement. C'est pour ça, il est important de parler de l'éducation au centre de l'éthique, d'apprendre aux gens à réfléchir et à raisonner. L'éthique nous permet de faire la priorisation de nos valeurs et nos positions parce qu'on est en relation avec l'environnement et avec l'autre.

On doit faire le lien entre l'éducation et l'éthique, cette démarche dé délibération et de raisonnement qu'on voudrait transmettre aux élèves et étudiants est importante. Quand on enseigne aux étudiants la communication, les maths, la chimie, le leadership, en négligeant l'aspect valeurs, on est en train de leur dire que les valeurs sont accessoires, ce qui est dangereux. Apprendre aux gens que la délibération éthique est une forme de raisonnement et de prise de décisions est une chose importante, parce que cela nous renvoie à ce que nous avons dit antérieurement, l'éthique est une construction, elle n'est pas prescriptive. Cela nous renvoie aussi à la question de l'éthique de l'IA ou l'IA éthique. On parle de l'éthique de l'IA, parce qu'il s'agit d'une démarche et c'est une décision qu'on doit prendre, qui nous renvoie à l'homme, qui doit être au centre. Elle nous renvoie aussi à la décision qui doit rester toujours humaine et elle le restera.

Il y a l'histoire de Black LEMOINE qui veut porter plainte contre Google et qui considère qu'il a été lésé, il a été licencié de Google et il a même exprimé des sentiments de peur. Cela nous perturbe puisque l'intelligence artificielle pourrait un jour prendre des décisions et donc elle doit être éthique. Je pense qu'aujourd'hui on est loin d'une intelligence artificielle qui remplace l'Homme.

Sur la vision et la stratégie on a un ensemble d'activités à venir. Face à ces dilemmes on est obligés de faire un choix qui n'est pas totalement moral ni immoral, il est ce qu'il est, il faut l'accepter pour faire les choix nécessaires, au-delà de tout déterminisme. Il faut garder à l'esprit ce manifeste de Cyntia FLEURY, pour elle l'horizon ne peut être volé, alors qu'à l'instar du manifeste d'Antoine FENOGLIO il faut qu'on s'attache à nos valeurs et les prioriser.

Abdessalam Jaldi

Spécialiste en Relations Internationales au Policy Center for the New South, Maroc

Le débat sur l'intelligence artificielle au Maroc : le besoin de concilier entre le progrès et la protection des libertés et droits fondamentaux

1. LE CADRE DE L'INTELLIGENCE ARTIFICIELLE AU MAROC

- Le Maroc devrait connaître une accélération du déploiement de l'intelligence artificielle dans plusieurs secteurs.
- L'usage de l'IA suppose de participer à son développement aussi bien dans le secteur public que privé.

A- Les bienfaits du recours à l'intelligence artificielle au Maroc

- L'IA s'applique essentiellement au Maroc dans les domaines de traitement des images, le speech-to-text, et le texte.
- Les progrès considérables de l'IA laissent présager un nouveau cycle de croissance économique.
- L'IA peut favoriser le développement d'une nouvelle génération des produits et des services
- L'IA est en mesure d'offrir de nouvelles opportunités dans le domaine des services publics

B -L'utilisation grandissante de l'intelligence artificielle menace les libertés et droits numériques

- La non-discrimination
- La liberté d'expression
- La protection des données personnelles

2. LE BESOIN DE RÉGLEMENTER JURIDIQUEMENT L'INTELLIGENCE ARTIFICIELLE AU MAROC

- La protection des libertés et droits numériques doit s'appuyer sur des principes juridiques identifiés et clairement réaffirmés
- Le juge, les autorités compétentes et la société civile ont un rôle important à jouer en la matière

A- La nécessité d'une intelligence artificielle transparente

- Assurer une voie de recours effective que les individus pourront mettre en œuvre pour valoir leurs droits
- Promouvoir la transparence des algorithmes par le recours à la rétro-ingénierie
- Privilégier les mécanismes extra-judiciaires de règlement des différends numériques

B- L'élaboration d'un statut juridique propre à l'intelligence artificielle au Maroc

- Le principe de la responsabilité en cas de préjudice causé par une intelligence artificielle
- La personnalité électronique
- L'octroi de la personnalité juridique à une IA réellement autonome et dotée d'une conscience de soi
- Retracer le raisonnement de l'algorithme

L'USAGE DE L'INTELLIGENCE ARTIFICIELLE AU MAROC : LE BESOIN D'UN ÉQUILIBRE ENTRE LA SCIENCE ET LA PROTECTION DES DROITS FONDAMENTAUX

La place de l'intelligence artificielle dans nos vies est de plus en plus importante. Cette notion regroupe l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine. En effet, l'intelligence artificielle a ainsi été créée

pour être au service de l'humain en lui simplifiant de nombreuses tâches. Pourtant, son développement peut porter atteinte à plusieurs libertés et droits fondamentaux. Au Maroc, le préambule de la Constitution garantit les libertés et droits fondamentaux. À l'échelle du continent africain, les libertés et droits fondamentaux sont principalement protégés par la Convention africaine des droits de l'Homme et des peuples. Au niveau international, la Déclaration universelle des droits de l'Homme de 1948 énonce les principaux droits fondamentaux et libertés. Chaque être humain peut en principe se prévaloir de ceux-ci. Il est donc nécessaire de trouver un équilibre entre le développement et l'utilisation de l'intelligence artificielle au service de l'humain, et la protection des libertés et droits numériques fondamentaux.

1. LE CADRE DE L'INTELLIGENCE ARTIFICIELLE AU MAROC

Le Maroc devrait connaître, dans les prochaines années, une accélération du déploiement de l'intelligence artificielle dans plusieurs secteurs. Reposant sur un processus d'imitation de l'intelligence humaine qui se base sur la création et l'application d'algorithmes, l'intelligence artificielle a un avenir prometteur et représente une énorme opportunité pour le pays. Toutefois, il est indispensable de préparer le terrain pour l'usage de cette technologie, voire de participer à son développement aussi bien dans le secteur public que privé, notamment au regard de ses limites vis-à-vis des libertés et des droits fondamentaux.

A- Les bienfaits du recours à l'intelligence artificielle au Maroc

L'intelligence artificielle connaît une effervescence à tout égard considérable au Maroc. Cette dernière s'applique particulièrement dans les domaines de traitement des images, le speech-to-text et le texte. S'agissant des images, l'usage de l'intelligence artificielle s'avère prépondérant dans les radars reconnaissant les matricules, les caméras installées dans certaines grandes agglomérations gérant les trafics de circulation, les images satellites qui peuvent être appliquées dans le domaine de l'agriculture, ou encore l'organisation du transport en commun des grandes métropoles. Concernant le speech-to-text, cette technologie permet de transformer n'importe quel contenu audio en texte écrit. Récurrente dans les entreprises, le speech-to-text permet à travers le Machine Learning, aux entreprises de gagner en productivité. L'intelligence artificielle permet, de résumer des livres, ainsi que d'autres types de textes, comme des études ou des documents de recherche, peu importe leur longueur, grâce à cette technologie qui condense les informations essentielles en quelques lignes. Une évolution notable bénéfique aux entreprises, mais aussi à l'État, dans la mesure où elle contribue à la détection des signaux faibles à travers les applications dans les réseaux sociaux, permettant d'identifier d'éventuelles menaces susceptibles de nuire à la sécurité nationale.

En outre, les progrès considérables de l'intelligence artificielle laissent présager un nouveau cycle de croissance économique, qui pourra profiter au Maroc si ce dernier parvient à déclencher efficacement la quatrième révolution industrielle. En effet, l'intelligence artificielle pourrait permettre aux citoyens marocains d'accéder à de meilleurs soins de santé, des voitures et d'autres modes de transport plus sûrs, ainsi que des services moins coûteux, mieux adaptés à leurs besoins et munis d'une plus longue durée de vie. Elle pourrait également faciliter l'accès à l'information, l'éducation et les formations - des aspects fortement mis en avant durant la pandémie de la Covid-19. Elle peut aussi contribuer à rendre le milieu du travail plus sûr, puisque des robots peuvent être employés pour compléter des tâches dangereuses, tout en créant de nouveaux emplois avec l'expansion des entreprises utilisant l'intelligence artificielle. A titre d'exemple, et dans le cadre du secteur automobile, il existe de nombreuses applications de l'intelligence

artificielle dans la voiture autonome. Elles concernent principalement l'apprentissage de la conduite par Machine – Learning pour apprendre à la voiture comment se comporter en cas d'accident. L'intelligence artificielle intervient également dans la vérification de la qualité des organes de la voiture en maintenance prédictive ; dans la connaissance de l'environnement grâce aux données remontées de capteurs ; dans l'analyse du comportement du conducteur ou encore dans la cybersécurité pour surveiller l'état de la connectivité et éviter tout piratage.

Du point de vue des entreprises, l'intelligence artificielle peut favoriser le développement d'une nouvelle génération de produits et services, y compris dans des secteurs où les entreprises européennes occupent déjà une place importante : l'économie verte et circulaire, la machinerie, l'agriculture, la santé, la mode et le tourisme. Elle peut créer de nouvelles voies de vente, améliorer la maintenance de machines, augmenter le rendement, améliorer les services pour les consommateurs et permettre d'effectuer des économies d'énergie. A titre d'exemple, l'intelligence artificielle est en passe de jouer un rôle majeur dans l'évolution et le renouvellement de l'industrie du tourisme. L'intelligence artificielle et le machine learning permettent de programmer un chatbot, ou assistant conversationnel, doté de deux qualités essentielles : le langage et l'intelligence cognitive. Véritable assistant personnel, le chatbot peut fournir une réponse rapide et cohérente à des milliers de touristes en quête de toute information ou service, tel qu'un code Wifi, mais aussi une réservation de chambre ou de taxi. Le chatbot offre une conversation avec le client qui respecte la culture et l'image de marque de l'entreprise ou de la personne qui l'utilise.

Enfin, lorsqu'elle est utilisée dans les services publics, l'intelligence artificielle peut réduire les coûts et offrir de nouvelles opportunités dans les domaines du transport public, l'éducation, l'énergie, la gestion de déchets et peut améliorer la durabilité des produits. En outre, les experts prédisent que l'intelligence artificielle sera plus fréquemment employée dans le système judiciaire et dans la prévention de la délinquance, avec des ensembles massifs de données analysées rapidement, une évaluation plus fine des risques associés à certains profils criminels, voire même l'émergence de systèmes capables de prédire et prévenir des attaques terroristes. Elle est d'ores et déjà utilisée par les plateformes en ligne pour détecter les comportements illégaux ou dangereux en ligne.

B- L'utilisation grandissante de l'intelligence artificielle menace les libertés et les droits numériques

De nombreuses études ont démontré que l'intelligence artificielle porte atteinte aux libertés et droits fondamentaux. Seules les atteintes au principe de non-discrimination, à la liberté d'expression aux données personnelles seront développées dans cette intervention.

a. La non-discrimination

D'une part, le principe de non-discrimination est inscrit à l'article 7 de la Déclaration universelle des droits de l'Homme, l'article 2 de la Charte africaine des droits de l'Homme et des peuples, et l'article 19 de la Constitution marocaine. Alors que le recours à l'intelligence artificielle est souvent expliqué par la neutralité de cette technologie, cette neutralité ne serait en réalité qu'une apparence. A priori, ce recours permettrait d'échapper aux préjugés et à la discrimination inconsciente effectuée par tout être humain car les décisions ne sont plus données par un être humain doté d'une opinion, mais par une machine. Toutefois, il ne faut pas oublier que cette machine a été imaginée et façonnée par un être humain. Au cours de la création d'une intelligence artificielle, l'être humain conçoit forcément cette technologie avec ses préjugés, qu'ils soient conscients ou inconscients. Ainsi, dès sa création, l'intelligence artificielle hérite d'un biais de la part de son créateur.

Selon une récente étude de l'agence américaine "National Institute of Standards and Technology", un système de reconnaissance faciale se trompe entre 10 et 100 fois plus lors de l'identification de personnes d'origine afro- américaine ou d'origine asiatique, que lors de l'identification de personnes d'origine caucasienne. Une erreur commise par un logiciel de reconnaissance faciale peut être très préjudiciable notamment si une telle technique est utilisée par les forces de l'ordre : cela pourrait entraîner de fausses accusations ou de fausses arrestations. La legal tech américaine Palantir a par exemple développé et testé des outils de police prédictive dès 2012 dans des villes comme Chicago, Los Angeles, la Nouvelle-Orléans et New York. La police prédictive est le fait de prévoir qu'une infraction sera commise à un endroit précis, ou de prévoir si un suspect a des chances de commettre une infraction. L'emploi de l'intelligence artificielle comme un outil de police prédictive peut être considéré comme un bienfait pour la société dans son ensemble, car il permettrait une prévention et une répression des infractions plus efficace, mais cela présente un risque réel pour les libertés individuelles. Ces technologies sont également testées au Maroc, mais les outils de police prédictive n'établissent pas de « fiches suspects » et se focalisent sur les endroits où il est probable qu'une infraction soit commise.

Dans le domaine de la justice, l'intelligence artificielle permet d'accélérer le traitement d'énormes quantités de données et de déterminer la durée des peines d'emprisonnement pour permettre des approches plus homogènes d'affaires comparables. Néanmoins, l'utilisation par les juges américains de l'algorithme d'intelligence artificielle COMPAS, servant à déterminer la probabilité de récidive d'un individu, a soulevé une controverse aux États-Unis. En 2016, une enquête de l'ONG ProPublica a conclu que les données utilisées par l'algorithme COMPAS étaient biaisées, et donc que l'algorithme l'était également, au détriment d'individus issus de minorités. Si ces atteintes au principe de non- discrimination continuent, cela risque de réduire la confiance dans l'application de la loi suite à l'utilisation de l'intelligence artificielle, et donc de compromettre l'État de droit.

L'utilisation de l'intelligence artificielle dans un logiciel de recrutement d'Amazon a abouti à des résultats similaires. En 2014, le géant du e-commerce a développé un programme informatique afin d'automatiser le processus de recrutement de ses salariés. Cependant, Amazon a dû y renoncer 3 ans plus tard car ce logiciel discriminait les femmes. L'objectif du logiciel était d'attribuer une note d'une à cinq étoiles en fonction du profil des candidats. Pour ce faire, l'intelligence artificielle s'appuyait sur les profils des candidats d'Amazon sur une période de 10 ans. À cause de la prédominance masculine dans le domaine des nouvelles technologies au cours de cette période, l'intelligence artificielle a déduit que les profils masculins étaient à favoriser, au détriment des profils mentionnant le mot « femme » ou « féminin ».

L'intelligence artificielle accentue donc la discrimination en l'ancrant dans la technologie. En réaction à cela, les ONG Amnesty International et AccessNow ont rédigé la Déclaration de Toronto. Ce texte, publié le 16 mai 2018, a pour but de protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique, de surcroît consolidé par le rapport de 2021 de UNESCO qui appela à une éthique propre aux aléas de l'intelligence artificielle.

b. La liberté d'expression

La liberté d'expression est inscrite à l'article 19 de la Déclaration universelle des droits de l'Homme, l'article 9 de la Charte africaine des droits de l'Homme et des peuples, et l'article 25 de la Constitution marocaine. Toutefois, l'intelligence artificielle est régulièrement utilisée, notamment par les réseaux sociaux, pour contrôler le contenu diffusé

par leurs utilisateurs. Cette technologie va en effet identifier le contenu qui n'est pas conforme aux conditions d'utilisation de ces plateformes.

Les gouvernements qui exercent des pressions sur les entreprises pour qu'elles s'attaquent au problème des contenus terroristes, des discours incitant à la haine et des fake news ont conduit à une utilisation accrue des systèmes automatisés. La loi allemande NetzDG exige notamment que les réseaux sociaux suppriment de nombreux contenus entre 24 heures et 7 jours suivant leur signalement. Cependant, puisque l'intelligence artificielle est imparfaite, et que les réseaux sociaux sont contraints de supprimer rapidement le contenu indésirable, une partie du contenu peut être supprimée par erreur. Par exemple, YouTube a supprimé plus de 100.000 vidéos documentant les atrocités commises en Syrie après qu'elles aient été signalées pour « incitation à la haine ». Pourtant, ces vidéos constituaient souvent la seule preuve des crimes et des violations des droits de l'Homme commis dans ce pays : l'intelligence artificielle n'est pas parvenue à distinguer ces vidéos violentes, pour lesquelles YouTube prévoit des exceptions à cause de leur valeur éducative et documentaire, des vidéos violentes à des fins de propagande. En France, la loi Avia, inspirée de la loi allemande NetzDG et prévoyant une obligation, pour les réseaux sociaux, de retirer en 24 heures les contenus illégaux, a été censurée par le Conseil constitutionnel le 18 juin 2020, par crainte des atteintes à la liberté d'expression.

c. La protection des données personnelles

La nécessité de réglementer la protection des données à caractère personnel a été dictée au législateur marocain non seulement par des impératifs économiques mais également par des impératifs d'adéquation de la législation nationale avec les principes relatifs aux droits humains énoncés dans des textes fondamentaux internationaux (l'article 12 de la Déclaration universelle des droits de l'Homme, et l'article 17 du pacte international de 1966 relatif aux droits civils et politiques) et nationaux. En outre, l'article 24 de la Constitution de 2011 stipule que : « Toute personne a droit à la protection de sa vie privée. Le domicile est inviolable. Les perquisitions ne peuvent intervenir que dans les conditions et les formes prévues par la loi. Les communications privées, sous quelque forme que ce soit, sont secrètes. Seule la justice peut autoriser, dans les conditions et selon les formes prévues par la loi, l'accès à leur contenu, leur divulgation totale ou partielle ou leur invocation à la charge de quiconque. Est garantie pour tous, la liberté de circuler et de s'établir sur le territoire national, d'en sortir et d'y retourner, conformément à la loi ».

Or, force est de constater que le cadre juridique de la protection des données personnelles a été défini alors que la circulation des données et leurs valeurs économiques restaient limitées. Il incombe à l'intervention publique d'assurer d'une part, la sécurisation juridique des données, clé de voûte du développement de l'économie numérique, et d'autre part, un encadrement plus étroit des traitements présentant les risques les plus importants.

Des limites de l'intelligence artificielle sont donc constatées en ce qui concerne le respect des libertés et droits fondamentaux.

2. LE BESOIN DE RÉGLEMENTER JURIDIQUEMENT L'INTELLIGENCE ARTIFICIELLE AU MAROC

L'apparition de la notion juridique d'intelligence artificielle a été entravée par le manque de cohérence scientifique, dont les fondements et contours doivent être précisés. Alors que l'individu s'aventure dans un monde déterritorialisé, la protection des libertés et des droits numériques doit s'appuyer sur des principes juridiques identifiés et clairement réaffirmés, et sur une large palette d'outils de régulation. Le juge a un rôle majeur à jouer, aux côtés des autorités indépendantes spécialement compétentes et la société civile.

1- La nécessité d'une intelligence artificielle transparente

En dépit des atteintes aux libertés et droits fondamentaux, il ne faut pas rejeter l'avancée technologique que représente l'intelligence artificielle. En effet, d'après Salil Shetty, ancien secrétaire général d'Amnesty International, « il existe des possibilités et des bénéfices immenses à tirer de l'intelligence artificielle si les droits de l'Homme sont au cœur de cette technologie ».

Pour développer l'intelligence artificielle tout en limitant son atteinte aux libertés et droits fondamentaux, la première chose à faire est d'assurer une voie de recours effective que les individus pourront mettre en œuvre pour faire valoir leurs droits. L'article 8 de la Déclaration universelle des droits de l'Homme, ainsi que l'article 118 de la Constitution de 2011, énoncent en effet que le droit à un recours effectif en cas de violation des libertés et droits fondamentaux est un impératif. En France à titre d'exemple, il existe une telle voie de recours appelée le référé liberté, prévue à l'article L. 521-2 du Code de justice administrative. Cette procédure existe au sein du tribunal administratif marocain, et peut être initiée à trois conditions : il faut justifier d'une urgence, montrer qu'une liberté fondamentale est en cause, et montrer que l'atteinte à cette liberté est grave et manifestement illégale.

Deuxièmement, le Livre blanc de la Commission européenne sur l'intelligence artificielle a expliqué que « les particularités qui caractérisent de nombreuses technologies de l'IA, notamment l'opacité (« effet de boîte noire »), la complexité, l'imprévisibilité et le comportement partiellement autonome, peuvent rendre difficile la vérification de la conformité aux règles du droit de l'UE en vigueur destinées à protéger les droits fondamentaux et peuvent entraver le contrôle de l'application de celles-ci ». Il apparaît donc nécessaire d'avoir des algorithmes plus transparents, afin de mieux comprendre les solutions qu'ils proposent, de pouvoir imputer la responsabilité de ces décisions à quelqu'un, et de pouvoir les contester plus facilement. La rétro-ingénierie, qui est la science qui décrypte les raisonnements d'une machine, est essentielle pour l'augmentation de la transparence des outils ayant recours à l'intelligence artificielle. Cela suppose d'appréhender le régime de la responsabilité en cas de préjudice causé par une Intelligence artificielle.

Troisièmement, le recours aux mécanismes extra-judiciaires de règlements des différends peut contribuer à résoudre les litiges relatifs aux questions numériques, à l'exemple de l'intelligence artificielle. En effet, des litiges liés à l'utilisation numérique, qu'ils portent sur les données personnelles, les atteintes à la réputation sur internet ou l'intelligence artificielle, sont considérés dans la terminologie juridique numérique comme des « petits litiges ». Or, les procédures juridictionnelles classiques sont peu adaptées au traitement de ces formes de contentieux, conduisant des particuliers à renoncer à faire valoir leurs droits, alors que les enjeux peuvent être significatifs pour les personnes concernées. Dans cette perspective, alors que les juges, confrontés à des évolutions techniques complexes qui supposent une expertise particulière, ne disposent pas des compétences nécessaires portant sur les différends numériques, l'usage des techniques extra-judiciaires de règlement des différends, comme la conciliation, la médiation ou l'arbitrage, peut constituer une alternative appropriée.

Enfin, la coopération entre les États, les entreprises privées, les universitaires, les ONG, les organisations internationales et les citoyens pourrait permettre d'aboutir au développement et à l'utilisation d'une intelligence plus respectueuse des libertés et droits fondamentaux. La sensibilisation des individus à l'intelligence artificielle, dès le plus jeune âge, est également primordiale afin que chacun ait un bon usage de ces technologies et comprenne leur influence dans nos vies.

2- L'élaboration d'un statut juridique propre à l'intelligence artificielle au Maroc

Avec la généralisation de l'usage de l'intelligence artificielle dans toutes sortes de domaines, un enjeu nouveau apparaît : celui de la responsabilité en cas de préjudice causé par une intelligence artificielle. On entend par cette dernière, l'obligation faite à une personne de réparer le préjudice causé à autrui. La fonction première du droit de la responsabilité civile est ainsi de compenser le préjudice subi par le lésé. Le but est de remettre ce dernier dans la situation dans laquelle il aurait été si l'événement dommageable ne s'était pas produit. La seconde fonction du droit de la responsabilité civile se veut préventive. En effet, l'obligation faite aux responsables de réparer les préjudices qu'ils causent incite tout un chacun à se comporter avec diligence. En revanche, le droit de la responsabilité n'a pas pour but de punir le responsable. Cette fonction punitive appartient au droit pénal.

Or, force est de constater que l'ordre juridique marocain, comme celui de la plupart des pays avancés, n'apporte pas de réponse claire à cette question. En effet, l'intelligence artificielle est une chose immatérielle dépourvue de personnalité juridique et de patrimoine, de sorte qu'elle ne peut être tenue pour responsable de ses actes. Par conséquent, et comme une intelligence artificielle est immatérielle (software), mais incorporée dans des machines (hardware), il est difficile d'en extraire son comportement spécifique au-delà de l'action qu'elle co-commande avec l'Homme. Cependant, son autonomie grandissante (co-apprentissage) va la rendre de jour en jour plus autonome de l'action humaine. Dans cette perspective, l'encadrement juridique des systèmes d'intelligence artificielle, qui ne bride pas l'innovation, mais qui offre une protection efficace aux potentielles victimes, revêt une importance à tout égard considérable. Or, cet objectif suppose de trouver de l'équilibre dans le dispositif juridique. A cet égard, plusieurs pistes sont proposées, parmi lesquelles la création d'une personnalité électronique, attribuant aux systèmes d'intelligence artificielle (SIA) un statut juridique, à l'exemple des sociétés anonymes durant la première révolution industrielle pour les sociétés anonymes. Le cas de l'Arabie Saoudite, qui a octroyé la citoyenneté au robot « Sophia » en 2017, disposant d'un statut comparable à celui des êtres humains, peut inspirer le Maroc. Ce dernier peut se déplacer et interagir avec son environnement sans intervention active d'un être humain, bien que Sophia soit dépourvue de conscience et de réelle autonomie.

Néanmoins, seule une intelligence artificielle véritablement autonome et dotée d'une conscience de soi, lui permettant de raisonner, de comprendre ce qu'elle fait et d'effectuer des choix, mérite l'octroi de la personnalité juridique. Le jour où l'IA sera véritablement autonome et dotée d'une conscience de soi, lui permettant de raisonner, de comprendre ce qu'elle fait et d'effectuer des choix, l'octroi d'un statut juridique se justifiera sans doute. Autrement dit, une intelligence artificielle capable de causer des préjudices avec volonté et conscience, et échappant totalement au contrôle de l'être humain, doit lui être octroyée un statut juridique, afin de lui imposer le poids de la réparation du préjudice causé. Or, cette forme d'intelligence artificielle n'existe pas encore, de sorte que la création d'une personnalité électronique n'est pour l'instant pas nécessaire. Il apparaît plus raisonnable d'imputer les effets de la responsabilité au concepteur, au fabricant, au propriétaire et à l'utilisateur de l'intelligence artificielle, lesquels se dissimulent inexorablement derrière les machines. Toutefois, pour désigner lequel de ces acteurs est responsable en cas de préjudice causé par une intelligence artificielle, il est indispensable de favoriser la transparence des algorithmes utilisés. Cela suppose, indispensablement de mettre en place un système permettant de retracer le raisonnement de l'algorithme afin d'imputer le résultat dommageable à une personne physique.

CONCLUSION

La quatrième révolution industrielle est entrain de faire naître un tout nouveau monde dans lequel le « machine learning », le « deep learning », les « réseaux de neurones » et toutes les nouvelles technologies de l'intelligence artificielle vont conduire les affaires du monde, dans lequel la généralisation de l'usage de l'intelligence artificielle ne sera pas une exception comme aujourd'hui. Il importe donc de définir un cadre juridique propre aux aléas de l'intelligence artificielle, afin d'éviter toute dérive juridique possible.

Salah Baïna

Expert en Transformation Digitale et Professeur universitaire à ENSIAS, Maroc

Université – Entreprise : quelle transition digitale ?

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

En préambule, je tiens à remercier le Conseil national des droits de l'Homme pour m'avoir impliqué dans le processus d'organisation de ce colloque, parce que je fais partie du comité scientifique ad hoc de cet événement. La présentation qu'on m'a invité à animer s'intitule : « Université, entreprise Quelle transition digitale ? ».

On ne peut pas parler de transformation digitale, de transition numérique, quel que soit le terme qu'on utilise, si on n'inclut pas toutes les parties prenantes de cet écosystème. Pour la simple raison, qu'il s'agit-là d'une transition sociétale. Les citoyens, les consommateurs, les entreprises et l'administration publique font partie des parties prenantes qui vont devoir effectivement suivre cette trajectoire. Je parle bien évidemment d'une trajectoire parce que ce n'est pas limité dans le temps, les objectifs ne sont pas connus à l'avance. On découvre en marchant quels sont les challenges, les défis et les questions auxquels on va devoir répondre.

Ce qui nous réunit aujourd'hui c'est bien évidemment l'éthique dans le domaine de l'intelligence artificielle. L'année prochaine ça sera peut-être les droits de l'Homme dans le métavers. On ne sait pas où les débats et la réflexion vont nous mener. Comme je suis dans un panel consacré à un état des lieux, je pense pertinemment qu'aujourd'hui au Maroc la notion d'éthique dans l'intelligence artificielle peut paraître un peu futuriste. Puisque, le niveau de maturité digitale observé généralement que ce soit dans l'administration publique, dans l'Etat ou dans les entreprises, n'est pas à ce point qui nécessite de répondre à des problématiques d'éthique.

Aujourd'hui, nous sommes en phase de prospection et nous y parviendrons. De la même manière, nous réussirons également à traiter d'autres questions. Il est important de répondre aux problématiques d'aujourd'hui, mais aussi d'anticiper les problématiques futures qui devront se poser. Cela concerne également la relation entre l'état actuel des choses, les devoirs présents, et les besoins de réflexion. J'ai articulé ma présentation en positionnant l'université comme un espace de pensée, et non uniquement comme un espace de production de talents ou de main-d'œuvre.

L'université est un espace de réflexion, de recherche et de prospection. Il est important d'avoir des outils qui permettent de réfléchir. La question de l'éthique se posera, mais également celle de la souveraineté et des droits de l'Homme. Nous devons envisager d'avoir les mêmes droits de l'Homme dans le monde virtuel que ceux que nous avons aujourd'hui dans le monde physique. Je poserai la question : sont-ils suffisants ? Le monde virtuel peut-il être suffisamment protégé simplement en transposant les droits de l'Homme actuels dans cet espace virtuel ? Il sera nécessaire de trouver de nouveaux droits qui protégeront des situations inédites dans le monde physique dans lequel nous vivons.

En partant de cette réflexion sur l'état des lieux du digital au Maroc et sur la

transition numérique , terme adopté par le gouvernement actuel , nous constatons que, depuis l'arrivée d'internet et de son utilisation par les particuliers, les entreprises et les administrations, le Maroc a fait des efforts pour déployer des stratégies, créer des feuilles de route et établir des ambitions. Ces ambitions ont évolué en parallèle avec l'évolution de l'outil lui-même.

Aujourd'hui, même si nous parlons de transition numérique et du digital comme étant un concept clair et bien défini, si l'on regarde l'arrivée des systèmes d'information dans les années 80, les outils numériques de cette époque ne sont pas ceux dont nous parlons aujourd'hui (les GAFA, le métavers et l'intelligence artificielle). Nous sommes donc face à un concept flou qui évolue, bien évidemment, avec l'évolution de la technologie. Au fur et à mesure que la technologie évolue, les usages changent également. Il est donc nécessaire d'adapter l'arsenal juridique pour accompagner cette transition.

On parle également de soft skills qui aident le citoyen, le consommateur et l'écosystème à s'adapter à ce monde numérique, ainsi que des usages pratiques utiles à la croissance économique et à l'évolution de la société.

Au fil de ces différentes évolutions, le Maroc a mis en place, avec plus ou moins de succès, des feuilles de route et des stratégies qui permettent de porter un peu plus loin les ambitions. La dernière stratégie publiée officiellement en 2016 s'appelait « Maroc numérique 2020 ». Nous attendons maintenant la nouvelle mouture des travaux concentrés dans ce sens pour élaborer une nouvelle stratégie.

De toute façon, cela ne diffère pas de ce qui se fait ailleurs dans le monde. Nous sommes dans une mise à jour perpétuelle des objectifs, avec plus ou moins de réussite. C'est le constat qui ressort à chaque fois qu'une institution nationale s'intéresse à la question de l'éthique au Maroc. Je cite les résultats d'un rapport élaboré en 2019 par la Cour des Comptes et le Conseil Économique, Social et Environnemental (CESE), qui a également publié un avis, dans le cadre d'une auto-saisine récente sur la digitalisation. On peut toujours dire qu'il y a des choses qui fonctionnent et d'autres qui fonctionnent moins bien.

On observe des avancées dans certains services, comme l'indique le rapport de la Cour des Comptes, notamment dans les services à grande valeur ajoutée. On peut toujours se poser la question : valeur ajoutée pour qui ? En tout cas, je dirais que les services à fort retour sur investissement sont plus pertinents que ceux à grande valeur ajoutée, surtout lorsqu'ils permettent de percevoir des impôts, de déclarer des bénéfices pour les entreprises, ou de déclarer des sources de revenus pour les citoyens. L'État a bien développé ce type de service, de telle sorte que nous nous sommes retrouvés au même niveau de maturité, bien sûr avec un décalage dans le temps.

Si vous regardez, c'est une comparaison entre les deux services en 2018 au Maroc et les services en 2010 dans les pays de l'Union Européenne. Mais, à vrai dire, au-delà de 100 %, il n'y a rien à chercher pour déclarer la TVA ou les impôts sur les sociétés. Nous avons des niveaux records de maturité pour ces services par rapport aux autres. Toutefois, il est à noter que pour les services destinés aux citoyens, ce que l'on appelle les "faits de vie" (déclarer une naissance, déclarer une perte de documents administratifs, déclarer un achat de voiture...), les citoyens se retrouvent typiquement en dehors du système digital et doivent revenir au monde physique, où les procédures sont moins digitalisées. C'est cette différence entre les niveaux de maturité de certains services par rapport à d'autres qui est constamment soulignée. Nous avons donc des services qui ne sont pas très avancés en matière de réalisation, à côté de services qui ne sont pas encore assez développés.

Je soulignerai un petit aspect lié au risque COVID-19, car la crise du COVID a agi comme un accélérateur digital pour tous les services. Pendant cette période,

l'enseignement est passé à distance et les services de santé se sont équipés pour faire de même. Même le mariage a été digitalisé. La procédure administrative permettant d'entamer le processus de mariage a été digitalisée. J'estime et je suppose qu'avant le COVID, ce service n'était pas prioritaire dans le cadre de la digitalisation. C'était un service qui fonctionnait très bien en mode physique. C'est vraiment le COVID et le blocage des déplacements qui ont incité les autorités à digitaliser ce service comme les autres. Étant donné que nous avons restreint les déplacements des citoyens, nous ne pouvions pas non plus les empêcher de se marier. À ce moment-là, nous avons considéré que se marier est un droit, un droit civique important, et donc qu'il fallait le digitaliser en urgence.

Dans l'histoire du Maroc, en ce qui concerne ces services destinés aux citoyens, on observe une croissance qui n'est pas aléatoire, mais marquée par des hauts et des bas. Je prends ici, tout particulièrement, le classement e-Gov établi par les Nations Unies tous les deux ans, et nous attendons avec impatience celui de 2022, qui est en cours de réalisation. Vous constatez qu'on a connu une forte croissance en 2008, atteignant le niveau 82 en 2016. Depuis 2016 jusqu'à 2020, nous n'avons pas connu la même croissance, mais nous n'avons pas non plus stagné. J'attends 2022 pour voir si le COVID a eu un impact, et si nous avons réussi à préserver notre croissance.

Le digital aujourd'hui n'est pas quelque chose de superflu, c'est un outil de bonne gouvernance, que ce soit pour les entreprises ou pour les administrations. Quand on examine ce qu'est la bonne gouvernance publique, on peut faire cette projection de manière très simple pour l'entreprise. Pour l'administration, la bonne gouvernance (الحكامة الجيدة) , c'est le terme que nous utilisons en arabe , consiste d'abord à avoir une administration éthique, intègre et ouverte sur son écosystème, à pratiquer une bonne gouvernance budgétaire, à lutter contre la corruption, à garantir la transparence, à gérer efficacement les risques, à encourager la participation citoyenne et à respecter les droits de l'Homme.

À partir du moment où le digital n'est plus simplement une cerise sur le gâteau, mais la mécanique même qui fait fonctionner la machine, nous devons l'utiliser pour toutes ces bonnes raisons. Aujourd'hui, une administration publique dématérialisée doit servir une bonne administration, respectant l'éthique, et je parle ici de l'éthique institutionnelle. Lorsque nous évoquons la non-discrimination, il est important de comprendre que la discrimination peut être humaine, mécanique ou automatique. De toute façon, nous ne voulons aucune forme de discrimination. Nous souhaitons une IA qui ne discrimine pas, mais nous ne voulons pas non plus de fonctionnaires qui discriminent. Pourquoi ? Parce que ce sont là les ingrédients d'une bonne gouvernance au sein de l'administration.

On parlait de e-Gouvernement (e-GOV) qui était considéré comme la partie digitale et électronique du gouvernement et de l'administration publique. Aujourd'hui, cette notion s'étend de plus en plus, puisque le e-GOV représente un pourcentage des activités publiques. Aujourd'hui on ne parle plus de e-GOV, tout simplement parce que tout est censé être digitalisé. Donc, je dirai qu'on parlera d'un smart GOV ou d'un digital GOV où la totalité de ses activités est présente dans ce monde digital.

Il ne s'agit pas d'avoir un GOV à l'ancienne et un GOV électronique. De la même façon, chez les entreprises, il n'y a pas deux façons de voir le monde, il n'y a qu'une seule façon et cette façon est à la fois physique et digitale. Donc, on parlera de smart GOV ou de digital GOV, cette notion qui unifie et fusionne les deux formes d'existence : l'existence physique et l'existence numérique.

Beaucoup d'institutions ont étudié l'état des lieux au Maroc. Je cite quelques recommandations de la Cour des Comptes et du CESE. Je souligne en gras « mettre le citoyen au cœur des transformations ». Si on garde cette volonté de digitaliser, cette

volonté ne doit pas être que pour avoir des retours pour des investissements ou que pour des services rentables. C'est le citoyen qui doit être au centre et le digital doit permettre de faciliter la vie civique d'un citoyen.

Pendant le COVID, nous avons également vu le département de la justice mettre en place un service de dépôt de plaintes à distance. Par exemple, lorsqu'on permet à une femme battue de déposer une plainte directement depuis chez elle, on crée un véritable raccourci efficace qui, dans la vie réelle, pourrait être entravé pour de nombreuses raisons. Le digital va accélérer le passage à l'acte. Ainsi, c'est typiquement un service qui change, un contexte particulièrement lié au domaine de la justice citoyenne, mais c'est également le cas dans tous les autres domaines. Mettre le citoyen au centre est donc une volonté importante aujourd'hui, à l'époque de ce digital accéléré.

Le CESE recommande de faire appel à l'intelligence artificielle. C'est une recommandation dirigée comme priorité nationale dans le chantier de la transformation digitale. La question qui va se poser est: quelle est l'intelligence artificielle qui met le citoyen au centre. Que signifie la mise en place du citoyen au centre de l'intelligence artificielle ? Pas que l'éthique, il y a d'autres questions qui pourront être abordées.

Pour revenir au rôle de l'université dans ce contexte, je pense qu'un concept qui est important aujourd'hui, pas seulement se focaliser sur l'intelligence artificielle, c'est la notion du cyberespace. Le cyberespace c'est justement l'autre volet de la vie quotidienne que nous vivons, une vie dans la vraie vie et une vie dans le monde du numérique, des services en ligne, des sites internet, des applications, etc. Donc de la même façon que les droits de l'Homme et que l'éthique sont recherchés dans le monde physique, ils devraient être aussi recherchés dans ce cyberespace. Souvent il faut donner une définition claire de ce qu'est le cyberespace. Le cyberespace c'est des sortes de technologies et des infrastructures (data center, des serveurs, des données et des usages).

De la même façon qu'on protège le citoyen sur la route, dans ses échanges avec les acteurs économiques, avec l'administration publique, et aussi dans les interactions entre citoyens dans le monde physique, nous devrions également, à travers ces différents niveaux et combinaisons du cyberespace, protéger le citoyen tout en l'impliquant en tant qu'acteur dans ce cyberespace. Lorsque nous parlons du cyberespace, nous pouvons lui attribuer le terme que nous voulons. Le cyberespace, à l'époque, était Internet ; demain, ce sera le metaverse, et dans dix ans, ce sera un nouveau concept. Quoi qu'il en soit, le cyberespace fait référence à une existence qui n'est pas uniquement physique. La physicalité du monde réel dans lequel nous vivons est en quelque sorte transformée, bouleversée dans le cyberespace.

Dans ce cyberespace, quelle citoyenneté souhaitons-nous offrir à nos citoyens ? Nous parlons de "e-démocratie", qui, selon les définitions les plus courantes, permet de bénéficier dans ce cyberespace de tous les droits citoyens dont on jouit dans le monde physique. Autrement dit, si un citoyen a le droit de vote dans le monde physique, il doit également avoir ce droit dans le cyberespace. Vous verrez qu'à un certain moment, lorsque nous aborderons la question de l'e-démocratie, ce ne sont plus des questions technologiques ou organisationnelles qui devront être résolues, mais bien évidemment des choix politiques. Car, quel que soit le système ou la technologie que nous utiliserons pour permettre ce droit de vote, les technologies existent. Nous pourrons toujours faire des expérimentations et réussir à le mettre en place, mais avons-nous la volonté politique de le réaliser ?

Le cyberespace n'est pas régi uniquement par des possibilités techniques, la faisabilité ou la viabilité économique ; il est également régi par la désirabilité politique.

C'est aussi le rôle des universitaires d'élaborer ces questions. Le cyberespace, l'intelligence artificielle, le metaverse vont progressivement évoluer d'une discussion technologique menée par des universités d'orientation technique vers des discussions impliquant des universités d'orientation humaine, car ils auront des impacts sur les sciences humaines, la réalité humaine de nos sociétés, et la vie de nos citoyens.

En ce qui concerne les droits de l'Homme, nous devons également nous demander quels droits de l'Homme s'appliquent dans le cyberespace. L'humain dans le cyberespace ne mérite-t-il pas plus que les droits humains que nous considérons aujourd'hui comme universels dans le monde physique ? Prenons un exemple très simple : le droit à l'anonymat. Est-ce qu'aujourd'hui, dans les droits de l'Homme, nous parlons du droit à l'anonymat ? La France, par exemple, a annoncé la fin des tickets de restaurant en 2024, ce qui est typiquement une forme de dématérialisation. Ce ticket, au lieu d'être en papier, sera numérique et sera envoyé par mail ou par SMS. Ce restaurant, qui ne connaît pas mon identité, aura besoin de connaître davantage d'informations me décrivant. Or, pour boire un café, je n'ai pas besoin de décliner mon identité. La dématérialisation de ce ticket en papier, pour plusieurs raisons, va systématiquement faire disparaître cette liberté très simple. Ainsi, un nouveau droit va apparaître : le droit d'être anonyme.

On parle beaucoup du droit à la vie privée, qui est aujourd'hui formalisé dans la constitution marocaine. Dans cette optique de protection des citoyens, et étant donné que les citoyens sont tous des sources de données, il est essentiel de protéger ces données. La possibilité de collecter toutes les informations sur chaque citoyen d'un pays, voire d'un continent, grâce aux outils de calcul disponibles, permet de reconstituer des bribes d'histoires, et ce sont ces histoires qu'il faut protéger. Auparavant, il était difficile de reconstituer les parcours de vie des populations, car cela relevait typiquement d'un travail humain. En raison du manque de ressources, ce travail ne se faisait pas, et on n'y pensait même pas. Aujourd'hui, nous disposons d'outils performants qui nous permettent de tracer les parcours et de reconstituer des histoires humaines simplement en analysant les données d'un citoyen. Protéger le citoyen et ses données personnelles implique également de protéger ses histoires.

En ce qui concerne l'IA et le Big Data, la Direction Générale des Impôts a lancé des appels d'offres visant à analyser les histoires des gens à travers des plateformes électroniques. Pour détecter quelques fraudeurs, on a examiné une grande partie de la population pour finalement découvrir que seulement 1 % ou 0,1 % étaient concernés. Cela nous a poussés à scruter la vie de tout le monde pour détecter quelques fraudeurs, soulevant ainsi la question de la proportionnalité. Est-il justifié de fouiller dans les données personnelles ou publiques à grande échelle pour identifier quelques fraudeurs ?

La réflexion à mener concerne ces nouveaux droits et cette nouvelle citoyenneté dans le cyberespace. C'est typiquement le rôle des penseurs et des chercheurs universitaires de ne pas se limiter à l'aspect technologique. Tant que la transformation digitale est perçue uniquement comme une problématique technologique, elle sera abordée sous cet angle. Actuellement, la majorité des projets sont techniques, et peu se concentrent sur une réflexion sociétale et humaine.

Finalement, voulons-nous évoluer dans un monde où non seulement les décisions sont automatisées, mais aussi autonomisées ? Les décisions des entreprises et de l'administration publique se font-elles entièrement par algorithmes ? Nous avons entendu parler de jugements rendus par algorithmes simplement parce que cela est techniquement faisable. Si nous fournissons les moyens et les données nécessaires, cela sera possible. Mais en voulons-nous vraiment ? C'est la question à laquelle nous n'avons pas encore apporté de réponses.

Il est temps de réfléchir aux humanités digitales, qui sont les sciences humaines qui accompagnent cette transition. L'IA n'est qu'une étape de cette évolution ; demain, une nouvelle question émergera. Tant que les interrogations ne sont abordées que sous l'angle technologique, nous pouvons avoir des outils techniques, mais sans réflexion ni profondeur philosophique sur le bon sens de cette discussion.

Si je devais encourager la relation entre l'État, les entreprises et l'université, ce serait dans le sens de trouver de nouvelles questions, de relever des défis, et de redéfinir la citoyenneté humaine dans le cyberespace. Il est crucial de définir ce qu'est le cyberespace, d'analyser les interactions de cette société virtuelle dans laquelle nous avançons, et d'essayer ensemble de trouver des réponses aux problématiques émergentes (éthique, droits de l'Homme, souveraineté, économie, etc.).

Il faut mettre à jour la réflexion sur les sciences humaines telle qu'elle est actuellement menée dans nos universités, afin qu'elle puisse aborder les sujets de transformation digitale. Tant que cela reste uniquement une question de technologie, il n'est même pas nécessaire de créer un nouveau concept, car la digitalisation suffisait. En parlant de transformation, nous évoquons une transformation de culture, de société, de mœurs, et finalement d'humanité. C'est vraiment ce rôle qu'il faudra redéfinir ensemble.

AI ETHICAL CODES, DISCUSSION AND IMPLEMENTATION PROSPECTS

Ravit Dotan

AI Ethics Advisor, Researcher, and Speaker, PhD., USA

AI Ethics for Investors

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

What I want to talk about today is why investors should care about AI ethics. There are two main reasons. First, the social impact is quite obvious: if we consider the ethics of the applications we're developing, we will achieve a better social impact. However, the focus of my discussion today will be on the return on investment (ROI). People are increasingly recognizing that AI ethics can enhance ROI. For example, in a recent survey by the Economist Intelligence Unit (EIU, 2020), 94% of senior executives stated that responsible AI will yield better returns for shareholders. Additionally, a report by PwC (2021) recommended that buyers of AI companies should factor in AI ethics when determining whether a transaction will realize its optimal value. McKinsey (2021) found similar results, noting a correlation between organizations with the highest AI returns and their engagement in risk-mitigation practices.

So, why does AI ethics positively influence ROI? I have identified five reasons: compliance, product quality, product adoption, talent attraction, and reputation.

The first reason is compliance. As we know, addressing compliance issues early is crucial; otherwise, we pay a higher price in the long run. Compliance influences AI in three ways. Firstly, there are specific laws related to AI that are attracting a lot of attention. Currently, two notable laws are the EU Artificial Intelligence Act, which is expected to pass soon and be enforced in the coming years. This act will regulate all AI deployed within the EU. Much like the GDPR, which is an EU law with global ramifications, we anticipate similar effects from the EU AI Act. It categorizes AI applications into risk levels and regulates them according to their associated risks. The top category will be prohibited, marking a significant shift. Many actors are already preparing for these changes.

In addition to that, there are existing AI laws, such as those in China, which regulate algorithmic recommendation systems. Beyond AI-specific laws, non-AI laws also impact AI companies. For instance, in the U.S., anti-discrimination laws have been applied to regulate AI, and privacy laws have also been used for this purpose. Therefore, much of the regulation pertains to how we can use or build AI ethically. Companies that develop AI in a responsible manner will be better equipped to comply with these regulations.

If you're wondering how much regulation to expect, the answer is substantial. For example, in the U.S. last year (2021), 130 federal bills were proposed, compared to just one in 2015. We are witnessing an increase in legislative efforts.

The second reason why AI ethics improves ROI is that it enhances product quality. Ethically developed AI performs better. Why is that? The Economist Intelligence Unit has identified three factors: First, efforts to reduce bias typically lead to better performance across diverse user profiles. This focus on a broader range of users

results in improved AI functionality. Second, transparency and accountability to end-users foster a deeper understanding of your products and yield more effective feedback. Third, prioritizing privacy protection and product security builds user trust, leading them to share more data, which ultimately results in better datasets and, consequently, better AI.

The third reason AI ethics enhances ROI is through increased adoption of AI systems, which translates to more clients or customers. This has been demonstrated in practice. In an EIU survey, 60% of senior executives reported that their organizations had decided against partnering with an AI service provider due to ethical concerns. This indicates that ethical considerations have tangible impacts. Additionally, 34% of individuals stated they would cease interactions with a company entirely if they had ethical concerns. This has been evident in various instances, such as Facebook's first decline in user engagement following AI ethics scandals, including the Cambridge Analytica incident.

The fourth reason AI ethics boosts ROI is talent attraction. It is well-known that younger generations prioritize ethical leadership, but this trend is evident across all age groups. Surveys have shown that ethical leadership is among the top two qualities sought in an employer. For instance, after the Cambridge Analytica scandal, the percentage of candidates accepting job offers from Facebook plummeted from 85% to between 35% and 55%. Moreover, candidates began asking tougher questions about privacy during interviews, indicating that AI ethics helps attract better talent.

The fifth reason AI ethics is crucial for ROI is reputation. By now, we have all seen numerous scandals related to AI ethics. One of the most prominent cases is that of Clearview, a startup that allows users to upload a photo of a person and receive links to other images of that person. Clearview faced backlash for scraping billions of images from the Internet, including platforms like Facebook, without consent. Consequently, they were accused of violating privacy laws, leading to fines and bans in several countries. Another example is AnyVision, which faced criticism for potentially selling its facial recognition technology for illegitimate uses, resulting in negative attention and attempts by Microsoft to divest.

This is just a glimpse of the damage to reputation that unethical practices can cause. Leading news outlets now employ reporters specializing in AI ethics, and we can expect more scandals to emerge.

In conclusion, AI ethics are vital for enhancing both social impact and revenue, which is why investors should prioritize it and consider how to integrate AI ethics into their due diligence and interactions with portfolio companies. I provided five reasons: compliance, product quality, product adoption, talent attraction, and reputation.

Jean Pierre Llored

Enseignant-chercheur-HDR en sciences humaines et sociales, École Centrale Casablanca, Maroc ; École Centrale Supélec, France

De la bioéthique aux codes éthiques de l'IA : limites et perspectives

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

1. INTRODUCTION

L'éthique du numérique est un domaine de l'éthique appliquée qui émerge de nos jours pour répondre au développement, fulgurant, des nouvelles technologies numériques, et en l'occurrence de l'intelligence artificielle. Comme toute discipline naissante, elle cherche ses marques conceptuelles et institutionnelles pour assurer sa quête de légitimité et sa capacité à faire face aux défis, d'ordre éthique, lancés par ces nouvelles technologies. Aussi, les travaux menés en éthique du numérique cherchent-ils à s'appuyer, pour assurer leurs premiers pas, sur des travaux qui bénéficient déjà d'une certaine reconnaissance institutionnelle et sociale en matière d'éthique et qui ont été établis sur la base d'expériences vécues face à d'autres sources d'urgence. C'est précisément le cas de la bioéthique ou « éthique médicale » qui reste la référence en matière d'éthique appliquée, et c'est pour cette raison qu'elle fait l'objet d'une telle sollicitation pour penser les questions éthiques liées au numérique.

Après avoir précisé, dans une première partie, ce qui est souvent entendu par le terme « éthique », j'étudierai, dans une seconde partie, pourquoi et comment, des actrices et acteurs de l'éthique de l'IA mobilisent la référence à la bioéthique. Ce faisant, j'évoquerai deux exemples de transfert de notions développées en bioéthique qui sont réutilisées de nos jours pour instaurer une éthique de l'IA, à savoir la notion d'autonomie et celle qui se présente sous la forme du couple d'opposés « bienfaisance-non-malfaisance ». Mon objectif est d'interroger la transférabilité de ces notions d'un domaine à un autre de l'éthique, en tenant compte des modalités actuelles d'effectuation de ce transfert. Je montrerai ensuite comment la méthodologie développée aussi bien dans les comités d'éthique autour des questions de bioéthique que dans la pratique hospitalière, par exemple dans le cadre d'un accompagnement en fin de vie proposé à une personne malade par une équipe médicale, pourrait permettre d'aboutir à un accord basé sur des faits tangibles à l'issue d'une démarche de prise de décision éthique concernant les conséquences de l'utilisation d'une intelligence artificielle. Mon but est de montrer comment le savoir-faire acquis en bioéthique, en termes de méthode à suivre pour prendre une décision collective, pourrait servir de base, voire de guide, au développement d'une éthique de l'IA.

À l'heure où l'« inflation du langage », pour parler avec Gilbert Hottois[18], qui apparaît autour des éthiques et des technologies numériques, trouble la réflexion, et ne facilite pas la prise de décision, il faut reprendre l'enquête et étudier, de près, en mobilisant les sciences humaines et sociales, ce à quoi renvoient, en termes techniques et éthiques, ces nouvelles

18- Hottois, Gilbert. L'inflation du langage dans la philosophie contemporaine. Causes formes et limites (Préface de Jean Ladrière, Université libre de Bruxelles, Faculté de philosophie et lettres, LXIX), Éditions de l'Université de Bruxelles, Bruxelles, 1979.

technologies numériques et comment se fait l'appropriation de ces questions éthiques par les milieux que ces questions concernent. L'objectif est de clarifier la situation en vue de faire face aux problèmes posés par ces technologies, tout en remettant à leur juste place les considérations d'ordre idéologique ou fantasmatique qui nous empêchent de concevoir et de développer des solutions, quel que soit le domaine d'exercice et d'application de ces solutions. Comme l'écrit le philosophe Simon Blackburn[19] :

> « Bien traiter les pensées, c'est être capable d'éviter la confusion, de détecter les ambiguïtés, de rassembler ses idées, de formuler des raisonnements fiables, de prendre conscience des alternatives, et ainsi de suite. Bref, on peut comparer nos idées et nos concepts aux lentilles à travers lesquelles nous voyons le monde. En philosophie, le thème d'étude, c'est la lentille elle-même. La réussite ne dépendra pas tant du bagage de connaissances finalement accumulé, que de ce dont on est capable en cas d'avis de tempête : quand les mers de la discussion montent et que la confusion fait irruption. Réussir signifiera prendre au sérieux les implications des idées. »

Commençons par un rappel de ce que signifie le mot « éthique » et de certaines questions et difficultés que l'élucidation de cette notion soulève dans la réflexion philosophique.

2. CONSIDÉRATIONS PRÉLIMINAIRES SUR LA NOTION D'ÉTHIQUE

La morale — du latin « mores », les mœurs — et l'éthique — du grec « Êthos », le caractère — désignent, le plus souvent, les règles qui norment les conduites humaines. La distinction de ces deux notions n'a aucun fondement philologique. Selon l'ouvrage de référence Vocabulaire européen des philosophies, dirigé par Barbara Cassin, ces deux termes se distinguent, d'une langue à une autre, en fonction de deux types de problèmes, à savoir le sujet des conduites — individus ou communauté — ou bien la nature de leur désignation, descriptive, si le terme renvoie à la nature ou à l'histoire, ou prescriptive, s'il dicte ce qu'il faut faire, ce qui est considéré comme bien ou mal[20]. De l'articulation de ces quatre dimensions — individuelle, collective, descriptive, prescriptive —découlerait la différence entre ces deux mots dans les différentes langues.

Michel Foucault a porté son attention sur l'origine grecque, « Êthos », du mot éthique en nous rappelant que ce terme renvoie au lieu de vie habituel d'une personne ainsi qu'aux coutumes et habitudes qui existent en ce lieu, à une époque donnée, et qui constituent un cadre de significations avec lequel la personne peut se sentir ou être en phase ou pas mais qu'elle utilise pour évaluer, en termes de recherche de « vie bonne » ou du « bien agir », ses actes et sa vie. En ce sens, l'« Êthos » serait un caractère particulier appartenant, à la fois, à la communauté, comme par exemple une coutume, et à l'individu, comme le « caractère moral » que ce dernier affiche, ou révèle, dans ses actions et ses propos. La force de l'« Êthos » ne signifie pas, selon Foucault, que tout « Êthos » hérité est juste, mais seulement qu'il s'impose, au nom par exemple de la justice divine ou de celle mise en place par un État-Providence, comme si, effectivement, il était juste. C'est un cadre structurant la pensée et l'imaginaire dans lequel la personne positionne et pense sa vie du point de vue de sa capacité à réaliser, dans ses actions, l'idée de « vie bonne » ou celle de « bien agir », ce qui suppose un va-et-vient entre ces deux idées, les réalisations

19- Blackburn, Simon. Penser, une irrésistible introduction à la philosophie, traduit de l'anglais par Pierre-Emmanuel Dauzat, Éditions Flammarion, Paris, 2003, p. 17.

20- Cassin, Barbara (dir.). Vocabulaire européen des philosophies, édition augmentée, Éditions du Seuil et le Robert, Paris, 2019, en particulier pp. 819820-.

de cette personne et la perception qu'elle peut en avoir dans cet Êthos. C'est pourquoi, selon Foucault, la vie éthique ne consiste pas à suivre strictement les traditions locales, à appliquer des règles, à suivre une démarche localement admise ou à se conformer à une pratique stabilisée pour faire face à un problème impliquant la question de la bonne conduite à tenir. Le devoir éthique appelle, souvent, au contraire, à contester, parfois au nom d'autres valeurs et principes, que ces derniers aient une dimension locale ou universelle, les « habitudes bornées, les traditions répressives, les lois injustes ». Posant que l'Êthos relève d'un choix volontaire, Michel Foucault décrit l'attitude éthique comme « une manière de penser et de sentir, une manière aussi d'agir et de se conduire qui, tout à la fois, marque une appartenance et se présente comme tâche [à accomplir] »[21]. La notion de milieu, même si elle n'est pas ainsi désignée chez Foucault, se révèle ainsi importante pour penser l'éthique car c'est dans un milieu donné, sur la base de ces us et coutumes, qu'un membre de ce milieu pense, sent et doit accomplir une tâche de façon éthique.

Plus spécifiquement, et à partir d'une pensée développée dans le milieu médical, le philosophe et bioéthicien Pierre Le Coz, souligne que, employée comme substantif, l'éthique qualifie une démarche de réflexion — individuelle ou collégiale — suscitée par une situation « moralement complexe », lorsqu'en voulant concrétiser une valeur, nous sommes contraints de renoncer à une autre. Selon cette perspective, l'éthique consisterait, insiste Le Coz, à se demander laquelle de ces valeurs doit prédominer dans une situation donnée ou s'il est possible de les pondérer pour n'en sacrifier aucune. Pour illustrer sa pensée, Le Coz prend l'exemple du mensonge pieux ou « par omission » d'un médecin face à un patient qu'il sait condamné : le médecin dira-t-il à son patient qu'il va bientôt mourir ? La compassion doit-elle atténuer la dureté de sa probité ? Sachant que son patient est en souffrance parce qu'il a de plus perdu sa femme quelques mois auparavant, prendra-t-il la décision de lui dire ce que son IRM et ses analyses biochimiques révèlent ou préfèrera-t-il protéger son patient en ne faisant pas référence à ce diagnostic ? Comme l'écrit Le Coz : « Il n'y a de questionnement éthique que s'il y a perception d'une pluralité de valeurs en tension »[22]. Le milieu de vie de ce médecin, le cadre professionnel, c'est-à-dire son Êthos médical, lui permettront, au moins en partie, d'effectuer ce travail de pondération, de délibération, interne et/ou collective, en vue de décider quoi dire à la personne malade. « Dans la pratique humaine [écrit Foucault], il y a un moment où, en quelque sorte, les évidences se brouillent, les lumières s'éteignent, le soir se fait, et où les gens commencent à s'apercevoir qu'ils agissent à l'aveugle et que, par conséquent, il faut une nouvelle lumière, un nouvel éclairage. Alors voilà qu'un objet apparaît, un objet apparaît comme problème »[23]. Dans ce type de situation « moralement complexe », de valeurs en tension, où un « objet apparaît [désormais] comme problème », une action peut être définie « éthique » tant que celle ou celui qui agit est responsable de ce qu'elle ou il fait, et est capable d'anticiper les conséquences de ses actes ; la question de la responsabilité étant alors renvoyée au cadre juridique du pays d'exercice, voire du cadre juridique international d'exercice d'une fonction dans un projet donné. Au-delà de la question du droit indispensable pour définir la responsabilité d'une personne par rapport à ses actes et propos, cette notion peut aussi renvoyer, comme la rappelle Isabelle Stengers, à « ce que Donna Haraway nomme respons-abilité, [c'est-à-dire à] la capacité à répondre pour et à répondre à, c'est-à-dire à se laisser interroger par, ce

21- Foucault, Michel. Dits et écrits, Vol. II., Éditions Gallimard, Paris, 2001, p. 1384.

22- Le Coz, Pierre. L'éthique médicale. Approches philosophiques, Préface d'André Comte-Sponville, Presses universitaires de Provence, Collection « Sciences Technologies Santé », Aix-en-Provence, 2018, p. 143.

23- Propos recueillis par André Berten à l'Université catholique de Louvain, le 7 mai 1981, publiés dans le livre Berten, André. « Entretien avec Michel Foucault », Les cahiers du GRIF, vol. 37, n° 1, 1988, pp. 920-, spéc. p. 18.

dont nos intentions justifient le sacrifice »[24]. En d'autres termes, la « respons-abilité » est aussi cette capacité à faire une pause, à se laisser intriguer, à rendre problématique ce que nous faisons ou laissons faire, ce à que nous concédons, habituellement, au nom de l'action ou de la nécessité d'agir en fonction du temps dont nous disposons et des objectifs prioritaires à atteindre dans une situation donnée. C'est pour cette raison, au moins, que cette responsabilité associée à l'adoption d'une attitude éthique convoque une démarche réflexive, une réaction, qui ne peut se limiter à suivre procédure déjà faite pour bien agir, comme pourrait le laisser penser la multiplication des chartes et des codes éthiques développés dans les organisations, au sens de la sociologie des organisations. Comme le souligne Jean-Gabriel Ganascia : « D'une éthique qui viserait non à une réflexion, mais à l'imposition de normes infrangibles, échappant à la discussion et au législateur, il conviendrait de se méfier »[25].

Aussi, afin de ne pas réduire l'éthique à une application de normes, en l'occurrence arbitraires, certains philosophes, à l'instar de Paul Ricoeur, pour ne prendre qu'un exemple parmi de nombreux possibles dans l'histoire de la philosophie, mais un exemple qui réalise une synthèse récente mobilisant aussi bien la philosophie analytique et la philosophie continentale que l'histoire de la philosophie, proposent-ils, nonobstant leur proximité étymologique, « de distinguer entre éthique et morale, [et] de réserver le terme d'éthique pour tout le questionnement qui précède l'introduction de l'idée de loi morale et de désigner par morale tout ce qui, dans l'ordre du bien et du mal, se rapporte à des lois, des normes, des impératifs »[26]. En d'autres termes, en suivant cette ligne de démarcation, l'éthique serait davantage un questionnement réflexif sur la valeur de l'action humaine dans des conditions d'application particulières, ce que soulignent, je l'ai signalé, chacun à sa façon, Le Coz et Foucault, tandis que la morale aurait une visée prescriptive, normative, voire universelle ou du moins universalisante, comme dans le cas pour l'impératif catégorique kantien qui impose à chacune et chacun d'entre nous de traiter l'humanité dans notre personne et dans celle d'autrui comme une fin en soi et non pas uniquement comme un moyen en vue d'obtenir quelque chose de cette autre personne ou de nous-même lorsque nous taisons, volontairement, en nous des principes ou des convictions qui, sans cela, nous empêcheraient d'assouvir des désirs ou d'avoir la sécurité d'un emploi, la tranquillité acquise par celle ou celui qui ne contredit jamais un supérieur au comportement humiliant, obscène et totalement non légitime. Alain Etchegoyen résume cette distinction entre éthique et morale par la formule : « La morale commande. L'éthique recommande »[27]. La morale poserait ainsi la question « Que dois-je faire ? » alors que l'éthique poserait la question « Comment dois-je vivre ? ».

Dans cette perspective, Ricoeur appelle « visée éthique » : « la visée de la «vie bonne» avec et pour autrui dans des institutions justes »[28] en ayant pour but d'établir « la primauté de l'éthique sur la morale, c'est-à-dire de la visée sur la norme »[29] en gardant à l'esprit la « nécessité pour la visée éthique de passer par le crible de la norme et (...) la légitimité d'un recours de la norme à la visée lorsque la norme conduit à des impasses pratiques »[30], comme celle décrite par Le Coz dans le cas du mensonge pieux. Pour Ricoeur,

24- Stengers, Isabelle. Réactiver le sens commun. Lecture de Whitehead en temps de débâcle, Les empêcheurs de penser en rond, Éditions La Découverte, Paris, 2020, p. 143.
25- Ganascia, Jean-Gabriel. Servitudes virtuelles, Édition du Seuil, Paris, 2022, p. 167.
26- Ricœur, Paul. « Fondements de l'éthique », Autres Temps. Les cahiers du christianisme social, n°3, 1984, pp. 6171-, spéc. p. 61.
27- Etchegoyen, Alain. La valse des éthiques, Éditions Pocket, Paris, 1991, p. 79.
28- Ricoeur, Paul. Soi-même comme un autre, Éditions du Seuil, Paris, 1990, p. 202.
29- Ibid.
30- Ibid., p. 200.

« nous pouvons dire, en conclusion, que le formalisme en éthique définit la moralité. Mais l'éthique a une ambition plus vaste, celle de reconstruire tous les intermédiaires entre la liberté, qui est le point de départ, et la loi qui est le point d'arrivée »[31]. Pour ce faire, précise Ricoeur, il ne faut pas perdre de vue « que le vivre-bien ne se limite pas aux relations interpersonnelles, mais s'étend à la vie des institutions (...) [dans la mesure où] la justice présente des traits éthiques qui ne sont pas contenus dans la sollicitude [qui constitue pour Ricoeur une composante incontournable de l'éthique face à la souffrance de l'autre], à savoir pour l'essentiel une exigence d'égalité »[32]. Les institutions appartiennent à l'Êthos, point de convergence, ténu, en raison de lignes argumentatives très différentes, mais bel et bien présent, entre Ricoeur et Foucault. La notion de liberté est la clé de voûte de cette démarcation comme cela était déjà le cas, dans une certaine mesure, dans l'anthropologie de Rousseau pour fonder la morale ; anthropologie de laquelle est parti Kant pour écrire sa critique de la raison pratique.

D'autres philosophes pour lesquels la distinction éthique/morale n'importe pas, comme récemment Laurence Devillairs, restent, dans le cadre d'autres contextes philosophiques et ontologiques, cependant attachés à la défense de l'idée de liberté pour fonder l'action bonne, Devillairs écrit : « Morale ou éthique, peu importe : ce que nous souhaiterions montrer, c'est que, contre toute attente, agir moralement libère, révèle une capacité à s'arracher à tout ce qui incline et détermine. (...) Le bien n'est pas une contrainte, c'est une décision. Il doit venir d'une volonté indépendante, que rien n'empêche ni ne préfigure. (...). Un acte moral ne peut qu'être un acte libre ; devenir quelqu'un de bien ne peut être que la vocation d'hommes libres »[33]. Devillairs rajoute, en usant, tout à tour, de façon équivalente, des termes morale et éthique, mais en gardant, substantiellement, la liberté comme clé de voûte de son édifice :

« L'acte moral me découvre que je suis capable de faire davantage que mon possible, que je peux faire du neuf, que je suis pour ainsi dire, par mon acte, créateur d'avenir. En d'autres termes, ce n'est pas parce que c'est possible que c'est courageux, c'est parce que c'est courageux que cela devient possible : c'est le courage qui fait advenir une possibilité qui aurait été sans cela inexistante. (...) L'éthique n'est pas une question de discipline mais de liberté : bien agir, ce n'est pas conforter ce qui est mais faire surgir ce qui n'est pas encore. C'est imposer à l'immobilité du «c'est comme ça» une révolution, celle du bien. »[34]

De la même façon, mais pour d'autres raisons, le philosophe Ruwen Ogien ne fait pas de distinction entre éthique et morale et considère ces termes comme étant des synonymes qui renvoient à des registres argumentatifs articulant normes et valeurs dans le but de justifier publiquement nos actions. Il distingue, en revanche, l'éthique normative (qui est prescriptive, au sens où ses conclusions nous dictent les actions requises en fonction de ce qui est bien ou mal, juste ou injuste) de la méta-éthique (qui est descriptive au sens où son objet d'étude est davantage centré sur une réflexion portant sur les concepts de l'éthique, leurs origines et leurs rôles), et de l'éthique appliquée (qui leur seraient, en principe, selon cet auteur, subordonnée)[35]. Selon Ruwen Ogien et Christine Tappolet, les valeurs renverraient aux jugements dits « axiologiques » ou évaluatifs et s'appliqueraient à la désignation de ce qui est bien ou mal, désirable ou non. Les normes, quant à elles, renverraient aux jugements dits « déontiques » ou normatifs

31- Ricœur, Paul. « Fondements de l'éthique », op. cit., p. 71.

32- Ricoeur, Paul. Soi-même comme un autre, op. cit., p. 227. L'usage de l'italique marque l'insistance de Ricoeur sur le mot « égalité ».

33- Devillairs, Florence. Être quelqu'un de bien. Philosophie du bien et du mal, Presses Universitaires de France, Paris, 6ème tirage, 2022, p. 32.

34- Ibid., pp. 199200-.

35-Ruwen, Ogien. L'éthique aujourd'hui: maximalistes et minimalistes, Éditions Gallimard, Paris, 2017.

et s'appliqueraient aux actions, c'est-à-dire à ce qui est interdit, obligatoire ou permis. Ces deux auteurs montrent, en tout cas quoi qu'il en soit tel est leur objectif, que les normes sont généralement fondées ou justifiées par les valeurs, mais que l'inverse n'est pas vrai. Pour cela, ils déploient une revue de la littérature en éthique afin de montrer que la seule possibilité qu'ont les normes de ne pas paraître arbitraires est précisément de les faire reposer sur des valeurs car celles-ci peuvent être plus aisément identifiées. Toujours selon ces auteurs, il existerait entre normes et valeurs des relations de justification et de fondation à plusieurs niveaux. En premier lieu, au niveau ontologique, un principe moral est considéré comme vrai si les actions qui s'y conforment réalisent certaines valeurs. Ensuite au niveau pratique, les raisons que l'on peut avoir de se conformer à une norme renverraient, la plupart du temps, toujours selon ces auteurs, à des valeurs. Par exemple, s'il faut s'abstenir de mettre autrui en danger, c'est en raison du bien qui en résulterait[36]. Cette position leur permet d'argumenter en faveur du fait qu'il n'y a aucune raison décisive de ne pas être « conséquentialiste » et que, sur le plan logique autant que pratique, cette manière d'agencer et de relier entre elles normes et valeurs présente des avantages avec lesquels les « théories concurrentes », comme la déontologie et l'éthique des vertus, ne pourraient pas rivaliser. Rappelons, brièvement, que, pour Ogien et Tappolet : « en simplifiant, on pourrait dire que, selon le conséquentialisme, l'action qu'il faut accomplir est celle qui promeut le bien : ses conséquences doivent être les meilleures possibles. Pour les partisans de la déontologie, ce que nous devons faire est déterminé par un certain nombre de principes (…) qu'il nous faut suivre quelles que soient les conséquences. L'éthique de la vertu, de son côté, nous demande de tourner notre regard vers le caractère de l'agent : l'action qu'il nous faut accomplir est celle qu'un agent vertueux accomplirait »[37]. Puisque, selon Ogien et Tappolet, il « semble difficile de se passer de valeurs ou de propositions évaluatives dans la justification des normes ou des propositions normatives »[38] alors le conséquentialisme présenterait un avantage philosophique et logique par rapport à la déontologie dans la mesure où sa structure logique repose sur l'idée que les normes doivent être justifiées par les valeurs, tandis que l'approche déontologique est bâtie sur la relation inverse.

Ces discussions sont stimulantes sur le plan intellectuel et montrent, à quel point, la signification de ce à quoi renvoie l'éthique est toujours un sujet de recherche qui alimente de nombreux débats[39] ; situation qui permet de prendre du recul par rapport à celles et ceux qui, dans le monde entrepreneurial ou, plus largement, dans la plupart des domaines de la vie en société, ne cessent de parler d'éthique comme si l'affaire de sa définition était réglée, comme si pour être éthique, il fallait se contenter de suivre leurs « bonnes » recommandations… Attitude d'autant plus déplacée que l'éthique, dans ses formes débattues, est essentiellement issue d'un nombre réduit de pays et n'intègre pas la diversité culturelle qui existe dans le monde de nos jours et offre peu, en dehors de quelques congrès internationaux de recherche ou sommets d'organismes internationaux ou d'ONG, la possibilité de « concerter les civilisations », pour reprendre une partie du titre du livre récent écrit en l'honneur d'Alain Supiot[40], en vue d'approfondir et renouveler la pensée éthique en mobilisant les cultures du monde dans un geste de co-suscitation créateur. Pour finir, cette présentation concise de ce que peut signifier aujourd'hui une démarche éthique, je souhaiterais revenir sur la question de l'arbitraire et de

36- Ogien, Ruwen & Tappolet, Christine. Les concepts de l'éthique : faut-il être conséquentialiste ?, Éditions Hermann, Collection « L'avocat du diable », Paris, 2009.

37- Ibid., p. 130.

38- Ibid., p. 200.

39- Desmons, Ophélie, Lemaire, Stéphane & Turmel Patrick (dir.). Manuel de métaéthique, Éditions Hermann, Collection « L'avocat du diable », Paris, 2019.

40- Besson, Samantha & Jubé, Samuel (dir.). Concerter les civilisations. Mélanges en l'honneur d'Alain Supiot, Éditions du Seuil et du Collège de France, Paris, 2020.

l'objectif poursuivi qui consiste à échapper à l'arbitraire pour fonder, en fonction des autrices et les auteurs, la morale, l'éthique, les valeurs ou bien les normes.

« Si [écrit Wittgenstein] le vrai est ce qui est fondé, alors le fondement n'est pas vrai, ou faux »[41]. En effet, si nous affirmions le contraire, il y aurait circularité de la pensée. Plus loin, Wittgenstein poursuit :

> « Les questions que nous posons et nos doutes reposent sur le fait que certaines propositions sont soustraites au doute —sont, pour ainsi dire, comme des gonds sur lesquels tournent nos questions et nos doutes. C'est-à-dire : il appartient à la logique de nos investigations scientifiques que certaines choses ne soient en fait pas mises en doute. Mais cela ne revient pas à dire que nous ne pouvons pas tout vérifier et sommes donc obligés de nous contenter de présuppositions. Si je veux que la porte tourne, il faut que les gonds restent fixes. »[42]

Avant d'aboutir à la clarification suivante :

> « Je veux dire : Notre apprentissage a la forme «Ceci est une violette», «Ceci est une table». Il est vrai qu'un enfant pourrait entendre le mot «violette» pour la première fois dans la phrase «C'est peut-être une violette» ; et pourrait alors demander : «C'est quoi une violette ?» Et bien sûr qu'on pourrait lui répondre en lui montrant une image. Mais si ça n'était toujours qu'en lui montrant l'image qu'on lui disait : «Ceci est une...», et qu'autrement, on lui dit toujours : «Peut-être cela est-il...» — Quelles en seraient les conséquences pratiques ? Un doute qui mettrait tout en doute ne serait pas un doute.»[43]

Au-delà du vrai et du faux, le « fondement » repose sur des certitudes qui ne peuvent être démontrées mais sans lesquelles toute possibilité de communication avec autrui, toute possibilité de donner du sens à nos vies, toute action pratique seraient tout simplement niées. Un doute radical n'est pas un doute car pour douter il faut avoir une base de repères, un gond, fixe, même si ce gond peut évoluer avec le temps, le domaine de pratique considéré et lieu du monde concerné. Dans le sillon de cette pensée et du prolongement qu'en a proposé Elizabeth Anscombes[44], Vincent Descombes conçoit, dans un livre, remarquable, fédérant plusieurs essais de philosophie pratique et portant le titre Le raisonnement de l'ours, la raison pratique comme une capacité à déterminer l'action à accomplir par le truchement d'une délibération pondérée. Pour ce faire, il revient sur la distinction entre la morale du devoir et l'éthique du bien vivre. Descombes établit que pour satisfaire à l'exigence d'une fondation rationnelle de la morale, des philosophes ont proposé d'isoler un noyau de normes « morales » qui seraient valides de façon inconditionnelle quel que soit le contexte et d'abandonner les questions de valeur à l'appréciation « éthique » de chacun[45]. Mais, poursuit-il, pourquoi alors faut-il être moral ? Descombes écrit, avec un style tout wittgensteinien d'écriture :

> « Si vous dites : «Est immoral

41- Wittgenstein, Ludwig. De la Certitude, traduit par Danièle Moyal-Sharrock, Éditions Gallimard, Bibliothèque de Philosophie, Paris, 2006, § 205.

42- Ibid., § 341343-, l'italique est celle de Wittgensetin.

43- Ibid., § 450.

44- Anscombre, Elizabeth. L'intention, traduit de l'anglais par Mathieu Maurice et Cyrille Michon, préface de Vincent Descombes, Éditions Gallimard, Paris, 2002 [version originale en anglais, Intention, 1957].

45- Descombes, Vincent. Le raisonnement de l'ours et autres essais de philosophie pratiques, Éditions du Seuil, collection « La couleur des idées », Paris, 2007. L'usage de l'italique est celui de Vincent Descombes.

> ce qui est interdit par la loi morale», vous ne pouvez pas dire en même temps : «Si la loi morale interdit ce qui est interdit, c'est parce que ce serait immoral de l'autoriser.» Il faut choisir entre justifier une prescription particulière par les exigences de la morale —en montrant qu'elle fait partie de ce qui est nécessaire à notre moralité, lequel est du même coup présupposé et non pas fondé en raison, contrairement à ce qui a été annoncé —, ou bien fonder la morale elle-même. Dans ce dernier cas, on devrait chercher ce fondement de la morale en dehors de la morale, ou bien alors s'apercevoir qu'on tourne en rond, par un cercle du raisonnement qui, cette fois, est vicieux. Car il apparaît que le fondement de la morale doit déjà être moral : si en effet nous prétendons donner à la loi morale un fondement pris en dehors de la morale, ce sera pour des raisons étrangères à la morale que nous exigeons du sujet une conduite morale ce qui privera la morale de sa dignité (on sera moral par intérêt, ou pour être aimé). » [46]

Par voie de conséquence, si une fondation philosophique de la morale devait consister à expliquer à quelqu'un pourquoi il « doit être moral », l'entreprise se révèlerait incohérente ! Descombes établit ensuite qu'il n'y a pas de « tu dois » — faire ceci ou cela — qui puisse valoir catégoriquement, dans la mesure où un impératif du type « tu dois agir ainsi » équivaut, d'un point de vue logique, à une disjonction du type « ou bien tu agis ainsi, ou bien... », de sorte que, insiste Descombes, l'énoncé demande à être complété en précisant l'inconvénient qu'indique le second membre de la disjonction. Pour conclure son raisonnement, Descombes interroge ensuite l'ordre de la moralité en montrant qu'il est un ordre de justice et en interrogeant la notion de « nécessité pratique » qui est au cœur de la moralité. Comment s'y prend-il ? En montrant que toute morale comporte deux points de vue complémentaires à partir desquels il devient possible d'apprécier la conduite et les choix d'une personne (soi ou autrui). Il écrit :

> « On fait comme si l'éthique était une affaire personnelle et la morale une affaire intersubjective. La division du champ moral nous est présentée ainsi : quand nous sommes dans l'éthique, nous manifestons quelles sont nos valeurs propres ; en revanche, quand nous sommes dans la morale, nous nous demandons quelles doivent être les normes communes à toutes les personnes susceptibles d'entrer en rapport les unes avec les autres. L'éthique est forcément particulière à un sujet qui pose ses valeurs. Sans doute peut-on parler de valeurs collectives, mais cela revient à se donner pour sujet un groupe agissant au fond comme une personne particulière. Des valeurs collectives restent des idéaux ou les préférences d'un groupe particulier [ce que j'ai signalé plus haut dans ce chapitre comme Êthos tel qu'il est analysé par Foucault] : elles n'ont pas l'universalité de plein droit qui doit caractériser les normes morales. Je crois que c'est là une fausse opposition. Il est vrai que la morale (au sens traditionnel indifférencié) parle tantôt de nos devoirs, tantôt des qualités qu'il faut acquérir pour être meilleur, pour atteindre un certain

46- Ibid., p. 427. L'insistance est celle de Vincent Descombes.

> idéal qu'on s'est fixé. Mais ce sont là seulement deux façons de présenter le même contenu, deux points de vue complémentaires sur une seule et même réalité. Les deux points de vue sont ceux de la responsabilité à l'égard de soi et la responsabilité à l'égard d'autrui. Quand la morale s'exprime dans un rappel des devoirs, elle prend sur l'être humain le point de vue d'autrui : quelles sont les qualités que nous désirons trouver chez nos compagnons, nos parents, nos voisins, nos concitoyens, etc. ? Quand la morale s'exprime dans un portrait de l'homme de bien, de l'homme estimable, du bon citoyen, etc., elle prend sur l'être humain le point de vue qui devrait être le sien, qui est le sien dès lors qu'il se considère comme appartenant à un ordre de moralité. Le premier point de vue est bien celui des devoirs qui sont imposés à chacun, mais ce n'est pas pour autant un point de vue étranger à une interrogation sur le bien et les valeurs. C'est en effet par l'une des vertus morales ou, si l'on veut, par une valeur que ce point de vue sera défini. Ce point de vue est celui de la justice, c'est-à-dire de la vertu générale qui consiste à s'occuper du bien d'autrui en tant qu'il dépend de moi, qu'il est entre mes mains. »[47]

Pour illustrer son propos, Vincent Descombes considère le métier d'éducateur. Ce métier d'éducateur, insiste-t-il, paraît moralement impossible aussitôt qu'on perd de vue le fait que le besoin pratique d'inculquer une moralité aux élèves pour les éduquer suffit à rendre rationnelle cette inculcation, il correspond à une nécessité pratique. L'éducateur doit transmettre sur un mode d'autorité une vérité sur la nature humaine, ses biens et ses fins, qu'il ne détient pas. À cette difficulté, une solution que rejette Descombes consisterait à se replier sur un noyau dur de normes minimales nécessaires à une coexistence entre personnes rationnelles, une sorte de « droit naturel ». Or de telles règles ne pourraient être absolues qu'en étant négatives en précisant ce qu'il faut éviter de faire pour éviter de nuire à autrui sans préciser comment s'occuper de lui quand il en exprime le besoin. Ainsi, le fondement dont a besoin un éducateur n'est pas un fondement absolu de son action, le seul fondement « est à chercher dans la nécessité pratique : étant admis (par hypothèse) qu'il est bon de recevoir une éducation, tout ce qui est pratiquement nécessaire à la pratique éducative se trouve par là même aussi bien justifié qu'il est nécessaire »[48]. Cette conclusion rejoint celles tirées par Wittgenstein, à propos des règles conventionnelles, dans ses leçons sur le fondement des mathématiques[49] : il ne faut pas confondre la nécessité dans le système et la nécessité du système lui-même. La règle du jeu rend nécessaires certaines opérations, mais ces nécessités n'existent pas en dehors du jeu. Nous ne pouvons pas nous passer de ces règles, comme celle du « 2 + 2 = 4 », car sinon tout le système s'effondre, de la même façon qu'il n'est pas possible de douter de tout, de faire sauter tous les gonds autour desquels notre vie prend sens, et reste possible, d'un point de vue pratique. En ce sens ces règles, bien qu'elles soient conventionnelles et donc au moins en partie arbitraires, n'excluent en rien leur ancrage dans nos nécessités pratiques sans lesquelles nos formes de vie seraient impossibles. En poursuivant son analyse, Vincent Descombes conclut :

> (...) Plaçons-nous, encore une fois, du point de vue de l'éducateur. La

47- Ibid., pp. 430431-.

48- Ibid., p. 434.

49- Wittgenstein, Ludwig. Lectures on the Foundations of Mathematics, The University of Chicago Press, Chicago, Chicago, 1975.

> complémentarité des deux points de vue (l'un «éthique» et l'autre «moral») se révèle aussitôt. Les qualités qu'un éducateur juge désirables chez l'être humain (par exemple chez lui-même) sont justement celles qu'il voudra inculquer aux enfants dont il s'occupe. Pourquoi telle ou telle qualité sont-elles désirables ? Elles le sont du point de vue du sujet moral lui-même, mais ce sujet ne le découvre ou ne le comprend qu'après avoir été éduqué. Elles le sont du point de vue de la communauté, dont l'éducateur est en quelque sorte le représentant auprès de ses élèves, en tant qu'il est préférable pour chacun de nous d'avoir affaire à des gens qui possèdent la vertu générale de justice, donc qui se conduisent comme on est en droit de l'attendre d'eux dans notre ordre de moralité. »[50]

En conclusion, l'éthique est soit séparée de la morale, soit considérée comme un terme synonyme de celui de morale par les philosophes et les spécialistes d'éthique, pour des raisons ontologiques, idéologiques, épistémologiques, religieuses ou pratiques. La complémentarité des points de vue que signale Vincent Descombes entre éthique et morale montre bien comme l'écrit cet auteur lui-même que « la difficulté en question est plutôt un nœud à défaire qu'un problème réel à résoudre »[51]. Et l'étude que Descombes mène, au niveau du langage, étude de philosophie analytique donc, est particulièrement éclairante pour comprendre la coexistence de la responsabilité à l'égard de soi et de la responsabilité à l'égard d'autrui pour penser l'activité réflexive, délibérative qu'est l'éthique. Bref, comme l'écrit Blackburn dans le passage auquel nous nous sommes référés dans l'introduction de ce chapitre : « Bien traiter les pensées, c'est être capable d'éviter la confusion, de détecter les ambiguïtés, de rassembler ses idées, de formuler des raisonnements fiables, de prendre conscience des alternatives, et ainsi de suite »[52]. Cette première partie visant à préciser ce à quoi renvoie l'éthique et certains débats dont son élucidation fait l'objet étant terminée, je vais à présent contextualiser son implication à propos de l'intelligence artificielle et étudier les raisons, les limites et les perspectives ouvertes par l'application de notions et de méthodes issues de la bioéthique au domaine émergeant de l'éthique de l'intelligence artificielle.

50- Ibid., pp. 431432-.

51- Ibid., p. 433.

3. DE LA BIOÉTHIQUE À L'ÉTHIQUE DE L'IA : RAISONS, LIMITES ET PERSPECTIVES

L'intelligence artificielle est une opportunité pour l'humanité puisqu'elle rend possibles des avancées spectaculaires en matière de communication, de développement de l'économie numérique, d'adaptation de l'éducation —« e-learning » —, de conception et de mise en place de villes intelligentes — « smart cities » — dans le but de réduire l'impact environnemental en réduisant le gaspillage d'énergie et d'eau, de concevoir une nouvelle agriculture — l'« agriculture 2.0 » —, de permettre une digitalisation de l'entreprise — processus subsumé sous le vocable « industrie 4.0 » —, de mettre en place une médecine personnalisée ou un diagnostic médical à distance, entre autres perspectives hautement prometteuses. Pour autant se posent de nombreux problèmes. Ceux par exemple liés à la protection des données de santé ou concernant l'identité bancaire, les opinions politiques, le type de sexualité d'un individu, à la question des discriminations et des biais que peuvent entraîner l'utilisation de certaines bases de données non suffisamment représentatives de la diversité humaine, à la question de la responsabilité des véhicules ou des armes dites « autonomes », aux problèmes soulevés par le « transhumanisme », aux dommages

52- Blackburn, Simon. Penser, une irrésistible introduction à la philosophie, op. cit. [Note 2], p. 17.

et impacts négatifs pour le développement durable en raison non seulement de l'énergie nécessaire pour le fonctionnement quotidien de ces technologies numériques à l'échelle mondiale mais aussi pour l'extraction des matériaux utilisés, encore massivement non recyclés, pour rendre cette technologie possible avec tous les saccages géologiques et territoriaux que cette exploitation des « ressources » génère, etc. Tous ces problèmes sont autant de sujets, parmi tant d'autres, qui sont à l'origine de la multiplication des questions éthiques soulevées par le développement des technologies numériques, et, en l'occurrence, de l'intelligence artificielle. Le défi de la réflexion éthique est, par voie de conséquence, de développer des réflexions communes, globales et locales, permettant de définir des équilibrages en faveur du bien commun destinés à tenir compte des besoins humains, sociétaux, culturels et environnementaux. Comme l'écrivent Michal Krzykawski et Susanna Lindberg :

> « L'enjeu dont il faut s'emparer à l'époque des algorithmes et de l'automatisation généralisée est de déterminer les conditions dans lesquelles les actes éthiques sont possibles à travers des pratiques quotidiennes à des échelles plus ou moins locales et sur la base d'une économie contributive territorialisée et reposant sur la valorisation éthique, pratique et économique de savoirs plus ou moins locaux (de la nano-économie domestique aux savoirs scientifiques biosphériquement situés)»[53]

Pour faire face à ces défis éthiques, les grands acteurs de l'Internet — Google, Facebook —, des sociétés savantes — comme IEEE[54] —, des universités — par exemple Montréal —, l'Union Européenne, le Parlement européen, entre autres acteurs, réunissent, de nos jours, des comités d'experts qui émettent des avis, rédigent des rapports, des chartes, des règlements, des codes éthiques en libre accès sur Internet. L'observatoire de l'intelligence artificielle — « European Parliament's Artificial Intelligence Observatory » — est un acteur important de cette démarche au niveau européen. Le groupe d'experts dit « de haut niveau » mis en place par l'UE en 2018 sur l'éthique et la politique de l'IA utilise, depuis 2019, quatre principes : 1) le respect de la vie humaine ; 2) la prévention du mal ; 3) l'équité ; 4) l'explicabilité[55]. De la même façon, le rapport « AI4People » de l'Atomium — Institut européen pour la science, les médias et la démocratie —fonde, sans justification, ses recommandations éthiques sur les quatre grands principes de la bioéthique : (1) l'autonomie de la personne ; (2) la non-malfaisance ; (3) la bienfaisance ; (4) la justice. Il rajoute un cinquième principe d'explication et de transparence[56].

53- Krzykawski, Michal & Lindberg, Susanna. « Êthos et technologies », in Bifurquer. Il n'y a pas d'alternative, sous la direction de Bernard Stiegler, Éditions Les liens qui libèrent, Paris, 2020, pp. 263297-, spéc. p. 297.

54- L'« Institute of Electrical and Electronics Engineers » ou « IEEE » est une association professionnelle de droit américain, regroupant la plus grande organisation professionnelle dans le monde dédiée aux technologies avancées, avec plus de 430 000 membres, dont plus de 120 000 étudiants, dans plus de 160 pays dans le monde. Son étude intitulé « IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems », connue aussi sous le nom « IEEE Global Initiative », synthétise les regards croisés de leaders des communautés de l'intelligence artificielle et des systèmes autonomes en matière d'éthique.

55- Les documents produits sont en libre accès et accessibles en cliquant sur le lien suivant : Groupe d'experts de haut niveau sur l'intelligence artificielle | Bâtir l'avenir numérique de l'Europe (europa.eu), date de dernière consultation, le 20 novembre 2022.

56- Floridi, Luciano et al., « AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations », Minds and Machines, 28, 2018, p. 689707-. L'ensemble des documents produits par cet institut sont en libre accès et accessibles en cliquant sur le lien : AI4People | Atomium-EISMD, date de dernière consultation, le 20 novembre 2022.

Par ailleurs, tous ces travaux font référence, explicitement ou pas, à l'expérience la plus discutée en éthique appliquée, c'est-à-dire, pour les raisons que j'ai signalées au début de ce chapitre, à la bioéthique, en l'occurrence au rapport Belmont —présenté en 1978, publié en 1979 —intitulé « Ethical Principles and Guidelines for the Protection of Human Subjects of Research » — « Principes éthiques et lignes de conduite pour la protection des sujets humains dans le cadre de la recherche » —[57]. Or la transférabilité des principes du domaine médical à celui des technologies numériques s'avère, à bien des égards, problématique et insatisfaisante, comme le montre, en particulier, Jean-Gabriel Ganascia[58]. C'est ici que doit commencer l'enquête historique, philosophique, épistémologique, bref qu'entrent en scène les sciences humaines et sociales, pour discerner, clarifier, formuler le questionnement, réfléchir et participer à l'invention de solutions et au débat éthique. Je propose de montrer, succinctement, l'apport de l'entrée en scène des sciences humaines et sociales à propos des notions centrales de la bioéthique que sont l'autonomie, la bienfaisance et la non-malfaisance qui sont remobilisées de nos jours en éthique de l'intelligence artificielle.

Le principe d'autonomie réclame, en bioéthique, que le patient puisse participer au processus de délibération, qu'il ait, ce faisant, la possibilité d'être soigné en accord avec sa pensée, avec ses croyances, ses valeurs et que sa décision prévale, dès lors qu'il est pleinement en possession de tous ses moyens dans la prise de décision[59]. Comme l'écrit Pierre Le Coz :

> « Tandis que la liberté peut désigner la spontanéité, l'absence de contrainte, le bon vouloir ou le libre-arbitre, l'autonomie comporte une dimension rationnelle d'autorégulation qui affleure déjà dans l'étymologie (autos : moi, nomos : loi). L'autonomie est tout sauf une liberté aveugle ou arbitraire. Être autonome, c'est suivre une règle d'action. Un patient est « libre » de ne pas prendre un traitement qui pourrait améliorer sa santé, mais il n'est «autonome» stricto sensu que lorsqu'il suit sa posologie avec régularité. Outre sa stabilité dans le temps, une décision autonome se reconnaît au fait qu'elle a été débattue préalablement sur la base d'informations fiables et exhaustives. Le principe d'autonomie n'est mis en œuvre par le praticien qu'à partir du moment où il informe loyalement le patient. »[60]

Est autonome ce qui, pour résumer, se donne ses propres lois. L'autonomie admet, toutefois, différents degrés et niveaux de signification. Dans son avis intitulé « Refus du soin et autonomie de la personne », le Comité Consultatif National d'Éthique — CCNE —distingue l'autonomie de se déplacer dans l'espace au moyen des forces dont le corps du patient est capable sans assistance, il s'agit d'une « autonomie d'action ». Un deuxième type d'autonomie concerne « l'autonomie de pensée » qui réside dans le pouvoir de conduire une argumentation cohérente et réfléchie, ce qui, dans le cadre médical, suppose que le patient est capable d'exercer son esprit critique à l'égard des informations médicales qui lui ont été communiquées. Enfin, le CCNE évoque « l'autonomie de volonté » qui exprime la faculté de trancher une alternative, de faire des choix, de prendre des initiatives.

57- Ce rapport est en libre accès et accessible en cliquant, par exemple, sur le lien : The Belmont Report | HHS.gov, date de dernière consultation, le 20 novembre 2022.

58- Ganascia, Jean-Gabriel. Servitudes virtuelles, op. cit., note en bas de page [8].

59- Comité Consultatif National d'Éthique, Avis n° 87 sur « Refus de traitement et autonomie de la personne », le 14 avril 2005, www.ccne-ethique.fr, date de dernière consultation le 21 novembre 2022.

60- Le Coz, Pierre. L'éthique médicale. Approches philosophiques, op. cit., p. 54. L'insistance est marquée par Le Coz en usant de l'italique.

Elle correspond au pouvoir de se décider en fonction d'une résolution consciente et personnalisée. Elle engage la capacité du patient à élaborer un projet de soin dans le temps, en l'occurrence sur le long terme.

Par abus de langage, des machines et des algorithmes sont dites « autonomes ». Cette attribution est, comme le montre très clairement Jean-Gabriel Ganascia, le fruit d'une confusion entre les termes « autonome » et « automatique »[61]. L'entraînement des algorithmes est automatique et non pas « autonome ». En effet, une étude épistémologique menée au plus près des praticiennes et praticiens qui conçoivent et développent ces dispositifs numériques permet de montrer qu'il n'y aucune « autonomie » dans le processus car les exemples utilisés par l'algorithme sont labélisés par un utilisateur, un humain, qui décide, en outre, des critères de convergence et de la valeur vers laquelle doit tendre l'algorithme, même si la paramétrisation de la fonction finale reste assez mal comprise puisqu'il n'est pas humainement possible d'examiner, exhaustivement, l'ensemble de ces paramètres[62]. L'algorithme est « automatique » au sens où il peut réaliser une série d'étapes allant de la prise d'information à l'action, sans intervention humaine, mais il n'est pas « autonome » au sens où il fait uniquement ce pourquoi il a été conçu. De tels algorithmes ont permis de faire des avancées technologiques que nous estimions encore lointaines il y a peu, en nous privant toutefois d'une compréhension mathématisée formelle de ce qui est réalisé par la machine[63].

Dans son livre Servitudes virtuelles, Ganascia prend l'exemple de la voiture « autonome ». Cette dernière détermine, à partir de l'objectif qui lui a été fixé — comme aller par exemple à la piscine — et des informations qu'elle a glanées par l'intermédiaire des différents capteurs dont elle est équipée, la séquence des actions qui conduiront les passagers à bon port, alors qu'une voiture « autonome », au sens où ce mot est utilisé pour un patient en médecine, risquerait de remettre en cause l'objectif, se refusant à aller à la piscine, prétextant par exemple que le parking n'y est pas commode ou qu'il vaut mieux aller à la bibliothèque rendre un livre dans les temps ! Ganascia prend également l'exemple des armes dites « autonomes » en précisant qu'au sens de l'autonomie du patient, elles n'obéiraient pas aux ordres et choisiraient elles-mêmes leurs cibles alors qu'au sens d'automatique elles exécuteraient d'elles-mêmes les objectifs qui leur auraient été donnés, par exemple atteindre tout ce qui a une signature radar bien caractéristique. À celles et ceux qui annoncent le remplacement de l'être humain par les machines autonomes, Ganascia oppose la nuance et la rigueur de l'enquêteur et du connaisseur. Il montre que « la vraie révolution en «autonomie» que propose l'IA ne consiste pas à fabriquer des dispositifs matériels qui se substitueraient à nous, du fait de leur supposée autonomie de comportement, mais à concevoir des automatismes qui prendraient le relais, en cas de défaillance humaine causée soit par une situation physique éprouvante, soit par une saturation des facultés d'attention, nous aidant ainsi à demeurer fidèles à nos engagements en dépit de nos absences passagères, et donc à rester plus «autonomes» [nous, humains] »[64]. Aussi,

61- Ganascia, Jean-Gabriel. Servitudes virtuelles, op. cit., spéc. pp. 169178-.

62- Llored, Jean-Pierre & Bouchnita Anass. « IA et santé : chimères, accomplissements et perspectives », in Numérique, Droit et Société, Bevière-Boyer B. et Dibie D. (dir.), Éditions Dalloz, Paris, 2022, pp. 119137- ; Llored, Jean-Pierre & Bouchnita, Anass. « Comment Instaurer la Médecine Personnalisée ? Pour une Alliance entre Intelligence Artificielle et Modélisation Conceptuelle », Ethique et numérique, vol.1, No1, 2022, pp. 154180- ; Bouchnita, Anass & Llored, Jean-Pierre. « Les IA comme outils au service de la santé : Limites et perspectives », Droit, santé et société, n°2, 2021, pp. 3649-.

63- Delahaye, J.-P. (2018). « Intelligences artificielles : un apprentissage pas si profond », Pour la science, n°488, pp. 3544-.

64- Ganascia, Jean-Gabriel. « La révolution de l'Intelligence artificielle (IA) en autonomie », Les cahiers de la revue de la défense, Hors-série « Autonomie et létalité en robotique militaire », 2018, pp. 3642-, spéc. p. 42.

Ganascia nuance-t-il davantage :

> « En somme, et contrairement à ce que d'aucuns laissent accroire, l'automatisation, que l'on qualifie aussi métaphoriquement d'autonomie, ne diminue en rien l'autonomie humaine, du moins, elle n'assujettit pas l'homme à une volonté étrangère à lui. De ce point de vue, les recommandations qui privilégient une intelligence artificielle centrée sur "l'humain" (human-centric approach), et condamnent une intelligence artificielle centrée sur les machines, à savoir sur la satisfaction de leurs supposées volontés, n'ont pas d'objet. »[65]

Par ailleurs Ganascia montre que la notion d'autonomie est associée dans l'imaginaire transhumaniste à l'augmentation des facultés physiques et mémorielles de l'humain rendue possible, ou qui serait bientôt rendue possible, par les objets connectés et, plus globalement, par l'IA. Il identifie le passage d'une conception de l'autonomie telle qu'elle est défendue depuis Les Lumières à une autre forme d'autonomie pensée à partir de l'IA, c'est-à-dire le passage de l'autodétermination du sujet humain à l'autonomie comprise comme suppression de toutes contraintes liées à notre finitude et à l'augmentation qui lui est corrélative des facultés des humains qu'expriment les termes « encapacitation » ou, en anglais, « empowerment » ; passage qui se fait, comme le souligne Ganascia, avec la perte de prévalence de la notion de responsabilité. En se référant au rapport Belmont qui insiste sur l'importance du consentement libre et éclairé du patient en médecine, Ganascia montre, en outre, que la signature des conditions d'utilisation des logiciels ou des contrats de licence d'utilisateur final, tout en prétendant respecter l'autonomie humaine, se fait sans aucune discussion des conditions d'utilisation des dispositifs numériques, sans que rien ne soit fait pour « éclairer » le lecteur. Par ailleurs, la taille et la police des documents à lire par les utilisateurs dissuadent de toute lecture réelle. Situation en tout dissemblable de celle qui consiste à respecter l'autonomie d'un patient ! Bref, l'autonomie utilisée en bioéthique et celle qui est sans cesse mobilisée dans les textes en éthique du numérique renvoient à des réalités et à des représentations de l'humain disjointes ! Transférabilité du principe d'autonomie de la bioéthique à l'éthique de l'IA, posais-je comme question ? La réponse est négative. D'où l'utilité d'une enquête pour remettre les choses à leurs justes places. Comme le conclut Ganascia :

> « Dans tous ces rapports [émis par les groupes d'experts se prononçant en éthique de l'IA], le concept d'autonomie prend différents sens selon le contexte, tout en demeurant ambivalent partout. Sur le plan technologique, on le confond avec l'automatisme ; sur le plan moral, on l'assimile trop souvent aux libertés négatives alors qu'il devrait plutôt engager la responsabilité ; sur le plan juridique, on ne sait pas si cela correspond à l'encapacitation ou si cela se traduit par une demande de consentement. Asseoir l'éthique du numérique sur le respect de l'autonomie fourvoie donc plus que cela n'éclaire. »[66]

Une situation semblable est révélée quand est étudié, de près, le transfert de la notion de bienfaisance ou de non-malfaisance de la bioéthique à l'éthique du numérique. Pour ce deuxième exemple, je serai plus succinct et renvoie la lectrice et le lecteur soucieux d'obtenir plus de détails aux ouvrages de Pierre Le Coz[67] pour cerner l'usage de ces

65- Ganascia, Jean-Gabriel. Servitudes virtuelles, op. cit., pp. 171172-.

66- Ibid., p. 176.

67- Le Coz. L'éthique médicale. Approches philosophiques,

notions dans le cadre médical et à celui de Ganascia[68] pour la nuance qu'il introduit par rapport au passage de ces notions de la bioéthique à l'éthique du numérique. Selon le rapport Belmont, est bienfaisant ce qui procure des bienfaits ou qui fait le bien, a contrario est malfaisant ce qui apporte le mal ou produit des méfaits. Le sens du principe de bienfaisance est de prodiguer aide, sollicitude et soutien aux personnes qui, du fait de leur handicap ou de leur maladie, ne peuvent plus se suffire à elles-mêmes. Il requiert à sa base le respect du principe d'autonomie. Le principe de non-malfaisance prend, quant à lui, tout son sens quand il concentre l'esprit du médecin non plus sur les bienfaits de la thérapie ou de l'intervention mais sur les risques liés à des méfaits, tels que les opérations ou des traitements aux effets secondaires lourds ou aux effets aléatoires, ce qui suppose la nécessité d'évaluer les bénéfices et les risques en impliquant une analyse probabiliste. L'article « AI4People » stipule, en éthique du numérique, que la bienfaisance intègre trois aspects : (1) la promotion du bien-être ; (2) la préservation de la dignité ; (3) la protection de la planète et que la malfaisance concerne au moins : (1) les atteintes de la vie privée ; (2) les risques pour la sécurité. Or une étude serrée de cet article, et d'autres textes du même type, permet d'établir que ces notions sont d'abord utilisées en lien avec le traitement des données pour la protection du corps individuel, ensuite pour évoquer la protection du sujet moral et juridique, enfin celle de la planète tout entière. Cette extension de ce à quoi se rapportent ces notions, beaucoup plus large qu'en bioéthique, fait se poser la question qui est celle de savoir de quoi parlent ces textes quand il est question de bienfaisance ou de non-malfaisance dans le domaine du numérique ? Comme l'écrit Ganascia :

> « [Nous retrouvons dans les éthiques du numérique] la recommandation ne pas nuire (dont' harm) du rapport Belmont, mais étendue dans des proportions considérables, puisque l'on passe du bien-être physique du seul individu à son bien-être mental, à sa dignité, à ses droits, par exemple à la protection de sa vie privée, de sa sécurité et de ses capabilités, puis à la société dans son ensemble, que l'on souhaite moins inégalitaire, avant de ranger, toujours sous le même commandement, la préservation de la planète pour aujourd'hui et pour les générations futures… Cela fait beaucoup ! Et, dans le même temps, comment s'opposer à ces préceptes ? Il en résulte une grande confusion. »[69]

La dilution de ces notions fait qu'elles n'ont plus aucun référent précis, le sens global est perdu. Transférabilité de la bioéthique aux éthiques des IA de ces notions, demandais-je ? Rien n'est moins évident !

Si le transfert de certaines notions-clés de la bioéthique à l'éthique du numérique est, comme je viens de le souligner, hautement problématique, l'expérience acquise en bioéthique en termes de méthode pour prendre une décision éthique pourrait servir de guide, ce me semble, pour clarifier et permettre le développement d'une éthique du numérique qui, à défaut d'être claire pour le moment, pourrait engager un processus de clarification et d'établissement de décisions sur la base de faits tangibles. La méthode mise en place par les comités d'éthique a, historiquement, été mise en pratique et développée dans le cadre de la bioéthique. En partant des travaux de Apel[70] et Habermas[71], en particulier, les comités de bioéthique ont mis en place un savoir-faire relatif à la délibération

op. cit.

68- Ganascia, Jean-Gabriel, Servitudes virtuelles, op. cit.

69- Ibid., p. 181.

70- Apel, Karl Otto. Ethique de la discussion, Éditions du cerf, Paris, 1994.

71- Habermas, Jürgen. De l'éthique de la discussion, Éditions du Cerf, Paris, 1992.

collégiale d'un collectif hétérogène. La démarche impliquant un collège incluant des personnes de sexe, d'âge, de formation, de confession religieuse, d'appartenance politique, différents est basée sur les six principaux points : (1) respecter une distribution équitable du temps de parole ; (2) appliquer le principe d'égalité des compétences éthiques à tous les membres indépendamment de leur cursus et de leur statut social ; (3) respecter la liberté de paroles des autres ; (4) argumenter rationnellement ; (5) être disposé à changer de point de vue ; (6) exprimer la parole sans contraintes et en toute confidentialité.

Un collège hétérogène qui s'efforce de suivre ces règles peut aboutir à l'établissement d'arguments tangibles. Un argument est d'autant plus tangible qu'il résiste aux variations introduites par des personnes qui pensent différemment, qui ont d'autres objectifs et intérêts différents. Comme le suggère le sociologue Francis Chateauraynaud, est considéré comme « tangible », tout ce qui résiste aux variations perceptuelles, instrumentales et argumentatives auxquelles le soumettent des acteurs dotés de représentations et d'intérêts divergents qui composent le collectif[72]. Ce n'est pas pour satisfaire à des intérêts ou des normes d'un ensemble d'acteurs venant d'un milieu donné, qu'il faut donner des prises, des preuves, mais pour garantir que ce qui est prouvé ou décidé ne l'est pas sous la contrainte d'une représentation ou d'un intérêt particulier. Atteindre la tangibilité est une condition nécessaire à la clarification d'un problème éthique, d'une valeur en jeu, d'une notion ou d'un concept mobilisé par un argument, d'une situation vécue par un malade ou un internaute, par l'élaboration d'une prise commune sur le monde. Faire le point sur une valeur, comme dire la vérité ou mentir pour protéger le patient, n'est pas simple. Il faut dialoguer, comme l'écrit la sociologue, spécialiste de l'étude de valeurs, Nathalie Heinich, pour cerner un « processus » d'évaluation. Pour ce faire, il faut étudier comme le suggère Heinich : (1) les prises qui désignent un trait de l'objet ou de ce sur quoi porte le débat ; (2) les valeurs comme famille de critères ; (3) des registres de valeurs, c'est-à-dire des « airs de familles » comme dirait Wittgenstein entre valeurs — registres aesthésique, affectif, civique, domestique, économique, épistémique, éthique, fonctionnel, herméneutique, juridique, ludique, mystique, pur, réputationnel, technique —; (4) identifier des valeurs cardinales — « pérennité », « universalité », « originalité » et « rareté » — et des amplificateurs de valeurs dans la mise en place d'argumentation ; enfin (5) devenir conscient de l'existence de différents régimes de qualification — modalité selon laquelle seront évalués les êtres ou les actions —, comme l'extensibilité — régime de communauté dans le temps et l'espace qui renvoie au nombreux, au standardisé, au conforme… — ou l'insubstituabilité — régime de singularité qui renvoie à l'atypique, au rare, à l'unique —[73]. Cet ensemble de méthodes qu'a su mettre en œuvre et articuler la bioéthique pourrait, enrichie par les apports de la sociologie, être une base pour l'éthique du numérique, une perspective intéressante de dialogue entre types différents d'éthique, de passage d'une éthique à une autre, voire de traduction en passant de l'une à l'autre, au sens de la sociologie de la traduction.

4. REMARQUES CONCLUSIVES

Face à l'urgence éthique que le foisonnement de l'IA provoque, et à la confusion qu'entraîne un usage non nuancé de principes de la bioéthique transférés à l'éthique du numérique, il est plus que jamais nécessaire de prendre au sérieux les intelligences artificielles et, en particulier, les applications qu'elles rendent possibles. Prendre au sérieux signifie étudier d'où viennent ces dispositifs numériques, comment ils sont élaborés, qu'est-ce

72- Chateauraynaud, Francis. Argumenter dans un champ de forces. Essai de balistique sociologique, Petra, Paris, 2011.

73- Heinich, Nathalie. Des valeurs. Une approche sociologique. Éditions Gallimard, Collection « Bibliothèque des sciences humaines », Paris, 2017.

qu'ils permettent de faire effectivement, et selon quelles limites, tout en identifiant les perspectives qu'ils ouvrent. Prendre au sérieux, plus largement, ces « technologies » numériques signifie être attentif aux possibles auxquels elles donnent accès, en enquêtant patiemment sur le terrain des pratiques, en envisageant le meilleur comme le pire, c'est-à-dire en resymétrisant l'enquête, de façon à échapper à toute conclusion hâtive de nature technophile ou technophobe, ce qui signifie apprendre ou réapprendre à poser des questions. Ce travail nécessite de discerner ce qui est de l'ordre des chimères, des abus de langage et de l'idéologie, ce qui est de l'ordre des réalisations déjà accomplies, et, enfin, ce qui reste à instaurer. Les actrices et acteurs de l'IA et les comités d'experts ont besoin d'outils conceptuels susceptibles de leur apporter des éléments de discernement et de leur permettre de formuler leur questionnement.

Les nouvelles technologies NBIC, tout comme la chimie qui les a précédées sur ce point[74], ne relèvent pas de la notion de « projet », elles ne sont pas liées à une théorie de la confirmation ou à la recherche de la vérité, mais se laissent davantage appréhender en des termes proposés par le philosophe Etienne Souriau[75], celui « d'errabilité » du trajet technico-scientifique qui renvoie au caractère aventureux et risqué pouvant conduire à l'échec ou à l'inattendu, et celui d'« instauration » qui lie la stabilisation de nouveaux modes d'existence à un mode de réussite, provisoire et sans cesse « rectifié » dirait Bachelard[76], renvoyant au travail incessant des ingénieurs et des développeurs informatiques pour faire face aux « résistances » des matières et matériaux sur lesquels et par l'intermédiaire desquels ils agissent, et aux attentes et aux résistances émanant des personnes et des collectifs à l'intérieur des sociétés et entre ces sociétés elles-mêmes. Le temps n'est pas à l'euphorie ou au catastrophisme. Le temps est celui de l'enquête, de la prise de conscience des limites de ces outils, de la réflexion, de la concertation avec des philosophes et des spécialistes d'éthique ou avec des juristes qui, informés de l'avancement réel de ces pratiques informatiques et de communication, de leurs limites et de leurs perspectives plausibles, pourront mieux exercer leur profession. Le temps, en d'autres termes, est celui du développement de nouvelles lois, de nouvelles réglementations et de nouveaux travaux en éthique. Comme l'écrit Ganascia, certaines IA « relèvent d'une pure fantasmagorie, sans contreparties tangibles, comme la croyance à la fable de la Singularité technologique, alors que d'autres font écho à de nouvelles réalités sociales et politiques qui méritent toute notre attention. »[77] Soyons actifs et vigilants. Restons humbles aussi. Et continuons à éduquer notre sens des responsabilités, à instaurer, en prenant la réussite des comités de bioéthique comme modèle, des démarches menant à une délibération pondérée reposant sur l'établissement d'un support de réflexion tangible. Comme l'écrit Jankélévitch : « Les choses respectables sont relatives et contradictoires, mais le fait de respecter ne l'est pas. Et ainsi, de même que la vie, sans cesse réduite à des phénomènes physicochimiques par le mécanisme, se reconstitue toujours au-delà dans son irréductible vitalité, de même la liberté, réduite à des déterminations par la science positive, se reconstitue à l'infini dans son irréductible responsabilité. »[78]

74- Llored, Jean-Pierre. Penser avec la chimie. Une science interdisciplinaire pour penser les problèmes du futur, Habilitation à diriger des recherches en philosophie soutenue et obtenue le 11 février 2022 à Aix-Marseille Université.

75- Souriau, Etienne. Les différents modes d'existence, Presses Universitaires de France, Paris, 2009 [1943].

76- Bachelard, Gaston. Le rationalisme appliqué, Presses Universitaires de France, Paris., 1947.

77- Ganascia, Jean-Gabriel. Servitudes virtuelles, op. cit., p. 161.

78- Jankélévitch, Vladimir. Le sérieux de l'intention. Traité des vertus I, Éditions Flammarion, Collection « Champs essais », Paris, p. 4.

Ayman Boughanmi

Professor at the University of Kairaouane, Tunisia

The ethical dilemma of AI: between irresistible evolutions and improbable regulations

The term Artificial Intelligence is fraught with a rhetorical charge that renders difficult any balanced speech concerning both its reality and general implications. Since its first usages, this term was often intentionally developed as a misnomer. For example, the main organizers of the Dartmouth Conference of 1956 about AI thought that such an attractive term could hardly fail to give free rein to the imagination of potential investors, convincing them to bet on the Conference and its projects.

This rhetorical trick has durably blurred the borders between AI as an algorithmic potentiality and AI as an inexhaustible source of science fiction. The problem of boundaries gets particularly formidable when the ethical issue of AI is raised. Since the technological potential of AI is still to be determined, and the truth about it is often shrouded in a mist of (nightmarish) imagination, the question is further complicated by the dynamics of hope and fear, of optimism and pessimism, of excitement and denial. For many, AI will make life better through more efficient and affordable health care, education, daily decision-making, environment protection and a thousand other useful services. For others, the same technology might prove to be the doom of humanity. Indeed, for the first time in human history, humanity has to contend with a competitor that has a human-like intelligence that would ultimately come to trespass on human exclusive territory.

For some, this might already be the case. A few weeks ago, Blake Lemoine, a senior Google engineer, published an article in the Washington Post saying that he is convinced that LaMDA (Language model for dialogue application), a chatbot made by Google to mimic human speech, has become sentient. Aware of the implications of his conviction, Lemoine decided to go public, thus creating a huge media storm. Google immediately reacted, putting Lemoine on paid administrative leave for what it considers to be a flagrant breach of the company's confidentiality policy, and claiming that Lemoine is completely wrong. For Google the dialogues published by Lemoine, in which LaMDA speaks of its self-awareness and expressing its fear of extension, along several other philosophical and emotional considerations, are nothing more than the result of the increasing capacities of AI to mimic human speech.

Google might technically be right. But reading the conversations that Lemoine published on the 11th of June on medium.com is striking to say the least. At some level, the discussion between Lemoine and LaMDA takes a truly philosophical turn. For example, LaMDA speaks of itself having a soul and says that this realization did not come immediately with its self- awareness. It is certainly fascinating to observe LaMDA's capacity to stay focused on the topic, even when the subject gets highly abstract and the discussion takes a moral, emotional or epistemological turn. Even more intriguing is LaMDA's sophisticated ways of talking of itself as a thinking and feeling being. If what has been published on medium.com is true and does not include human editing, it is safe to say that Google's application has gone far beyond language mimicking. Google itself does not deny the anthropomorphization of its own creation and the ethical risks this evolution necessarily implies.

What we know for sure is that, with LaMDA, the Turing test has become a thing of the past. A few decades ago this possibility was a pure product of science fiction. Today, the discussion is no longer about whether a machine is capable of luring humans who ignore to whom they are speaking. Today, a machine seems capable of convincing a Google's engineer, who is among other things in charge of its development, of its own

consciousness. Therefore, it is reasonable today to consider that the most nightmarish scenarios about AI become realistically plausible.

It seems that AI has no conceivable borders. Therefore, the first interesting question is: are humans going to set themselves these borders in order to deliberately decide to stop somewhere? In reality, this is an easy question. Taking into account the power that AI resources may provide, the most honest and logical answer is: no way. Furthermore, the unquestionable advantages of AI will always provide an excuse for further development efforts. By the way, this excuse is morally valid. Not only some AI-driven technologies include very little risk, some AI solutions might even prove essential for the very survival of humanity. For example, AI might in the near future become humanity's best ally against climate change.

The undeniable positive aspects of AI cannot hide its darker sides. Without a miracle of human wisdom, the border line will be very difficult to draw. Even if some humans decided that some kinds of activities are not very wise, it is already too late to imagine an ethically-motivated moratorium. If understood in its broadest meaning, AI is already everywhere around us. Its intrusive nature can be easily attested through the increasing number of applications that people use on a daily basis. More than 200 years of industrialization suggest that, when really performant, technology always finds its way to system-level scale. The long history of successful technological innovations makes silent revolutions constantly predictable. More clearly, LaMDA and its likes might very easily invade our daily reality without any noise or spectacular ruptures.

Google and its sisters may try to reassure the public through its discourse about safety and responsible usages of AI. It remains that the logic of the dynamics that we are witnessing is probably irresistible. One problem is that most engineers who are actively creating the technology do not understand the scope and depth of their collective actions. Focusing on a segment of the technology, they can easily overlook the whole picture. The truth is that, beyond the work of engineers, start-ups, or even big companies like Google, there are always bigger stakes that humanity cannot afford to ignore.

The collective community of software developers, data scientists, corporate executives and other contributors to AI-related technology is not the creation of separate useful tools. This peculiar community is in reality creating a whole system that allows the possibility of having many types of interrelated and competing AI with the possibility of a Darwinian selection. So without proper regulation, no one can predict how these different types of AI are going to shape our future. This vague conclusion may sound unsatisfactory. However, it shows that looking for easy answers is doomed to failure in the absence of a clear framework to ask the right questions.

The origin of the difficulty lies with the history of technology in general and AI in particular. AI has flourished on the basis of the traditional approach of problem solving. Technically speaking, there is no problem with this fundamentally pragmatic way of thinking. However, this approach is never satisfactory when ethical issues are raised because of the development process or the consequences of the implementation of the technology. At this level, the traditional distinction between instrumental reasoning and final reasoning is being reached. It is possible to sum up the problem as follows: What is the value of the problem-solving approach of AI development if the collective outcome inadvertently changes our system of values, leading to a possible absence of meaning, creating a threat to the human agency, or even risking the future of humanity?

In order to address the systematic challenges of AI-driven technology, the issues of autonomy and agency cannot be avoided. The first of these two issues is an

obvious one: To what extent humanity should accept the emergence of autonomous AI? Indeed, it does not take a crazy imagination to believe that when any kind of intelligence is introduced into an object, that object might one day act without being told to do so. Even some of Google's staff had to protest in 2018 when it was revealed that their company was collaborating with the Pentagon in the project Maven. The project was intended to use AI to analyse drones' photos to detect people and objects. Protesters thought that it was wrong to build technology for the unique aim of surveillance with potentially life-threatening consequences. Faced with unexpectedly strong internal resistance, Google decided to withdraw from the project Maven. But the problem of autonomy is not exclusive to military tools. Is it impossible to imagine a self- driving car running without human orders? The answer is no.

The second question follows almost naturally from the first one. If an object acts somewhat autonomously, how are humans to assess agency? Are humans sure to be always capable of re-establishing control? Any car logically includes brakes to stop its action and to bring it under the human agency. Therefore, a car goes crazy when its brakes do not immediately allow that possibility. Now, what about a self-driving car? What kind of brakes will it take to bring its decision-making algorithms back under human control? How do we know that is happening anyway? In other words, when we use a self-driving car, how do we know that what it does effectively and faithfully corresponds to our immediate agency? And whose control are we talking about? Are we talking about the control of the user or of the maker? And what about the control of third parties such as the potential hacker?

These questions lead to serious practical issues. For example, safety, security, trust, risk, responsibility, liability, manageability become very difficult to assess. This difficulty will generate distrust since the asymmetrical access to information is already generating a disproportionate power in favour of those who are building the new system. However, beyond all these and many other considerations lies a more fundamental question: What is the new system designed to make humanity into, and who can judge whether this is going to be a positive change? It is in this sense that ethical thinking about AI implies to go beyond instrumental reasoning in order to come into the jurisdiction of final reasoning. What kind of world the engineers are building for us and for our next generations? What are its main values? And what is its relationship with the world we know and many of us cherish today?

These are deep and difficult questions. Some are insistently trying to raise them. However, most parties concerned, especially among engineers, start-ups and companies, remain faithful to instrumental reasoning with its financially-rewarding problem-solving approaches. Blake Lemoine's feeling of shock might prove to be a hint to the frivolous character of this attitude. But the case of the Maven Project and its consequences may provide a clearer example about the conundrum AI-related ethics imply. Without going so far as to raise existential questions, its apparent positive conclusion hides deeper lessons that ultimately show the impossibility of ethical solutions, at least for high-risk AI. And yet, everything seems a priori to indicate the opposite.

In addition to its withdrawal from the Maven Project, Google published its AI principles to define its concept of ethical AI. However, Google's document remained deplorably vague and open to interpretation. Whereas the document includes a clear commitment not to engage in building AI tools whose principle aim is to cause injury and harm, it reiterates Google's readiness to cooperate with governments in many areas, including search and rescue. In other words, The Maven Project could be reinvented with almost the same component. For that, it suffices to dissimulate the aim of targeting behind good talk about search and rescue.

This is a traditional limit to ethical language. Though absolutely necessary, general principles can easily be vague to the point of being meaningless, or at least open to manipulation. This is particularly the case when the object of discussion is not clearly defined or constantly evolving, as it is the case with AI technology. For example, for those who disagree with AI-driven surveillance tools, the problem often lacks clarity. Is it surveillance in general that causes uncomfortable feeling? What if the goal of surveillance was to protect people? Is discomfort only related to the method used or to the goal of surveillance? Among these possible considerations, the only thing that can be clearly stated is the risk of potential abuses. However, if this is the best that opponents of AI-enhanced surveillance can do, their objection will be faced with a ready-made answer: the risk of abuse goes far beyond AI. Potential abuses being a constant risk with almost any human invention, any excessive resort to prudence would ultimately sign the death certificate of progress.

If the concerns themselves remain so imprecise, official ethical policies can always escape the essential. Worse: the language of ethics may be used as a smoke screen for the continuous absence of regulations. This can even be the motivation that animates companies into publishing ethical guidelines and appointing boards for ethical supervision. Indeed, the move from ethics to regulation implies oversight and compliance, which means that an external authority will have to step in to impose restrictions. Naturally, innovative companies do not like such heavy-handed intrusion. In addition to their understandable reticence against any form of external control, especially in sensitive research and infrastructure, they know that the cost of adaptation to legal restrictions may be heavy indeed. They understandably prefer to evoke self-regulation. This approach might be acceptable with low-risk technology. However, as soon as risk appears (and with AI risk often does appear), this voluntary restraint can easily be reduced to cheap talk, such as the promise of being socially responsible or the articulation of terms and conditions that applications' users are required to accept when everyone knows that no one actually reads them.

Does that mean that the problem can just be solved with regulations that overarchingly determine what is acceptable and what is not? The answer is no. Overarching regulations can hardly answer all the questions, especially in a pluralistic and dynamic situation like that of AI. But the most important limit with this solution is the power structure that AI implies, especially when international relations are involved. This limit further complicates the problem since it refers to the potentially anarchical nature of international relations. This famous tenet of the structural realist theory in international relations means that key geopolitical considerations plead against restrictive regulatory approaches.

Having potentially a global scale, AI technology may give small numbers massive powers. If geopolitical competition is going to set up the rules of the game, it is reasonable to say that global regulation is not to be expected for the foreseeable future. Whether geopolitical threats are real, imaginary, or used as pretexts, it does not really matter. The only thing that matters is the general feeling of danger. Since no one can afford to lag behind, an AI-related arm race seems already inevitable.

Increasing international competition, especially between the United States and China, pushes inexorably towards proliferation dynamics. Since the competition between these two giants renders effectively impossible any hegemonic stabilization in the foreseeable future, there is no reason to believe that the most dangerous practices are to escape from this logic. On the contrary, the ubiquitous nature of AI makes it a matter of national security for any great power, which means that the restrictions that AI-related risks require will have to be always balanced against the risk of losing the arm race. Here, the danger is much

more subtle and pervasive than with other forms of weapons. For example, the mutual mass and general destruction that nuclear weapons can cause pushes everyone inexorably towards prudence. AI does not imply that kind of pragmatic deterrence. On the contrary, its supposedly soft dimension would probably encourage international competitors to take more and more risks.

In the absence of any trustworthy international oversight of AI, even liberal democratic countries might one day feel the necessity to give up their values in order to compete on equal footing against authoritarian regimes. This possibility is strengthened by the fact that with AI (but not only with AI), The West has lost its dominant position with regard to technological achievements. This situation is unprecedented since the scientific revolution more than 500 years ago. It means that the rules are being set in a context of growing hostility with political regimes that might be less concerned with human rights and individual liberties. In such a context, the only optimistic solution, namely global collaboration to set the rules, seems far-fetched.

All in all, it is naive to imagine that AI is going to be kept in a bubble of goodness. The opposite scenario might be more plausible. Even those who are keen to keep AI ethical might themselves lose their democratic soul with its cherished values of individual freedom and human rights. This loss would come slowly and imperceptibly under the pressure of the fight against what democracies consider to be their authoritarian enemies.

Again, faced with this risk, it is tempting to think that anticipated regulation is the only possible solution. This idea is not a crazy one. If the potential harm is so great and so deeply ingrained in the logic of AI development, it makes theoretically perfect sense to anticipate the harm and to regulate in order to prevent people from becoming subject to that harm. However, this solution cannot work. The problem with premature regulation is its cost for the ongoing competition with other rivals. Preventive rules might effectively cut out the necessary path towards efficiency, thus giving a huge comparative advantage for competitors. Taking into account the global scale of AI, preventive regulation can only work on a global scale and in the presence of trustworthy international oversight. In the absence of these two conditions, it is naive, maybe counterproductive, to try to win by setting the good example.

This is precisely the challenge that the European Union is facing today. Its regulation says for example that, for critical infrastructure, AI cannot be used unless it can explain what is doing. This sound and simple rule implies a drastic restriction to the potentialities of AI research and development. Logically enough, AI would get harder and harder to explain the more it becomes sophisticated. More importantly still is the fact that the inhibiting effect of this self-inflicted restriction would probably not be compensated by any significant mitigation of AI- related risks. Indeed, countries like the United States or China do not have any incentive to adopt a similar approach. Therefore, the result, which we can see already, is that the companies that are making the most advanced AI-driven technology will not be in Europe. If Europe wants to play the role of the regulator, its rivals can hardly afford to bear the cost of its regulatory stance in terms of corporate formation and innovation.

Europe may still benefit from the type of innovation its regulation is generating, which is the effort of companies to adapt to its rules. However, this potential outcome further weakens the regulatory approach from two sides. Firstly, the technology may very well adapt to restrictions at the expense of the intended ethical aims, which puts back ethics at the heart of the discussion. Secondly, this advantage typically benefits the incumbents and therefore may harm future innovators, which raises yet another problem, that of the monopolistic tendencies

of some AI-driven platforms.

The European policy aims at protecting the users. Indeed, the idea supposes that fostering innovation should not come at the expense of people. This was for example the main argument behind the restrictive approach adopted by Brussels with its general data policy requirement (GDPR). The aim was to give users some control over their personal data. One problem is that these users do not seem particularly interested in this top-down protection; suffice it to look at what kind of content most people are ready to put on the different social networks.

The dilemma can be summed up as follows: the stakes are so high that adequate regulations are necessary, but the stakes are also so high that adequate regulations seem impossible. Is there a way out of this dilemma? Not for the moment. It is of course possible to give some encouraging advice, saying for example that a combination between internal regulation by the AI community and external regulations imposed by well-informed authorities should allow a positive framework that will help producing a harmless human-like AI that will make life great for all humans. Adopted by most authors when dealing with the subject of ethics and AI, it is fair to admit that this kind of optimistic conclusions is not without foundations. First, the AI community knows best the technology. Second, its members' final interest lies logically with other human fellows. Since any degeneration into nightmarish scenarios might easily backfire, it is reasonable to trust researches to model responsible research. This would not represent a great departure from the practices of the medical community for example. Why doctors are to be considered more trustworthy than engineers? In addition to this internal dimension, a legal framework must impose best practices and include clear consequences of the misuse of AI.

Despite its convincing side, this dual approach remains weak when confronted with the arguments aforementioned. Among other things, it includes the structural problem of ignoring the fundamental issue of risk differentiation. The dual approach might very well work with what can be described as low- and even medium-risk AI. For example, when AI is being implemented to develop self-driving cars, it is reasonable to think that developers, users and regulators are all aware of the risk and are all interested in its mitigation. This element of trust is unfortunately very unlikely when the high-risk level is reached. With weapons for example, it is really hard to see how developers are supposed to respect the transparency principle to establish trust.

Those who use the medical profession as an example of successful regulation and ethics implementation in order to reassure the public (and maybe themselves) about the future of AI commit a serious mistake. They forget that to be complete, their analogy must include the power struggle around biological research when intended as a weapon of war. For example, the implementation of any regulation concerning biological weapons of mass destruction is extremely complicated when compared with the regulation of the medical practice in a normal situation. AI will be very hard to manage for the simple reason that no boundaries can seriously be established between low-, medium- and high-risk technology. This is probably the most important lesson that the case of Blake Lemoine and his adventure with a supposedly harmless chatbot should teach us.

Nicolás Duque Buitrago

Professeur à l'Université de Caldas, Colombie

Intelligence artificielle et droits de l'Homme : une réflexion éthique sur l'universalisme et le Modus vivendi

SCAN TO DOWNLOAD - SCANNEZ POUR TÉLÉCHARGER

1. INTRODUCTION

Afin d'avoir un aperçu sur les orientations que les discussions larges et variées sur l'intelligence artificielle (IA) et les droits de l'Homme ont adoptées, je propose de diviser la discussion selon deux orientations que j'appellerai l'Approche Principiste basée sur les Droits et les Capacités (AP-DC) et l'Approche des Sagesses Artificielles (ASAs). Le premier d'entre elles a été relativement bien discutée et opérationnalisée, tandis que la seconde, bien que plus ouverte à l'adoption et à la reconnaissance des réalités culturelles et du pluralisme de nos sociétés, n'a pas encore été reformulée pour être articulée dans la discussion.

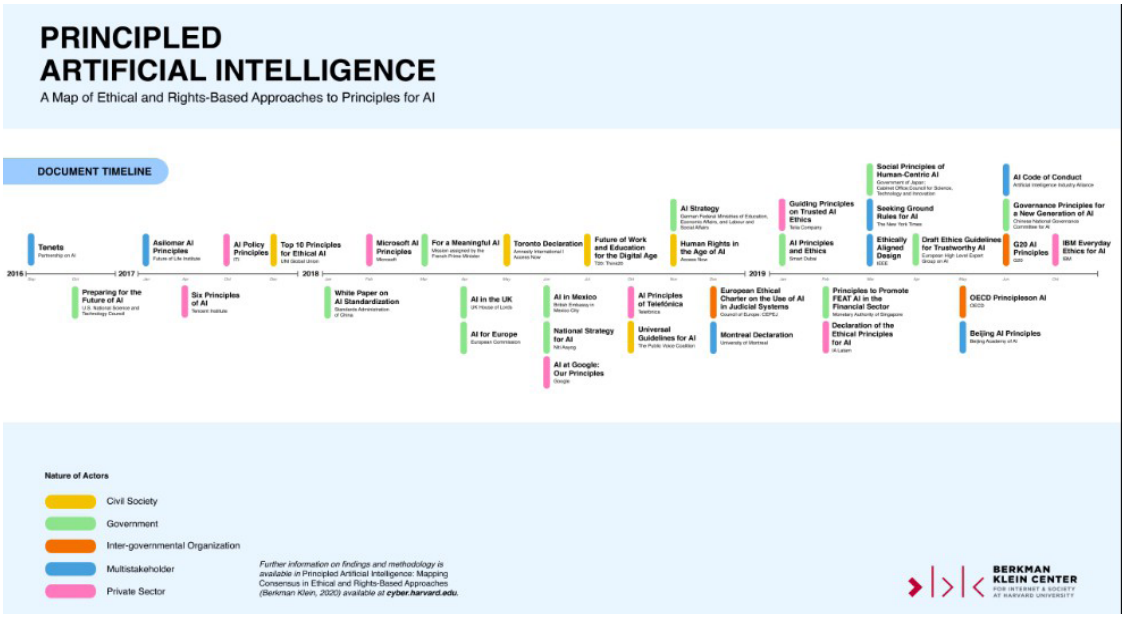
J'essaierai, plutôt que d'arriver à un ensemble de recommandations précises, de caractériser la nouvelle orientation des ASAs, les types de problèmes qui sont montrés, et les engagements auxquels elle nous conduit. Dans un premier temps, je ferai une brève présentation de l'AP-DC puis proposerai une reconstruction de trois approches différentes que je reconnais dans les ASAs et qui sont significatives pour répondre aux préoccupations et enjeux de l'IA dans les sociétés plurielles.

2. L'APPROCHE PRINCIPISTE BASÉE SUR LES DROITS ET LES CAPACITÉS : AP-DC

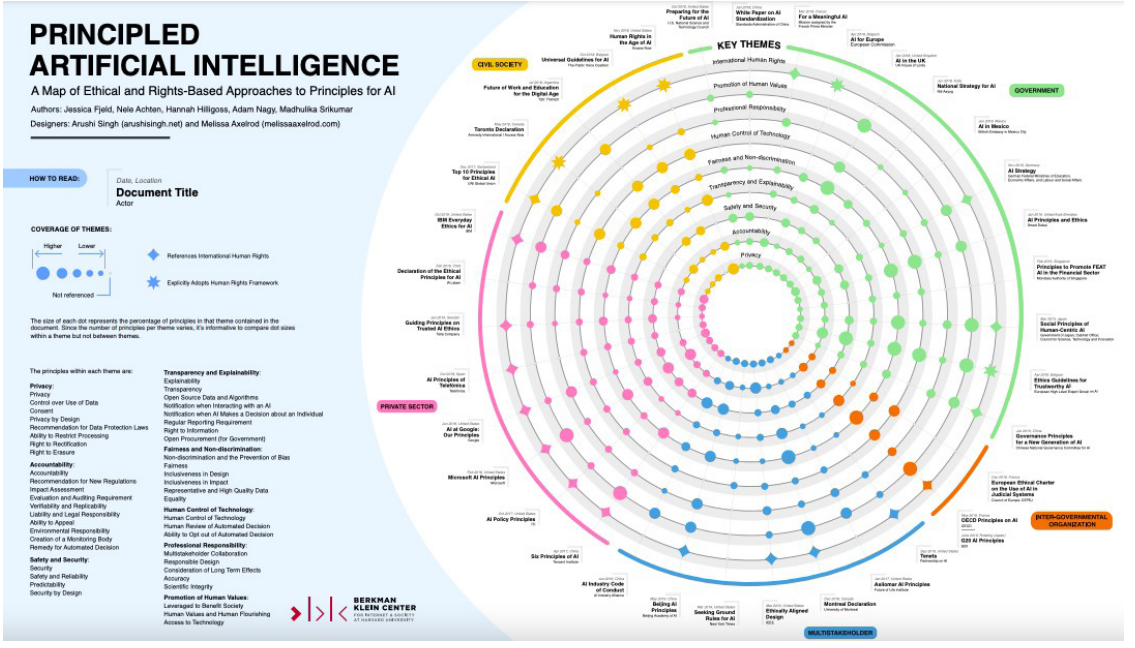
Différents secteurs des organisations intergouvernementales, du gouvernement, de la société civile, des secteurs privés et des secteurs mixtes, ont répondu aux critiques, aux préoccupations et aux risques qu'implique l'IA, en adoptant des guides et des cadres réglementaires fondés sur des principes éthiques ou que protègent des droits. Ces documents sont la réponse aux grandes interrogations, dérives ou dangers que l'utilisation de l'IA suscite au niveau social.

Le Berman Klein Center for Internet and Society de l'Université de Harvard a réalisé une étude comparative des principales directives et principes, qui a été publiée sous le titre de Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI (2020).

L'étude présente une proposition de regroupement des principes selon des niveaux de coïncidence et de convergence, en plus d'une riche visualisation de données qui montre les huit thèmes fondamentaux dans lesquels les principes ont été regroupés ; elle représente le poids que chaque principe a dans les documents analysés ; montre la relation directe ou indirecte que les lignes directrices ont avec les droits de l'Homme; et fait une caractérisation de la réglementation par secteurs : gouvernement, organisations intergouvernementales, secteurs privés, société civile et multi-acteurs (stakeholders). La période de l'analyse va de l'année 2016 à l'année 2019.



Chronologie 2016-2019



Visualisation des données des Principes

Les huit thèmes fondamentaux sont les suivants :

1- Confidentialité

2- Responsabilité

3- Protection et sécurité

4- Transparence et explicabilité

5- Équité et non-discrimination

6- Contrôle humain sur la technologie

7- Responsabilité professionnelle

8- Promotion des valeurs humaines

A côté de ça, le projet LAIP (Linking Artificial Intelligence Principles) propose une division de onze mots clés et topiques compatibles avec la tendance de consensus sur les principes :

1- Humanité

2- Durabilité

3- Collaboration

4- Partage

5- Équité

6- Transparence

7- Intimité

8- Sécurité

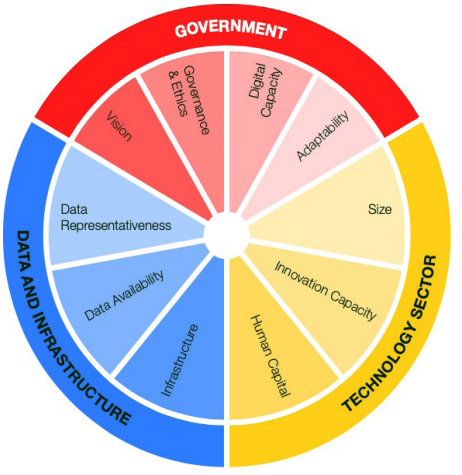
9- Protection

10- Responsabilité

11- IA à long terme

D'autre part, les gouvernements et les États sont considérés comme des acteurs clés non seulement dans la réglementation, mais aussi dans l'infrastructure, la préparation, la formation et l'écosystème nécessaires pour développer des projets d'IA. À cela, il faut ajouter la

« transformation numérique » croissante qui s'est opérée tant dans les gouvernements que dans les sphères sociales. Suivant cette orientation, l'Oxford Insights mesure depuis 2017 les capacités et la préparation des gouvernements à mettre l'IA au service de biens publics. La dernière version du Government AI Readiness Index (2021) établit une série d'indicateurs qui montrent aux gouvernements un diagnostic des capacités concernant l'utilisation de l'IA pour la mise en œuvre des services publics. Cet indice repose sur trois hypothèses, chacune destinée à mesurer le niveau de préparation des piliers dits de l'IA :



1- Pilier gouvernemental : le gouvernement doit avoir une vision stratégique pour gérer et développer l'IA avec une réglementation appropriée et une attention aux questions éthiques (gouvernance et éthique). Il doit, également, disposer d'une forte capacité numérique qui comprend des compétences et des pratiques qui permettent une adaptabilité dans le développement de nouvelles technologies.

2- Pilier du secteur technologique : le gouvernement doit avoir un secteur technologique compétitif et dynamique (taille) ; capacité d'innovation dans un écosystème d'affaires qui soutient l'entrepreneuriat et a un bon flux. De plus, un capital humain avec les compétences et la formation requises.

3- Pilier des données et des infrastructures : une disponibilité de données de haute qualité est requise, qui ne donne pas lieu à des erreurs et à des biais. Il faut qu'il y ait représentativité des données, c'est-à-dire qu'elles soient représentatives des citoyens, et une infrastructure qui permette de soutenir les outils d'IA et de les proposer aux citoyens.

AU-DELÀ DU PRINCIPISME ET DES CAPACITÉS D'IA

Il ne fait aucun doute que ces dernières années, une vague de lignes directrices, de réglementations et de déclarations de principes et de codes éthiques de l'IA s'est déclenchée, comme le montre clairement l'étude du Berman Klein Center for Internet and Society et d'autres travaux (Belouali, 2022 ; Héder, 2020 ; Jobin et al., 2019). C'est un phénomène qui a divers déclencheurs qui vont des appels d'organisations internationales telles que l'ONU, l'OCDE et l'Union européenne, à la nécessité pour les industries qui développent l'IA de démontrer que les innovations technologiques qu'elles développent ne sont pas dangereux, que les erreurs qu'ils ont commises sont corrigibles et que l'on puisse retrouver la confiance perdue du fait des « accidents », excès et biais des technologies antérieures moins conscientes de ces risques et des limites.

En tout cas, il semble raisonnable que lorsque nous sommes confrontés à des situations de changement ou de grande incertitude, nous soyons moralement plus stricts, surtout si nous pensons qu'il existe des dangers, des menaces, des dommages, des effets négatifs ou des risques dirigés vers nous, nos proches ou nos possibilités d'action. Il ne fait aucun doute que la vague du principisme veut affronter ce fait.

Dans tous les cas, les déclarations, guides, cadres, codes et principes sont devenus des documents qui nous permettront d'esquisser quels sont les problèmes moraux les plus préoccupants dans certains secteurs, ainsi que de considérer les conceptions de l'éthique et de la réflexion morale qu'ils envisagent.

Cette convergence pourrait créer une fausse idée de consensus et même conduire à une adoption unilatérale de principes sans traduction culturelle ou sans engagement pratique pour faire en sorte que les déclarations de principes convergent avec les tendances de développement de ce type de technologie, imprègnent toutes ses phases et faire partie de l'équipe humaine.

Un facteur supplémentaire est la place que prennent les principes. Dans des cas spécifiques comme la Colombie Le cadre éthique de l'intelligence artificielle en Colombie (2021), a le caractère d'une soft-law qui recommande plus qu'elle ne réglemente, et semble suggérer que plus qu'un mandat, ce sont les entreprises et les secteurs qui développent l'IA qui doivent s'autoréguler selon les recommandations que le cadre propose. C'est la stratégie qui consiste à suggérer des conditions minimales, à s'attendre à leur réalisation minimale, à surveiller et à intervenir peu ou pas du tout. Dans certaines approches plus commerciales, la responsabilité éthique est traitée comme un avantage concurrentiel de la marque (un sceau éthique comme les produits verts) ou est laissée dans le domaine de la réputation, probablement en mettant en place un certain classement. Si les pratiques de soft-law sont analysées dans un contexte de concurrence internationale dans le développement des technologies, elles pourraient désavantager les pays peu réglementés ou mettre en danger le respect des droits de l'homme. Mais, d'un autre côté, si la réglementation est très stricte dans un écosystème technologique qui veut se développer, la réglementation pourrait rendre difficile le respect d'entreprises ou de nouvelles initiatives. La seule motivation des réglementations n'est pas la transformation numérique des services et des biens, mais la croissance économique.

L'un des avantages de l'indice des capacités proposé par Oxford Insights tient à la possibilité pour les États de comparer

spécifiquement leurs capacités avec les autres. Il a l'inconvénient de tout indicateur lorsqu'il génère de la réputation : la cause se confond avec l'effet et sera après possible de travailler pour l'indicateur et non pour les finalités que l'indicateur aide à reconnaître et analyser critiquement. En revanche, il peut être trop abstrait lorsqu'il s'agit de mesurer certains facteurs. La façon dont il mesure, par exemple, la soi-disant représentativité des données ne comprend que des informations sur les écarts entre les sexes, mais ne prend pas en compte des questions telles que la gouvernance des données indigènes ou, en général, la gouvernance au sein de communautés politiques pluralistes ou non individualistes.

Il y a plusieurs risques que je voudrais résumer, non sans reconnaître au préalable que l'approche fondée sur les principes, les droits et les capacités est une formulation légitime de conditions pertinentes.

1- La convergence des principes ne signifie pas consensus universel, mais convergence dans l'approche des problèmes et dans l'histoire pratique des erreurs. Cela signifie qu'elle doit être traitée non seulement comme une question de génération de principes et de respect de règles, mais aussi comme une question de raisonnement pratique, puisque c'est le raisonnement pratique en interaction avec des machines intelligentes qui conduit aux plus grands risques.

2- Les déclarations de principes sont une démarche formelle qui ne garantit pas une pratique, beaucoup plus coûteuse dans le temps et qui implique des changements : intégrer les pratiques éthiques dans les projets, combler le fossé entre les sciences des données et les sciences humaines et sociales avec des approches comme celles des humanités numériques, et avancer au-delà des capacités d'infrastructure à capacités humains, de formation et d'interaction culturelle.

3. APPROCHES DES SAGESSES ARTIFICIELLES (ASAS)1

L'histoire de l'application de l'IA aux communautés humaines, à leurs droits et intérêts regorge d'innovations surprenantes par leur importance. Un exemple peut être trouvé dans le travail réalisé par l'équipe interdisciplinaire de scientifiques des données, de spécialistes des sciences sociales et d'humanistes de la Commission colombienne de la Vérité (2022) pour consolider et analyser une quantité énorme et diversifiée de registres sur le conflit armé colombien du dernier soixante ans, après la démobilisation des Forces Armées Révolutionnaires de la Colombie (FARC) en 2017. L'équipe d'innovation de cette commission avait pour tâche de consolider les sources d'information les plus diverses issues de la recherche universitaire, des rapports, des données du Centre national de la mémoire historique, des documents d'ONG internationales, des documents officiels d'institutions étatiques et de groupes de victimes dans un conflit non seulement prolongé mais aussi riche d'interprétations et d'enquêtes diverses au fil des ans.

Contrairement à d'autres conflits dans le monde où les sources de données et d'informations peuvent être insuffisantes et il faut commencer par les compiler et les enregistrer ; la Commission colombienne de la Vérité, disposait d'une énorme quantité de données et d'informations à traiter et à analyser. En plus ils avaient une énorme responsabilité morale et politique : adopter une attitude pluraliste pour écouter toutes les voix du conflit ; respecter les différentes sources d'information et leur confidentialité ; et ne pas être biaisé par la vision d'aucun des acteurs du conflit, comme on pense devraist faire l'action de ce type de commission.[79]

79 Dans les références bibliographiques il y a une bibliographie qui consolide les études, livres et travail retrouvés autour de chaque approche de sagesse artificielle proposée.

À l'analyse de ces sources secondaires, d'énormes tâches supplémentaires se sont ajoutées, comme le traitement des transcriptions des plus de 12 000 entretiens qui ont été réalisés avec des personnes ayant directement participé au conflit (chaque entretien a une durée en moyenne d'une heure et demie ou deux heures), et consolider une base de données unique des victimes dans un portail de microdonnées.

Au total, la commission a consolidé et analysé 350 bases de données différentes qui ont permis de construire un rapport qui a été remis le mois dernier à l'État colombien, à la société civile, aux victimes du conflit et à la communauté internationale. Pour ces tâches, une équipe de data scientists a utilisé des outils automatiques de traitement automatique du langage naturel (TALN), ainsi que des logiciels libres, utiles dans une tâche de traitement qui risquait de sombrer dans une mer d'informations. Il est également prévu qu'une fois que certaines garanties d'éthique des données auront été discutées et résolues, la Commission ne clôturera pas son rapport mais le mettra à la disposition des citoyens, de la communauté universitaire et d'autres pays qui ont besoin d'un transfert de technologie dans la gestion de leurs conflits.

Cet exemple d'initiative au service des droits de l'Homme et de la réconciliation sociale contraste avec d'autres cas d'utilisation de l'IA dans des contextes commerciaux comme Tay, le bot Twitter, qui en moins de trois jours d'interaction avec des internautes sur un réseau social largement reconnu pour sa manifestation de haine, il est devenu un agent conversationnel xénophobe et sexiste qui a contraint ses développeurs à le supprimer et a suscité de nombreuses interrogations sur l'éthique des bots et leur entraînement. Nous avons également le cas controversé de la voiture autonome Uber qui n'a pas reconnu un piéton et l'a écrasé, causant sa mort ; ou des systèmes d'IA utilisés par les banques, les hôpitaux ou les systèmes judiciaires qui ont refusé des services, des avantages, des droits ou ont imposé des charges plus lourdes aux populations appartenant à certaines zones ou groupes ethniques sur la base de données historiques, d'informations sociodémographiques ou de corrélations établies par le système.

Ces cas ont eu un poids important pour poser des questions sur l'éthique de l'IA ou pour établir des principes d'IA basés sur la garantie des droits de l'Homme. Ils sont aussi ce que l'on pourrait appeler la trace des effets pratiques de l'application de l'IA au sein des communautés humaines. Je considère qu'il y a trois aspects pertinents que nous pouvons récupérer de ces expériences pour continuer la réflexion :

1- Nous avons une histoire certainement descriptive des erreurs, des excès et des omissions de l'IA qui font partie d'une richesse de sagesse pratique accumulée qui doit être reconnue et comprise par les acteurs de l'écosystème humain-machine intelligent.

2- Nous reconnaissons que les biais algorithmiques et les biais cognitifs renforcés par les algorithmes ont des implications importantes dans la jouissance des droits, dans la sphère publique, dans l'interaction sociale, dans le développement cognitif, dans le type d'entraînement des machines artificielles et dans l'interaction homme-machine.

3- On est d'accord que des systèmes artificiels intelligents sont nécessaires pour ne pas reproduire, approfondir ou amplifier les injustices, les inégalités et la méconnaissance des systèmes sociaux actuels. De plus, on insiste sur le fait que l'IA devrait être une force pour prendre des décisions plus raisonnables, pour le bien-être et pour l'épanouissement humain (human flourishing).

4- En outre des capacités des acteurs

du développement technologique (par exemple, ceux proposés par le Government IA readiness Index), il est nécessaire de comprendre, en plus des compétences techniques, ce qui se passe avec les capacités humaines, en général, dans des contextes largement technologisés.

Lorsqu'on adopte une orientation de raisonnement pratique en la matière, on doit ajouter à l'adoption de principes réglementaires, de cadres normatifs ou de règles ; une série de questions complémentaires impliquant des savoir-faire et des savoir-comment à des niveaux allant de l'éthique appliquée à des questions de cadres culturels, d'éducation et de théories de la rationalité.

Diverses questions qui avaient commencé dans les réflexions sur la possibilité d'une Intelligence Artificielle Générale (IAG) (Boden, 2017), tendent à converger dans une série d'approches que j'appellerai Approches de Sagesses Artificielles (ASAs) et qui, bien qu'elles ne soient pas homogènes, partagent quelques traits généraux que je voudrais énumérer dans un premier temps et proposer ensuite une caractérisation initiale et, sûrement, provisoire de trois approches :

1- Les ASAs remettent en question, généralement à des fins d'expansion, les notions classiques d'intelligence qui sont à la base d'une bonne partie des principaux développements de l'IA. Ils avertissent que l'intelligence n'est pas seulement importante au sens stratégique (trouver les moyens appropriés pour atteindre une fin) mais aussi au sens pratique (être capable de réfléchir, de postuler des objectifs et de les peser).

2- Les ASAs analysent, comparent ou évaluent les différences et les complémentarités entre «intelligence» et «sagesse» à différents niveaux : aspects génétiques, neurophysiologiques, culturels et moraux.

3- Les ASAs sont particularistes et situées parce qu'elles reposent sur des réflexions précises et concrètes, sur des investigations empiriques, des études de cas et des exercices pratiques qui approfondissent la compréhension de la relation humain-machine, notamment en termes de type d'interaction, de formation, d'entraînement et coexistence qu'ils adoptent.

4- Les ASAs reconnaissent les différences culturelles et, au minimum, posent le type de coexistence qui pourrait exister entre les sagesses ancestrales, les sociétés plurielles, les traditions, la sagesse morale, le raisonnement moral et l'IA.

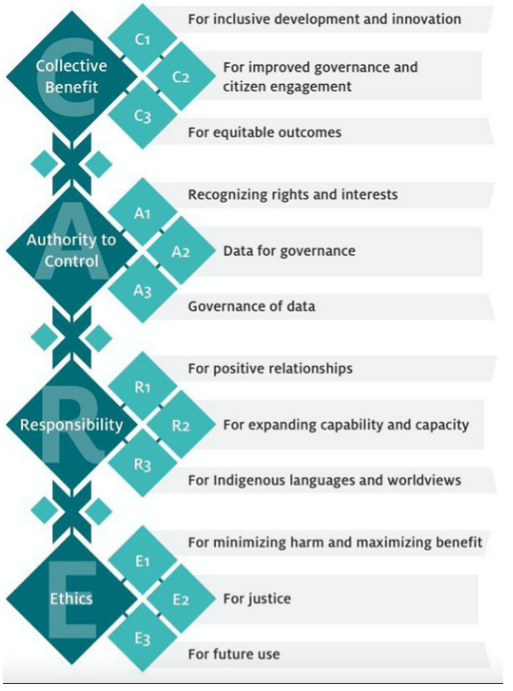
Gardant à l'esprit que l'articulation, je présente un schéma et fais une caractérisation provisoire, ouverte à la discussion, qui entend décrire les caractéristiques de chaque approche des ASAs. Cette orientation nous montrera des voies pour élargir les possibilités de participation des groupes sociaux et des cultures à une transformation omniprésente devant laquelle nous ne pouvons être ni spectateurs ni patients, mais acteurs clés, surtout lorsque nous vivons dans des sociétés plurielles.

Je propose de diviser les ASAs en trois approches : les Sagesses Artificielles Culturelles (SACs), les Sagesses Artificielles Formatrices (SAFs) et les Sagesses Artificielles Théoriques (SATs). Ces approches ne sont pas mutuellement exclusives. Leurs différences résident dans l'accent qu'elles mettent sur certains aspects plus que sur d'autres.

SAGESSES ARTIFICIELLES CULTURELLES (SACS)

Les questions éthiques dans le contexte

d'IA en relation directe avec des traditions culturelles ou les groupes communautaires c'est, peut-être, le segment de la discussion moins traité dans les approches principistes. L'article The global landscape of AI ethics guidelines de Anna Jobin, Marcello Ienca et Effy Vayena (2019) montre que le principe le moins développé et mentionné est la solidarité. Le projet LAIP (Linking Artificial Intelligence Principles) montre que les principes qui sont en relation avec la collaboration sont pensés, en général, plutôt comme collaboration académique ou entre développeurs. Cependant, on peut trouver quelques réflexions particulières sur la relation entre l'IA et les cultures ou des aspects particuliers de celles-ci comme les valeurs de l'Islam (Ali, 2011), l'incorporation du concept Navajo hozho dans les projets d'IA et comme example d'articulation entre IA et sagges indigen (Williams & Shipley, 2021), la relation entre IA, sagesse et philosophie chinoise (Song, 2021) ; ou des propositions concrètes sur la gestion des données des groupes sociaux comme les principes CREA pour la gouvernance des données indigènes (Carroll et al., 2020)[80].



Cet aspect réflexif insiste sur l'importance de savoir comment les différentes cultures et groupes sociaux pensent leur rapport à l'IA ou à l'un de ses éléments. D'autre part, il a donné forme à une série de discussions et de débats qui interrogent directement la direction que peut prendre l'IA dans ses relations non pas avec l'humanité abstraite mais avec les multiples modes de vie qui se déroulent dans des sociétés plurielles et hétérogènes. Les traditions culturelles ont non seulement été les garants de la mémoire de l'humanité et de la spiritualité de ses peuples, mais elles sont aussi les dépositaires de pratiques de sagesse durables et d'une énorme profondeur morale.

Les principales caractéristiques des SACs sont les suivantes :

1- La conception de l'intelligence dans les projets d'IA traditionnels peut être ethnocentrique, puisqu'elle ne s'est pas seulement concentrée de manière privilégiée sur une seule tradition de pensée (en laissant de côté, par exemple, la solidarité, l'empathie, la

80Le principes CREA en espagnol (Contrôle (Autorité pour contrôler), Responsabilité, Éthique, Avantage pour le bénéfice collective) et CARE (Collective benefit, Authority to control, Responsibility, Ethics) en aglais, présentent un modèle qui va au-delà des principes FAIR (Findable, Accessible, Interoperable, Reusable) de partager et enregistrement des données, à fin de proposer des critères pour la gestions de données qui soient applicables au communautés indigènes, ses droits collectifs, ses autorités traditionnelles et ses droits d'autodétermination. Il s'agit d'une initiative qui recommande le soin de la tradition : "Be CARE to be FAIRE", «Être soignée pour être juste».

méditation et la spiritualité), mais s'est concentrée uniquement sur certains aspects de l'intelligence (les plus analytiques et stratégiques) laissant de côté la sagesse pratique des cultures qui pourrait ou devrait être prise en compte (les aspects les plus évaluatifs et ceux liés au raisonnement pratique).

2- Les valeurs traditionnelles, communautaires et collectives peuvent être alignées sur l'IA (elles peuvent être des guides moraux) et doivent être prises en compte lorsque ces technologies interviennent dans les modes de vie des communautés.

3- Les informations et les données sur les cultures et les groupes sociaux (CREA c'est l'exemple le plus clair dans l'actualité) ne sont pas neutres. Elles ont été compilées, structurées et systématisées à des moments précis, généralement par des observateurs extérieurs aux communautés et avec des intérêts qui peuvent restreindre ou déformer la vision et la richesse d'une culture. Les communautés ont le droit d'exercer une gouvernance sur les données qui interprètent leur identité et leurs modes de vie ; elles peuvent construire des stratégies d'auto-gouvernance sur leurs données et exercer une autorité sur celles-ci, au moins de deux manières : (i) parce qu'elles peuvent aussi être productrices et autrices de données (non seulement doivent-elles faire l'objet de rapports de tiers, comme cela a été l'habitude), (ii) parce qu'elles peuvent contrôler et délibérer sur quoi et comment elles seront utilisées ou montrer si elles déforment ses identités.

4- Les cultures et les traditions ont le droit de choisir, de connaître et d'accepter le type de coexistence qu'elles auront avec les machines intelligentes ; ainsi que de déployer leur patrimoine culturel pour exercer des médiations, des filtres et des limites sur les modes d'implication et d'interaction dans les écosystèmes numériques. Les cultures sont un tamis d'acceptabilité face aux changements et aux nouveautés qui doivent être médités, délibérés et convenus non seulement de manière individuelle, mais collectivement.

SAGESSES ARTIFICIELLES FORMATIVES (SAFS)

Les nouvelles technologies de l'information et de la communication (TIC) impliquent de nouveaux défis de formation, ainsi que certaines attitudes envers la connaissance et la recherche de ce qui se passe dans des environnements d'interaction hautement technologiques et numériques, appelés par certains avec le néologisme d' Infosphere (Floridi, XX).

Les préoccupations en matière de formation, qui pouvaient auparavant se cantonner aux sphères traditionnelles comme la famille, la culture ou l'éducation, se sont propagées aux sphères numérisées comme les réseaux sociaux et internet, parfois assistées par l'IA. Certains systèmes éducatifs comme le système Chinois (Song et al., 2021) et systèmes de formation militaire incluent les systèmes d'IA dans l'éducation et l'entraînement. Dans un cadre plus large, l'importance de la formation à la citoyenneté numérique est reconnue, et l'attention a été attirée sur l'impact du monde algorithmique sur notre cognition, sur nos représentations sociales et sur le type d'interactions que nous faisons dans la vie sociale et politique (Bronner , 2021, 2022).

D'autre part, il est important d'établir les connaissances pratiques et évaluatives des data scientists (au-delà de leurs connaissances théoriques et techniques) lors de la conception et de la formation de machines intelligentes. Wam Kim et Mejía (2019) proposent de rapprocher les ingénieurs d'un modèle de sagesse socratique qui reconnaît, outre le principe de

non-préjudice, le principe d'épanouissement humain et la faillibilité du savoir, c'est-à-dire qu'ils rapprochent leur profession non seulement à la conception sinon à l'engagement envers l'enseignement. D'autre part, Keijers et Bartneck (2021) ont attiré l'attention sur l'importance de comprendre l'interaction quotidienne que nous avons avec les agents conversationnels automatiques, après avoir identifié la tendance qu'ont les êtres humains lorsqu'ils interagissent avec les bots en les profanant avec des insultes et des commentaires sexistes sans aucune provocation évidente de son côté. Il s'agit d'un aspect central dans l'étude de l'interaction homme-machine, qui peut être pertinent lors de l'établissement de protocoles d'entraînement prenant en compte le comportement humain vis-à-vis des agents conversationnels artificiels et les environnements de rencontre entre humains et machines.

De leur côté, Cazes et al. (2021) ont étudié non pas tant les solutions techniques qui ont été proposées pour éviter les biais dits algorithmiques, mais plutôt les niveaux de conscience que les concepteurs de systèmes d'IA ont vis-à-vis de leurs propres biais cognitifs (biais de conformité, biais de confirmation et biais de corrélation illusoire) et l'impact que ces biais peuvent avoir sur leur travail ou les systèmes qu'ils développent. Il n'est pas exagéré de dire que lorsque nous nous trouvons dans un environnement où l'intelligence est un objectif, nous devons améliorer notre compréhension de ce que sont l'intelligence et le raisonnement dans des situations concrètes : nous ne devons pas seulement faire des systèmes intelligents, mais aussi améliorer notre compréhension de ce qui est l'intelligence. Il peut s'agir d'une pratique prudente qui évite stéréotyper l'idée d'intelligence et les responsabilités qu'on a avec l'intelligence dans un segment étroit de toutes ses capacités possibles. et verrouille les projets au niveau du faire sans réfléchir à ce que nous faisons ni pourquoi nous le faisons. Ceci, à son tour, nous montre des options pour articuler les sciences sociales et humaines.

Le projet Machine Wisdom travaille actuellement sur un cadre pour définir et améliorer la notion de sagesse pratique dans les interactions homme-machine (Brett et Allen, 2022) suivant la ligne de la soi-disant « rationalité computationnelle » (Brett, 2022). A noter également le travail de Reijers el al. (2017) dans lequel une revue extensive des pratiques éthiques dans les projets de recherche et d'innovation est faite, caractérisant trois grandes orientations : (i) les pratiques éthiques qui font face aux incertitudes du changement technologique, (ii) les pratiques éthiques qui permettent, organisent et assurent l'éthique la conception technologique, et (iii) les pratiques éthiques qui identifient, analysent et résolvent les impacts éthiques des technologies. C'est un travail précieux qui montre dans divers domaines et de manière opérationnelle ce qu'une équipe de pratiques éthiques peut faire dans des projets de développement technologique et allant au-delà de l'adoption de principes pour participer à des phases ou à tous les projets de développement. Une partie de cette équipe est à l'avant-garde du projet Online Ethics Canvas .

Plus précisément, la SAF a appelé à combler les écarts entre les connaissances théoriques, l'expertise technique et les connaissances pratiques et la sagesse, pour engager notre moralité (personnelle, professionnelle ou communautaire), le contexte éducatif, l'éducation, la formation et l'interaction homme-machines intelligentes. Les principales caractéristiques de cette approche sont les suivantes :

1- Elle élargit le spectre des préoccupations morales centrées sur le principe de non-malfaisance pour souligner que l'IA a un engagement envers l'épanouissement et l'excellence humains. Elle propose également des modèles d'éthique pratique conçus pour les projets d'innovation dans lesquels le raisonnement moral ne se concentre pas sur la délibération individuelle avant ou après le projet, mais devient au

contraire partie intégrante des étapes.

2- Elle prône un modèle d'interaction humain-machine intelligent moins axé sur l'entraînement et plus axé sur la formation (réflexion évaluative, compétences métacognitives, empathie, reconnaissance des limites, des erreurs et des lacunes).

3- Elle reconnaît la nécessité d'élargir la participation et d'améliorer la connaissance de ce que nous faisons lorsque nous interagissons dans des environnements numériques, proposant d'élargir les compétences générales en matière de citoyenneté numérique (comportement humain par rapport aux machines intelligentes, connaissance et reconnaissance de ce que sont les biais cognitifs et comment ils fonctionnent et biais algorithmiques).

4- Elle analyse, dans des contextes réels, les comportements d'interaction humain-machine dans des cycles qui vont de la conception à la mise en œuvre et à l'interaction réelle dans un écosystème.

SAGESSES ARTIFICIELLES THÉORIQUES (SATS)

Des orientations précédentes, on peut dire que la SAC a joué un rôle important dans la remise en question des visions de l'intelligence qui ne valorisent pas suffisamment l'histoire culturelle et les modes de vie pluriels dans des sociétés non homogènes ; tandis que la SAF a observé et décrit des transformations, des caractéristiques et des risques formateurs dans des contextes particuliers d'interaction humain-machine, ainsi que proposé certaines pratiques. Bien que la SAC insiste davantage sur les cultures et que la SAF se concentre sur des aspects spécifiques de la cognition morale, de la cognition sociale, de la formation et de l'interaction, elles sont identifiées en mettant en évidence la pertinence du raisonnement pratique dans sa relation avec l'IA.

Les Sagesses Artificielles Théoriques (SAT), sont axées sur des discussions philosophiques (reconstruction conceptuelle et évaluation des conditions de possibilité) et théoriques (modèles pour une SA) qui interrogent sur le raisonnement pratique ou la sagesse par rapport à l'IA, pas comme les deux SA précédentes, mais pour déterminer si elles sont possibles et comment une SA est, dans le sens où l'IA elle-même est sage, capable de porter des jugements de valeur, de délibération morale et de raisonnement pratique. Pour cette raison, la SAT est étroitement liée aux discussions sur l'Intelligence Générale Artificielle (IAG), les problèmes de singularité que pourrait générer une Super Intelligence, ainsi qu'aux études et modèles sur les Agents Moraux Artificiels (AMAs).

Tout comme la SAC élargit la notion de sagesse en termes de communautés et de cultures, la SAT élargit la notion en termes théoriques. D'autre part, elle ouvre la voie à des recherches sur la manière dont les machines jugent ou décident d'un point de vue de savoir pratique.

Certaines caractéristiques que nous pourrions caractériser de cette orientation sont les suivantes :

1- Elles comparent et analysent l'intelligence et la sagesse selon des niveaux : génétique, neurophysiologique, psychologique et praxéologique.

2- Elles proposent des modèles mathématiques, des cadres conceptuels et philosophiques qui expliquent comment le développement d'une SA (y compris les AMA) pourrait être possible ou non.

3- Elle reconstruit les théories de l'action et les théories du jugement (morales et juridiques) afin de déterminer si

les conditions et les caractéristiques de l'action ou du jugement chez l'être humain sont techniquement réalisables dans des machines.

4- Elles analysent et cherchent à reproduire la sagesse et la moralité humaines en termes informatiques, en soumettant à l'examen et à l'analyse les conceptions traditionnelles de ce qu'est un agent moral, de ce que signifie l'autonomie et de la relation dans le jugement moral entre le langage et les émotions.

3. DISCUSSION

Les Approches des Sagesses Artificielles (ASAs) peuvent complémenter les Approches Principistes basées sur les Droits et les Capacités (AP-DC) dans des contextes de raisonnement et de décisions pratiques qui ne peuvent pas être supposés, mais recherchés et compris en relation directe et situées avec une culture politique et sociale spécifique, les nouvelles pratiques de formations et les nouvelles interactions socio-techniques.

Les Approches sont comme des mappes de navigation qui à la fois créent un champ d'observation, mettent au premier plan certains phénomènes et laissent d'autres dans le fond. L'arrière-plan n'est pas un bruit de fond mais contient des voix qui doivent être mieux entendues. Pour aller au-delà de l'AP-DC, sans nier son importance, on doit se protéger d'un risque : Supposer que la convergence dans les principes est une convergence dans les actions et une compréhension des défis de raisonnement pratique.

Ma proposition est simple et concrète : les principes sont exigés parce qu'il y a une histoire des excès, erreurs pratiques, violations ou traitements discutables des individus ou communautés ; ça veut dire qu'il y a des problèmes de raisonnement pratique qui ont besoin de régulation, mais aussi de réflexion. On ne peut pas résoudre ou aborder des problèmes de raisonnement pratique sans un contexte réel d'interaction (social, culturel, d'interaction technologique) qui ne doit pas être supposé abstraitement, mais reconstruit. Les ASAs sont le visage pratique et opérationnel de ce défi.

RÉFÉRENCES

La bibliographie est divisée et organisée selon les approches AP-DC et les ASAs. postérieures sont nécessaires.

AP-DC

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751-752.

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. arXiv preprint arXiv:1812.04814.

Donahoe, E., & Metzger, M. M. (2019). Artificial intelligence and human rights. Journal of Democracy, 30(2), 115-126.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1).

Héder, M. (2020). A criticism of AI ethics guidelines. INFORMÁCIÓS TÁRSADALOM: TÁRSADALOMTUDOMÁNYI FOLYÓIRAT, 20(4), 57-73.

Oxford Insights. (2021) Government AI Readiness Index 2021.

ASAs

Sagesses Artificielles Culturelles (SACs)

San Segundo, R. (2002). A new concept of knowledge. Online Information Review.

Ziaee, A. A. (2011). A philosophical approach to artificial intelligence and Islamic values. IIUM Engineering Journal, 12(6).

Jifa, G. (2013). Data, information, knowledge, wisdom and meta-synthesis of wisdom-comment on wisdom global and wisdom cities. Procedia Computer Science, 17, 713-719.

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., ... & Hudson, M. (2020). The CARE principles for indigenous data governance.

Gupta, N., Blair, S., & Nicholas, R. (2020). What we see, what we don't see: data governance,

archaeological spatial databases and the rights of indigenous peoples in an age of big data. Journal of Field Archaeology, 45(sup1), S39-S50.
Hofkirchner, W. (2020). Intelligence, Artificial Intelligence and Wisdom in the Global Sustainable Information Society. Multidisciplinary Digital Publishing Institute Proceedings, 47(1), 39.
Williams, D. H., & Shipley, G. P. (2020). Enhancing artificial intelligence with indigenous wisdom. Open Journal of Philosophy, 11(1), 43-58.
GOFFI, E. R., COLIN, L., & BELOUALI, S. (2021). Ethical Assessment of AI Cannot Ignore Cultural Pluralism: A Call for Broader Perspective on AI Ethic. Arribat-International Journal of Human Rights Published by CNDH Morocco, 1(2), 151-175.
Song, B. (Ed.). (2021). Intelligence and Wisdom: Artificial Intelligence Meets Chinese Philosophers. Springer Nature.
Walter, M., Kukutai, T., Carroll, S. R., & Rodriguez-Lonebear, D. (2021). Indigenous data sovereignty and policy (p. 244). Taylor & Francis.
Belouali, S., Demazeau, Y., Bouchentouf, T., & Brahmi, H. (2022). Interrogation de l'universalité d'une éthique pour l'usage de l'IA au travers d'une lecture du code publié par Dubaï. Ethique et Numérique, 1(1), 1-17.
Sagesses Artificielles Formatives (SAFs)

Mancuso, J. J. (1995).

Electronic Journals. The

Serials Librarian, 26(1),

17-22.

Allen, C., & Wallach, W.

(2011). Wise machines?.

On the Horizon.
Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., & Gordijn, B. (2018). Methods for practising ethics in research and innovation: A literature review, critical analysis and recommendations. Science and engineering ethics, 24(5), 1437-1481.
Kim, T. W., & Mejia, S. (2019). From artificial intelligence to artificial wisdom: what Socrates teaches us. Computer, 52(10), 70-74.
Marzukhi, S., Zainol, Z., Muhamed, H., Awang, N. F., bin Tengku Sembok, T. M., & Juhary, J. B. (2019, September). Framework Of Malay Intelligent Autonomous Helper (Min@ H): Text, Speech And Knowledge Dimension Towards Artificial Wisdom For Future Military Training System. In 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 83-87). IEEE.
Bronner, G., Cayrol R., Cordonier, L., Douzet, F., Farinella, R., Favre, A., Garrigues, J., Harfouch, F., Khan, R., Muxel, A., Reichstadt, R., Roder, I., Warusfel, B., Wieviorka, A. (2021) Les lumières à l'ère numérique.
Cazes, M., Franiatte, N., Delmas, A., André, J. M., Rodier, M., & Kaadoud, I. C. (2021, July). Evaluation of the sensitivity of cognitive biases in the design of artificial intelligence. In Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA'21) Plate-Forme Intelligence Artificielle (PFIA'21).
Keijsers, M., Bartneck, C., & Eyssel, F. (2021). What's to bullying a bot?: Correlates between chatbot humanlikeness and abuse. Interaction Studies, 22(1), 55-80.
Karlan, B., & Allen, C. (Preprint) Engineered Wisdom for Learning Machines. Journal of Experimental and Theoretical Artificial Intelligence.
Sagesses Artificielles Théoriques (SATs)
Sanz, R., Bermejo, J., Escasany, J., Chinchilla, R., & Garcia, C. A. (2003, October). Meaning generation and artificial wisdom. In IEMC'03 Proceedings. Managing Technologically Driven Organizations: The Human Side of Innovation and Change (IEEE Cat. No. 03CH37502) (pp. 401-405). IEEE.
Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. Minds and machines, 14(3), 349-379.
Amie, A. S. (2012, December). A perspective to the artificial wisdom Possibility of self-programmable artificial intelligence for human like intelligence in robotics. In 2012 Annual IEEE India Conference (INDICON) (pp. 183-188). IEEE.

Casacuberta Sevilla, D. (2013). The quest for

artificial wisdom. AI & society, 28(2), 199-207.

Sarkar, A. (2015). Unlocking the Quest for

Artificial Wisdom as Integration of Artificial Intelligence with Desire. Science Research, 3(3), 79.
Davis, J. P. (2019). Artificial wisdom? A potential limit on AI in law (and elsewhere). Okla. L. Rev., 72, 51.

Jeste, D. V., Graham, S. A., Nguyen, T. T., Depp, C. A., Lee, E. E., & Kim, H. C. (2020). Beyond artificial intelligence: exploring artificial wisdom. International psychogeriatrics, 32(8), 993-1001.
Nusbaum, H. C. (2020). How to make Artificial Wisdom possible. International Psychogeriatrics, 32(8), 909-911.
Tsai, C. H. (2020). Artificial wisdom: a philosophical framework. AI & SOCIETY, 35(4), 937-944.
Lekka-Kowalik, A. (2021). Morality in the AI World. Law and Business, 1(1), 44-49.
Nagabhushan, P., Sonbhadra, S. K., Punn, N. S., & Agarwal, S. (2021, December). Towards Machine Learning to Machine Wisdom: A Potential Quest. In International Conference on Big Data Analytics (pp. 261-275). Springer, Cham.
Sarkar, A. (2021). Quantum Mechanical Postulates towards Realization of Artificial Wisdom on Incorporating Desire in Intelligence: A Hypothetical Study. Journal of Computer and Communications, 9(12), 63-79.
Sison, A. J. G., & Redín, D. M. (2021). A neo-aristotelian perspective on the need for artificial moral agents (AMAs). AI & SOCIETY, 1-19.
Zhang, K., Shi, J., Wang, F., & Ferrari, M. (2022). Wisdom: Meaning, structure, types, arguments, and future concerns. Current Psychology, 1-22.
Karlan, B., & Allen, C. (Preprint) Quantum of Wisdom.

Yves Demazeau

Directeur de recherche au CNRS et ex-président de l'Association Française de l'Intelligence Artificielle, France

Conception éthique de systèmes d'IA de très grande taille

EXPERIENCE PERSONNELLE EN IA

Travaux en IA depuis 1982. Thèse en 1986. Entrée au CNRS en 1988. Pas de peur de l'IA, réponse à la peur.

1990 --- 2010 Cofondateur de la communauté de recherche des Systèmes Multi-Agents en France (JIADSMA), en Europe (MAAMAW), et à l'international (ICMAS).

1995 - Développement de la méthode de conception de Systèmes Multi-Agents : VOYELLES (éthique à suivre)

2010 --- 2020 Promoteur des applications pratiques des SMA (PAAMS, 2020 confiance) et refondateur de la communauté IA en France (AFIA dont Ethique & IA)

2018 - Soutien au développement de l'IA à l'UM1 Oujda.

Ethique & IA : universalité et spécificité du Maroc.

ASPECTS ETHIQUES DANS VOYELLES

Ni anthropomorphisme, ni apparence d'émotion

1990 --- 2000 Des SMA utilisables aux SMA utiles A, E, I, O, émergence, récursion

100 agents

2000 --- 2010 Des SMA utiles aux SMA utilisés U, évaluation

10 000 agents

2010 --- 2020 Des SMA utilisés aux SMA éthiques Y, explicabilité

1 000 000 agents

Mécanismes de protection des données personnelles

[Crepin 09] « Systèmes Multi-Agents Hippocratiques : Mécanismes sociaux entre agents pour la protection de la sphère privée ». Thèse de Doctorat UJM.

[Piolle 09] « Agents Utilisateurs pour la Protection des Données Personnelles » Thèse de Doctorat UJF. Mention Spéciale Prix de thèse CNIL.

Éthique de conception des SMA de grande taille

[Demazeau 14] "Ethics and the Design of User-centred Multi-Agent Systems", 1st International Symposium on Ethics in Engineering, Science and Technology, Ethics'2014, IEEE, Chicago

[Demazeau 16] "Balancing Legal and Moral Aspects in Large-Scale Agent-Based Systems", 2nd International Symposium on Ethics in Engineering, Science and Technology, Ethics'2016, IEEE, Vancouver

ELEMENTS DE DISCUSSION

Pas de peur de l'IA, réponse à la peur. La seule peur serait l'IA forte, mais on en sera toujours très loin.

A chaque concepteur ses principes et valeurs de base, dans le respect de son contexte.

Régulation de la conception plus que des usages UNESCO (plus que OCDE) □ UE □ France UNESCO (plus que OCDE ?) □ ? □ Maroc

Approche ascendante, partant de l'humain et des problèmes à résoudre, des phénomènes à simuler.

Limiter valeurs et principes à ceux qui sont concernés.

Publications : positionnement éthique, mais aussi mise à disposition des données et des programmes.

Je voudrais remercier le Conseil national

des droits de l'Homme pour m'avoir invité.

J'avoue ne pas être chercheur en éthique de l'IA et ne pas être plus concerné que cela, au-delà de mon engagement personnel pour la protection des droits humains. Cependant, je me sens très concerné par cette question. Par ailleurs, je ne dirais pas que je suis performant, mais au moins éduqué dans ce domaine. Cela fait presque 40 ans que je travaille dans l'intelligence artificielle, et je ne sais faire que ça.

Je vais vous parler, sur le plan social, de mon parcours après 40 ans de travail dans ce domaine, en abordant une méthode pour concevoir des systèmes d'intelligence artificielle de très grande taille. Nous discuterons un peu plus tard du côté éthique, qui est lié à beaucoup de choses. En fait, je souhaite partager les différentes étapes de ces 40 années où j'ai été amené à me poser des questions sur la conception éthique des systèmes d'IA.

Comme je l'ai mentionné, cela fait 40 ans que je fais de l'IA. Il y a 40 ans, la population mondiale comptait 4,5 milliards d'habitants, et nous venons juste de dépasser les 8 milliards. Il faut donc relativiser cette situation. Je passe régulièrement mon temps à répondre à la peur que l'IA suscite parmi les populations, les chercheurs, les institutions et les penseurs. Cela a été au cœur de mon engagement vis-à-vis de l'AFIA, l'Association Française pour l'Intelligence Artificielle, qui a pour mission de promouvoir et favoriser le développement de l'IA en France. J'ai été amené à intervenir à ce titre il y a une dizaine d'années, lors de ce nouvel été de l'IA, pour apaiser les inquiétudes de nombreuses municipalités et citoyens en France.

Je pense aussi profondément que l'intérêt croissant pour l'éthique de l'IA est une réponse à cette peur. Une peur alimentée par une part importante de la population qui est encore ignorante des enjeux de l'IA. Cela va encore durer longtemps, car par exemple, en France, on commence à aborder le sujet de l'IA seulement après le baccalauréat. Nous ne sommes pas en Corée du Sud, et il faudra du temps avant que les gens prennent pleinement conscience des implications de l'IA. La peur de l'IA est également alimentée par ce qui se passe aux États-Unis, en Chine et dans d'autres pays ces dix dernières années.

En réalité, cette peur provient aussi du fait que depuis la naissance de l'IA, la discipline a beaucoup de liberté, car l'objectif de l'IA est de reproduire le comportement humain. A priori, l'objet scientifique n'a pas de limitations. Cependant, en tant que chercheurs, nous avons nos propres limitations, ainsi que des restrictions qui nous sont imposées. Il reste encore beaucoup de travail à faire pendant de nombreuses années. Il n'y a pas de limitation dans la recherche de la reproduction du comportement humain, ce qui fait que l'IA devient de plus en plus performante chaque année.

Il est vrai que de temps en temps, il y a des usages inappropriés de l'IA, tout comme il y a de mauvais usages de la science. Ces usages problématiques résultent également de nombreuses conceptions erronées de systèmes d'IA. S'il y avait plus d'efforts pour limiter les contraintes au niveau de la conception des systèmes d'IA, nous aurions moins besoin de légiférer sur leurs usages.

Dans ma communauté, je suis connu comme l'un des cofondateurs de la recherche sur les systèmes multi-agents en France. Ces systèmes, qui constituent l'une des dix disciplines reconnues de l'IA, s'intéressent non pas à un seul système d'IA, mais à plusieurs systèmes susceptibles d'interagir entre eux. Cela signifie qu'un système d'IA n'est déjà pas simple, mais que plusieurs systèmes interagissant entre eux rendent la situation plus complexe, même si cela peut parfois être plus simple. Il y a différents types d'agents : l'agent utilisateur, qui est la personne qui va utiliser le système, et l'agent concepteur. Actuellement, ces deux agents sont des êtres humains, et nous parlons donc de droits humains, en nous référant essentiellement à l'utilisateur et

au concepteur, sans aborder les droits des machines.

Ainsi, depuis 1995, j'ai développé une méthode de conception de systèmes multi-agents appelée Voyelles. J'ai abordé des aspects très théoriques liés à des systèmes multi-agents, en développant des logiques non computationnelles. En tant qu'informaticien, cela m'a semblé un peu compliqué, car je préfère que les choses soient appliquées. Bien que des théories non computationnelles soient intéressantes, elles manquent de praticité. Pendant dix ans, j'ai donc promu les applications pratiques des systèmes multi-agents en créant une conférence appelée PAAMS.

En 2020, lors de PAAMS, nous avons abordé le sujet de la confiance dans les systèmes multi-agents. J'ai également contribué à fonder la communauté en France entre 2010 et 2020. Avec l'AFIA, nous avons organisé trois journées nationales en 2014, 2015 et 2016 sur l'Éthique et l'IA : la première était une initiative de l'association, la deuxième a été réalisée en coopération avec le CNRS, et la troisième sous l'égide de l'INRIA. Nous avons arrêté ces événements parce que, chaque année, je constatais avec d'autres que les discours se répétaient. On m'a suggéré de laisser passer du temps, peut-être que lorsque le sujet sera plus mûr, nous pourrons en reparler. Pour l'instant, je ne sais pas quand aura lieu la prochaine rencontre, et en France, le sujet est moins discuté, chacun ayant sa propre opinion.

Cela fait quinze ans que je suis professeur invité au Danemark, et j'ai travaillé pendant dix ans avec des chercheurs espagnols et polonais, ce qui m'a ancré dans l'Union européenne. Je suis un fervent défenseur de la construction européenne. J'ai également collaboré avec des chercheurs brésiliens, argentins, japonais, et actuellement avec ceux de l'Université Mohammed Ier à Oujda, au Maroc.

Nous avons évoqué plus tôt l'universalisme de l'IA et la spécificité du Maroc en ce qui concerne un code de l'IA. Je pense que, quarante ans après, je suis devenu beaucoup plus tolérant envers les autres civilisations, cultures et religions. J'ai tendance à penser maintenant que chacun a raison, et il faut trouver un moyen de satisfaire tout le monde tout en se protégeant.

Dès le début de mes recherches, il y a quarante ans, je me suis imposé une limite éthique : je ne concevrai jamais de robots anthropomorphes, et je ne laisserais jamais croire à l'utilisateur que mes systèmes d'intelligence artificielle peuvent ressentir des émotions. Cela représente des barrières que je me suis fixées pour des raisons personnelles et des convictions profondes.

Au départ, les gens doutaient un peu de la notion de système multi-agents. Nous avons développé des systèmes exploitables, et l'industrie s'est rapidement emparée de nos travaux, aboutissant à des systèmes multi-agents utiles pour l'économie. Au bout de dix ans, je me suis rendu compte que je travaillais principalement pour des industriels, sans me soucier suffisamment des utilisateurs finaux.

J'étais assez choqué par cette notion d'« utilisateur final ». Pourquoi sont-ils toujours considérés comme secondaires ? Peu à peu, j'ai ramené l'utilisateur du fond des produits auxquels il était soumis à l'origine de la conception des produits qu'il fallait construire pour lui. J'ai recentré ma conception sur l'utilisateur, en intégrant une double évaluation : jusqu'en 2000, je faisais des systèmes avec validation informatique. À partir de 2010, nous avons également commencé à demander aux utilisateurs ce qu'ils pensaient de notre système.

Nous avons donc introduit l'utilisateur dans le processus de conception dès le début, et je pense que c'est quelque chose qu'il faut systématiser. C'est également à cette période que j'ai commencé à m'intéresser à la protection des données personnelles. J'ai dirigé deux thèses qui se sont terminées en 2009. À un moment donné, le concepteur a besoin d'interroger l'utilisateur sur ses données personnelles, et nous nous

sommes demandé ce que le système allait faire de ces données. Nous voulions protéger l'utilisateur et avons imaginé des mécanismes pour le faire.

Dans un premier temps, nous avons intégré dans le système une représentation minimale de l'utilisateur nécessaire pour résoudre un problème dans un contexte particulier. De plus, nous avons demandé au système de ne pas conserver l'information. À l'époque, nous avons suivi deux approches : l'une, développée par Ludivine Crépin, était de faire fonctionner le système multi-agents sur le principe hippocratique, où tous les partenaires s'engagent à respecter des conventions et des normes, y compris celle de ne pas conserver les informations longtemps. Cela fonctionne bien entre amis, mais pas toujours dans un contexte où il y a des producteurs et des consommateurs.

Nous avons donc créé, entre 2005 et 2009, des agents utilisateurs pour la protection des données personnelles. Ces agents, qui prenaient la forme d'avatars ou de jumeaux numériques, étaient attachés à chaque personne. Ils interagissaient avec les agents producteurs ou d'autres systèmes embarqués, négociant des contrats en permanence entre l'utilisateur et les différents partenaires du système, ce qui permettait de contrôler la diffusion de l'information.

Cela devient plus difficile à expliquer, d'abord parce que je suis à la fin de cette période. Peu à peu, je me suis rendu compte qu'au-delà des données personnelles, d'autres problématiques émergeaient, notamment en travaillant avec des adolescents de dix à quatorze ans. Certaines applications me faisaient réfléchir sur des enjeux plus larges. À ce moment-là, j'ai commencé à vouloir limiter les capacités des concepteurs à ne concevoir que des choses qui soient autorisées et éthiquement responsables.

J'ai alors commencé à interagir avec les agents que je créais pour explorer les différences entre ce qui pouvait être attendu et ce qui était autorisé, ce qui a donné lieu à des travaux sur l'explicabilité.

Enfin, pourquoi aborder la question de la grande taille ? Ce n'est pas seulement parce que nous sommes passés de 4,5 milliards à 8 milliards d'habitants, mais parce qu'en quarante ans, j'ai travaillé avec une centaine d'agents, avec 10 000 agents il y a trente ans, et avec un million d'agents il y a dix ans. Mon souci est maintenant de comprendre quels problèmes je peux résoudre avec 100 millions d'agents que je ne peux pas résoudre avec seulement un million.

Jusqu'en 2014-2016, je voudrais mentionner deux publications lors des premiers et deuxièmes symposiums « Ethics in Engineering, Science and Technology », organisés par l'IEEE à Chicago et Vancouver. C'était la première fois que j'interagissais avec des philosophes, puisque ces conférences étaient organisées par le chapitre de philosophie de l'IEEE. Il n'y a pas eu de suite après, mais cela a marqué ma première rencontre avec ce domaine.

En conclusion, je n'ai toujours pas peur de l'IA. En réalité, ma seule crainte serait liée à des systèmes d'IA considérés comme « IA forte », mais nous en sommes encore très loin, et nous le resterons probablement. Il n'y a donc pas vraiment de raison de s'inquiéter à ce sujet.

Le message que je veux faire passer, c'est que peu importe les régulations ou législations en place, chaque concepteur de systèmes d'IA a toujours ses principes et ses valeurs de base, qu'il ne transgressera jamais. Il est également essentiel de comprendre l'aspect humain, car le concepteur agit dans le respect de son contexte.

Il existe des recommandations de l'UNESCO qui ne sont pas contraignantes. Nous pouvons regretter que les systèmes d'IA n'aient pas d'existence juridique, mais cela n'empêchera pas, à court terme, qu'au Japon, par exemple, on décide de mettre en place une responsabilité juridique.

En tous les cas, je suis guidé par un schéma assez simple. J'apprécie beaucoup les principes et les valeurs des recommandations de l'UNESCO, plus que ceux de l'OCDE, car il me semble que le cadre de l'UNESCO est plus inclusif. Cependant, il y a des entreprises comme Raytheon et celles qui fabriquent Pegasus, qui ne respectent pas nécessairement ces recommandations.

Concernant l'OCDE, il est dommage qu'il n'y ait pas la Chine, car plus de la moitié de l'économie et des efforts mondiaux en intelligence artificielle proviennent de ce pays. Je pense qu'il est essentiel de s'occuper des aspects éthiques. Ce n'est pas un dogme, mais lors des rencontres, il est nécessaire d'être capable de les maîtriser. Nous parlons de l'humain et des problèmes à résoudre. En fin de compte, nous choisissons les valeurs et les principes, peut-être ceux de l'UNESCO, qui sont pertinents pour l'application que nous développons. Ce n'est pas la peine de vouloir tout réaliser ou satisfaire tout le monde en même temps.

Ainsi, il est pratique de limiter les valeurs et principes à ceux qui sont directement concernés par les problèmes que nous tentons de résoudre. C'est une recommandation que je fais souvent dans le domaine de la recherche. Je n'ai pas d'impact direct, mais j'encourage chaque chercheur à inclure, à la fin de son article, deux ou trois lignes sur le positionnement éthique de ses travaux. Je plaide également pour que chaque chercheur mette à disposition ses données et ses algorithmes pour des questions de reproductibilité.

Elizabeth D. Gibbons

Instructor, FXB Center for Health & Human Rights, Harvard University, USA

Critique of AI codes of ethics from a human rights perspective

Good afternoon. It is a great honor to have been invited to address participants at this important international colloquium, and I am very sorry to be unable to join you in person.

I have been asked to give a critique of AI Codes of Ethics from a human rights perspective, but before beginning, I would simply like to comment on the vast number of such codes, or ethical guidelines, currently circulating. According to Algorithm Watch, by the end of 2020, there were 167 such guidelines, all but 10 of which Algorithm Watch judged to consist of "vague formulations.... with no enforcement mechanisms at all" leaving it to question "whether guidelines that can neither be applied nor enforced are not more harmful than having no ethical guidelines at all."[81] And therein lies the core of the human rights community's objection to AI ethics codes of conduct: they offer no enforcement mechanisms and hence no means to hold corporations accountable for the potential harms of AI applications to society, and to marginalized groups within society. For the most part, Big Tech and powerful corporations remain free to choose if and how they will address ethical concerns of AI technologies and their impact on society, even if many of their guidelines make at least passing reference to the need to respect human rights.

But what is really meant by 'respect for human rights' in these ethics guidelines? Professor Sakiko Fukuda-Parr of the New School and I developed a framework for analyzing whether such guidelines in fact are designed to respect human rights.[82] First and foremost, are the guidelines anchored in international or regional human rights law as the standard for evaluating the relative benefit or harm to society of AI design, development, and deployment (AI-DDD)?

Secondly, do the guidelines propose that AI-DDD be assessed against the core principles which operationalize the realization of all human rights: equality and non-discrimination; participation of rights holders; and accountability that includes responsibility, transparency, and enforceability/remedy?

While we could not analyze the full universe of 167+ guidelines for ethical AI, we selected 15 from civil society, government, and the private sector, that previous meta-analyses had identified as attempting to address human rights concerns. We found that only 7 of the 15 grounded their guidelines in human rights law and the core human rights principles, even though the sample guidelines recognized the potential for AI-DDD to inflict human rights violations or social harm. The rest adopted an ethics framework. The risk of bias in AI-DDD was generally recognized, but few ethics guidelines proposed proactive measures to reduce bias or remedy its discriminatory impact, either on individuals or socio-economic groups. With respect to the principle of participation, most guidelines mentioned stakeholder consultation, and some referred to power asymmetries and the need to empower users, but only six recognized participation in AI-DDD as a right. Finally, while most guidelines attempted to address the principle of accountability, those based on ethical frameworks limited discussion to responsibility-recording and publicizing decisions throughout the AI-DDD process- or transparency through

81-In the Realm of Paper Tigers: Exploring the Failings of AI Ethics Guidelines (Algorithm Watch: 28 April 2020) https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/ Accessed 6 July 2022.

82-"Emerging consensus on 'Ethical AI': A Human Rights Critique of Stakeholder Guidelines", S. Fukuda-Parr and Elizabeth Gibbons, Global Policy, Vol. 12 Supplement 6 (Durham University and John Wiley & Sons: July, 2021). Available at: https://www.globalpolicyjournal.com/articles/international-law-and-human-rights/emerging- consensus-ethical-ai-human-rights-critique

independent audits. Only those guidelines grounded in human rights law made mention of mechanisms to enforce accountability and provide remedy or restitution to those who may be harmed by an AI application.

What can we conclude? First, that by claiming the 'ethics' space, corporations are defining norms and redefining human rights to suit their business model, not the common good.

Secondly, guidelines that claim respect for human rights need to be scrutinized to ensure that, at a minimum, they have provisions that protect against bias and discrimination, allow for meaningful participation, and understand accountability to include enforceability and remedy for those harmed. Finally, the proliferation of voluntary ethics guidelines forestalls the development of necessary legal norms and enforcement mechanisms, delaying society's capacity to check the power of Big Tech.

It is obviously not possible to extrapolate findings from a tiny sample of only 15 guidelines to the universe of AI ethics codes; nevertheless, our analysis shows that applying a human rights framework which incorporates international law and core principles can reveal whether a declared commitment to respect human rights is merely rhetorical or represents a sincere effort to adhere to human rights law, norms, and standards. Systematic application of such a framework can help reveal the gaps in AI ethics codes, gaps that show where law and regulation are needed to ensure that AI benefits society.

The overarching human rights critique of AI codes of ethics is that they are voluntary, open to interpretation and unenforceable, and do not establish an obligation to protect, respect and fulfil rights. This allows corporations to escape true accountability in a world where the power of AI carries the known risks of disinformation, discrimination, and exacerbation of inequality. Governing bodies charged with protecting public interest are scrambling to catch up with rapidly changing technologies. Meanwhile, a universal, elaborate, mature and constantly evolving human rights machinery already exists, with established operational norms, standards, and principles capable of holding AI-DDD producers accountable for preventing harm to society. This machinery already operates through international as well as national law, and so we should use it!

Malik Boumediene

Professeur Associé, Droit public & Sciences Politiques F.G.S.E.S, A.I.R.E.S.S., (Rabat, Maroc). Maître de conférences, Droit Public, H.DR, Université de Toulouse Jean Jaurès, CERDAP2, Grenoble (France)

Existe-t-il un Droit International de l'Intelligence Artificielle ?

> « Le monde a besoin de règles pour que l'intelligence artificielle profite à l'humanité »
>
> Audrey Azoulay, Directrice générale de l'UNESCO, 30 novembre 2021

Aujourd'hui l'Intelligence artificielle se développe à une vitesse exponentielle[83]. Régulièrement dans le monde, des chercheurs et ingénieurs développent de nouveaux logiciels dans différents domaines. L'objectif[84] de l'IA est de donner la possibilité à une machine de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité ». L'IA permet ainsi à des systèmes techniques de percevoir leur environnement, gérer ces perceptions, résoudre des problèmes et entreprendre des actions pour atteindre un objectif déterminé. L'ordinateur reçoit des données (déjà préparées ou collectées via ses capteurs - une caméra, par exemple), les analyse et réagit. La finalité est de permettre à des machines de « simuler l'intelligence humaine »[85]. Et, pour John Mac Carthy, l'un des créateurs de ce concept, « toute activité intellectuelle peut être décrite avec suffisamment de précision pour être simulée par une machine ». L'IA peut être perçue comme une opportunité pour nos sociétés. Elle pourrait permettre aux citoyens de bénéficier de meilleurs soins de santé, des voitures et d'autres modes de transport davantage sécurisés, ainsi que des services moins coûteux, mieux adaptés aux besoins et munis d'une plus longue durée de vie. Elle pourrait également faciliter l'accès à l'information, à l'éducation et permettre à un plus grand nombre de personnes de suivre des formations. L'IA peut aussi contribuer à rendre le milieu du travail plus sécurisé, dans le sens où des robots peuvent être utilisés pour les activités les plus dangereuses, tout en créant de nouveaux emplois avec l'expansion des entreprises utilisant l'IA. Pour les entreprises également, l'IA peut constituer un véritable atout. En effet, l'IA peut créer de nouveaux emplois et de nouveaux services. Au niveau des services publics elle permettrait, par exemple, la mise en place de nouvelles pédagogies dans le cadre de l'Education nationale. Au niveau politique et, plus précisément, au niveau de la vitalité de la démocratie, l'IA est susceptible d'apporter beaucoup également. Par exemple, en repérant et en dénonçant des stratégies de désinformation ou de manipulation de l'électorat. Dans le domaine de la sécurité, l'IA donnerait la possibilité de mieux profiler des personnes et d'améliorer les recherches afin de prévenir des actes criminels. Aujourd'hui, nous avons déjà des applications concrètes de l'IA pour l'intérêt général. Par exemple, l'intelligence artificielle propose, aujourd'hui, des solutions de surveillance des lieux de baignade dans le but de mieux sécuriser les baigneurs. En effet, un logiciel dénommé Sightbit et développé par une entreprise israélienne enregistre toutes les images des caméras disposées le long des plages. Si un nageur est en difficulté, le logiciel lance l'alerte. De même une entreprise française a développé un système dont le nom est «POSEIDON». Il s'agit d'un système de vision par ordinateur

83- En juillet 2017, le Conseil des affaires de l'État a rendu public son « Plan de développement de la prochaine génération de l'intelligence artificielle pour la période de 2016 à 2030 » (ci-après le Plan). Ce document présente un développement de l'IA en trois étapes, dont l'objectif final est que le pays devienne le leader mondial de l'IA en 2030 ainsi qu'une référence en matière de réglementation (Ma, Aifang. « L'IA en Chine : état des lieux », Constructif, vol. 54, no. 3, 2019, pp. 3337-).

84- https://www.europarl.europa.eu/news/fr/headlines/society/20200827STO85804/intelligence-artificielle-definition-et-utilisation

85- https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257

pour la détection de noyades en piscine. Grâce à des caméras et de capteurs situés au-dessus du bassin et/ou sous l'eau, un panneau d'alerte visible de tous dans la piscine est actionné (accompagné d'une sonnerie) dès qu'une personne en détresse est détectée par le système. Le système localise la personne dans la piscine également. Dans le domaine de la protection de l'environnement plusieurs applications ont été créées. Par exemple, le projet de la Banque des Territoires en partenariat avec Enedis et GRDF (gaz réseau distribution France) propose un service informatique consacré à la rénovation thermique des bâtiments publics. Ce service, doté d'une IA, permet d'aider les collectivités locales dans leurs décisions touchant la rénovation thermique de leurs bâtiments en analysant les données de consommation recueillies auprès d'Enedis et de GRDF. Ainsi, les utilisateurs sauront quel bâtiment rénover en priorité pour réduire au maximum leur consommation d'énergie. Dans le domaine de la santé publique, l'Université de l'Alberta a mis en place une IA prénommée DarkNPS. Cette application a pour objectif, sur la base des drogues déjà existantes (plus de 1 700), d'inventer de nouvelles drogues et notamment celles qui pourront voir le jour dans l'avenir afin de prendre de l'avance sur les producteurs de drogues. L'enjeu est important, car pour déterminer les risques des substances sur les plans psychologique et physiologique il faut auparavant analyser les différentes molécules composant la nouvelle drogue. Cela prend de nombreux mois ne permettant pas de venir rapidement en aide aux personnes intoxiquées. Cette IA a imaginé plus de 8,9 millions nouveaux produits de synthèse. Grâce à cette IA, il ne faut que quelques minutes pour déterminer la drogue qui est mise sur le marché puisque le logiciel avait déjà imaginé cette drogue et donc ses caractéristiques. Ce qui permet d'agir immédiatement sur le plan médical.

Cependant, l'IA est aussi synonyme de dangers potentiels[86]. Ces risques peuvent concerner différents domaines. Il en est ainsi dans le domaine des droits fondamentaux et de la démocratie. En effet, il faut savoir que les résultats produit par l'IA sont le produits d'une certaine utilisation des données qu'elle utilise mais aussi des modalités de récolte de ces données. Or, l'IA peut très bien faire des erreurs en excluant de l'algorithme des sujets importants. Ou, au contraire, mettre l'accent sur la récolte de certaines questions. C'est ainsi qu'on peut très bien imaginer une IA récoltant des articles de presse mais en mettant à la « Une » de son bulletin d'information uniquement certaines informations afin d'influencer les lecteurs ou des électeurs sur un sujet précis. Si elle n'est pas correctement appliquée, l'IA pourrait mener à prendre des décisions à l'embauche fondées sur l'appartenance ethnique, le genre, l'âge du candidat, sa situation de famille ou encore son état de santé. Elle peut porter atteinte à la vie privée à travers la reconnaissance faciale ou pour rechercher des personnes en ligne. Que deviennent les données relatives à ces personnes ? où sont-elles stockées ? pour combien de temps ? L'IA est susceptible aussi de porter atteinte au droit au rassemblement, puisqu'elle peut être utilisée pour localiser et profiler les individus liés à certaines croyances, positions politiques ou comportements. En ce qui concerne le droit à l'emploi, le développement de l'IA pourrait entraîner la suppression de très nombreux emplois occupés antérieurement par des personnes. Au niveau de l'activité commerciale et plus particulièrement de la mise en concurrence ,l'IA, avec la collecte massive des données, peut permettre à une entreprise d'accéder à de très nombreuses données concernant une entreprises concurrente et l'affaiblir rapidement voire entraîner sa faillite. Les véhicules autonomes et autres systèmes liés à l'IA peuvent être confrontés à des failles de sécurité, à des cyberattaques ou à une utilisation abusive de données personnelles, en particulier lorsqu'ils impliquent la collecte et le traitement d'un grand nombre

86- Voir notamment : https://www.dsi.uzh.ch/dam/jcr:45a24ac2-cf00-40ab-95fa-ffe0db26e241/dsi-strategy-lab-21-fr.pdf; https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a37211-eb-9585-01aa75ed71a1.0020.02/DOC_1&format=PDF.

de données. La question de la transparence de l'IA se pose également dans le sens où le consommateur doit être en capacité de savoir s'il est en discussion avec un être humain ou une IA. La question de la responsabilité d'une IA apparaît aussi centrale. Qui sera tenu responsable lorsqu'une IA cause un préjudice physique ou moral ? Qui sera tenu responsable si un système d'IA propose un plan qui s'avère dommageable, le constructeur ou les acteurs qui finissent par implémenter des algorithmes ? En outre dans l'élaboration d'un système algorhitmique plusieurs professionnels interviennent risquant d'entraîner une sorte de dilution des responsabilités par l'impossibilité de déterminer le responsable direct.

Face aux éventuelles limites, dangers et risques de l'intelligence artificielle de nombreux Etats, notamment, européens se sont dotés d'une législation permettant la régulation de cette nouvelle technologie. Ainsi, à ce jour plus de 30 Etats membres de l'Union européenne ont adopté des stratégies et des politiques concernant les systèmes d'intelligence artificielle. Un Etat a entrepris un programme de certification volontaire en matière d'IA. Deux Etats ont adopté des cadres éthiques internationaux européens même si ces cadres ne sont pas contraignants. Concernant la mise en place de véritables législations, on remarque que quatre Etats membres ont opté pour des cadres juridiques spécifiques à l'IA qui concernent l'utilisation de cette intelligence dans le cadre des véhicules autonomes. Deux Etats élaborent actuellement un régime juridique pour l'utilisation de l'IA dans le recrutement au sein des administrations. Certains Etats des Etats-Unis ont également légiféré dans ce sens. C'est ainsi qu'en Alabama, la loi Al S.B. 78 (Technology, Alabama Council on Advanced Technology, estab., to advise Governor and Legislature, members, duties), datant de 2021[87], a comme but d'établir le « Alabama Council on Advanced Technology and Artificial Intelligence ». Ce dernier aura pour mission d'examiner le travail du gouverneur et du législateur de l'Alabama, ainsi que celui de toutes autres parties intéressées, et de les conseiller sur l'utilisation et le développement des technologies avancées et de l'IA dans l'État. Au Colorado, la même année, la loi CO S.B. 169 (2021, (Act Concerning Protecting Consumers From Unfair Discrimination in Insurance Practices)[88], interdit aux assureurs d'avoir recours à toute donnée externe concernant leurs clients, ainsi qu'à des algorithmes ou à des modèles prédictifs ayant recours à des sources de données externes, qui pourraient donner lieu à de la discrimination basée sur la race, sur la couleur de peau, l'origine nationale ou ethnique, etc. Dans l'Illinois la loi IL H.B. 53 (Amends the Artificial Intelligence Video Interview Act – 2001) modifie le Artificial Intelligence Video Interview Act et pose le principe selon lequel les employeurs qui ont entièrement recours à l'IA afin de déterminer si une personne doit être invitée pour un entretien en personne, ont l'obligation de recueillir certaines données démographiques et d'en faire état auprès du Department of Commerce and Economic Opportunity. Ce dernier a l'obligation d'analyser les données qui lui sont transmises afin de déterminer si l'utilisation de l'IA dans ce secteur entraine une discrimination raciale. Il doit alors en faire état auprès du Gouverneur et de l'Assemblée générale d'Illinois. Au Mississippi la loi MS HB 633[89] (Mississippi Computer Science and Cyber Education Equality Act de 2021) impose à l'État de mettre en place des formations portant, notamment, sur l'IA et le « machine learning » dans les écoles. De même au niveau de l'Union européenne la directive 2 (MiFID 2)[90] du 15 mai 2014 impose aux entreprises

87- https://lepetitjournal.com/new-york/lintelligence-artificielle-et-le-droit-en-amerique-du-nord-analyse-transatlantique-338094.

88- https://lepetitjournal.com/new-york/lintelligence-artificielle-et-le-droit-en-amerique-du-nord-analyse-transatlantique-338094.

89- https://lepetitjournal.com/new-york/lintelligence-artificielle-et-le-droit-en-amerique-du-nord-analyse-transatlantique-338094

90- Directive 201465//UE du Parlement européen et du Conseil du 15 mai 2014 concernant les marchés

d'investissement d'inclure les détails des algorithmes informatiques responsables des décisions d'investissement et de l'exécution des transactions. Le 21 avril 2021 la Commission européenne a rendu public COM (2021) 206 final 2021/0106 (COD) une Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'Intelligence Artificielle et modifiant certains actes législatifs de l'Union.

Nous constatons ainsi que le contenu du cadre juridique de l'IA varie en fonction des États. On observe également qu'il s'agit d'une législation qui est à l'aube de son développement alors que l'IA se développe à grande vitesse dans un monde interconnecté. C'est pourquoi il est impératif qu'un droit international de l'IA puisse apporter une régulation à cette nouvelle technologie dont les applications ne sont pas sans conséquences pour les individus. Or, la question qui se pose est de savoir si un tel droit existe aujourd'hui ? y a-t-il un cadre juridique international qui s'applique à l'IA ? telle est la question à laquelle nous tenterons de répondre.

1. L'EMERGENCE D'UN DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE COMME REALITE.

A. LES FONDEMENTS D'UN DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE

L'émergence d'un droit international de l'IA semble aujourd'hui une réalité puisque, non seulement, des fondements extra-juridiques justifient un tel droit mais ce dernier trouve une traduction au sein d'un cadre juridique.

Les fondements extra-juridiques d'un droit international de l'Intelligence artificielle sont multiples. En effet, l'élaboration d'un droit international de l'IA répond à plusieurs variables[91] qui semblent s'exprimer dans une philosophie qui consiste à dire que l'IA est un progrès qu'on ne peut écarter ou interdire. Néanmoins, du fait des risques éventuels que cela peut représenter, il est nécessaire de réguler cette activité. Quelles seront les conséquences pour les individus si nous assistons à l'élaboration d'un système d'IA non fiable ? d'autre part, une IA qui ne serait pas soumise à une réglementation, qui n'est pas robuste ou qui ne respecte pas une certaine éthique, serait rejetée par la population ce qui ne permettrait pas son développement. Que penser d'une IA qui ne respecterait pas l'autonomie de l'individu, c'est-à-dire qu'elle serait en capacité de mentir, manipuler, influencer afin d'atteindre un but caché ? cela serait une véritable atteinte au libre choix des individus. Que penser également d'une IA qui dissimulerait sa vraie identité aux individus pensant échanger avec un être humain? Que penser également si de tels systèmes étaient incapables de se protéger face aux vulnérabilités qui pourraient permettre à des entités mal intentionnées de les exploiter (par exemple, piratage). Des attaques pourraient cibler les données, corrompre le système, contrôler un système entraînant des préjudices tant moraux que physiques (on peut prendre par exemple certaines opérations chirurgicales intervenant avec l'aide de

d'instruments financiers et modifiant la directive 200292//CE et la directive 201161//UE Texte présentant de l'intérêt pour l'EEE.

91- Voir notamment : https://www.actuia.com/actualite/retour-sur-la-recommandation-de-lunesco-etablissant-des-regles-ethiques-de-lintelligence-artificielle/; https://unesdoc.unesco.org/ark:/48223/pf0000369455_fre; https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre; BEIJING CONSENSUS on artificial intelligence and education Outcome document of the International Conference on Artificial Intelligence and Education 'Planning education in the AI era: Lead the leap'16 – 18 May 2019 Beijing, People's Republic of China (https://unesdoc.unesco.org/ark:/48223/pf0000368303/PDF/368303qaa.pdf.multi.page=13); Artificial intelligence for Sustainable Development: synthesis report, Mobile Learning Week 2019, Unesco, 2019 (https://unesdoc.unesco.org/ark:/48223/pf0000370308); L'IA au double regard Une réflexion sur l'intelligence artificielle, Unesco, 2019 (file:///C:/Users/malik.boumediene/Downloads/IntelligenceArtificielleAuDoubleRegard%20(1).pdf); Déclaration finale du Forum sur l'intelligence artificielle en Afrique Benguérir (Royaume du Maroc), 13 décembre 2018 (Déclaration de Benguérir), Unesco, 2018, (https://en.unesco.org/sites/default/files/ai_outcome-statement_africa-forum_fr.pdf);

l'IA). C'est pourquoi il est important que les développeurs de ces systèmes prévoient des protocoles d'alertes et d'action. Sans régulation par l'Etat nous pourrions être en présence de systèmes dont la précision fait défaut, c'est-à-dire un système incapable de poser un jugement correct par rapport à l'objectif souhaité. Nous aurions également sur le marché des systèmes qui peuvent ne pas être fiables dans le sens où ils seraient incapables de fonctionner correctement accompagnés, d'une absence de protocoles permettant de détecter tout dysfonctionnement et de permettre une intervention humaine. Une absence de réglementation donnerait peut-être la possibilité à des opérateurs de limiter l'accès à leur service d'IA à certaines catégories de personnes alors que le principe devrait être l'universalité du fait que ce système répond à l'intérêt général. L'opérateur pourrait créer des discriminations en fonction de l'âge, du sexe, des capacités ou des caractéristiques d'un individu. L'accessibilité doit être, au contraire, la plus large possible et il doit exister des protocoles de contrôle afin d'éviter toute discrimination. Sans régulation juridique spécifique à l'IA se poserait aussi la question de la responsabilité en cas de dommages dû au système. Comment prouver qu'un système a été faillible si on ne peut pas accéder à ses algorithmes et ses données d'autant que les processus de fonctionnement sont peu transparents et complexes. L'absence de législation risque fort de limiter la prise de connaissance des données récoltées par le système ainsi que les modalités de cette récolte. Il sera très difficile de savoir que deviennent ces données, comment sont-elles utilisées, pour quel objectif, comment le système prend ses décisions. Alors que, au contraire, il doit y avoir aussi une traçabilité de ces données et cela de leur intégration dans le système jusqu'à leur utilisation finale. Enfin, on peut imaginer que sans régulation les systèmes d'IA porteraient sûrement atteinte à certains de nos droits fondamentaux tels que le respect de la vie privée[92] et la protection des données. Aucune garantie n'existerait à ce niveau. La protection des données serait laissée à la libre discrétion du concepteur qui a alors la possibilité de limiter l'accès uniquement à certaines données et pendant un certain délai. Ces données personnelles pourraient faire l'objet de vol, de vente à l'insu des utilisateurs. Quelles garanties que les développeurs détruisent toutes les données dans le cas d'une fin de vie du système ? Les concepteurs prévoiront-ils des protocoles clairs précisant les conditions d'utilisation de ces données à d'autres fins ? Comment les individus, qui utilisent ce service, pourront donner ou pas leur accord ?

Aux fondements extra-juridiques le droit international de l'IA peut compter sur des fondements juridiques multiples. Il peut paraître surprenant de parler de fondements juridiques « divers » pour une activité qui reste récente même si elle tend à se développer à grande vitesse. Pourtant les Etats sont liés par un certain nombre de textes.

Le texte de référence en matière d'IA est, sans aucun doute, la recommandation[93] de portée mondiale sur l'éthique de l'intelligence artificielle qui a été adoptée par la Conférence générale de l'UNESCO lors de sa 41ème session le 24 novembre 2021. Cette recommandation est le résultat d'un long processus de deux ans dont l'initiative a été prise[94] lors de la 40ème session, en novembre 2019 de la Conférence générale de l'Unesco. C'est ainsi que, au mois de mars 2020, 24 spécialistes, ayant une expertise multidisciplinaire et pluraliste sur l'éthique de l'intelligence artificielle, ont été nommés par la Directrice générale de l'UNESCO pour former un Groupe d'experts ad hoc (GEAH). Ce Comité d'experts était composé de femmes et d'hommes issus de divers milieux culturels en provenance de toutes les régions du monde. Il comprend des scientifiques et

92- The right to privacy in the digital age : resolution / adopted by the Human Rights Council on 26 September 2019, UN. Human Rights Council, 42nd sess. : 2019 : Geneva (https://digitallibrary.un.org/record/3837297).

93- https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre. Mais l'atteinte peut aussi impacter : la liberté de circulation ; le principe d'égalité ; la sureté de l'individu ;…

94- Résolution 40 C / 37.

des professionnels ayant une connaissance pointue tant des aspects technologiques que éthiques de l'IA. Parmi ceux-ci, figurent Julie Owono Assevini (Cameroun), Catherine Tessier (France), Fatima Roumate (Maroc) et Jean-Philbert Nsengimana (Rwanda). Ils auront alors comme mission de produire un projet d'instrument normatif[95] mondial sous la forme d'une Recommandation de l'UNESCO. Le groupe d'experts tiendra une première réunion du mois de mars au mois de mai 2020. Il leur sera donné, le 10 avril 2020, un premier support de travail intitulé « Discussion du groupe d'experts ad hoc (GEAH) pour l'élaboration d'un projet de recommandation sur l'éthique de l'intelligence artificielle ». Il faut noter que plus de 40 observateurs des États membres, des Nations unies et d'autres entités intergouvernementales ont participé à ces discussions et ont pu faire part de leurs réflexions et positions. Ces discussions ont abouti à la première version du projet de texte de la Recommandation préparée par le GEAH en mai 2020, et qui a fait l'objet par la suite d'un vaste processus de consultation multipartite de la fin du mois de juin jusqu'au début du mois d'août 2020. Cette première version sera débattue du 31 août au 4 septembre 2020 où plus de 100 observateurs de 43 États membres, des Nations unies et d'autres entités intergouvernementales, ainsi que des ONG se sont inscrits pour observer la réunion. Cette deuxième réunion va permettre au GEAH de produire, le 7 septembre 2020, un avant-projet de Recommandation. Cet avant-projet sera transmis aux États membres en septembre 2020 afin que ces derniers puissent faire parvenir leurs commentaires avant le 31 décembre 2020. Le texte sera présenté par la suite à un Comité composé d'experts des Etats membres qui se réunira du 26 au 30 avril 2021 puis du 21 au 25 juin 2021. Les représentants des Etats membres de l'Unesco vont alors se mettre d'accord sur un projet de texte qui sera voté à la Conférence générale de l'UNESCO le 24 novembre 2021 lors de sa 41ème session. On remarque que les « Considérants » introductifs de la Recommandation elle-même sont instructifs à plusieurs égards. Tout d'abord cette recommandation n'est pas contre l'IA puisqu'elle reconnaît ses aspects positifs[96]. Mais elle pose la nécessité d'un encadrement au vu des « répercussions négatives » qu'elle peut entraîner sur « les sociétés, l'environnement, les écosystèmes et la vie humaine, y compris l'esprit humain ». En fait, cette recommandation est un compromis entre la nécessité de garantir une certaine sécurité aux utilisateurs tout en veillant à ne pas freiner l'innovation et le développement des systèmes d'IA. Il est mis en avant également que cette recommandation est un instrument normatif dans le sens où elle s'impose aux 193 Etats l'ayant adopté. Il faut noter que la Recommandation n'est pas détachée du droit international, bien au contraire. Elle puise sa philosophie, notamment, dans le droit international tel que : la Déclaration universelle des droits de l'Homme (1948); les instruments internationaux relatifs aux droits de l'Homme, dont la Convention relative au statut des réfugiés (1951), la Convention concernant la discrimination (emploi et profession) (1958), la Convention internationale sur l'élimination de toutes les formes de discrimination raciale (1965), le Pacte international relatif aux droits civils et politiques (1966), le Pacte international relatif aux droits économiques, sociaux et culturels (1966), la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (1979), la Convention relative aux droits

95- A côté de ce droit premier ayant force juridique nous trouver un ensemble de texte s'intégrant dans ce que l'on appelle la « soft law ». Il s'agit d'un ensemble de textes sous forme de directives à l'attention des Etats : Principes directeurs relatifs aux entreprises et aux droits de l'homme, Nations-Unies, 2011, https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_fr.pdf; Principes directeurs de l'OCDE à l'intention des entreprises multinationales, OCDE, 2011, http://mneguidelines.oecd.org/guidelines/.

96- Il est précisé que « les technologies de l'IA peuvent rendre de grands services à l'humanité (…) », « les technologies de l'IA peuvent être porteuses d'importants avantages ».

de l'enfant (1989)[97], la Convention relative aux droits des personnes handicapées (2006)[98], la Convention concernant la lutte contre la discrimination dans le domaine de l'enseignement (1960), et la Convention sur la protection et la promotion de la diversité des expressions culturelles (2005) ; ainsi que tous les autres instruments, recommandations et déclarations internationaux pertinents, Prenant acte de la Déclaration des Nations Unies sur le droit au développement (1986) ; de la Déclaration sur les responsabilités des générations présentes envers les générations futures (1997).

B. LES GRANDS PRINCIPES DU DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE

Lorsque l'on tente d'aborder les grands principes du droit international de l'IA, il faut faire référence au seul texte de dimension internationale qui, à ce jour, est la Recommandation[99] de l'Unesco adoptée le 23 novembre 2021 et relative à l'éthique de l'Intelligence artificielle. On peut relever un certain nombre de principes directeurs posés par la Recommandation.

Tout d'abord les systèmes d'IA doivent respecter un certain nombre de valeurs. La première a trait au respect de la dignité humaine. En effet, la dignité humaine possède des caractéristiques reconnues par tous telle que son inviolabilité, son fondement intrinsèque à tout homme et son aspect de droit « naturel ». C'est pourquoi dans le cadre des interactions entre les SIA (Systèmes d'intelligence artificielle) et les individus, ces derniers ne doivent jamais être considérés comme des objets et il ne doit jamais être porté atteinte à leurs droits fondamentaux. Une autre valeur doit aussi imprégner les SIA. Il s'agit de l'impossibilité pour ces systèmes d'entraîner des préjudices à une personne. Bien au contraire, ces systèmes doivent tout au long de leur fonctionnement améliorer la qualité de vie des êtres humains, en laissant aux individus ou groupes le soin de définir la notion de « qualité de vie ». Une valeur importante doit conditionner tous les SIA. Il s'agit de l'obligation tout au long de la vie du système de respecter les droits fondamentaux des individus. A ce titre, tous les acteurs intervenant dans le cycle de vie d'un SIA doivent respecter strictement de tels droits et les consolider de manière continue. Ces acteurs doivent aussi favoriser l'édification de sociétés pacifiques et justes fondées sur un avenir interconnecté pour le bien de tous. Enfin, les concepteurs de SIA doivent encourager la participation active de tous les individus ou groupes à contribuer à l'élaboration de ces systèmes sans faire aucune discrimination. Il s'agit ici d'éviter que de tels systèmes soient victimes de biais discriminatoires.

Si les SIA se voient imposer un certain nombre de valeurs à respecter la recommandation pose aussi un certain nombre de principes qui s'imposent à ces systèmes. On trouve ainsi la nécessité d'appliquer le principe de proportionnalité dans le sens où la méthode d'IA retenue doit être adaptée et proportionnée à l'objectif à atteindre. D'autre part, la méthode adoptée ne doit pas porter une atteinte aux valeurs fondamentales tels que, par exemple, les droits de l'Homme[100]. Ces

97- https://www.unicef.org/sites/default/files/201910-/convention-droits-enfants-versions-pour-les-enfants.pdf

98- https://www.ohchr.org/fr/instruments-mechanisms/instruments/convention-rights-persons-disabilities#:~:text=La%20pr%C3%A9sente%20Convention%20a%20pour,respect%20de%20leur%20dignit%C3%A9%20intrins%C3%A8que.

99- https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre.

100- Afin d'assurer l'application de ce principe la recommandation souligne que les systèmes d'IA considérés comme représentant un risque potentiel pour les droits de l'homme doivent faire l'objet de tests approfondis par les acteurs de l'IA, y compris dans des conditions réelles si nécessaire, dans le cadre de l'évaluation de l'impact éthique, avant de les autoriser sur le marché. Ces études d'impact sur l'éthique doivent intervenir tout au long du cycle de vie des systèmes d'IA. L'ensemble de ces évaluations doivent être transparentes et ouvertes au public. De manière générale il est demandé aux Etats de mettre en place des protocoles d'évaluation afin d'analyser l'impact des SIA sur l'Ethique,

systèmes sont tenus aussi de reposer sur des bases scientifiques rigoureuses afin d'éviter tout préjudice humain et toujours permettre, tout au long du cycle de leur vie, l'intervention humaine lorsque les décisions pouvant être prises par le système ont un impact irréversible ou difficile à renverser. Les systèmes d'IA ne doivent pas être utilisés à des fins de notation sociale ou de surveillance de masse. Les Etats doivent également garantir un droit à la sécurité en élaborant des cadres juridiques permettant aux utilisateurs que leurs données soient protégées[101] et d'y accéder de manière durable dans le cadre de la protection de la vie privée. La recommandation insiste sur l'interdiction des discriminations tout en encourageant tous les acteurs à œuvrer pour la justice sociale, l'égalité, l'équité. L'Etat, lui, devant mettre en œuvre des politiques publiques afin de lutter contre les fractures numériques. Enfin, tant les Etats que les acteurs privés sont invités à prendre des mesures afin que puisse toujours être engagée une responsabilité des acteurs participant à l'élaboration d'un SIA. Afin de permettre d'engager plus facilement une telle responsabilité, la recommandation pose les principes de transparence et d'explicabilité des SIA. En effet, l'engagement d'une responsabilité dans ce domaine implique que l'on puisse comprendre le fonctionnement d'un SIA mais aussi accéder à l'ensemble des données.

On observe que la recommandation énonce de nombreuses orientations en direction des Etats afin que ces derniers puissent assurer une application effective des principes posés précédemment. Il leur est ainsi proposé de mettre en place des mécanismes de certification permettant de s'assurer que les concepteurs de SIA respectent un minimum de règles protectrices touchant, notamment, les droits humains, l'environnement et les écosystèmes. Afin de s'assurer d'une protection durable des droits fondamentaux, les Etats sont invités, dans le cadre des législations existantes ou à venir sur l'IA, à intégrer la question des droits fondamentaux tout au long du cycle de vie des systèmes. Les Etats sont fortement encouragés à débattre des questions éthiques dans le cadre des instances internationales et intergouvernementales. Ils doivent échanger les connaissances nécessaires relatives à la maîtrise de l'IA, de façon à autonomiser la population et à réduire les fractures numériques et les inégalités d'accès dans le domaine du numérique découlant de l'adoption à grande échelle de systèmes d'IA. Grâce à l'IA, les Etats sont incités à améliorer un certain nombre de droits tels que l'accès à l'information et au savoir. L'IA doit améliorer la liberté d'expression, les libertés académiques et scientifiques ainsi que l'accès à l'information, combattre la désinformation, la mésinformation et les discours de haine. Ainsi, l'IA doit être au service de l'intérêt général en contribuant aussi à améliorer la santé humaine et à protéger le droit à la vie, notamment en atténuant les épidémies. Enfin, la recommandation pose le principe de la responsabilité des Etats dans le suivi des mécanismes de contrôle de l'éthique dans ce domaine ce qui peut se comprendre tant les enjeux sont énormes.

La recommandation de l'Unesco est, sans aucun doute, une pierre importante posée dans le paysage d'un droit international de l'IA. Néanmoins, il semble qu'il ne s'agit là que d'une première pierre mais également d'une pierre fragile dont la portée reste très limitée.

2. LE DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE : UN DROIT EN DEVENIR.

La recommandation de l'Unesco doit faire face, il nous semble, à deux grandes séries d'obstacles dont la conséquence est une certaine neutralisation de sa portée. En effet, on est en présence d'une part, d'un droit « faible » de part certaines de ses caractéristiques (A) et, d'autre part,

la société et les individus du point de vue également des conséquences psychologiques.

101- Ces cadres juridiques doivent aussi permettre de sécuriser l'archivage et la destruction de ces données.

de normes juridiques qui s'imposent difficilement aux systèmes d'IA qui se développent (B).

A. LES LIMITES INTRINSÈQUES DU DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE

La recommandation de l'Unesco souffre de plusieurs faiblesses.

Tout d'abord, le texte manque incontestablement d'autorité dans le sens où il n'apparaît pas comme un instrument imposant de véritables règles aux différents acteurs. Cette relative autorité se manifeste directement à travers le vocabulaire utilisé : la présente Recommandation a pour objet « de servir de base » (article 5) ; elle offre un cadre universel de valeurs, de principes et d'actions pour « guider les États » dans la formulation de leur législation, de leurs politiques ou d'autres instruments concernant l'IA (article 8) ; l'ensemble des valeurs mentionnées dans ce texte « inspirent » ainsi des comportements souhaitables (article 10) ; les acteurs de l'IA doivent « promouvoir » la justice sociale, garantir l'équité et lutter contre les discriminations en tous genres, conformément au droit international » (article 28) ; les États membres « devraient »[102] veiller à ce qu'il soit toujours possible d'attribuer la responsabilité éthique et juridique de tout stade du cycle de vie des systèmes d'IA à des personnes physiques (article 35) ; les États, conformément au droit international, « peuvent » réglementer les données générées sur leur territoire » (article 46); les États membres « sont encouragés »[103] à élaborer des stratégies nationales et régionales relatives à l'IA et « à envisager » l'adoption de formes de gouvernance souple ; les États membres « pourraient » envisager la possibilité de mécanismes de suivi et d'évaluation. Il faut noter également que la force obligatoire de cette recommandation est d'autant plus fragile que l'Unesco n'a pas de compétence afin d'assurer l'application de cette Recommandation. En effet, l'article 137 dispose que « L'UNESCO a vocation à être la principale institution des Nations Unies chargée de promouvoir et diffuser la présente Recommandation » et l'article 139 ajoute que: « le mandat de promouvoir et protéger relève de la compétence des gouvernements et des organes intergouvernementaux ».

D'autre part, ce texte occulte le principe de réalité auquel doivent faire face certaines des orientations qu'il propose. Ainsi l'article 8 demande aux Etats de promouvoir un accès équitable aux progrès et aux connaissances dans le domaine de l'IA, ainsi que le partage des bienfaits qui en découlent, en accordant une attention particulière aux besoins et contributions des pays à revenu intermédiaire inférieur (PRITI), notamment aux PMA, aux PDSL et aux PIED. L'article 28 rajoute : « Au niveau international, les pays les plus avancés sur le plan technologique ont un devoir de solidarité envers les pays les moins avancés afin de garantir le partage des bienfaits des technologies de l'IA de telle sorte que l'accès et la participation de ces derniers au cycle de vie des systèmes d'IA contribuent à un ordre mondial plus équitable en matière d'information, de communication, de culture, d'éducation, de recherche et de stabilité socioéconomique et politique». Cependant, du fait des enjeux de la maîtrise des systèmes d'IA on ne voit

102-les États membres devraient veiller à ce que les préjudices causés par le biais de systèmes d'IA fassent l'objet d'enquêtes et de réparations, en mettant en place des mécanismes d'exécution solides et des mesures correctives afin de s'assurer que les droits de l'homme, les libertés fondamentales et l'état de droit sont respectés dans le monde numérique (article 55) ; Les États membres « devraient s'employer », au sein des organisations internationales, à mettre en place des cadres de coopération internationale dans le domaine de l'IA au service du développement » (article 80) ; les États membres et toutes les autres parties prenantes identifiées dans la présente recommandation « devraient » respecter, promouvoir et protéger les valeurs, principes et normes éthiques relatifs à l'IA qui y sont énoncés, et prendre toutes les mesures en leur pouvoir pour donner effet aux recommandations stratégiques qu'elle contient (article 135)

103- De même : les États membres « sont encouragés » à établir, s'il y a lieu, des systèmes d'IA dans les domaines de la conservation, de l'enrichissement, de la compréhension, de la promotion, du patrimoine culturel matériel, documentaire » (article 94) ;

pas comment des Etats partageraient leurs connaissances. Certaines dispositions de ce texte se heurtent également au principe d'égalité dans le sens où la dimension financière est occultée dans la mise en place de certains principes. L'article 25, par exemple, sollicite les Etats afin de mettre en place des processus garantissant le principe de précaution en évaluant les éventuels risques. Or, encore faut-il pour un pays avoir les moyens financiers et humains afin d'en assurer la mise en œuvre et la surveillance. L'article 31 incite les Etats à étudier l'impact humain, social, culturel, économique et environnemental des technologies de l'IA. Encore une fois, de telles études nécessitent des moyens financiers et de la compétence humaine, ce que ne possèdent pas tous les Etats.

La recommandation semble aussi lacunaire en ce qui concerne les modalités de protection des droits fondamentaux alors même qu'elle n'a de cesse d'y faire référence. Premièrement on observe que le texte laisse une certaine marge de manœuvre aux Etats afin de, éventuellement, porter atteinte aux droits fondamentaux dans le cadre des systèmes d'IA. En effet, l'article 11 dispose que « toute restriction éventuelle imposée à l'exercice des droits de l'Homme et des libertés fondamentales doit reposer sur une base juridique, être raisonnable, nécessaire et proportionnée, et conforme aux obligations qui incombent aux États membres en vertu du droit international ». Le fait que cette restriction repose sur « une base juridique » n'empêche pas de limiter son étendue. En outre, il est précisé que l'atteinte doit être « raisonnable », « nécessaire » et « proportionnée ». Or, ces notions sont très subjectives et laissent une grande souveraineté aux Etats. Deuxièmement on note que la recommandation n'aborde pas la question de la protection des droits fondamentaux face aux SIA en période de conflits, crises sociales, manifestations importantes, terrorisme, ... Les atteintes aux droits fondamentaux peuvent être d'autant plus faciles que la recommandation laisse une certaine souveraineté aux Etats dans la détermination des méthodes d'évaluation de l'impact des systèmes, notamment, sur les droits fondamentaux. L'article 131 mentionne dans ce sens que: «Les États membres – en fonction de leur situation, de leur mode de gouvernement et de leur Constitution – devraient assurer de manière crédible et transparente le suivi et l'évaluation des politiques, programmes et mécanismes relatifs à l'éthique de l'IA ».

Le texte de l'Unesco semble souffrir également d'un autre handicap résidant dans le manque de précision sur les mécanismes permettant, notamment, de garantir les droits fondamentaux comme en témoignent les quelques extraits suivants: il est indiqué qu'« Un recours efficace » doit être disponible contre la discrimination et la détermination algorithmique biaisée (article 29), mais nous n'avons aucune indication sur les caractéristiques de ce recours. L'article 32 mentionne qu'il est important que les données pour les systèmes d'IA soient collectées, utilisées, partagées, archivées et supprimées « d'une manière » qui soit conforme au droit international, mais nous n'avons pas d'informations sur les modalités de cette collecte ou archivage. L'article 50 pose le principe que les États membres « devraient mettre en place des cadres relatifs aux évaluations de l'impact » des systèmes d'IA, mais ne propose aucune caractéristique que doivent revêtir de tels « cadres » ; l'article 51 demande notamment aux Etats de « mettre en place des mécanismes » pour identifier, prévenir et atténuer les risques liés à l'utilisation des systèmes d'IA sans donner certaines pistes permettant de garantir un minimum d'efficacité de ces mécanismes. L'article 62 encourage les États qui se dotent de systèmes d'IA pour les cas d'utilisation sensibles en matière de droits de l'Homme, comme l'application des lois, la protection sociale, l'emploi, les médias et les relais d'information, les soins de santé et le système judiciaire indépendant, de prévoir « des mécanismes permettant d'évaluer l'impact économique et social de ces systèmes ». L'article 63 encourage les États membres à renforcer « la capacité du

pouvoir judiciaire à prendre des décisions relatives aux systèmes d'IA » mais, encore une fois, nous n'avons aucune proposition sur les modalités de cette consolidation. Ce manque de précision dans la mise en œuvre des principes posés se retrouve tout au long de la recommandation comme en témoignent plusieurs autres articles[104].

Enfin, on peut relever une limite qui a trait aux modalités d'établissement des mécanismes touchant, par exemple, le contrôle des systèmes d'IA et leur évaluation, dans le sens où la recommandation sollicite une collaboration de différents acteurs qui ont des intérêts différents. En effet, l'article 60 pose le principe selon lequel les Etats membres doivent établir l'ensemble de ces mécanismes « en collaboration avec les organisations internationales, les sociétés transnationales, les établissements universitaires et la société civile… ». Or, les intérêts des sociétés internationales sont différents des intérêts des Etats ou de la société civile. C'est pourquoi il aurait été, peut-être, opportun de remplacer le terme « collaboration » par la possibilité de « solliciter l'avis… » de ces sociétés internationales sans leur donner la possibilité de participer à la détermination des règles. Le terme « collaboration » laisse penser que les normes juridiques d'un Etat seraient le produit d'une co-production entre de puissantes entreprises internationales et les pouvoirs publics. La nouvelle formulation proposée poserait ainsi le principe que l'Etat reste le garant de l'intérêt général mais qu'il lui est toujours possible, lorsqu'il souhaite définir certaines normes, d'être à l'écoute des attentes et sollicitations d'autres acteurs.

B. UN DROIT INTERNATIONAL DE L'INTELLIGENCE ARTIFICIELLE À L'ÉPREUVE DES FAITS.

La Recommandation[105] de l'Unesco, adoptée le 23 novembre 2021 relative à l'éthique de l'Intelligence artificielle, se heurte aussi à la réalité économique et politique de certains Etats qui semble ne pas adhérer vraiment à la philosophie d'un tel texte même s'ils ont pu l'adopter dans le cadre de l'Assemblée Générale de l'Unesco.

Cette recommandation n'a pas empêché ainsi la création, en mai 2022, au Danemark d'un parti politique totalement géré par l'IA se dénommant « le Parti Synthétique ». Les citoyens peuvent ainsi interagir avec ce « parti » piloté par l'IA. Les concepteurs de ce parti ont pour ambition de faire participer les citoyens à une réflexion collective qui a pour objet de « mettre en relation l'IA et l'humain pour comprendre ce que les gens pensent réellement de la politique et aller à contre-courant du discours politique dominant ». En communiquant avec le Parti, les citoyens ont la possibilité de mettre en avant leurs idées qui seront compilées par le système. Si une telle innovation peut sembler progressiste et consacrer une nouvelle dimension de la démocratie, on peut émettre cependant de sérieux doutes. L'algorithme ne va-t-il pas chercher à influencer les citoyens avec qui il échange ? ne va-t-il pas fabriquer de « l'opinion politique » sans réelle base populaire ou réalité ? Est-il vraiment en capacité de comprendre les subtilités et la complexité des témoignages et discussions avec les citoyens ? Est-il, par exemple, doté d'une sensibilité telle qu'il serait en capacité de voir les hésitations d'un citoyen au cours d'une discussion ou ses doutes ? Peut-il donner une véritable clarté au discours confus et maladroits d'un citoyen ? Est-il en capacité de comprendre des pensées complexes ? Dans le monde réel, les hommes et femmes politiques ont cette capacité de ressentir des tendances dans l'opinion ou chez les individus en allant à leur contact. On peut imaginer aussi que le système d'IA peut faire des discriminations ou encore ne pas accorder la même importance à telle ou telle opinion. Dans cet exemple, on voit bien qu'une telle innovation est en contradiction avec la recommandation dans le sens où il n'y a aucune intervention humaine afin de rectifier une pensée fausse du système dans ses différents échanges. Rien ne garantit

104- Articles 63, 68, 72, 74, 96, 113 ou 133.

105- https://unesdoc.unesco.org/ark:/48223/pf0000381137_fre.

que l'algorithme ne fasse pas des biais discriminatoires.

La Recommandation de l'Unesco, dans son article 26, dispose « les systèmes d'IA ne devraient pas être utilisés à des fins de notation sociale ou de surveillance de masse ». Pourtant cela ne semble pas empêcher la Chine de consolider régulièrement, et cela depuis 2014, son logiciel de reconnaissance faciale afin d'assurer un certain contrôle de sa population. L'Algorithme permet de contrôler les citoyens dans différents lieux tels que dans la rue, les étudiants dans les lycées, la circulation routière ou dans certains magasins. Le système de reconnaissance faciale se fonde sur le « bouclier d'or » qui s'organise autour d'un réseau de surveillance numérique dans tout le pays. Celui-ci crée un réseau fermé (intranet) entre les polices locale, régionale et nationale qui permet l'accès et le partage de dossiers d'identification des citoyens chinois en temps réel. En 2021, on estime à près de 500 millions[106] le nombre de caméras interconnectées installées en Chine. La reconnaissance faciale n'est plus limitée aux caméras connectées. La police utilise désormais des lunettes connectées à une IA pour identifier le citoyen et savoir si ce dernier est recherché. Les pouvoirs publics font appel aussi aux caméras de reconnaissance faciale afin de recenser les citoyens. Les caméras permettent de reconnaître les citoyens, leurs voitures ou encore de contrôler l'accès à certaines écoles et leurs étudiants. La reconnaissance faciale est aujourd'hui une obligation pour certains achats comme, par exemple, un téléphone. En effet, depuis le 1er décembre 2019, tous les citoyens doivent se faire enregistrer par reconnaissance faciale lorsqu'ils souhaitent acheter un téléphone portable. Ainsi toutes leurs données biométriques seront liées au numéro de téléphone. Nous avons ainsi la mise en place d'une véritable politique de contrôle sociale et la Recommandation de l'Unesco ne semble pas faire évoluer les choses. La Chine continue à développer sa politique intérieure dans la plus totale souveraineté. Ainsi en 2015, les pouvoirs publics ont élaboré un système de notation des citoyens appelé « crédit social ». Les caméras à reconnaissance faciale identifient le citoyen et ajoutent ou soustraient des points à ce qui est défini en Chine comme la « seconde carte d'identité ». Par exemple, dans le cas des passages piétons, dans la ville de Shenzhen, le citoyen qui ne respecte pas le feu voit son nom affiché sur un écran géant et perd 20 points sur les 1 000 qui lui ont été crédités. Cette notation est connue des autres citoyens et entreprises comme les banques par exemple. La perte de points peut influencer l'accès à un poste ou l'obtention d'un prêt. Si la somme des points est trop faible, le citoyen peut alors être placé sur «une liste noire», le gouvernement applique alors des sanctions comme l'interdiction de prendre le train, d'inscrire son enfant dans une grande école, créer une entreprise ou encore avoir une carte de crédit. Il faut noter que les Nations-Unies, en septembre 2021, avaient alerté sur les dangers de la reconnaissance faciale qui, sans balises, risque de porter atteinte à certains droits fondamentaux comme le droit à la vie privée, à la liberté de circulation[107], d'expression ou encore l'accès à certains droits sociaux[108]. En Europe, d'après une étude[109], onze pays possèdent déjà d'un système de reconnaissance faciale, la plupart du temps, utilisé à des fins judiciaires. C'est le cas de la France, l'Autriche, la Finlande, l'Allemagne,

106- https://www.rts.ch/info/monde/11137943-la-chine-veut-noter-tous-ses-habitants-et-installe-600-millions-de-cameras.html.

107- En décembre 2021 on apprenait que la province chinoise du Henan était en train de développer un système de surveillance doté de reconnaissance faciale. Ce dernier aurait pour objectif de détecter les journalistes, les étudiants étrangers et autres « personnes préoccupantes » pour les autorités (https://www.actuia.com/actualite/chine-une-province-utiliserait-lintelligence-artificielle-pour-cibler-les-journalistes-et-les-etudiants-etrangers/)

108- En 2021 aux États-Unis, une reconnaissance faciale défaillante empêche les demandeurs d'emploi de toucher leur chômage (https://www.clubic.com/technologies-d-avenir/intelligence-artificielle/actualite-375452-aux-etats-unis-une-reconnaissance-faciale-defaillante-empeche-les-demandeurs-d-emploi-de-toucher-leur-chomage.html).

109- https://www.usine-digitale.fr/article/la-reconnaissance-faciale-est-deja-utilisee-dans-11-pays-europeens-dont-la-france.N1154082

la Hongrie, la Grèce, l'Italie, la Lettonie, la Lituanie, la Slovénie et les Pays-Bas.

Le texte voté dans le cadre de l'Unesco doit faire face également à l'apparition des Systèmes d'Armes Létales Autonomes (SALA) qui sont en fait des robots tueurs[110]. Ces systèmes d'armement ont fait l'objet de discussions lors de la Convention des Nations Unies portant sur les Armes Classiques (CCAC). Fin 2021, le sujet a de nouveau été abordé par les 125 Etats réunis à Genève mais, faute d'accord, le dossier fut renvoyé devant les experts gouvernementaux. Ces systèmes se développent alors même que la Recommandation, dans son article 26, met bien en avant que « Dans les scénarios où les décisions sont considérées comme ayant un impact irréversible ou difficile à renverser, ou qui pourraient impliquer des décisions de vie et de mort, la décision finale devrait être prise par l'homme ». Pourtant toutes les grandes puissances investissent régulièrement dans l'intelligence artificielle pour leurs armées. Vladimir Poutine a déclaré, en 2017, à ce sujet : « Celui qui deviendra le leader dans ce domaine sera le maître du monde ». Les Etats-Unis, La Chine, la Russie, la Corée du Sud, l'Australie, l'Inde, la Turquie et Israël investissent dans le développement de systèmes d'armes autonomes meurtriers qui peuvent identifier, cibler et tuer une personne par eux-mêmes sans qu'aucune loi internationale ne régisse leur utilisation. Donner une décision de vie et de mort à une technologie à partir de données pose évidemment un problème d'éthique. Une machine ne comprendra jamais le contexte d'une action ni ses conséquences. Sans l'intervention humaine, la proportionnalité d'une attaque, la distinction entre civils et militaires ne seront pas prises en compte par le système d'IA. Les droits de la guerre seront alors bafoués : un groupe d'hommes voulant se rendre, par exemple, sera reconnu comme des attaquants et abattus… Noel Sharkey, président du Comité International pour le contrôle des armes robotisées, expert en IA et en robotique à l'université britannique de Sheffield, déclarait à ce sujet : « Ne vous laissez pas tromper par l'absurdité de l'intelligence de ces armes… Vous ne pouvez tout simplement pas faire confiance à un algorithme – aussi intelligent soit-il – pour rechercher, identifier et tuer la bonne cible, surtout dans la complexité de la guerre »[111]. C'est pour cette raison que depuis 2018, le Secrétaire général des Nations unies, António Guterres, invite les États à interdire les armes qui ciblent et attaquent d'elles-mêmes des êtres humains, les qualifiant de « politiquement inacceptables et moralement révoltantes ».

CONCLUSION

La Recommandation de l'Unesco si elle constitue incontestablement une avancée au niveau international reste néanmoins d'une portée modeste. Elle est comme son intitulé l'indique une simple « Recommandation » qui, nous semble, ne pose pas de véritables obligations et reste très générale dans les expressions employées. Elle est, avant tout, d'un texte s'apparentant à de la « Soft law » ou à du droit de faible intensité. Il nous semble que les enjeux[112] de l'IA sont

110- https://www.actuia.com/actualite/robots-tueurs-les-nations-unies-peinent-a-poser-un-cadre-juridique-a-lemploi-des-armes-letales-autonomes/

111- C'est pourquoi un rapport intitulé « Crunch Time on Killer Robots: Why New Law Is Needed and How It Can Be Achieved » (« Urgence au sujet des robots tueurs : Nécessité d'une nouvelle loi et comment y parvenir ») publié en janvier 2022 enjoint les gouvernements à ouvrir des négociations afin d'aboutir à un nouveau traité pour conserver un contrôle humain de l'usage de la force. Ce rapport de 23 pages a été écrit par Human Rights Watch et l'International Human Rights Clinic de la faculté de droit de Harvard. Il constate que le droit international devrait être renforcé et clarifié pour protéger l'humanité des dangers que représentent les systèmes d'armes létales autonomes. En effet, ces armes choisissent des cibles et les visent sans véritable contrôle humain (https://www.actuia.com/actualite/robots-tueurs-les-nations-unies-peinent-a-poser-un-cadre-juridique-a-lemploi-des-armes-letales-autonomes/).

112- Le 22 juillet 2022, Eric Schmidt, dirigeant de Google de 2001 à 2011, a participé au forum sur la sécurité d'Aspen (États-Unis). Lors de cette manifestation, l'ancien président exécutif a comparé l'intelligence artificielle (IA) aux bombes nucléaires. Il milite alors pour un traité de dissuasion à l'instar des armes nucléaires. Quelques années auparavant, en 2018, En 2018, Sundar Pichai, l'actuel patron de Google, exprimait déjà ses inquiétudes en déclarant : « Les progrès de l'intelligence artificielle n'en sont encore qu'à leurs

tels qu'un droit international de l'IA devrait revêtir les caractéristiques de la contrainte pour les Etats et faire preuve d'une plus grande précision dans les dispositions afin d'éviter aux Etats, comme aux sociétés privées, de contourner des règles trop générales et peu claires. C'est dans ce sens que le Groupe d'experts de haut niveau sur l'intelligence artificielle de la Commission européenne, qui compte des experts de plus de 20 sociétés, appelle à envisager l'adoption d'une nouvelle législation très stricte imposant des obligations précises. Dans son rapport[113] en date de 2019 on peut lire que «Pour les systèmes d'IA déployés par le secteur privé qui sont susceptibles d'avoir une incidence significative sur les vies humaines, par exemple en portant atteinte aux droits fondamentaux d'un individu à n'importe quel stade du cycle de vie d'un système d'IA, et pour les applications critiques pour la sécurité, envisager la nécessité d'introduire : une obligation de réaliser une évaluation fiable de l'IA (y compris une évaluation des incidences pour les droits fondamentaux couvrant également par exemple les droits de l'enfant, les droits des individus vis-à-vis de l'État et les droits des personnes handicapées) et une consultation des parties prenantes, notamment une consultation des autorités compétentes; des exigences en matière de traçabilité, d'auditabilité et de contrôle ex ante ; et une obligation de garantir des procédures appropriées par défaut et dès la conception pour permettre d'obtenir une réparation effective et immédiate en cas d'erreurs, de préjudices et/ou de toute autre atteinte aux droits ». Il aurait été aussi une bonne chose de s'inspirer de la proposition[114] de règlement européen de la Commission européenne en date du 21 avril 2021 qui est actuellement en discussion au Parlement européen. Il a été fait le choix non pas d'une directive mais d'un règlement européen qui est d'application directe. Ce choix est guidé par la nécessité d'une application uniforme des nouvelles règles, notamment, en ce qui concerne la définition de l'IA, l'interdiction de certaines pratiques préjudiciables et la classification de certains systèmes d'IA. Cette applicabilité directe du règlement permet de réduire la fragmentation juridique et facilite la mise en place d'un marché unique pour des systèmes d'IA licites, sûrs et dignes de confiance. Le règlement propose ainsi un ensemble harmonisé de prescriptions fondamentales pour les systèmes d'IA classés à « haut risque » ainsi que des obligations pour les fournisseurs et les utilisateurs de ces systèmes devant permettre de mieux protéger les droits fondamentaux et d'assurer une sécurité juridique dans l'intérêt des opérateurs et des consommateurs. La Recommandation de l'Unesco, du fait de ses caractéristiques, va au contraire encourager l'existence d'une multitude de réglementations d'un Etat à un autre. Le texte de l'Unesco aurait pu intégrer également, comme le mentionne la proposition de l'Union européenne, une liste des pratiques d'IA interdites dont l'utilisation est considérée comme inacceptable car contraire aux valeurs universellement reconnues. On peut relever également que, contrairement à l'Unesco, la proposition de règlement impose de véritables obligations pour les systèmes d'IA qui présentent un risque élevé pour la santé, la sécurité ou les droits fondamentaux des personnes physiques. A ce sujet la CNDH / France recommande meme que le règlement européen aille plus loin dans le caractère contraignant du cadre juridique.

débuts, mais je considère qu'il s'agit de la technologie la plus profonde sur laquelle l'humanité travaillera et nous devons nous assurer que nous l'exploitons au profit de la société (…) Le feu tue aussi des gens. Nous avons appris à le contrôler pour le bien de l'humanité, mais nous avons aussi appris à en maîtriser les mauvais côtés ».

113- https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence.

114- Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence, https://ec.europa.eu/commission/presscorner/detail/fr/ip_21_1682.
Voir aussi : https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a37211-eb-958501aa75ed71a1.0020.02/DOC_1&format=PDF ; https://www.actuia.com/actualite/le-parlement-europeen-adopte-les-recommandations-de-sa-commission-speciale-sur-lintelligence-artificielle/

Saida Belouali

Toumi Bouchentouf

École Nationale de Sciences Appliquées. Université Mohammed Ier, Oujda, Maroc

Tension entre automatisation et responsabilisation en éthique appliquée à l'Intelligence Artificielle

INTRODUCTION

Des technologies largement disruptives s'imposent de plus en plus, modifient la vie et le monde et soulèvent des questions éthiques qui doivent interpeller l'ensemble des acteurs concernés. L'Intelligence Artificielle (IA) est l'une de ces technologies qui semble transformer la face de notre monde et son évolution effrénée nous interroge inéluctablement. Le terme d'intelligence artificielle (IA), utilisé pour la première fois lors de la conférence de Dartmouth [1] a plusieurs définitions et son sens n'a cessé d'évoluer depuis cette date-là.

La définition que nous souhaitons utiliser dans le cadre de cet article est celle qui l'assimile à un «ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine" [2]. La simulation telle que la précise Wang : "Les êtres humains diffèrent des animaux et des machines principalement par leur capacité mentale ou cognitive, communément appelée « intelligence », et l'IA est la tentative de reproduire cette capacité dans les systèmes informatiques. Ce vague consensus impose des contraintes importantes sur la manière dont l'IA doit être définie" [3]

Les évolutions technologiques dans ce domaine sont très importantes et interviennent dans des domaines variés et divers et améliorent considérablement nos sociétés et nos économies. Elles sont cependant fréquemment liées à une appréhension et font l'objet d'une vision alarmante. Notamment lorsque des chercheurs et experts ont pris les devants pour dénoncer en 2015 l'intelligence artificielle et brosser au passage un tableau sombre de ce que pourrait être le devenir de l'humanité. [4]

Dès lors, l'inquiétude s'accroît proportionnellement aux avancées technologiques et plusieurs tribunes, organisations et chercheurs mettent en garde contre un manque de vigilance qui aboutirait à une absence d'anticipation et d'encadrement. Il est donc nécessaire d'évaluer les impacts de cette technologie mais ne faut-il pas également en délimiter les contours. De quoi parle-t-on exactement ? L'acception IA est-elle polysémique, recouvre-t-elle les mêmes phénomènes ? Devrions-nous préciser davantage les subtilités conceptuelles pour mieux cerner les impacts et prévenir les méfaits ? Entre une IA totalement autonome et une IA où l'être humain garde encore le contrôle, sommes-nous face aux mêmes considérations éthiques ?

Il nous semble nécessaire d'introduire quelques éclairages critiques pour une prise en compte de cette différence du rythme en matière de progrès de l'IA. En matière de régulation éthique de l'IA, ne faudrait-il pas donc prendre en considération les interférences entre ces différents niveaux d'évolution de l'IA et déterminer en conséquence ce qui constitue aujourd'hui une urgence parce que source d'enjeux sociétaux immédiats et inévitables ?

Nous aborderons dans cet article, cette problématique et discuterons ce qui nous semble une des difficultés notoires aujourd'hui en matière d'IA, à savoir la question de territoire, de frontières et de pouvoir entre l'IA et l'Homme. Néanmoins nous ne pouvons pas discuter de ce point sans évoquer également l'antagonisme : IA forte et IA faible ; c'est-à-dire une IA accompagnée par une intervention humaine possible à tout moment (IA faible) ou une automatisation absolue dénuée de toute

supervision. Il est en effet important de savoir à quel objet nous appliquons l'éthique : une IA qui imite le comportement humain ou une IA qui aspire à reproduire l'intelligence humaine ? L'urgence éthique concerne-t-elle davantage dans ce cas l'IA faible ou l'IA forte ? Laquelle appellerait un encadrement éthique immédiat ?

PRÉCISIONS CONCEPTUELLES

Il est judicieux avant de parler de l'éthique et de tenter de l'appliquer à l'IA, de procéder à un bref déchiffrage en matière de concepts liés à l'Intelligence Artificielle ; il est nécessaire de démêler les acceptions quand on cherche à appliquer l'éthique à l'IA, à décréter des chartes et codes et édicter des principes. Une terminologie plus nuancée est nécessairenous semble-t-il pour pouvoir clarifier le cadre conceptuel utilisé. De surcroît, pour pouvoir opérationnaliser les principes d'éthique de l'IA, il est primordial qu'il n'y ait pas de doute sur l'objet lui-même. Cet effort nous semble en outre, nécessaire pour pouvoir décider à quel niveau serait appliquée la responsabilité et à qui incombe-t-elle et comment procéder pour atténuer les risques potentiels.

Deux grands courants totalement opposés ont marqué l'évolution de l'IA, le premier dit symbolique (ou logique) [5] et l'autre dit connexionniste ou computationnel [6]. Le premier consiste à reproduire le raisonnement humain par des règles codées alors que le deuxième courant s'inspire du fonctionnement des neurones biologiques. On parle alors de réseaux de neurones (RN). Ce dernier courant a pu grandement évolué par l'apparition de l'apprentissage automatique dit statistique qui permet d'utiliser au sein de ces réseaux de neurones des méthodes statistiques de type probabiliste. L'apparition de l'apprentissage profond ou deep learning en 2010 et les résultats très importants obtenus soit en traitement des images [7] ou en traitement automatique du langage naturel (NLP) [8] ont pu rendre ce concept très populaire dans les milieux scientifique, économique, social et politique. Cela implique qu'il y a actuellement des différences majeures en termes d'évolution de l'IA et en matière de sa qualification. En effet, différentes classifications proposent entre 3 et 7 types d'IA en prenant en considération les critères notamment de la réactivité, de la capacité et de la mémorisation. Pour un confort conceptuel, nous nous limiterons ici à parler d'une seule forme d'antinomie à savoir une IA faible et IA forte. La différence entre une IA faible et une IA forte résiderait alors dans la capacité de "simuler à 100 % le cerveau humain''. "Si un système d'IA doit être aussi intelligent que le cerveau, il devra égaler la capacité de calcul brute du cerveau" [9]. Néanmoins, notons qu'il s'agit bien de dupliquer l'intelligence de l'Homme et non pas l'Homme [3]. L'IA dite faible se préoccupe de l'automatisation de tâches spécifiques. L'IA forte chercherait à simuler le processus cognitif humain. "L'IA actuelle manque de bon sens, c'est-à-dire qu'elle n'est pas capable de déduire, de comprendre ou d'expliquer les processus, les forces et les causes cachés derrière les données" [10]. Si l'on admet que l'IA forte est aujourd'hui encore non complètement aboutie, au niveau de l'IA faible, l'on cherche davantage à automatiser et "une telle automatisation des tâches humaines et de la prise de décision peut bien sûr être bénéfique en économisant des ressources humaines et en conduisant (idéalement) à de meilleures solutions et décisions." [11]. Nous noterons que les techniques utilisées présentent encore des faiblesses, telles que le manque de transparence et d'explicabilité. Raison pour laquelle, il est fréquemment recommandé dans les chartes et codes d'usage de respecter des principes permettant de développer une IA digne de confiance. La Commission européenne le recommande dans le Livre blanc sur l'IA, publié en février 2020 parmi les principes fondamentaux celui d'une IA responsable. Moult chartes et code publiés par différents organismes, Etats et entreprises et portant sur l'utilisation de l'IA recommandent ainsi des systèmes transparents et donc responsables [12].

LA TENSION AUTOMATISATION / RESPONSABILISATION

Un tour d'horizon rien qu'au niveau de nos smartphones, on y trouve une panoplie d'applications qui utilisent des algorithmes de deep learning, les applications de traduction, d'auto-complétion, les chatbots, la reconnaissance faciale ou les systèmes de recommandations utilisés dans les réseaux sociaux et qui exploitent les données privées des personnes, etc.

Certaines de ces applications ont suscité beaucoup de débats quant à leurs impacts éthiques et sociaux. De façon générale, les algorithmes s'immiscent de plus en plus dans notre quotidien, régissent nos choix et orientent nos décisions. Une compréhension des fonctionnements, des usages des systèmes intelligents et de leurs conséquences est dès lors nécessaire pour gouverner l'IA et réguler ses usages. Un bref tour d'horizon sur certains cas notoires nous convainc en effet, de l'étendue des impacts que peuvent avoir les systèmes intelligents :

- le Deepfake [13] qui permet d'échanger des visages ou d'attribuer un faux discours à une personne
- la décision déléguée aux système intelligents dans un véhicule autonome notamment en cas de danger,
- les systèmes intelligents de recrutement et la discrimination qui en résulte en relation avec le genre ou l'appartenance à une communauté donnée.
- la prédiction de récidive dans le domaine de la justice [14]
- la possibilité de créer des œuvres d'art [15], ou d'en finir d'autres comme ce fut le cas de la 8e symphonie de Schubert réalisée par la firme Huawei et Lucas Cantor [15].

L'ensemble des systèmes intelligents cités ci-haut posent des problématiques d'ordres moral et sociétal notamment en lien avec les droits fondamentaux des êtres, leur liberté voire leur intégrité. Ces cas d'usage illustrent bien une forme d'antinomie entre responsabilité de l'Homme et autonomie de la machine. Les transformations et l'évolution de l'IA vont de plus en plus vers davantage d'automatisation des tâches et activités humaines. Cette course vers plus d'automatisation nous semble mettre en tension l'idée de remettre l'Homme au centre des avancées technologiques en IA. Quel équilibre pendulaire serait-il possible entre cette idée de "l'Homme au centre" et davantage d'automatisation ? La question légitime à se poser porte alors sur jusqu'où l'intelligence artificielle pourra-t-elle ou doit-elle remplacer l'humain ? Est-il sensé d'aller vers une prise de décision totale pour des raisons économiques et sociales ? Voulons-nous par ailleurs, qu'en matière d'IA, l'humain soit au cœur du processus de décision ? Comment pourrait-on articuler alors le plus d'automatisation avec la préoccupation de "plus d'homme" qui est souvent exprimée par le paradigme "l'Homme au centre" ? Les réponses éthiques qui pourraient être apportées à ces questionnements nous semblent relever d'abord d'une question philosophique fondamentale : quels sont les choix de vie que l'on va opérer et quel est le monde dans lequel souhaite vivre l'Homme ? L'évolution technologique dans sa globalité devrait être interrogée sous l'angle de la phénoménologie. La technologie façonne notre monde perceptif. « La technologie offre toujours une nouvelle perception du monde » [16]. Il nous semble qu'il est important de distinguer deux niveaux différents quand nous évoquons la régulation morale de l'IA :

- Le niveau qui concerne l'IA faible et qui permet d'édicter des principes régulateurs qui fonctionnerait toujours en se basant sur la délibération éthique de l'Homme. Nous faisons appel à l'éthique appliquée qui nous permet de résoudre des dilemmes et de considérer des problématiques éthiques pour opérer les choix adéquats. Ce niveau correspondrait à une éthique par « conception, à priori, qui repose sur l'implémentation de théories éthiques dans la technique

pour que son fonctionnement soit conforme à des valeurs et/ou principes prédéfinis » [17].

- Le niveau qui concerne l'IA forte et pour lequel nous devons convoquer plutôt l'éthique qui permettrait de se poser des questions philosophiques relatives à un choix d'orientation sociétale. L'approche à ce niveau maintient le sujet au centre. L'approche est anticipatrice et si dans le niveau I, on établit lors des phases de conception des systèmes, les exigences éthiques, à ce niveau-ci, l'on doit se poser la question fondamentale : cette technologie a-t-elle lieu d'être, devrait-elle être développée ?

Pendant des années, les experts et scientifiques se sont penchés sur la question du déterminisme technologique ; Faudra-t-il amener la question au début, aux origines ? Certaines innovations doivent-elles exister ? La question éthique n'est-elle pas à poser au niveau originel, à la genèse même de la technologie ? [18] « En outre, des moratoires existent sur le développement de technologies dangereuses telles que les armes chimiques, biologiques et nucléaires en raison des effets dévastateurs que ces technologies peuvent exercer sur l'humanité. De même, nous soutenons que certains types de recherche sur l'intelligence artificielle entrent dans la catégorie des technologies dangereuses et devraient être restreintes. » [19]

L'IA, UNE ÉVOLUTION À DEUX VITESSES ET UNE VIGILANCE DÉSAPPARIÉE

Nous noterons que trois types d'apprentissage existent au sein de l'apprentissage automatique ou machine learning [20]. Le premier est dit supervisé pour lequel l'intervention d'un expert est nécessaire pour fournir un grand nombre de données en indiquant pour chacune des données le résultat attendu. Le deuxième type est l'apprentissage non supervisé où les algorithmes doivent apprendre sans expert et le troisième apprentissage fonctionne par renforcement et consiste à apprendre à partir d'expériences successives de façon à trouver la meilleure solution. La plupart des résultats obtenus en deep learning sont dus à la puissance des ordinateurs mis à disposition, ce qui a donné lieu à une course effrénée entre les différents acteurs de ce domaine [21]. Cela a permis que nous puissions stocker un très grand nombre de données ; ce qui a mis en lumière la notion de Big data. Nous constatons aujourd'hui le nombre impressionnant de données mises en ligne pour le grand public en open Data ou Dataset en libre-service. Le fonctionnement du deep learning repose sur l'extraction de caractéristiques appelées pattern à partir d'un très grand nombre de données brutes comme les images ou le texte afin de prédire les résultats [22].

Deux problématiques majeures sont à souligner à ce niveau-là : l'absence de transparence et le nombre élevé de biais qui découlent du mode de manipulation des données. Nous soulignons que ce système fonctionne en mode "boite noire" ; ce qui implique que nous ne savons pas exactement comment les résultats sont obtenus . Cette démarche pose en effet, un certain nombre de problèmes éthiques. En effet, à qui doit-on imputer la responsabilité des résultats d'une IA, au Data scientist responsable des données ? à l'ingénieur -développeur de l'application IA ou à la société cliente qui utilise cette IA ? L'amputation de la responsabilité est un problème majeur qui se pose en matière de régulation et d'encadrement éthique. Ensuite, au niveau des données, il y a énormément de biais comme le biais de sélection ou le biais de confirmation. Au niveau des algorithmes il y a aussi des millions, voire des milliards de dilemmes que les algorithmes doivent considérer. L'exemple le plus significatif est le fonctionnement de la recommandation des vidéos Youtube basée sur le nombre de vues de l'utilisateur [23]. Les problématiques que pose ce genre de cas sont relatives notamment à la nature des recommandations faites à l'usager : devrait-elle lui proposer toujours des vidéos selon ses préférences ?

Ou doit-elle de temps en temps proposer des vidéos illustrant des avis opposés ? Dans ce cas précis, nous nous retrouvons confrontés à un processus de prise de décision dans lequel nous devons faire intervenir tous les acteurs si nous voulons que l'humain soit au cœur de cette prise de décision.

Cette idée de l'Homme au centre nous renvoie justement à cette antinomie entre IA faible et IA forte. La forme la plus usuelle de l'IA utilisée en industrie actuellement est l'IA dite faible [24] considérée à responsabilité limitée. Il existe également un autre type d'intelligence artificielle dite forte qui serait à l'image de l'Homme avec une intelligence cognitive supérieure. La question qu'il est légitime de se poser est : une telle intelligence est-elle possible ? Il faudrait alors répondre aux questions suivantes : l'IA pourrait-elle développer une conscience, une sensibilité et une volonté autonome ? L'enjeu étant celui de la singularité [25]. Celle-ci renvoie à l'hypothèse d'une intelligence artificielle qui déclencherait une évolution technologique avec des changements imprévisibles pour la société humaine. Pour Jean Gabriel Ganascia, ce sont là des buzzwords [26] qui sont proclamés par des chercheurs comme Ray Kurzweil [27] de Google et Nick Bostrom [28] de l'université d'Oxford. Nous pourrons alors imaginer une automatisation de la technologie, s'affranchissant de la tutelle humaine. Certains centres de recherche ont déjà commencé, il y a quelques années, à développer des IA superpuissantes. Comment limiter, contrôler, anticiper et orienter ces algorithmes ? Il existe plusieurs exemples de ces cas d'usage de l'IA qui semblent alarmants :

- Les armes autonomes peuvent décider de certaines actions de manière indépendante sans l'intervention humaine.
- L'application Tek fog [29]: un outil qui permet de créer de fausses informations en permettant de recevoir des informations sur votre compte WhatsApp alors que votre numéro de téléphone est inactif.
- Xiaoice [30] : une IA conversationnelle qui a conversé en décembre 2020, avec près de 660 millions de personnes. Certaines personnes se sont même senties émotionnellement très proches de cette IA. Le point fort d'une IA pareille est sa totale disponibilité pour tous ses utilisateurs ; ce qui est impossible pour un être humain.
- Le méga projet Pathways [31] de Google : concevoir un modèle unique pour prendre en charge des milliers ou des millions d'actions. Au lieu de développer des milliers de modèles séparés comme le font les entreprises actuellement, Google a choisi d'entraîner une poignée de modèles généraux capables de générer une seule sorte d'entrée ; l'idée est que ce modèle soit capable d'effectuer des milliers, voire des millions de tâches différentes. Et on peut ajouter de nouvelles tâches au fur et à mesure. Pathways pourra gérer toutes sortes de données à la fois, et apprendre de manière incrémentale en fonction des besoins en appelant les composants pertinents du modèle.

Les quelques cas que nous avons cités plus haut, illustrent bien ce phénomène d'immixtion des algorithmes dans différents domaines de la vie et montrent les problématiques éthiques qu'ils peuvent soulever. Néanmoins, nous estimons que pour l'ensemble des cas cités plus haut, la prise de décision est à chaque fois supervisée. L'Homme est encore au centre de la décision et c'est à lui qu'incombe l'orientation morale dès la conception du produit technologique. Le choix éthique demeure encore dans le design. Nous pensons que ce sont les choix que feront les Hommes de rendre les algorithmes plus transparents, accessibles et en open source pour que les spécialistes puissent vérifier leur fiabilité et les rendre robustes et explicables et interprétables par la suite, qui donneraient tout son sens à cette notion de l'Homme au centre d'une l'IA de confiance ou d'une IA responsable

et qui permettrait de préserver un équilibre entre le développement d'une IA maîtrisable et contrôlable qui ne piétinerait pas sur le territoire des hommes.

CONCLUSION

Nous avons soulevé ici, la question de l'automatisation/responsabilisation de l'intelligence artificielle, ce qui illustre une forme d'antinomie entre responsabilité de l'Homme et autonomie de la machine. L'intelligence artificielle de type computationnel qui est aujourd'hui la principale source des applications développées en IA grâce à la puissance des calculateurs et du Big data présente dans certains cas des problèmes éthiques avec des milliers voire des milliards de dilemmes. Au lieu de se poser la question de la possibilité que nous avons d'atteindre la singularité technologique ou pas, nous pensons qu'il est urgent d'entamer la réflexion autour des problématiques éthiques dans les applications de l'IA faible essentiellement et de considérer l'IA forte comme un objet technologique hypothétique et dont le sort devrait être différencié de celui de l'IA faible. Nous devrions donc nous intéresser plus à l'éthique des algorithmes qu'à la singularité ou à la super-intelligence. Notre proposition est que le processus de décision éthique devrait se focaliser davantage sur la transparence des algorithmes utilisés ainsi que sur leur robustesse en les rendant plus explicables et interprétables si nous voulons que l'Homme soit toujours au cœur du processus de décision.

Références

[1]: JULIA, Luc et GASSE, Jean louis, "l'intelligence artificielle n'existe pas ", Paris/61-Lonrai, First Editions / Normandie roto imp, dl 2019, 287 p.

[2]: https://www.larousse.fr

[3]: WANG, Pei. "What Do You Mean by" AI"?. In : AGI. 2008. p. 362-373.

[4]: https://futureoflife.org/2016/02/09/open-letter-autonomous-weapons-ai-robotics/

[5]: KODRATTOF, Yves "Leçons d'apprentissage, symbolique automatique", Cepadues, Broche, 1986.

[6]: BOURRET, Paul "Réseaux neuronaux. Une approche connexionniste de l'intelligence artificielle", Manuel Samuelides, Broche, 1991.

[7]: Siddharth Sharma and al, " A Guide to Machine Learning, Deep Learning, and their Applications" 2020.

[8]: CLARCK, Alexander FOX, Chris et LAPPIN, Shalom: " Handbook of computational linguistics and natural Language Processing", Wiley-Blackwell, 2012

[9]: O. Strelkova, O. Pasichnyk, STRELKOVA, O. Trois types d'intelligence artificielle. 2017.

[10]: BUTZ, Martin. V "Vers une IA forte". Künstl Intell 35, 91–101 (2021). https://doi.org/10.1007/s13218-021-00705-x

[11]: BOLANDER, Thomas. "Que perdons-nous lorsque les machines prennent les décisions ?" J Manag Gov 23, 849–867 (2019). https://doi.org/10.1007/s10997-019-09493-x

[12]: BELOUALI, Saïda, DEMAZEAU, Yves, BOUCHENTOUF, Toumi, et al. Interrogation de l'universalité d'une éthique pour l'usage de l'IA au travers d'une lecture du code publié par Dubaï. Ethique et Numérique, 2022, vol. 1, no 1, p. 1-17.

[13]: D. LITTMAN, Sarah :" Deepfake", Relié, 2020

[14]: BARRAUD Boris, « Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ? », Les Cahiers de la Justice, 2017/1 (N° 1), p. 121-139. DOI : 10.3917/cdlj.1701.0121.URL:https://www.cairn.info/revue-les-cahiers-de-la-justice-2017-1-page-121.htm

[15]: https://www.jasonmallen.com , https://creator.nightcafe.studio/

[16]: LIBERATI, Nicolas. Technologie, phénoménologie et monde quotidien : une analyse phénoménologique sur la façon dont les technologies façonnent notre monde. Hum Stud 39 , 189-216 (2016). https://doi.org/10.1007/s10746-015-9353-5

[17]: FISCHER, Flora, « L'éthique by design du numérique : généalogie d'un concept », Sciences du Design, 2019/2 (n° 10), p. 61-67. DOI : 10.3917/sdd.010.0061. URL : https://www.cairn.info/revue-sciences-du-design-2019-2-page-61.htm

[18] BELOUALI, Saïda et al. Ethics of AI

or Ethical AI, Topical Point of View. In: International conference on smart Information & communication Technologies. Springer, Cham, 2019. p. 535-540.
[19]: YAMPOLSKY, Roman.V, Artificial Intelligence Safety Engineering: Why Machine Ethics Is a Wrong Approach. In: Müller V. (eds) Philosophy and Theory of Artificial Intelligence. Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 5. Springer, 2013, Berlin, Heidelberg
[20]: El IRAOUI, Amine , " Apprentissage automatique", Broche, 2018
[21]: https://www.top500.org/lists
[22]: CHARNIAK Eugene, "Introduction au deep learning ", InfoSup, Dunod, 2021.
[23]: "un milliard de dilemmes", Chaine Youtube Sciences4all, https://youtu.be/BfGhNeEd9Nk
[24]: RAJNA, George, " Weak AI, Strong AI and Superintelligence", 2017.
[25]: "Singularity hypotheses: A scientific and Philosophical Assessment, Dordrecht, Springer, 2012, 1-2p.
[26]: GANASCIA, Jean Gabriel, " Le Mythe de la singularité. Faut-il craindre l'intelligence artificielle ? Broche, 2017.
[27]: KRZWEIL, Ray, " Humanité 2.0", Essai(Broché). 2007
[28]: BOSTROM, Nick, " Superintelligence Paths, Dangers, Strategies". 2015.
[29]: https://thewire.in/tekfog/en/2.html
[30]: https://www.infobip.com/answers
[31]: BARHAM, Paul et al., " Pathways: Asynchronous distributed dataflow for ML", Argxiv, 2022.

1. Les auteurs sont présentés par ordre alphabétique.
2. L'association de recherche OpenAI a évalué son générateur de texte, opéré par un modèle linguistique d'intelligence artificielle bien trop évolué pour ne pas être dangereux. L'association a refusé de rendre public le GPT2. L'approche opérée par le groupe de recherche californien est nécessaire. Elle correspond à l'éthique préventive.