



*ÉCOLE DOCTORALE MATHÉMATIQUE,
SCIENCE DE L'INFORMATION ET DE L'INGÉNIEUR- ED269*

THÈSE présentée par

Lalla Aicha KONÉ

soutenue le 15 octobre 2024 à l'Université de Strasbourg pour obtenir le grade de :

Docteur de L'université de Strasbourg

Discipline/Spécialité : INFORMATIQUE

IA et éthique pour l'amélioration du service public

Soutenue à Strasbourg, le 15 octobre 2024

THÈSE dirigée par :

Dr Aline DERUYVER
Pr Ahmedou HAOUBA

MCF HDR, université de Strasbourg, France
Professeur, université de Nouakchott, Mauritanie

ENCADRANT :

Pr Pierre COLLET

Professeur, université Andrés Bello, Chili

RAPPORTEURS :

Jean SALLANTIN
Pr Juliette VION-DURY

DR émérite, université de Montpellier, France
Professeur, université de Sorbonne Paris-Nord

EXAMINATEURS :

Pr Farouk NANNE
Dr Mohamed Lemine SALIHI
Dr Mamadou Tourad DIALLO
(co-encadrant)

Professeur, université de Nouakchott, Mauritanie
MCF, université de Nouakchott, Mauritanie
Docteur, université de Nouakchott, Mauritanie

INVITÉE :

Dr Anna OUSKOVA LEONTEVA Co-encadrante

Dédicace

Je dédie ce travail, de tout mon cœur :

À mes parents, pour leur amour inconditionnel, leur soutien indéfectible et leurs sacrifices sans limites. Leur foi en moi a été ma lumière dans les moments sombres de cette aventure académique.

A Pathé SECK, NDack SECK, Mohamed El Hady SECK et à tous mes parents, cousins et cousines du Sénégal.

À mes professeurs et mentors, pour leur aide précieuse, leur expertise généreusement partagée, et leur inspiration constante. Leurs conseils éclairés ont façonné ma pensée et nourri ma passion pour la recherche.

À mes collègues de laboratoire, pour leur collaboration, leurs conseils et les moments de détente qui ont rendu ce parcours plus agréable.

À mes amis et proches, pour leur compréhension et leur encouragement tout au long de ce parcours exigeant.

À tous ceux qui ont croisé ma route et contribué, de près ou de loin, à la réalisation de cette thèse, je vous adresse ma profonde gratitude. Vos contributions ont enrichi mon travail et ont rendu cette aventure académique encore plus significative.

Enfin, à tous les esprits curieux et les âmes passionnées qui cherchent inlassablement à repousser les frontières de la connaissance, cette thèse vous est dédiée.

Remerciements

Je souhaite d'abord exprimer toute ma profonde reconnaissance envers mes directeurs de thèse à qui je dois tout, Pr A. HAOUBA et Pr P. COLLET, mes encadrants qui m'ont accompagnée du mieux qu'ils le pouvaient, Dr T. DIALLO et Dr Mohamed Lemine SALIHI et, enfin, Dr. Anna LEONTEVA qui fit son apparition dans ce parcours doctoral à un moment où j'avais le plus besoin d'un compagnon avec qui échanger au quotidien. A Tous, je dis du fond du cœur : Un grand Merci pour votre aide, votre expertise et votre soutien constant. Vos conseils éclairés et votre mentorat ont été d'une valeur inestimable pour la réussite de ce travail de recherche.

A la Direction de l'Orientation et des Bourses du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, je tiens à exprimer mes remerciements, pour nous avoir autorisé l'accès à leurs données qui ont permis de réaliser les tests dans le cadre de cette thèse. Un grand Merci à Moussa DIA, chef du service de l'Orientation, pour les échanges enrichissants.

Dr Ahmed Salem et Mr Moussa Diarra du Nic-Mauritanie m'ont accueillie dans leur espace de travail et mise dans de bonnes conditions pour finaliser la rédaction de cette thèse. Qu'ils en soient chaleureusement remerciés.

Lors de mon parcours doctoral, j'ai eu des échanges amicaux et des discussions techniques avec des collègues du laboratoire et des collègues de l'Ecole Doctorale, des discussions et des échanges qui m'ont enrichie à titre personnel et à titre professionnel. Je leur dis Merci pour votre soutien.

Je tiens à remercier le Service de Coopération et d'Action Culturelle (SCAC), de m'avoir accordé une mobilité afin de réaliser mes travaux de recherche en France, à Strasbourg, sur un sujet passionnant. Mes remerciements vont également à l'endroit de tous les agents qui ont eu à gérer ma mobilité.

Je remercie Dr Tourad Diallo, Dr ELveth Sidi et Dr Ahmed Sejad, pour m'avoir aidé à assurer des missions d'enseignements, me donnant ainsi la possibilité de valider un module obligatoire aux yeux de l'école doctorale de la FST de l'Université de Nouakchott.

Que ceux qui président aux destinées de notre institution ou qui m'ont enseignée, trouvent, ici, l'expression forte de notre gratitude. Je pense plus particulièrement aux Professeurs : Sidya (Doyen de la FST), Diagana (Vice Doyen), Beddy (Responsable de la formation doctorale), Mohamed Vall (Directeur de l'école doctorale), et Isselmou Babbah.

Enfin, je tiens à remercier ma famille et surtout mon cher papa pour leur soutien inconditionnel et leur patience tout au long de cette aventure académique. Merci d'avoir cru en moi.

Résumé

À une époque où l'intelligence artificielle (IA) transforme nos habitudes et les industries mondiales, l'intégration de l'IA dans divers domaines est primordiale, notamment pour aider les gouvernements à penser des choix structuraux d'autant plus importants dans le cas de pays en développement et ce, tout en respectant des principes éthiques.

Le premier défi de cette thèse était de tester le potentiel de l'IA pour la prédiction et la prise de décision à partir de séries temporelles courtes, à défaut des données massives, celles-ci faisant typiquement défaut pour de tels objectifs. En effet, l'évolution d'un indicateur annuel pendant 100 ans ne fournit que 100 points de données, mais même cela n'est pas disponible dans les pays en de développement.

La première contribution porte sur la prédiction des statistiques d'admission au baccalauréat en Mauritanie. Les données historiques étant limitées, il est essentiel de disposer d'une technique de prédiction adaptée aux séries temporelles courtes. Cette thèse propose un outil basé sur la programmation génétique et le filtre de Kalman, améliorant la précision des prédictions à court terme. La validation sur divers ensembles de données démontre la supériorité de l'outil sur les techniques de la littérature, promettant ainsi des améliorations significatives dans le système éducatif mauritanien.

Par ailleurs, compte tenu des risques et des menaces qu'il pose, le recours à l'IA interpelle et appelle à tenir compte de principes éthiques, qui doivent nécessairement s'aligner sur les normes socioculturelles locales, ce qui constitue un autre défi de taille.

La deuxième contribution présente une solution intuitive utilisant l'IA générative visant à permettre aux gouvernements d'établir des principes éthiques locaux pour les logiciels d'IA. Cette solution consiste en deux applications web : l'une pour les gouvernements et l'autre pour les développeurs de logiciels. L'application centrée sur le gouvernement calibre de manière dynamique les poids éthiques dans différents domaines en fonction du contexte socioculturel, créant ainsi des plans éthiques sur mesure pour chaque domaine. Quant à l'application destinée aux développeurs, elle évalue activement la conformité des logiciels avec les principes éthiques établis par le gouvernement et fournit un retour d'information pour leur recalibration éventuelle.

Ensemble, ces contributions ouvrent la voie à un déploiement éthique de l'IA dans le respect des valeurs locales et à l'amélioration des prévisions pour l'amélioration du service public des pays en développement.

Abstract

At a time when artificial intelligence (AI) is transforming our habits and the world's industries, the integration of AI in various fields is essential, particularly to help governments think through structural choices, which is especially important in the case of developing countries, while at the same time trying to respect ethical principles.

The first challenge of this thesis was to test the potential of AI for prediction and decision-making based on short time series, in the absence of massive data, which is typically lacking for such purposes. Indeed, the evolution of an annual indicator over 100 years provides just 100 data points, but even this is not available in developing countries.

The first contribution concerns the prediction of baccalaureate admission statistics in Mauritania. As historical data are limited, it is essential to have a prediction technique adapted to short time series. This thesis proposes a tool based on genetic programming and Kalman filter, improving the accuracy of short-term predictions. Validation on various data sets demonstrates the tool's superiority over literature techniques, promising significant improvements in the Mauritanian education system.

On the other hand, given the risks and threats it poses, the use of AI raises questions and calls for ethical principles to be taken into account, which must necessarily be aligned with local socio-cultural norms — another major challenge.

The second contribution presents an intuitive solution using generative AI to enable governments to establish local ethical principles for AI software. This solution consists of two web applications: one for governments and the other for software developers. The government-focused application dynamically calibrates ethical weights in different domains according to socio-cultural context, creating tailor-made ethical plans for each domain. As for the developer application, it actively assesses the compliance of software with the ethical principles established by the government and provides feedback for their possible recalibration.

Together, these contributions pave the way for the ethical deployment of AI in line with local values, and for improved forecasting for public service improvement in developing countries.

Table des Matières

Liste des figures	iv
Liste des tables	iv
1 Introduction	1
1.1 Contexte	1
1.1.1 État des services gouvernementaux	2
1.1.2 État des lieux de l'Intelligence Artificielle (IA)	3
1.2 Problématiques et objectifs	4
1.2.1 Services gouvernementaux et prise de décision	4
1.2.2 Principes éthiques de l'Intelligence Artificielle (IA)	6
1.3 Pertinence et importance	13
1.4 Contributions	14
1.5 Structure de la thèse	15
I Revue de la littérature philosophique et scientifique	17
2 Revue de la littérature philosophique sur l'éthique	18
2.1 Éthique et méta-éthique	20
2.1.1 Qu'est-ce que l'éthique?	20
2.1.2 Champs de l'éthique	21
2.2 Éthique déontologique	23
2.2.1 Théories déontologiques centrées sur l'agent	24
2.2.2 Théories déontologiques centrées sur le droit des personnes	25
2.2.3 Théories déontologiques Contractualistes	25
2.2.4 Théories déontologiques	26
2.2.5 Avantages des Théories déontologiques	26
2.2.6 Inconvénients des Théories Déontologiques	27
2.2.7 Déontologie et incertitude des issues	27
2.2.8 Théories déontologiques et méta-éthique	28
2.2.9 Applications pratiques de la déontologie et de l'éthique	28
2.3 Apports des philosophes au progrès de l'éthique	29
2.3.1 Philosophie morale et politique de Thomas Hobbes (1588-1679)	29
2.3.2 David Hume et l'utilité (1711-1776)	29

2.3.3	Éthique déontologique d’Immanuel Kant (1724-1804)	30
2.3.4	Éthique conséquentialiste de Jeremy Bentham (1748-1832)	32
2.3.5	Éthique utilitariste de John Stuart Mill (1806-1873)	34
2.3.6	Henry Sidgwick et l’utilitarisme (1838-1900)	36
2.4	Éthiques modernes	40
2.4.1	Principes de justice de John Rawls (1921-2002)	40
2.4.2	Éthique du Care de Carol Gilligan (1936—)	43
2.4.3	Éthique du Discours de Jürgen Habermas (1929—)	44
2.4.4	Éthique des données (Data Ethics) de Luciano Floridi	45
2.5	Philosophie “occidentale”	46
2.5.1	Philosophie analytique	46
2.5.2	Éthique de l’intérêt	47
2.5.3	Philosophie continentale	48
2.5.4	Éthique contractualiste, éthique du désintérêt	51
2.5.5	Conclusion sur les éthiques occidentales	53
2.6	Méta-éthiques universalistes	53
2.6.1	Introduction à l’utilité d’une éthique universelle	53
2.6.2	Claude Lévi-Strauss (1908 - 2009) et Jürgen Habermas (1929—), pour une éthique universelle	54
2.6.3	Proposition d’éthique minimaliste, par Ruwen Ogien (1947 - 2017)	57
2.7	Méta-éthiques relativistes	59
2.8	Philosophie et éthique africaine	61
2.8.1	Philosophie africaine	61
2.8.2	Éthique africaine	62
2.9	Justice et droit des animaux et des robots	63
2.9.1	Justice et droit des animaux non-humains	64
2.9.2	Justice et droit des robots	65
2.10	Conclusion sur la revue de la littérature philosophique sur l’éthique	70
3	Revue de la littérature scientifique sur l’IA	71
3.1	Intelligence Artificielle (IA)	71
3.1.1	Définitions et applications de l’IA	71
3.1.2	Technologies et Domaines d’applications de L’IA	74
3.1.3	Prédiction pour l’aide à la décision	79
3.1.4	Travaux connexes lié à la prédiction de courtes séries temporelles	85
3.1.5	Stratégies nationales en matière d’IA	87
3.1.6	IA générative dans le secteur public	92
3.2	Ingénierie de prompt	97

4	Revue de la notion d'éthique appliquée à l'IA	99
4.1	IA et démarches pour une éthique réfléchie	99
4.1.1	Questionnements sur l'IA éthique	99
4.1.2	Revue des documents publiés par des chercheurs indépendants	100
4.1.3	IA et initiatives nationales, régionales et internationales	107
4.2	Cadre national d'éthique en matière d'IA	111
4.2.1	Résumé de l'évolution du Numérique en Mauritanie	113
4.2.2	Cadre national mauritanien de l'éthique de l'IA (résumé par Chat-GPT 4.0)	115
4.2.3	Cadre national de l'éthique de l'IA (version originale)	118
II	Contributions scientifiques	122
5	Système de prédiction de courte séries temporelle	123
5.1	Problématique de disponibilité des séries de données temporelles	127
5.2	Matériels et méthodes	128
5.2.1	Données utilisées	128
5.2.2	Techniques utilisées dans la méthode proposée	131
5.2.3	Description de la méthode proposée	138
5.2.4	Données utilisées	139
5.3	Expériences et résultats	142
5.3.1	Données d'expérimentation	142
5.3.2	Paramètres de l'algorithme de programmation génétique	143
5.3.3	Paramètres du Filtre de Kalman	143
5.3.4	Modèles de référence	143
5.3.5	Métriques de performance	144
5.3.6	Résultat et Analyse	144
5.3.7	Conclusion	145
6	Éthique relativiste gouvernementale	149
6.1	Introduction	149
6.2	Problématique	152
6.3	Approche proposée	155
6.3.1	Principes éthiques choisis dans le cadre de cette contribution	155
6.3.2	Détails techniques de l'implémentation	155
6.3.3	Vue d'ensemble des outils Ethiconorm et Ethicotest	159
6.4	Résultats expérimentaux	168

6.5	Discussion	187
6.6	Conclusion	189
7	Conclusion et Perspectives	201
7.1	Conclusion	201
7.2	Perspectives	202
7.3	Communications et participations effectuées	204
	Bibliographie	205
8	Annexes	226

Liste des figures

3.1	termes utilisés dans la recherche en texte intégral dans les archives du SSIS/IJIM. [Duan 2019].	74
3.2	L’IA et ses sous domaines.	76
3.3	Problèmes de qualité des données relatives aux séries temporelles.[Gitzel 2016]	85
3.4	Les auteurs visualisent les stratégies nationales en matière d’IA de 34 pays, regroupés par ordre décroissant, de gauche à droite, de ceux qui sont les plus performants dans six attributs spécifiques, collectivement, à ceux qui sont les plus faibles. [Denford 2023].	92
4.1	Proportion de documents dans lesquels le principe éthique apparaît[Jobin 2019].	101
4.2	Niveau d’automatisation (robotisation) de certains pays africains : comparaison avec le reste du monde. https://static.latribune.fr/article_body/689839/automatisation.png	106
5.1	Les différentes composantes d’une série temporelle [Jose 2022].	130
5.2	Fonctionnement de la Programmation Génétique. L’illustration est empruntée à [Kommenda 2009].	132
5.3	Représentation d’un programme (ou Individu).	133
5.4	Kalman filter equations and instruction. The illustration is adapted from [Welch 1995]	136
5.5	La structure simplifiée de l’outil proposé.	139
5.6	Evolution des indicateurs(séries temporelles).	147
5.7	Graphique en violon de l’erreur moyenne en pourcentage absolu (MAPE) pour tous les cas de test.	148
6.1	Schémas des applications Ethiconorm et Ethicotest	160
6.2	Menu de l’application destinée aux instances gouvernementales.	161
6.3	Capture d’écran de la section “Chatbot”.	162
6.4	Étape 1 du processus de formulation des principes éthiques (sélection du domaine).	163
6.5	Étape 2 du processus de formulation des principes éthiques	164
6.6	Affichage de la description du principe “Non-malfaisance”	165
6.7	Étape 3 du processus de formulation des principes éthiques	165

6.8	Le résumé (les pondérations précédemment saisies et un graphique en radar)	166
6.9	Le résumé (normes éthiques proposées par le LLM)	190
6.10	La synthèse (corrélations entre les différents principes éthiques, proposés par le LLM)	191
6.11	Test de conformité aux principes éthiques	192
6.12	Menu de l'application de diagnostic des développeurs	192
6.13	Étape 1 du formulaire (nom du logiciel, domaine d'application et description)	193
6.14	Étape 2 du formulaire (une série de principes éthiques avec un bouton permettant de générer cinq questions connexes pour chacun d'entre eux). 194	
6.15	Étape 2 du formulaire (Questions proposées pour le principe de transparence)	195
6.16	Synthèse des réponses des développeurs	196
6.17	Section "Conformité aux principes éthiques"	197
6.18	Détails sur la réponse du développeur au formulaire de conformité aux principes éthiques	197
6.19	prompt qui propose des questions au développeur	198
6.20	prompt qui propose une liste de normes	199
6.21	prompt pour les corrélations existantes entre les principes éthiques	199
6.22	prompt pour identifier les principes éthiques les plus essentiels	200

Liste des tables

2.1	Comparaison entre éthique (en France) et déontologie	23
2.2	Dix objections des contemporains de Mill et ses réponses à ces objections.	37
4.1	Principes éthiques identifiés dans les directives existantes en matière d'IA[Jobin 2019].	102
5.1	Séries temporelle utilisées dans la contribution	141
5.2	Méthode de calcul des indicateurs utilisés dans la contribution	141
5.3	Rapports statistiques (RMSE) de la GP sur les ensembles d'entraînement.	144
5.4	Erreurs de prédiction (RMSE).	146
5.5	Erreurs de prédiction (MAPE)	146
1	Indicateurs généraux (4 indicateurs)	226
2	Indicateurs relatifs aux baccalauréats scientifique et technique (9 indicateurs)	226
3	Indicateurs relatifs aux baccalauréats en lettre (6 indicateurs)	227
4	Indicateurs relatifs au genre (3 indicateurs)	227
5	Liste des séries temporelles	228

Introduction

Sommaire

1.1 Contexte	1
1.1.1 État des services gouvernementaux	2
1.1.2 État des lieux de l'Intelligence Artificielle (IA)	3
1.2 Problématiques et objectifs	4
1.2.1 Services gouvernementaux et prise de décision	4
1.2.2 Principes éthiques de l'Intelligence Artificielle (IA)	6
1.3 Pertinence et importance	13
1.4 Contributions	14
1.5 Structure de la thèse	15

1.1 Contexte

L'objet de cette thèse, en collaboration entre l'Université de Nouakchott (UN) et l'Université de Strasbourg (UNISTRA), est de répondre à un problème majeur posé par le gouvernement mauritanien : au moment où l'Intelligence Artificielle est de plus en plus efficace, de plus en plus interactive et facile d'accès grâce aux grands modèles de langage (*Large Language Models* ou LLMs), comment en faire bénéficier le pays d'un point de vue gouvernemental ?

Pour répondre à cette grande question, nous avons identifié deux points qui nous ont semblé importants et qui seront adressés dans cette thèse : comment utiliser l'IA pour aider les administrations publiques dans leurs prises de décision et comment s'assurer que les applications utilisant l'IA respectent des principes éthiques acceptables pour la société, si l'on considère que les méta-éthiques sont relatives aux contextes sociétaux.

Étant donné que les pays en développement comme la Mauritanie n'ont pas beaucoup de données ou d'informations historiques, ni de cadres éthiques spécifiques bien décrits, comment peut-on aider les administrations (publiques ou privées) dans leur procédure de prise de décision ?

En Mauritanie, à l’instar des pays en développement, l’état des services gouvernementaux (notamment, en matière d’infrastructures administratives comme dans le cas de l’éducation, la santé, mais aussi l’élevage, l’agriculture, etc) pour la prédiction et la prise de décision, requiert une amélioration en vue d’en assurer l’efficacité et la pertinence. Le présent travail tente de répondre à ce besoin en tenant compte de critères éthiques, non seulement appropriés à la culture et au contexte mauritanien, mais en pensant un système permettant de s’adapter à d’autres cultures dans le monde, élargissant ainsi le périmètre de cette thèse.

1.1.1 État des services gouvernementaux

De nos jours, les impératifs de développement, la pression démographique de plus en plus forte et les besoins de prestations en matière de services gouvernementaux, amènent les états à recourir aux Technologies de l’Information et de la Communications (TIC), dans le souci de répondre à ces besoins et d’assurer la qualité des services offerts, d’où la naissance des concepts d’e-gouvernement et d’e-gouvernance où le préfixe e- signifiant “électronique” se réfère aux technologies de l’information et à tout ce qui est transmis de manière numérique sur l’Internet. l’e-gouvernement étant défini comme “l’utilisation des TICs dans les administrations publiques, associée à des changements au niveau de l’organisation et de nouvelles aptitudes du personnel” (définition de la communauté européenne présentée par [Nawel 2019]). L’objectif est d’améliorer les services publics, renforcer les processus démocratiques et de soutien aux politiques publiques.

L’e-gouvernement se réfère au gouvernement en ligne (les structures organisationnelles de l’administration publique, l’information, les systèmes d’information, les outils technologiques, etc.). Quant à l’e-gouvernance, elle se réfère aux interactions possibles entre tous les acteurs, quel que soit leur niveau d’implication. Les deux concepts de l’e-gouvernement et de l’e-gouvernance sont apparus au cours du 20ème siècle et sont jugés comme des dispositions utiles par les organisations internationales, pour garantir la bonne gouvernance et le développement socio-économique. Les besoins en matière d’e-gouvernement peuvent couvrir plusieurs axes [Almarabeh 2016] : gouvernement-citoyens (G2C), gouvernement-gouvernement (G2G), gouvernement-business (G2B) et gouvernement-employés (G2E). Ainsi donc, les opportunités de la transition numérique ont tracé aux états avancés, émergents et mêmes ceux en développement, le chemin de la modernisation technologique. L’e-Gouvernement a connu une évolution qui peut être déclinée en trois générations, d’après [Charalabidis 2019]. L’examen des différentes expériences présentées dans la littérature scientifique, en matière de pratique e-Gouvernement, montre des niveaux d’avancement différents selon les pays. Certains sont même suffisamment avancés dans la course mondiale en matière d’IA. C’est le cas notamment de la Russie, des États-Unis, de la Chine, de l’Estonie, du Japon, d’Israël, du Canada, du Royaume-Uni, de Singapour, de la Corée du Sud,

etc[[Institut-Sapiens 2018](#)].

En Europe, l'État Estonien est considéré comme une réussite en matière de Gouvernement électronique (e-gouvernement). Il s'est constitué en un État plateforme, selon un processus qui a couvert les années 1990 et le début des années 2000 [[Institut-Sapiens 2018](#)]. Ce modèle de e-gouvernement destiné à l'amélioration de l'action publique fournit près de 1200 e-services. Il repose sur un système standardisé d'échange de données, alimenté par une multitude de systèmes d'information de près de 900 institutions et entreprises publiques et privées. Ce système, dit X-Road, est la colonne vertébrale de l'administration estonienne. Les enjeux de cette dématérialisation sont un gain de temps dans les démarches administratives, des économies de dépenses publiques, une plus grande implication des usagers et une amélioration des services numériques.

Dans le cas des pays en développement (PED) et des pays les moins avancés (PMA), on peut noter que dans leur grande majorité, l'intérêt pour les TIC a commencé vers le milieu des années quatre-vingt-dix. Sous l'impulsion des grands organismes internationaux (Banque Mondiale, Programme des Nations Unis pour le Développement), ces pays ont élaboré des stratégies nationales pour les TIC mais restent confrontés à des problèmes pour les mettre en œuvre : un accès physique insuffisant à la technologie, un faible niveau de préparation des compétences humaines et, enfin, le manque de motivation et de volonté réelle de conduire le changement. Les dispositifs électroniques se limitent à de simples portails Web sans valeur ajoutée et sans interactions avec les utilisateurs (services gouvernementaux, citoyens, entreprises).

Le projet e-Fez, au Maroc, débuté en 2006, mérite d'être cité, compte tenu de l'approche pragmatique et volontariste de ses auteurs [[Kettani 2014](#)]. Ce projet pilote envisageait de développer "un système en ligne de prestations de services aux citoyens pour le gouvernement municipal de la ville de Fès".

En Mauritanie, l'État a élaboré en 2002 une Stratégie Nationale et un Plan Directeur de développement des TIC pour la période 2002-2006 pour la mise en place de l'intranet gouvernemental, l'installation de la fibre optique et le développement des systèmes d'information et de collaboration. Cette stratégie a été révisée, par la suite, en 2011, pour la période 2012 – 2016 [[MEFPTN 2011](#)], pour mieux affirmer les TIC comme moyen de modernisation de l'Administration et comme vecteur du développement économique.

1.1.2 État des lieux de l'Intelligence Artificielle (IA)

Née depuis une soixantaine d'années, l'IA a connu un développement important depuis 2010, avec l'accès aux données massives et l'augmentation de la puissance

de calcul pour leur traitement, qui ont permis de grandes avancées technologiques [Zouinar 2020]. Tout comme dans le reste du monde, les exemples d'utilisation de l'IA sont nombreux sur le continent africain : [Brandusescu 2017] et [Smith 2018], au Kenya, au Nigeria et en Afrique du Sud “pour répondre aux besoins en matière de santé, d'agriculture, de technologie financière, de transports publics et de traduction linguistique” ; [Smith 2018] en Ouganda et en Éthiopie, “dans les diagnostics aux points de service, la prestation de services gouvernementaux, la conservation de la faune, la surveillance des cultures, la gestion de l'eau, le développement d'entreprises et celui des services financiers.”

Le potentiel qu'offre à l'humanité la quatrième révolution industrielle (l'internet des objets et les technologies du numérique) et l'IA constituent une aubaine pour l'Afrique dans de nombreux domaines sensibles (prestations administratives, information, santé, élevage, agriculture, éducation, etc.) avec, comme avantages, l'amélioration des services, la prédiction, la prise de décision, etc.

Cependant, les menaces et les risques de l'IA sont nombreux : discriminations, renforcement et exploitation des préjugés, désinformation, etc., pour ne citer que les menaces et les risques de plus faible intensité. D'où, la nécessité d'une IA répondant à des lignes directrices et des principes éthiques.

1.2 Problématiques et objectifs

L'amélioration du service public suppose des possibilités de prédiction et de prise de décision factuelle. Elle repose donc, nécessairement, sur la disponibilité (qualitative et quantitative) et le traitement de l'information collectée. Or, étant donné que la Mauritanie est un pays en développement, le premier problème auquel s'est trouvé confronté ce travail est celui des données disponibles. Le deuxième problème relève de l'éthique de l'Intelligence Artificielle (IA), compte tenu des menaces qu'elle pose. Ces deux problèmes sont abordés ci-dessous.

Dans le contexte défini, il s'agit du défi suivant : améliorer les services publics à l'aide de l'IA en tentant de respecter les principes éthiques de la population concernée.

Dans le cadre de cette thèse, nous travaillerons dans un domaine où nous avons pu obtenir des données gouvernementales, à savoir le domaine de l'éducation.

1.2.1 Services gouvernementaux et prise de décision

Il est techniquement admis, en général, que le recours à l'IA dans la démarche de prédiction et la prise de décision, requiert des données massives (Big Data). La modélisation des séries temporelles (ensemble de données qui représente l'évolution d'un phénomène

au cours du temps), à partir de telles données, a acquis une popularité spectaculaire, avec des applications comprenant la gestion et l'ingénierie du monde réel.

Si dans les pays émergents et les pays développés, on note un besoin grandissant et une disponibilité croissante d'approches big data, pour répondre aux besoins de changements, la tendance est différente dans les pays en développement. Ceux-ci font face à une disponibilité limitée des données et doivent recourir à des séries temporelles courtes ou, plus simplement, à quelques observations. Cependant, les méthodes de prédiction existantes ne fonctionnent pas toutes avec des séries temporelles courtes. Certaines parmi ces méthodes requièrent l'identification et l'optimisation de plusieurs paramètres, ce qui exige nécessairement la mobilisation d'une masse importante de données. Les auteurs s'accordent à reconnaître que l'étude des séries temporelles courtes, générées à partir d'une quantité limitée d'informations statistiques historiques, pose un certain nombre de difficultés [Hyndman 2007].

Cependant de manière plus générale, et indépendamment de l'état de développement des sociétés, certains domaines doivent nécessairement utiliser des données inadéquates et incomplètes provenant d'une courte période, pour prédire des observations inconnues [Cui 2021].

À propos des difficultés que pose le caractère limité des informations statistiques [Hyndman 2007], et à titre d'exemple (en l'absence d'informations supplémentaires), il est souvent impossible de conclure que le processus est déterminé ou indéterminé, ce qui a un effet significatif sur la construction du modèle. Dans le cas déterminé, il est évident que pour tout ensemble de points, il existe de nombreuses courbes qui passent par ces mêmes points, et il est difficile de soutenir qu'une seule courbe (modèle) est précisément la loi (matérialisée par une équation) qui décrit de manière exhaustive le phénomène et qui permettra de prédire efficacement son comportement dans le futur [Turbal 2019]. Autrement dit, comment améliorer la prédiction et donner plus de chance et d'efficacité à la prise de décision lorsque très peu de points de données sont disponibles ? Pour résoudre ce problème, quelques approches ont été proposées dans la littérature [Cruz-Nájera 2022, Cui 2021, Kirshners 2012].

Parmi les objectifs de priorité nationale de l'état mauritanien, dans le cadre de ses politiques d'ajustement structurel engagées depuis les années 1980 et compte tenu des constats posés, figurent la qualité de l'enseignement, le développement des ressources humaines (en qualifications et en compétences), la promotion de l'enseignement technique [UNIES 2009].

Tenant compte de ces objectifs, l'état a un fort besoin de données statistiques fiables à analyser pour réaliser des prédictions et faire face aux besoins de développement du pays, en termes de financements (infrastructures scolaires et universitaires), de création de filières et de gestion des flux. Le défaut de disponibilité des données, de leur cohérence et de leur diffusion, nous a conduits à fixer, comme objectif, de réaliser nos tests sur les fichiers disponibles du baccalauréat (données disponibles de 2013 à 2022).

En utilisant ces données, nous cherchons à explorer les possibilités d'utiliser les outils technologiques et l'Intelligence Artificielle avec, comme question fondamentale : quels dispositifs, modèles expérimentaux, peuvent être proposés dans notre contexte, pour contribuer à l'évaluation, la prédiction, et la prise de décision au sein de gouvernements de pays en développement et plus particulièrement du gouvernement mauritanien tout en considérant le faible historique de données ?

1.2.2 Principes éthiques de l'Intelligence Artificielle (IA)

Si les applications de l'IA, évoquées plus haut, ont suscité de grands espoirs et montré leurs avantages, le développement de cette technologie a aussi été source d'inquiétudes, de menaces et de risques. D'où, plusieurs questions d'ordre éthique :

1. Quel est le rôle de l'éthique dans la conception, le développement et l'application de l'IA ?
2. Comment l'éthique aidera-t-elle à maximiser les avantages de l'IA pour accroître le bien-être des citoyens et le bien commun ?
3. Comment garantir que les systèmes d'IA servent le bien public au lieu d'exacerber les inégalités existantes ?
4. Comment s'assurer que l'usage des algorithmes ne soit source d'injustices et d'inégalités ?

Quoi de mieux qu'un exemple pour montrer l'importance de l'éthique pour l'intelligence artificielle ? Dans sa "Critique de la Raison Pratique", Kant (qui a inspiré la révolution française) prône l'action désintéressée. À l'inverse, Jeremy Bentham dans "Introduction aux principes de morale et de législation" et John Stuart Mill dans "L'Utilitarisme", basent le comportement de l'être humain sur la maximisation de l'intérêt, personnel et global.

Si un gouvernement s'aide d'une IA pour développer l'infrastructure éducative d'un pays, cette IA doit-elle faire des propositions en se basant sur le désintérêt (avec une éducation gratuite pour tous comme en France, qui suit historiquement l'éthique continentale de Kant) ou sur l'intérêt avec la commercialisation des universités (comme cela se passe aux USA, qui suivent historiquement l'éthique utilitariste de Bentham et Mill) ? [Masseys-Bertoneche 2011]

Et que penser d'une IA qui permettrait de mettre en œuvre les principes du "Panoptique", du même Jeremy Bentham?

Face à toutes ces questions sur les avantages, mais aussi, sur les risques et les incertitudes liées à l'IA, un consensus sociétal s'impose, consensus auquel s'attellent les organisations internationales (UNESCO, Conseil de l'Europe, Commission européenne,

Agence des droits fondamentaux de l'Union Européenne, Banque inter-américaine de développement, OCDE et OECD-AI, Organisation des Nations Unis, et Groupe de la Banque Mondiale).

De nombreux auteurs se sont penchés sur ces questions d'éthique. L'examen de cette littérature révèle que l'IA impacte fortement la jouissance des droits de l'homme et des libertés fondamentales, notamment : la désinformation, les bulles de filtre (ou chambres d'écho), le profilage, la discrimination, et la radicalisation, impacts résumés dans [Bensalah 2021].

L'IA, à travers l'apprentissage automatique, peut faire courir le risque d'entretenir certaines exclusions traditionnellement implantées dans les sociétés patriarcales, notamment en Afrique [Gwagwa 2019] : exclusions des jeunes (des jeunes filles en particulier) et des femmes. D'autres menaces sont également pointées par l'auteur : menaces sur la vie privée, les droits individuels et collectifs, compte tenu de l'utilisation croissante de l'identité numérique.

Compte tenu des risques et de menaces associées aux usages de l'IA, une régulation s'impose, qui passe par une démarche de. Bonne gouvernance de cette technologie. D'après [Bell 2023], les différentes approches de gouvernance de l'IA peuvent être classées en deux catégories principales : les mesures réglementaires prises par les gouvernements à travers la législation ou les directives des régulateurs d'une part, et l'autorégulation et les normes volontaires d'autre part. Ces politiques et réglementations, ont pour objectif de demander à toute personne qui développe et déploie l'IA de repérer, réduire, surveiller et gérer les risques de préjudice ou d'utilisation abusive[Bell 2023]. Nous nous proposons de présenter, dans les propos qui suivent, les démarches de régulation, aux échelles continentales et nationales.

1.2.2.1 Régulation de l'IA : Textes législatifs, vision globale et spécificités

La législation européenne sur l'IA

Tout récemment (depuis 2022), l'apparition et le développement de deux systèmes d'IA, ChatGPT développée par OpenA (2022) et Gemini par Google (plus récemment encore, en 2023), capables (tous les deux) de traiter et de générer du texte et réaliser diverses tâches liées aux langages [Rane 2024], a conduit l'Union Européenne (UE) à proposer et adopter, en février 2024, un règlement sur l'intelligence artificielle, pour encadrer le développement de ces systèmes et de l'IA, en général ¹.

Présentons maintenant la législation européenne sur l'IA. Ce cadre juridique sur l'intelligence artificielle (Baptisé "AI Act") constitue la première loi sur l'intelligence artificielle dans le monde ¹. En cohérence avec "la Charte des droits fondamentaux de l'UE, le droit dérivé de l'UE en matière de protection des données, de protection des

¹<https://www.vie-publique.fr/questions-reponses/292157-intelligence-artificielle-le-cadre-juridique-europeen-de-lia>

consommateurs, de non-discrimination et d'égalité entre les femmes et les hommes", le règlement vise de nombreux objectifs ¹, notamment :

- veiller à la sûreté des systèmes d'IA, au respect de la législation en vigueur en matière de droits fondamentaux, des valeurs de l'UE, de l'État de droit et la durabilité environnementale ;
- garantir la sécurité juridique et l'innovation dans le domaine de l'IA, en renforçant la gouvernance et l'application effective de la législation existante en matière d'exigences de sécurité applicables aux systèmes d'IA et de droits fondamentaux ;
- interdire certaines pratiques et faire respecter des exigences spécifiques applicables aux systèmes d'IA à haut risque et des règles harmonisées en matière de transparence applicables i/- aux systèmes d'IA destinés à interagir avec des personnes, ii/- aux systèmes de reconnaissance des émotions et de catégorisation biométrique, et iii/- aux systèmes d'IA générative utilisés pour générer ou manipuler des images ou des contenus audio ou vidéo.

Les pratiques d'IA interdites par la législation de l'UE

Les pratiques interdites par la législation de l'UE concernent notamment les systèmes d'IA qui :

- ont recours à des techniques subliminales qui peuvent causer un préjudice physique ou psychologique altérant le libre arbitre ;
- exploite les vulnérabilités dues à l'âge ou au handicap d'un individu pour altérer son comportement de manière à causer un préjudice physique ou psychologique ;
- réalisent une classification des personnes basée sur des caractéristiques sensibles personnelles (opinions politiques, religieuses, philosophiques, orientation sexuelle, etc.) conduisant une catégorisation biométrique et une notation sociale ;
- utilisent l'identification biométrique à distance, "en temps réel", dans des espaces publics et ce, à des fins répressives.

L'incitation au respect de ces interdictions (sous peine d'amendes) passe nécessairement par la possibilité offerte par la législation de l'UE, aux citoyens, de recevoir des explications sur les décisions fondées sur des systèmes d'IA à haut risque ayant une incidence sur leurs droits et de déposer des plaintes concernant les situations créées par les systèmes d'IA en cause.

Ce modèle de l'UE se démarque en distinguant les différentes situations d'utilisation de l'IA : il interdit les utilisations inacceptables et identifie d'autres utilisations comme

relevant de l'IA "à haut risque", pour lesquelles la conformité aux directives réglementaires, de même qu'une surveillance stricte et constante s'imposent. Ceci explique que la loi soit axée sur une démarche de régulation de l'utilisation des systèmes d'intelligence artificielle en fonction des niveaux de risque [Csernatonì 2024]. Le modèle impose des responsabilités spécifiques aux différents acteurs de la chaîne d'approvisionnement de l'IA : les fournisseurs, les prestataires, les importateurs et les utilisateurs. Le cadre juridique établi par la loi repose sur quatre niveaux de risque et introduit des règles spécifiques pour les modèles d'IA à usage général [Csernatonì 2024]. Il nous semble important de préciser, par ailleurs, qu'en vertu de cette loi sur l'IA ("Ai Act"), les systèmes d'IA à haut risque doivent obligatoirement faire l'objet d'évaluations d'impact sur les droits fondamentaux dans des secteurs sensibles tels que l'assurance, les soins de santé et la finance. En plus de ces secteurs sensibles, la loi établit des droits de plainte pour les citoyens touchés par les systèmes d'IA à haut risque, en particulier dans les cas qui pourraient influencer le comportement des électeurs et les résultats des élections. Quant aux systèmes et modèles d'IA à usage général, ils doivent au moins respecter des exigences de transparence, notamment en ce qui concerne la documentation technique, le respect de la législation européenne sur les droits d'auteur et la diffusion de résumés du contenu de l'entraînement [Csernatonì 2024].

Jusqu'à présent, seule une directive européenne, distincte de la "EU AI Act", traite de manière indirecte la question de la responsabilité des modèles à usage général. Le cadre proposé par l'UE pour régir l'IA pourrait ne pas englober l'ensemble des contextes dans lesquels elle peut être utilisée, et la question de savoir si les modèles à usage général pourraient être classés comme "à haut risque" est débattue [Bell 2023].

Le règlement a pour ambition de "garantir que les droits fondamentaux, la démocratie, l'État de droit et la durabilité environnementale sont protégés contre les risques liés à l'IA, tout en encourageant l'innovation et en faisant de l'Europe un des leaders dans ce domaine."².

Malgré l'accord politique conclu en décembre 2023 sur l'IA Act entre les institutions de l'UE, il existe des tendances et des spécificités exprimées par certains pays :

- la France semble rester ferme dans sa position consistant à vouloir imposer des conditions strictes pour s'assurer que la loi n'entrave pas le développement de modèles d'IA compétitifs. Plus précisément, elle affirme qu'il est important de trouver un juste équilibre entre la protection des secrets commerciaux et la transparence, d'éviter de surcharger les entreprises avec des obligations à haut risque et de réévaluer les seuils et les critères utilisés pour qualifier les modèles d'IA comme présentant des risques systémiques[Csernatonì 2024].
- le Canada suit la même trajectoire. Dans le secteur public, le Canada a déjà

²<https://www.europarl.europa.eu/news/fr/press-room/20231206IPR15699/loi-sur-l-intelligence-artificielle-accord-sur-des-regles-globales>

adopté une loi qui impose des évaluations d'impact à l'avenir pour l'utilisation de systèmes automatisés. En mars 2023, il a annoncé des mesures pour élargir la réglementation sur les risques et "s'harmoniser" avec la loi de l'UE. Le pays souhaite mettre l'accent principalement sur l'éducation et le perfectionnement. Cependant, la législation proposée (à savoir, la loi sur l'intelligence artificielle et les données) est structurée de manière similaire à la loi de l'UE et donne au régulateur le pouvoir d'exiger des audits et même de suspendre l'utilisation d'un système d'intelligence artificielle [Bell 2023].

- Les États-Unis se basent sur l'autorégulation axé sur le volontarisme basé sur la mise en œuvre des processus volontaires multipartites, mais pilotés par le secteur public, pour élaborer la gestion des risques et les normes techniques, des processus similaires dans des domaines particuliers (par exemple, les dispositifs médicaux), et la participation à des organismes de normalisation internationaux. Ils soulignent également les dangers de l'IA et la nécessité de réglementations telles que le "Blueprint for an AI Bill of Rights", projet de charte des droits de l'IA ³. Ajoutons que des outils spécifiques ont également été développés aux États-Unis (comme à Singapour), afin d'assister les développeurs et les utilisateurs d'IA dans la détection et la réduction des risques [Bell 2023].
- la Chine se concentre sur des réglementations ciblées avant l'adoption d'une législation nationale sur l'IA⁴. Sa méthode consiste en des collaborations entre le gouvernement et les entreprises privées sur les directives de réglementation de l'IA, ainsi qu'un soutien considérable du gouvernement pour le développement de la technologie et des entreprises locales [Bell 2023].
- en Australie, il existe un développement des technologies basé sur l'autorégulation et une approche des normes volontaires. Cependant, les démarches ont toujours été neutres sur le plan technologique. De plus, les développeurs n'ont pas l'obligation légale d'effectuer une évaluation des risques, sauf pour les entreprises qui fournissent le gouvernement de la Nouvelle-Galles du Sud, pour lesquelles l'obligation est contractuelle plutôt que légale [Bell 2023].

Ce qui précède suscite et inspire une petite comparaison concernant les priorités en matière d'application de l'IA. Cette comparaison pourrait être vue sous l'angle des similitudes culturelles, politiques et religieuses d'une part, et sous l'angle des définitions apportées à l'IA. Ces deux niveaux de comparaisons font naturellement ressortir deux blocs : le bloc occidental (avec les différences affichées par les pays qui le composent) et la Chine.

³<https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>

⁴https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm

La question des similitudes culturelles

Dans [Denford 2023], les auteurs souhaitent évaluer si des pays ayant une culture similaire adoptaient des méthodes similaires pour aborder l'intelligence artificielle. Ils ont constaté que les pays culturellement similaires présentent peu de similitudes et ont émis l'hypothèse que cela pourrait être attribué à l'immaturité relative de l'IA. Ils concluent à des disparités étonnantes entre la manière dont les pays occidentaux et la Chine gèrent l'IA.

- en Occident, les nations – les États-Unis en tête – mettent l'accent sur les risques de l'IA et s'efforcent de mettre en œuvre des “garde-fous” adéquats afin d'assurer une gestion adéquate de la technologie dès le départ. L'objectif de l'Union européenne est de consolider sa souveraineté technologique et de devenir un centre mondial d'IA en mettant l'accent sur une gouvernance responsable de l'IA et en élaborant un code de conduite en matière d'IA.
- quant à la Chine, elle se focalise presque uniquement sur la recherche et le développement (R&D), en cherchant à être le leader technique et politique dans ce domaine, et elle consacre très peu d'efforts à réduire les éventuels impacts néfastes du développement de l'intelligence artificielle. Cela génère un environnement où la Chine, déjà en tête de l'intelligence artificielle, peut encore accroître son avance, tandis que les États-Unis et d'autres pays mettent en place des “garde-fous”.

La question des définitions de l'IA

Les définitions apportées à l'IA peuvent être fonction des visions et des objectifs qui leur sont assignés. Elles peuvent être larges (cas de l'UE) ou étroites (cas de la Chine). Cette orientation n'est pas sans conséquences. Dans le cas de l'UE, le caractère assez large de l'éventail des systèmes d'IA, qu'ils soient, ou non, à haut risque, conduit à de réels problèmes quant à la définition précise des notions de risque, de sécurité, de transparence, d'équité et de non-discrimination, entraînant des difficultés quant à la mise en œuvre de la réglementation. Dans le cas de la Chine, le caractère étroit des définitions entraîne le risque que certains usagers puissent contourner les règles.

La législation africaine sur l'IA

Nos recherches ne nous ont pas permis de préciser l'existence d'une législation officielle destinée à réguler l'IA et à en assurer la gouvernance. A notre connaissance, une telle législation, unifiée à l'échelle du continent n'existe pas. Cependant, des pays africains isolés ont élaboré des stratégies et des textes législatifs et réglementaires portant sur l'IA (portant, notamment, sur l'éthique et la protection des données).

Les ministres africains des TIC et des communications ont approuvé une stratégie continentale en matière d'IA et un pacte numérique africain afin d'accélérer la transformation numérique de l'Afrique. Lors d'une réunion virtuelle tenue du 11 au 13 juin 2024, plus de 130 ministres et experts africains ont discuté de l'exploitation de l'IA pour

atteindre les objectifs de développement tout en garantissant une utilisation éthique et inclusive. La stratégie promeut les capacités de l'IA en matière d'infrastructures, de talents, d'ensembles de données, d'innovation et de partenariats, en mettant l'accent sur l'éducation, la santé, l'agriculture et la gouvernance ⁵.

Un livre blanc intitulé : Réglementation et adoption responsable de l'IA pour l'Afrique vers la réalisation de l'Agenda 2063 de l'UA "Regulation and Responsible Adoption of AI for Africa Towards Achievement of AU Agenda 2063", rédigé suite au plan directeur 2021 sur l'IA (2021 AI blueprint) et la résolution 473 de la Commission africaine des droits de l'homme et des peuples, appelle à une approche de la gouvernance de l'IA axée sur les droits de l'homme [Fokane 2024]. Ce livre examine en détail les mesures prises pour encourager l'adoption de l'IA tout en réduisant les dangers qui y sont liés en Afrique. En se basant sur l'Agenda 2063 de l'UA, il analyse attentivement le cadre actuel de l'IA à l'échelle nationale et internationale. En soulignant le potentiel de l'intelligence artificielle pour combattre les défis variés auxquels nos communautés font face en matière de pauvreté, d'éducation, de soins de santé [Etori 2024]. Il appelle également à une approche harmonisée de l'adoption de l'IA et invite les États membres à adopter des stratégies nationales en matière d'IA qui mettent l'accent sur la sûreté, la sécurité et la protection des données afin de promouvoir l'utilisation éthique de l'IA [Fokane 2024].

Il est structuré autour de cinq piliers essentiels. Chaque pilier constitue une étape vers un écosystème de l'IA plus dynamique et plus inclusif, allant du renforcement du pool de talents en IA en Afrique à la consolidation de l'infrastructure numérique et à la création d'environnements favorables au développement de l'IA [N'Guessan 2024].

Le premier pilier met l'accent sur le renforcement du capital humain en comblant les lacunes en matière de compétences et en favorisant des initiatives et des partenariats éducatifs innovants. Le deuxième pilier se concentre sur la création d'une infrastructure essentielle et sur l'exploitation efficace des données pour stimuler l'innovation en matière d'IA. Le troisième pilier a pour objectif de créer des conditions favorables au développement et à la mise en œuvre de l'IA, ce qui nécessite de stimuler les écosystèmes de soutien, de garantir un accès généralisé au haut débit et de développer des cadres politiques appropriés. Les piliers 4 et 5 ont pour objectif de stimuler l'économie de l'intelligence artificielle et de favoriser des partenariats durables, en défendant des politiques qui soutiennent les capacités locales et les écosystèmes socio-économiques [N'Guessan 2024].

Au niveau national, le Ghana et l'Ouganda, par exemple, ont participé au projet "Ethical Policy Frameworks for AI in the Global South", qui se concentre sur l'élaboration de cadres politiques locaux en matière d'intelligence artificielle [Walter 2024].

⁵<https://au.int/en/pressreleases/20240617/african-ministers-adopt-landmark-continental-artificial-intelligence-strategy>

La section 4.1.3.1 évoque en détail les stratégies d'IA publiées en 2023 (selon l'Indice de préparation à l'IA).

Notre modeste propos est de passer en revue ces usages problématiques de l'IA, particulièrement en Afrique et de voir les dispositions et les solutions qui pourraient être envisagées pour contribuer à assoir une IA éthique dans ce continent.

Dans ces conditions, et compte tenu de l'évolution actuelle des usages de ces technologies, il importe de se poser quelques questions préliminaires, d'ordre général : où en sont les gouvernements des pays en développement et plus particulièrement pour cette thèse, des pays africains ? Comment l'Afrique pourrait-elle, par ailleurs, tout en bénéficiant du potentiel offert par ces technologies, en limiter les menaces et les risques ? Enfin, les valeurs et les principes de l'Occident sont-ils tous applicables en Afrique sachant qu'ils sont loin d'être unifiés (*cf.* éthique continentale fondée sur le désintéret *vs* éthique conséquentialiste et utilitariste anglosaxonne fondée sur l'intérêt évoqués précédemment) ?

1.3 Pertinence et importance

En commençant notre introduction générale, nous avons donné les raisons de l'orientation de ce sujet de thèse. Celui-ci devait répondre aux thématiques de la recherche universitaire, en rapport avec les préoccupations nationales en matière de développement, d'où, à priori, son utilité et son importance. Il faut ajouter à cette nécessité universitaire, la pertinence du thème abordé. Celle-ci est d'ordre actuel, contextuel (contexte national et international) et scientifique (voir nos contributions scientifiques). L'actualité est marquée par de profonds bouleversements dus à la révolution numérique et au développement des applications pratiques de l'intelligence artificielle. L'IA est entrée dans notre vie, que cette intrusion soit vue sous l'angle individuel, social ou sociétal. Elle y restera pour toujours, sans aucun doute, avec ses avantages, ses menaces et ses risques. Cette intrusion définitive de l'IA peut être illustrée par des exemples pris dans de nombreux domaines : amélioration du service public, optimisation des prédictions, meilleure performance des prises de décision, assistants domestiques, assistants médicaux, etc. L'IA est un outil performant au service des stratégies de développement. Elle a permis le passage de la gouvernance publique, classique, au mode de gouvernance numérique. En adoptant, en 2012, le 2ème cycle du Plan National d'action Environnemental (PANE 2) qui représente l'outil de mise en œuvre de la stratégie Nationale de Développement Durable (SNDD), "la Mauritanie a exprimé sa volonté de placer le développement durable au centre de son action" [LE PREMIER MINISTRE 2012]. Le pays devra conduire un long processus de collecte et d'intégration de toutes les informations utiles (qualitatives et quantitatives) pour être en mesure d'évaluer, de faire des prédictions et de prendre les décisions pour l'atteinte progressive de ses objectifs stratégiques notamment en matière d'éducation, objet de notre travail.

1.4 Contributions

Dans cette partie, nous allons brièvement parler de chaque contribution scientifique et de la démarche méthodologique de chacune. Afin de répondre aux deux problèmes majeurs soulignés dans la section 1.2, les deux contributions scientifiques suivantes ont été réalisées :

1. Proposition d'une nouvelle méthode de prévision à partir de séries temporelles courtes, qui sont l'un des défis les plus importants des pays en développement comme la Mauritanie. Pour cette thèse, nous travaillerons sur le domaine de l'éducation et plus particulièrement sur la prédiction des statistiques d'admission au baccalauréat hors événement extérieur imprévisible. Conscients de l'importance de ce problème pour l'amélioration du système éducatif et sa capacité à répondre aux besoins économiques, nous avons proposé un outil innovant basé sur l'association entre programmation génétique et filtre de Kalman, pour permettre de réaliser des prédictions à court terme précises de séries temporelles courtes.

L'outil, validé sur divers ensembles de données, a démontré une robustesse et une fiabilité supérieures aux techniques existantes. Cela pourrait se traduire par une meilleure adaptation du système éducatif aux exigences économiques et sociales, améliorant ainsi la qualité de l'éducation.

2. Cadre éthique de l'IA : un outil d'aide au gouvernement utilisant l'IA générative. Notre seconde contribution est axée sur l'intégration d'une IA proposant une éthique adaptative pour respecter les normes socio-culturelles spécifiques à chaque pays ou région. Cette solution est basée sur une IA générative, spécifiquement les grands modèles de langage comme ChatGPT, pour aider les gouvernements à établir et à faire respecter des principes éthiques locaux dans le développement de logiciels. La solution se présente sous la forme de deux applications web : une pour le gouvernement et une pour les développeurs de logiciels. L'application gouvernementale permet une calibration dynamique des poids éthiques dans divers domaines, tandis que l'application destinée aux développeurs offre une évaluation en temps réel de la conformité éthique des logiciels. Ce travail illustre l'importance d'une synergie entre l'IA et les considérations éthiques et culturelles, posant un jalon essentiel pour un développement technologique responsable et aligné sur les valeurs sociétales.

1.5 Structure de la thèse

Le reste de ce document de thèse, qui complète cette partie introductive (Chapitre 1) est structuré en trois grandes parties : une revue de la littérature philosophique et scientifique sur le sujet, une revue de la notion d'éthique appliquée à l'IA et une présentation détaillée des contributions scientifiques réalisées.

1. La revue de la littérature philosophique et scientifique sur le sujet porte, d'une part sur la philosophie de l'éthique, d'autre part sur la littérature scientifique sur l'IA.

- A propos de notre essai sur l'état de l'art de l'éthique (Chapitre 2), nous abordons les apports des philosophes au progrès de l'éthique, les théories déontologiques, les éthiques modernes, la philosophie "occidentale" (y compris la philosophie analytique, l'éthique de l'intérêt, la philosophie continentale, l'éthique contractualiste ou éthique du désintérêt). Nous avons également étudié les méta-éthiques universalistes, les méta-éthiques relativistes, la philosophie et l'éthique africaine. L'intérêt que nous avons porté à la justice et au droit des animaux et des robots relève, d'une part de la passion exprimée dans la littérature philosophique par certains auteurs et, par ailleurs, du souci, de ne laisser de côté aucun champ d'intervention possible de l'IA, les animaux et les robots faisant partie intégrante de notre environnement.

L'éclairage fourni par cette littérature révèle une grande diversité des pratiques de l'éthique universaliste, ce qui nous a conduits à opter pour une éthique relativiste (l'éthique universaliste apparaissant comme une perspective lointaine).

- Dans la revue de la littérature scientifique sur l'IA (Chapitre 3), nous abordons les technologies de l'IA et leurs domaines d'applications, notamment les outils technologiques pour les applications fréquentes de l'IA et les notions d'apprentissage (machine learning et deep learning) ; la problématique des séries temporelles ; la prédiction pour l'aide à la décision, mettant l'accent sur la prédiction de courtes séries temporelles. Nous exposons, dans cette section, les méthodes les plus récentes pour aborder ce type de prédiction. Le recours à l'IA générative dans le secteur public et l'ingénierie de prompt a été, de notre part (nous le reconnaissons volontiers), une démarche innovante, une issue possible, voire la seule, pour élaborer une plateforme d'"éthique relativiste gouvernementale", objet de notre chapitre 6.

Nous avons cherché à bien comprendre le fonctionnement de ces technologies de l'IA, d'en mesurer les avantages et les limites, avant d'entamer une réflexion pour une utilisation optimale de ces technologies.

2. La revue de la notion d'éthique appliquée à l'IA (Chapitre 4), pose des questionnements sur l'IA face aux menaces et au risques possibles. Dans le cadre de ces travaux de thèse, il est logique de se demander quelles sont les démarches pour une éthique réfléchie, responsable, pertinente et adapté. Outre notre recherche documentaire sur notre cadre national d'éthique en matière d'IA et les initiatives nationales, régionales et internationales en la matière, nous nous sommes intéressés aux publications académiques de chercheurs indépendants et aux rapports d'experts d'organismes internationaux. Notre préoccupation centrale était l'application de l'éthique à l'IA.
3. La présentation détaillée des contributions réalisées : concerne nos deux contributions scientifiques : la première (Chapitre 5) intitulée "Short Time Series Forecasting Method Based on Genetic Programming and Kalman Filter.", parue dans "CCE'23, LNCS, Springer 2023.", et la seconde (Chapitre 6) "AI Ethical Framework: A Government-Centric Tool Using Generative AI.", publiée dans "International Journal of Advanced Computer Science and Applications (IJACSA)."

Partie I

Revue de la littérature philosophique et scientifique

Revue de la littérature philosophique sur l'éthique

Sommaire

2.1	Éthique et méta-éthique	20
2.1.1	Qu'est-ce que l'éthique?	20
2.1.2	Champs de l'éthique	21
2.2	Éthique déontologique	23
2.2.1	Théories déontologiques centrées sur l'agent	24
2.2.2	Théories déontologiques centrées sur le droit des personnes	25
2.2.3	Théories déontologiques Contractualistes	25
2.2.4	Théories déontologiques	26
2.2.5	Avantages des Théories déontologiques	26
2.2.6	Inconvénients des Théories Déontologiques	27
2.2.7	Déontologie et incertitude des issues	27
2.2.8	Théories déontologiques et méta-éthique	28
2.2.9	Applications pratiques de la déontologie et de l'éthique	28
2.3	Apports des philosophes au progrès de l'éthique	29
2.3.1	Philosophie morale et politique de Thomas Hobbes (1588-1679)	29
2.3.2	David Hume et l'utilité (1711-1776)	29
2.3.3	Éthique déontologique d'Immanuel Kant (1724-1804)	30
2.3.4	Éthique conséquentialiste de Jeremy Bentham (1748-1832)	32
2.3.5	Éthique utilitariste de John Stuart Mill (1806-1873)	34
2.3.6	Henry Sidgwick et l'utilitarisme (1838-1900)	36
2.4	Éthiques modernes	40
2.4.1	Principes de justice de John Rawls (1921-2002)	40
2.4.2	Éthique du Care de Carol Gilligan (1936—)	43
2.4.3	Éthique du Discours de Jürgen Habermas (1929—)	44
2.4.4	Éthique des données (Data Ethics) de Luciano Floridi	45
2.5	Philosophie "occidentale"	46

2.5.1	Philosophie analytique	46
2.5.2	Éthique de l'intérêt	47
2.5.3	Philosophie continentale	48
2.5.4	Éthique contractualiste, éthique du désintérêt	51
2.5.5	Conclusion sur les éthiques occidentales	53
2.6	Méta-éthiques universalistes	53
2.6.1	Introduction à l'utilité d'une éthique universelle	53
2.6.2	Claude Lévi-Strauss (1908 - 2009) et Jürgen Habermas (1929—), pour une éthique universelle	54
2.6.3	Proposition d'éthique minimaliste, par Ruwen Ogien (1947 - 2017)	57
2.7	Méta-éthiques relativistes	59
2.8	Philosophie et éthique africaine	61
2.8.1	Philosophie africaine	61
2.8.2	Éthique africaine	62
2.9	Justice et droit des animaux et des robots	63
2.9.1	Justice et droit des animaux non-humains	64
2.9.2	Justice et droit des robots	65
2.10	Conclusion sur la revue de la littérature philosophique sur l'éthique	70

Le présent travail de thèse consacré au domaine de l'informatique, de l'IA plus précisément, nous invite à parler de l'éthique, compte tenu des raisons exposées précédemment (menaces et risques associés à la technologie), mais ne nous donne pas toute la liberté d'aborder, dans le fond, les aspects philosophiques de la question.

Dans ce chapitre, nous nous contentons d'exposer les éléments utiles pour une bonne compréhension de la question : une définition de l'éthique, un bref historique, les différents types d'éthique et, enfin, les éthiques modernes.

Mais vu le nombre de travaux sur le sujet depuis les philosophes grecs anciens, il n'est pas possible de faire une recension exhaustive de tout ce qui a été proposé. Nous ne présenterons donc que quelques théories, celles qui nous ont semblé les plus intéressantes dans un cadre gouvernemental. Il s'agit par exemple de se poser la question : l'éthique prônée par un gouvernement devra-t-elle être basée sur l'intérêt ou sur le désintérêt ? Sur l'utilitarisme ou sur le contractualisme ? Il est donc important de connaître et de comprendre les différents courants éthiques dans une thèse dont le titre contient le mot "éthique", même si tous n'ont pas pu être décrits en 50 pages.

En effet, la partie “état de l’art” d’une thèse contient souvent une liste des autres travaux similaires effectués précédemment sur le sujet de la thèse en question. C’est le cas pour notre première contribution. Dans le cas de la deuxième contribution, nous avons cherché à développer une plateforme permettant de définir une méta-éthique “relativiste” (cf. section 2.7) en proposant la création d’une application appelée Ethiconorm permettant à un gouvernement de pondérer l’importance du respect de 10 principes éthiques consensuels, et la création d’une application appelée Ethicotest permettant à un développeur de comparer l’alignement de son application sur les principes éthiques demandés par le gouvernement.

N’ayant rien trouvé de similaire à ce que nous proposons dans la littérature, cet état de l’art sera un récapitulatif des différentes théories éthiques, aboutissant à la conclusion que suivant les civilisations et les parties du monde, on rencontre plusieurs visions de ce qui est considéré comme éthique ou non.

Par exemple, les sociétés occidentales sont souvent plus individualistes que les sociétés africaines, ce qui signifie qu’il est possible qu’un gouvernement occidental et un gouvernement africain pondèrent différemment l’importance du respect de l’individu par rapport au respect de l’institution.

C’est donc sur ce chapitre sur les différentes méta-éthiques élaborées au cours du temps que nous fonderons notre contribution “Façonnage par IA générative d’une éthique relativiste gouvernementale” décrite dans le chapitre 6.

2.1 Éthique et méta-éthique

2.1.1 Qu’est-ce que l’éthique?

Le mot éthique vient du grec *êthikê* (science morale), *êthos* (mœurs, habitude, caractère) et du qualificatif latin *ethica* (morale). En fait, les philosophes grecs de l’antiquité utilisaient le mot “morale” pour parler du concept d’éthique ¹. Si ce rapprochement peut effectivement être fait, on peut se poser la question de savoir pourquoi Baruch Spinoza (1632 - 1677), dans son œuvre sur l’Éthique, n’utilise pas une seule fois le mot morale.

Une raison pourrait être fournie par son héritier spirituel, Constantin Brunner (1862-1937) pour lequel *la morale consiste en un ensemble de règles “relatives” fictivement érigées en Bien et Mal absolus, alors que l’éthique est précisément la morale débarassée de ses croyances superstitieuses absolutisant le relatif et de ses condamnations moralisatrices utilisées comme une arme contre les Autres.*

Le philosophe Stéphane Sangral explicite cette dichotomie entre morale et éthique dans *Fatras du Soi, fracas de l’Autre* [Sangral 2015]: la morale est un “ensemble de

¹<https://fr.wikipedia.org/wiki/%C3%89thique>

valeurs et de règles arbitraires et propres à chaque groupe”, “un linceul crasseux qui recouvre le vivre-ensemble” et pour ce qui est de l’éthique, un “ensemble de valeurs et de règles universelles”, dont la morale peut s’inspirer (parfois, mais parfois seulement).

Contrairement à la morale, l’éthique fait donc appel à la recherche du “bien”, du “bon” ou du “juste”, par un raisonnement conscient et critique sur la moralité des actions, et à des réflexions théoriques portant sur la valeur des pratiques et sur les conditions de ces mêmes pratiques. Tout comme dans l’Éthique de Spinoza [De Spinoza 2005], l’homme, de par sa dignité, doit comprendre les causes de ses passions et agir selon la raison, en harmonie avec la nature dont tout découle, par nécessité. En effet, d’après Spinoza, l’éthique est une attitude pratique qui va de Dieu pour aboutir à la liberté et la béatitude (bonheur, état de satisfaction parfaite). L’homme est invité à dépasser l’état ordinaire de servitude vis-à-vis des affects, à se tourner vers la connaissance de Dieu (c’est-à-dire, pour Spinoza, la nature), pour s’émanciper et connaître le bonheur.

L’éthique est également appelée philosophie morale. Dans les pays laïcs, elle est souvent influencée par la philosophie politique qui la sous tend. Elle peut être liée, par ailleurs, à la culture, la religion, aux traditions, aux croyances, et à l’idéologie ¹.

2.1.2 Champs de l’éthique

Au-delà des aspects liés à la philosophie, à la politique et aux croyances, en général, des questionnements d’ordre pratique permettent de distinguer différents champs de l’éthique :

Comment les personnes devraient-elles agir ? on parlera d’éthique normative ou prescriptive. L’éthique normative vise à classer les actions comme plutôt justes ou plutôt injustes, si possible sans parti pris, en se basant sur un ensemble de normes relatives à un comportement. Elle traite, donc, de ce que la population devrait croire être juste ou injuste. Quant à l’éthique prescriptive (prescriptivisme), elle est une approche méta-éthique qui considère que les jugements moraux sont des prescriptions, une forme logique d’impératifs.

Parmi les nombreuses théories de l’éthique, on peut distinguer :

- L’éthique de la vertu, qui se focalise sur les actions qui rendent une personne bonne. Elle est inspirée d’Aristote.
- L’éthique déontologique (dérivée du grec *deon*, devoir, et de *logos*, science ou étude), qui se concentre sur les choix moralement requis, interdits ou permis [Alexander 2007]. Elle exige que les actes se conforment à un devoir individuel ou professionnel et est inspirée de différents courants philosophiques :
 - des philosophes tels que John Locke, Jean-Jacques Rousseau et Montesquieu qui pensent que les hommes ont des droits et des devoirs absolus ;

- de Immanuel Kant, partisan de l'impératif catégorique ; celui-ci voit l'origine de la moralité dans la capacité du raisonnement humain de créer certaines lois morales inviolables ;
- plus récemment de John Rawls, partisan du contractualisme, qui soutient que les actes moraux sont ceux sur lesquels nous sommes tous d'accord à la condition que nous ne soyons pas partiels.
- Le Conséquentialisme, qui soutient que la moralité d'une action est uniquement liée au résultat de l'action. Parmi les théories conséquentialistes, on peut citer :
 - l'égoïsme, qui soutient qu'une action est juste si elle conduit au bien-être de celui qui agit, et
 - l'utilitarisme qui soutient qu'une action est juste si elle conduit au plus grand bonheur pour le plus grand nombre. C'est une forme de conséquentialisme : la justesse d'une action est entièrement comprise en termes de conséquences produites. Contrairement à l'égoïsme, l'utilitarisme prend en compte le bien-être global, y compris celui des autres.

Comment mettre en pratique les connaissances morales? Cette question se réfère à l'application de principes éthiques, aussi appelée en français déontologie². Bien que ces deux expressions (éthique appliquée et déontologie) soient souvent utilisées en France de manière interchangeable, car se rapportant toutes deux à la conduite morale, elles présentent des différences significatives (cf. tableau 2.1)

Les deux peuvent entrer en conflit lorsque les règles déontologiques s'opposent aux règles éthiques. Par exemple, une éthique utilitariste cherchera à maximiser le bien du plus grand nombre, potentiellement au détriment du bien de quelques-uns. Si, dans un hôpital, six patients sont sur le point de mourir faute de donneurs d'organes vitaux différents (foie, coeur, poumon, reins, pancréas), l'éthique utilitariste voudrait qu'on sacrifie une personne en bonne santé pour prélever ses organes et ainsi, permettre à six autres personnes de vivre, au détriment de la vie unique du donneur. Mais la déontologie médicale (qui interdit aux médecins de tuer une personne en bonne santé) s'y oppose.

Le cas peut aussi se présenter lorsque l'on demande à des journalistes de donner des informations compromettantes pour leurs sources, ou lorsqu'on demande à un prêtre de révéler des éléments dont il aurait eu connaissance lors d'une confession.

Là où très souvent, l'application de règles déontologiques l'emporte sur l'éthique utilitariste sociétale, on observe tout de même qu'il y a des cas de torture, sous prétexte que torturer quelqu'un pourrait permettre d'obtenir des informations permettant de sauver des vies.

²En anglais, le mot "déontologie" s'applique à l'éthique de ce qui doit être fait d'après les règles éthiques, sans tenir compte des conséquences des actes.

Table 2.1: Comparaison entre éthique (en France) et déontologie

Éléments de comparaison	Éthique	Déontologie (en France)
Définitions	Ensemble de principes moraux souvent sujets à débat, qui guident le comportement d'un individu dans une communauté. Les principes peuvent varier selon les cultures ou les communautés. Une transgression de ces principes peut entraîner des représailles, voire une exclusion de la communauté.	Ensemble de règles ou de normes dictées par une profession ou un organisme. Ces règles régissent la conduite des membres de la profession ou de l'organisation. Une transgression de ces règles peut entraîner des sanctions ou des pénalités.
Différence au plan professionnel (codes de conduite vs Valeursvaleurs individuelles)	L'éthique relève souvent de la conscience individuelle, des valeurs personnelles de l'individu et de sa capacité à discerner le bien du mal	La déontologie se manifeste généralement sous la forme de codes de conduite établis par une profession ou une organisation.
Points communs (rôle du bien et du mal)	La déontologie et l'éthique sont très attachées à distinguer entre le bien et le mal. Elles visent à encourager la conduite morale et à prévenir le comportement immoral.	
Points communs (importance de l'intégrité)	L'intégrité est un concept fondamental dans la déontologie et l'éthique. Elle concerne l'adéquation entre les actions d'une personne et ses valeurs déclarées.	
Points communs (importance de l'intégrité)	Les deux sont "relativistes", au sens où elles sont "relatives" au contexte culturel ou professionnel, ce qui peut amener à des conflits entre éthique et déontologie.	

On observe aussi des cas d'emprisonnement (voir d'assassinat) d'opposants politiques, montrant que l'éthique égoïste du dirigeant (voire l'éthique utilitariste d'un état, dans le cas de l'assassinat de Ben Laden) l'emporte sur l'éthique de l'armée, qui devrait être de protéger les citoyens et la personne.

2.2 Éthique déontologique

Nous avons précédemment décrit la déontologie comme étant un ensemble de règles ou de normes dictées par une profession ou un organisme. Il s'agit là de l'acceptation française du terme "déontologie", mais en dehors de la France, la déontologie fait référence au fait de suivre des règles (du grec *ontos*, devoir) sans prendre en compte les conséquences des

actes effectués. Pour Kant, il faut toujours dire la vérité même si cela peut amener à ce qu'un innocent soit condamné. Ainsi, les théories déontologiques jugent la moralité des choix selon des critères différents des états de choses résultants. Les formes les plus connues de la déontologie stipulent que certaines actions ne peuvent pas être justifiées par leurs conséquences, même si elles mènent à un plus grand bien. Pour les personnes qui suivent une éthique déontologiste, ce qui rend une action juste est sa conformité à une norme morale, indépendamment des conséquences bonnes ou mauvaises qu'elle pourrait produire. Le Droit prend ainsi la priorité sur le Bien [Alexander 2007].

Nous allons beaucoup parler d'éthique déontologique, car c'est celle qui est le plus facilement implémentable en informatique, car il s'agit d'appliquer les règles sans se poser des questions quant à leurs conséquences futures. Les ordinateurs fonctionnant à base de règles logiques booléennes strictes, c'est ce type d'éthique qu'il est le plus facile de mettre en œuvre directement dans un ordinateur. Cette éthique déontologique, car il est difficile de "négocier" avec un ordinateur, mais peut-être les grands modèles de langage comme ChatGPT changeront cela à l'avenir.

Les déontologistes s'opposent aux conséquentialistes qui évaluent les actes uniquement selon les résultats qu'ils entraînent. Les conséquentialistes peuvent être monistes ou pluralistes quant à ce qui est considéré comme "le Bien". Cependant, les critiques font valoir que les déontologistes peuvent être jugés comme étant trop exigeants, n'autorisant aucun écart à la règle, là où les conséquentialistes peuvent être jugés comme étant trop permissifs, justifiant des actes moralement répréhensibles si leurs conséquences sont bénéfiques à toutes et à tous [Alexander 2007].

Les éthiques conséquentialistes sont plus difficiles à implémenter en informatique, car même si évaluer la conséquence de l'acte après-coup peut se faire avec des sommes pondérées et des formules mathématiques (conduisant à ce qu'on appelle la philosophie analytique anglosaxonne), il faut savoir combien de points donner (ou soustraire si l'action est nuisible à certaines personnes), quoi additionner, ce qui est difficile car des actions peuvent avoir des conséquences non anticipées et non attendues sur un long terme.

2.2.1 Théories déontologiques centrées sur l'agent

Elles se distinguent de celles centrées sur le droit des personnes. Selon les théories centrées sur l'agent, nous avons des permissions et des obligations qui nous fournissent des raisons d'agir spécifiques à l'agent en question. Par exemple, chaque parent a des obligations particulières envers ses enfants, non partagées par les autres [Alexander 2007].

Ces théories mettent l'accent sur l'agence morale de chaque individu, en insistant sur le fait que nos obligations catégoriques ne doivent pas se focaliser sur la manière dont nos actions permettent ou causent le mal par d'autres agents, mais sur le maintien de notre propre agence sans tâche morale.

Les théories centrées sur l'agent se divisent entre celles qui mettent l'accent sur l'intention et celles qui mettent l'accent sur les actions. Certaines versions combinent les deux aspects[Alexander 2007]:

1. Intention et volition : ces versions affirment que ce sont nos intentions qui définissent principalement notre agence morale. C'est donc l'intention qui compte, et la personne ne peut pas être réprimandée si la conséquence d'une action est négative, si elle partait d'une bonne intention.
2. Actions : d'autres versions se concentrent sur les actions plutôt que sur l'intention, et disent qu'il faut juger les gens par leurs actes, et pas par leurs intentions.
3. Combinaison des deux : une troisième version combine les deux aspects, stipulant que même si l'intention est bonne, il faut aussi tenir compte des conséquences des actions effectuées.

2.2.2 Théories déontologiques centrées sur le droit des personnes

Les théories centrées sur le droit des personnes sont aussi dites théories de la victime. Une version illustrative affirme le droit (au sens "droit de l'homme" amenant à la notion de "contrat social") contre l'utilisation d'une personne uniquement comme moyen sans son consentement. Cela interdit des actions telles que l'utilisation du corps, du travail ou des talents d'une personne sans son consentement.

Ces théories expliquent des intuitions courantes face à des cas hypothétiques comme ceux du *trolley* et de la transplantation. Par exemple, il est généralement considéré comme acceptable de dévier un trolley l'amenant à tuer une personne pour sauver cinq personnes, mais inacceptable de tuer une personne saine pour sauver cinq patients en attente d'un organe.

La distinction réside dans le fait que dans le cas du *trolley*, la personne sacrifiée n'est pas utilisée comme un moyen direct pour sauver les autres, alors que dans le cas de la transplantation (ou du *Fat Man*, en anglais), elle l'est.

2.2.3 Théories déontologiques Contractualistes

Les théories déontologiques contractualistes considèrent les actes moralement mauvais comme ceux qui seraient interdits par des principes acceptés dans un contrat social approprié. Ces théories peuvent être centrées sur le patient, se focalisant sur les obligations de ne faire aux autres que ce à quoi ils ont consenti.

Une application médicale utilisant l'IA pourra-t-elle (ou pas) violer le contrat social, par exemple au bénéfice d'un plus grand nombre ? L'exemple typique est celui d'Henrietta Lacks [Callaway 2013], une jeune femme noire dont les cellules cancéreuses "HeLa" (pour Henrietta Lacks) ont été utilisées (sans son consentement ni celle de sa

famille) pour développer toute la médecine moderne, car elles se sont révélées comme étant immortelles.

Une application basée sur l'IA pourrait-elle (comme l'ont fait les chercheurs humains) recommander l'utilisation de cellules "HeLa" obtenues illégalement pour leur particularité qui permettrait d'obtenir un résultat scientifique intéressant ?

2.2.4 Théories déontologiques

Immanuel Kant est central aux théories morales déontologiques. Chacune des branches de l'éthique déontologique, que ce soit centrée sur l'agent, centrée sur la victime ou contractualiste, peut revendiquer une base kantienne.

- Pour les déontologistes centrés sur l'agent, Kant situe la qualité morale des actes dans les principes ou maximes sur lesquels l'agent agit et non principalement dans leurs effets sur les autres.
- Pour les déontologistes centrés sur la victime, Kant stipule qu'il est interdit d'utiliser les autres comme simples moyens pour leur propre fin.
- Pour les contractualistes, Kant insiste sur le fait que les maximes doivent pouvoir être voulues comme une loi universelle, acceptée par tous les agents rationnels.

2.2.5 Avantages des Théories déontologiques

Par rapport aux théories conséquentialistes, on peut très facilement les mettre en œuvre dans un ordinateur, car il s'agit simplement de suivre les règles d'une éthique inscrite dans un programme d'ordinateur. On peut aussi évaluer facilement l'alignement d'une action avec les règles éthiques, en regardant simplement combien de règles elle viole ou elle met en œuvre, ce sera l'objet de la section 6.3.3.2 décrivant "Ethicotest".)

De plus, les théories déontologiques permettent de mettre en avant :

- Le respect des relations personnelles : contrairement aux théories conséquentialistes, les théories déontologiques permettent aux agents de donner une attention particulière à leurs familles, amis et projets, évitant ainsi les exigences excessives et aliénantes.
- Un espace pour le surrogatoire : les déontologistes peuvent accomplir plus que ce que la morale exige, car certaines actions peuvent être moralement louables sans être obligatoires, ce qui n'est pas possible dans une vision conséquentialiste pure.
- La conformité aux intuitions morales communes : les théories déontologiques expliquent mieux les réactions communes à des dilemmes moraux comme le *trolley*, la transplantation (aussi appelée *fat man*).

- Les devoirs envers des personnes spécifiques : les devoirs déontologiques sont souvent dirigés vers des personnes spécifiques (enfants, parents, conjoints, employés, etc), ouvrant la voie à des plaintes lorsque ces devoirs ne sont pas respectés ou violés.

2.2.6 Inconvénients des Théories Déontologiques

- Paradoxe de la rationalité : les déontologistes doivent justifier pourquoi certaines actions sont obligatoires même si elles sont maléfiques, défiant ainsi la rationalité conséquentialiste évidente de l’accomplissement de l’action. L’exemple standard (depuis l’échange entre Benjamin Constant et Immanuel Kant) est celui d’un homme venant chercher refuge chez une personne car il est poursuivi par un meurtrier. Pour Kant, le mensonge étant impossible, si le meurtrier vient à frapper à la porte de la personne ayant caché l’homme, la personne devra lui dire qu’elle cache l’homme pour ne pas mentir, même si la conséquence sera la mort de l’homme poursuivi par le meurtrier.

De par ses conséquences, une telle action (dire au meurtrier que la personne cache l’homme) pourra être considérée comme maléfique, notamment par les théories analytiques car la somme de “bien” dans le monde aura alors diminué.

- Maintien de la pureté morale : les obligations catégoriques peuvent exiger de préserver la pureté morale personnelle au détriment de l’ensemble moralement meilleur.
- Conflits entre devoirs : problèmes de conflit entre certains devoirs ou droits, nécessitant des solutions plausibles comme les devoirs *prima facie* ou une *déontologie du seuil*.
- Réponse aux catastrophes morales : la déontologie doit gérer des scénarios où le respect des normes entraîne des conséquences désastreuses. Si un terroriste confesse à un prêtre qu’il a placé dans un stade une bombe qui fera des centaines de morts, respecter le secret de la confession pourrait aboutir à la mort de centaines de personnes.

2.2.7 Déontologie et incertitude des issues

Les déontologues étudient la manière d’évaluer les actions lorsque la règle déontologique comporte des aspects stochastiques. Parmi les exemples, on peut citer le fait de faire exploser de la dynamite avec des résultats incertains et le fait d’utiliser sa voiture dans un environnement urbain sujet aux embouteillages, car il peut en résulter de manquer le rendez-vous. L’incertitude des résultats remet en question les contraintes déontologiques axées sur les intentions, les croyances ou l’utilisation des autres comme moyens. Des

chercheurs comme Aboodi, Borer, Alexander, Lazar, Smith, Tarsney et Tomlin se sont penchés sur cette question [Alexander 2007].

2.2.8 Théories déontologiques et méta-éthique

Les théories déontologiques de l'éthique normative sont indépendantes d'une ontologie ou d'une épistémologie morale spécifique. Les déontologues peuvent s'aligner sur le réalisme moral (naturel ou non), l'expressivisme, le constructivisme, le transcendantalisme, etc. en méta-éthique. Alors que la déontologie peut être agnostique d'un point de vue méta-éthique, certains points de vue comme le réalisme non naturel ou la théorie du commandement divin sont plus compatibles avec la déontologie. Certains soutiennent qu'une méta-éthique naturaliste-réaliste peut soutenir l'éthique déontologique [Alexander 2007].

2.2.9 Applications pratiques de la déontologie et de l'éthique

Les principes de la déontologie et de l'éthique sont appliqués dans de très nombreux secteurs comme :

- **Le secteur médical** : les professionnels sont tenus de respecter un code de déontologie, qui comprend des principes tels que le respect de la vie et de la dignité humaine. Ils doivent également prendre des décisions éthiques lorsqu'ils sont confrontés à des dilemmes moraux.
- **Le secteur des affaires** : la déontologie peut prendre la forme de normes de conduite pour une concurrence loyale, tandis que l'éthique peut concerner des questions telles que la responsabilité sociale des entreprises

Que pensent les gens de ce qui est juste ? On parlera, dans ce cas, d'éthique descriptive ou comparative. L'éthique descriptive (aussi appelée éthique comparative) a pour objet d'étudier les croyances sur la moralité. Elle se concentre sur la description et la comparaison des différents principes moraux observés au sein d'une société ou d'un groupe. Elle étudie les pratiques éthiques réelles, au lieu de formuler des règles, comme le fait l'éthique normative. Elle est également différente de la méta-éthique qui se concentre sur le sens des concepts éthiques comme "le bien", "le bon" ou "le juste".

Que signifient les mots "bien", "bon", "juste", etc. ? On parle de méta-éthique qui signifie "par-delà" l'éthique. En effet, la méta-éthique n'a pas pour but de définir de nouvelles normes ou lois morales. Elle ne dit pas par exemple "tu dois agir de cette manière" Elle a pour objet d'étudier la nature des énoncés éthiques impératifs, d'analyser les concepts fondamentaux de l'éthique et enfin, de s'interroger notamment sur la nature et la signification de concepts moraux comme le bon, le bien, le juste, le devoir.

A propos des jugements moraux, il existe deux courants : le non-cognitivism et le cognitivism. Le non-cognitivism considère que les jugements moraux ne peuvent être vrais ou faux, puisque le monde n'a pas de caractéristiques morales objectives. Quant au cognitivism, il considère que les jugements moraux se réfèrent à une réalité morale.

2.3 Apports des philosophes au progrès de l'éthique

Il est impossible de passer en revue dans un chapitre l'ensemble des développements de la philosophie depuis son apparition en Grèce antique. Dans cette section, nous allons choisir certains philosophes pour montrer leur apport et leur influence sur les théories actuelles de l'éthique, car pour créer une IA respectant "l'éthique", encore faut-il savoir ce que recouvre un tel terme.

2.3.1 Philosophie morale et politique de Thomas Hobbes (1588-1679)

La philosophie morale et politique de Thomas Hobbes (1588-1679) s'articule autour du principe fondamental de l'ordre social et politique³. Hobbes explique comment les hommes doivent vivre en paix sous l'autorité d'un pouvoir souverain (un État politiquement autoritaire) pour assurer le maintien de la paix et éviter les conflits au sein de l'"état de nature".

En effet, la philosophie morale de Hobbes mène à l'autoritarisme⁴. Sur le plan de la liberté, Hobbes distingue deux formes : la "liberté naturelle" (qui est de ne pas être empêché de faire ce qu'on veut faire) et la "liberté civile" (théorie du "silence de la loi" : tout ce qui n'est pas interdit est autorisé, et puissance absolue du souverain).

2.3.2 David Hume et l'utilité (1711-1776)

David Hume, un précurseur des utilitaristes classiques comme Jeremy Bentham et John Stuart Mill, a mis en avant l'importance de l'utilité dans les jugements moraux bien avant la formulation systématique de l'utilitarisme. Les moralistes britanniques tels que Cumberland, Shaftesbury, Hutcheson, Gay, et Hume ont tous contribué aux fondements de l'utilitarisme, mais c'est en particulier dans les écrits de Hume que l'idée d'utilité a pris une importance significative [Driver 2009].

Dans l'*Enquête sur les principes de la morale* (1751), Hume avance que l'utilité est une source primordiale de louange et d'approbation morale. Hume décrit l'utilité comme un principe qui unifie les vertus sociales telles que l'honneur, l'obéissance et la chasteté. Cependant, contrairement à ce que prônait son aîné Bentham, Hume ne réduisait pas la morale à un simple calcul de plaisirs et de douleurs. Pour Hume, l'utilité

³https://en.wikipedia.org/wiki/Hobbes's_moral_and_political_philosophy

⁴[https://fr.wikipedia.org/wiki/L%C3%A9viathan_\(Thomas_Hobbes\)](https://fr.wikipedia.org/wiki/L%C3%A9viathan_(Thomas_Hobbes))

était entrelacée avec la sympathie et d'autres sentiments, ce qui rend la notion moins rigide et plus complexe que celle que Bentham proposera plus tard [Cléro 2013].

Son importance dans le développement de la pensée contemporaine est considérable : Hume eut une influence profonde sur Kant, sur la philosophie analytique du début du xxe siècle et sur la phénoménologie ⁵.

La Critique de Bentham : Jeremy Bentham s'intéressait profondément aux textes de Hume, en les reconnaissant comme des précurseurs essentiels de ses propres idées utilitaristes. Cependant, Bentham critiquait Hume pour ne pas avoir placé suffisamment l'accent sur les plaisirs et les douleurs dans ses analyses. Bentham considérait les plaisirs et douleurs comme des unités élémentaires de calcul moral tandis que Hume percevait l'utilité à travers une perspective plus large incluant des adaptations de moyens aux fins et impliquant la sympathie [Cléro 2013].

Bentham reconnaît que Hume a fait une distinction cruciale entre les impressions et les idées, ce qui constitue une base importante de son propre système, malgré ses critiques sur l'absence de classification claire des plaisirs et des douleurs chez Hume. Bentham estimait aussi que Hume avait une vision trop étroite des relations entre la raison et les passions, ne prenant pas en compte les nuances nécessaires dans l'évaluation morale.

Justice et Utilité : La justice, pour Bentham et Hume, est l'une des vertus centrales discutées. Hume voit la justice comme une vertu dérivée principalement de l'intérêt privé et non de la sympathie ou de la bienveillance. La justice doit se comprendre par son rôle dans la préservation de la propriété, une vue que Bentham adapte et intègre dans son système utilitariste en mettant davantage l'accent sur la maximisation du bonheur collectif.

F. Rosen, un commentateur contemporain, a montré des affinités entre les conceptions de la justice chez Hume et Bentham, tout en soulignant leurs différences. Si Hume reconnaît que l'utilité apporte du bonheur et peut trouver une place dans le calcul moral, il ne la systématise pas par des calculs précis comme le fait Bentham. Pour Bentham, la justice et l'utilité doivent être intégrées dans une démarche systématique de maximisation du bonheur pour le plus grand nombre, tandis que Hume maintient une perspective plus nuancée, intégrant l'esthétique et la sympathie dans ses appréciations.

2.3.3 Éthique déontologique d'Immanuel Kant (1724-1804)

Le mot "déontologie" a été créé par Jeremy Bentham dans son livre *Chrestomathia* [Bentham 1816], provient du grec *deon* "ce qui doit être fait" *logos*, "doctrine".

Dans son éthique déontologique, Kant s'oppose à l'éthique d'Aristote, l'éthique de la vertu morale et du bonheur, centrée sur l'individu⁶.

⁵https://fr.wikipedia.org/wiki/David_Hume#Traductions_fran%C3%A7aises

⁶<https://www.houseofethics.lu/2021/02/26/5-modern-ethical-theories/>

En effet, il postule que les agents devraient se concentrer sur ce qu'ils doivent faire et le type d'actions à accomplir. L'action doit alors, selon Kant, découler de l'obligation morale, indépendamment de toute circonstance empirique de l'action ⁷.

Ainsi donc, l'éthique déontologique (*deon* : "ce qui doit être") affirme que chaque acte moral doit être jugé selon sa conformité (ou sa non-conformité) à certains devoirs.

Mais il faut distinguer entre plusieurs déontologismes : le déontologisme moniste postule que toutes les obligations doivent découler d'un principe unique, là où le déontologisme pluraliste refuse ce qu'il considère être un réductionnisme et affirme que les actions humaines sont à juger en fonction de plusieurs principes distincts comme le devoir de ne pas faire de mal à autrui inutilement (principe de non-malfaisance), le devoir de se conformer aux engagements librement donnés (principe de fidélité), le devoir de remercier ceux qui nous ont aidés (principe de gratitude), le devoir de compenser les personnes à qui nous avons fait un tort (principe de justice restaurative), etc.⁷.

L'ambition de Kant est de fonder la morale sur des principes aussi peu nombreux que possible. Il défend donc déontologisme moniste, dérivant les obligations d'un principe unique comme l'impératif catégorique.

D'après [Braz 2005] :

Le projet de la réflexion kantienne, en constituant une métaphysique des mœurs, vise à résoudre le conflit, à l'origine, entre l'individu et la communauté. À plusieurs reprises, l'auteur souligne que cette divergence naît de la priorité de l'égoïsme qui détermine la maxime en opposition avec la raison. Cette collision ne peut se dissiper qu'en conférant à la législation de la raison toute son autorité. La doctrine des mœurs œuvre précisément en ce sens dans la mesure où elle explicite un ensemble de devoirs pris comme système déduit de la raison normative. Cela permet de comprendre que les lois de liberté sont à la fois les principes *a priori* de la raison législatrice et les conditions de possibilité du passage de l'individu à la communauté. Le droit et l'éthique sont des législations normatives déduites de l'impératif catégorique, revendiquant par là une source commune et exprimant une autonomie de la volonté générale et une autonomie de la seule volonté. Pourtant, il existe un écart tant dans les structures que dans la finalité, entre ces lois morales, qu'il est nécessaire d'appréhender et de mesurer.

Il s'agit donc à présent de cerner cette hétérogénéité des législations, afin de rendre compte de la distinction entre la communauté juridique et la communauté éthique [Braz 2005].

À travers cette thèse [Braz 2005], Kant radicalise l'indépendance du droit par rapport à l'éthique, en excluant du domaine juridique l'équité et le droit de nécessité. Le premier terme désigne un droit sans contrainte à qui il manque les conditions suffisantes

⁷https://fr.wikipedia.org/wiki/%C3%89thique_d%C3%A9ontologique

pour que l'individu puisse, devant le juge d'un tribunal civil, se réclamer d'un tel droit. En effet, la devise de l'équité admet que le droit le plus strict est la plus grande injustice: par exemple, celui qui dans une société commerciale établie sur la base de l'égalité des profits en a pourtant fait plus que ses partenaires, mais a connu cependant davantage de pertes, peut selon l'équité exiger de cette société plus qu'une répartition à part égale avec les autres membres.

Dans le champ de la philosophie morale, le déontologisme désigne donc une famille de théories qui mettent l'accent sur la notion de devoir et qui peuvent être caractérisées par une opposition, dans un premier temps avec l'éthique de la vertu et, dans un second temps, avec le conséquentialisme.

Là où l'éthique de la vertu met l'accent sur les agents et sur le type de personne que nous devrions être, le déontologisme est une théorie principalement concernée par ce que nous devrions faire et par le type d'actions que nous devrions accomplir.

Cependant, le déontologisme n'est pas seul à occuper ce terrain : il doit y affronter son plus farouche adversaire, le conséquentialisme, et c'est d'ailleurs par opposition au conséquentialisme qu'il se définit la plupart du temps.

2.3.4 Éthique conséquentialiste de Jeremy Bentham (1748-1832)

Bentham définit la déontologie comme la science de l'éthique personnelle [Bentham 1834]. Il tente de concilier les intérêts privés et publics en fixant ce qui doit être guidé par le principe de l'utilité. Il propose ainsi dans une nouvelle approche de la morale, basée sur l'expérimentation et la recherche du plus grand plaisir pour soi-même et pour le groupe auquel on appartient.

Mais contrairement aux déontologistes, Jeremy Bentham (philosophe et juriconsulte) pense que les actes moraux, fondés sur l'utilitarisme, doivent être jugés en fonction de leurs conséquences.

Inventé par Elizabeth Anscombe [Anscombe 2020], le terme "conséquentialisme" fait partie des éthiques téléologiques.

D'après Wikipédia⁸, "il constitue l'ensemble des théories morales qui soutiennent que ce sont les conséquences d'une action donnée qui doivent constituer la base de tout jugement moral de ladite action. Ainsi, d'un point de vue conséquentialiste, une action moralement juste est une action dont les conséquences sont bonnes. Plus formellement, le conséquentialisme est le point de vue moral qui prend les conséquences pour seul critère normatif."

L'éthique conséquentialiste est opposée, d'une part à l'éthique déontologique, qui se concentre principalement sur ce que nous devrions faire et le type d'actions que nous devrions accomplir et, d'autre part, à l'éthique de la vertu laquelle se concentre sur le

⁸<https://fr.wikipedia.org/wiki/Cons%C3%A9quentialisme>

caractère et les motivations de l'agent.

Les conséquentialistes divergent sur la nature du bien. Les hédonistes limitent le bien au plaisir, tandis que les pluralistes reconnaissent plusieurs biens intrinsèques comme la liberté, la connaissance, et les relations humaines. Cette diversité rend le conséquentialisme adaptable mais aussi sujet à débat sur la manière de mesurer et comparer ces différents biens [Sinnott-Armstrong 2003].

Ce qui importe dans la théorie conséquentialiste, c'est que l'agent moral doit agir de manière à produire de bonnes conséquences, même s'il ne produit pas les meilleurs résultats possibles. Une action est bonne s'il en résulte un accroissement global de bonheur, et la meilleure action est celle qui donne lieu au maximum de plaisir et au minimum de peine.

Mais comme on va le voir avec l'utilitarisme, ce raisonnement, poussé à bout, peut aboutir à léser une minorité pour satisfaire une majorité.

Pour améliorer cette situation, certains conséquentialistes intègrent des éléments non-conséquentialistes ou introduisent des règles et des perspectives relatives aux agents dans leur évaluation [Sinnott-Armstrong 2003].

La contrainte et l'ignorance des circonstances sont deux excuses majeures qu'un agent peut invoquer (ou que l'on peut invoquer pour sa défense) afin de se décharger de tout ou partie de sa responsabilité morale à l'égard d'un état de choses dont il est néanmoins causalement responsable [Michon 2016]. Si j'ai été violemment poussé contre quelqu'un marchant au bord d'un précipice, ou si l'on m'a menacé de mort afin que je le pousse, ou encore si j'ignorais qu'il y avait un précipice alors que je voulais lui faire une mauvaise plaisanterie, la chute et ses conséquences, que j'ai causées, ne me seront pas moralement imputables comme si je l'avais fait exprès, de plein gré et consciemment [Michon 2016].

Ces observations introduisent la Doctrine du Double Effet (DDE), qui vient nous placer sur le registre opposé : celui qui consiste à réaliser un acte non désiré au départ mais apte à produire un bon effet (un effet salvateur).

Cette doctrine du double effet est initiée par Thomas d'Aquin (1225/26-1274) puis revue par Elizabeth Anscombe dans son ouvrage *L'intention* [Elizabeth 1957]. L'auteure explique dans quelles circonstances il est "permis" de commettre une action ayant à la fois de bonnes et de mauvaises conséquences (autrement dit, un double effet).

Certaines conditions doivent être nécessairement remplies pour qu'une action puisse être moralement justifiée alors même qu'elle comporte de mauvais effets :

- l'action elle-même doit être bonne ou moralement neutre,
- le bon effet doit résulter de l'acte et non du mauvais effet,
- le mauvais effet ne doit pas être directement voulu, mais doit être prévu et toléré,

- le bon effet doit être plus fort que le mauvais effet, ou bien les deux doivent être égaux (au sens où ce serait un mal plus grand de l'éviter sans produire le bon effet).

A titre d'exemples, bombarder une ville —et faire de nombreuses victimes collatérales— pour tuer un chef terroriste, semble une action disproportionnée entre le mal réalisé et le bien recherché, mais bombarder des villes pour détruire des ponts, des usines ou des moyens de production a été effectué pendant la deuxième guerre mondiale, parfois de manière exagérée comme le bombardement de Portel les 8 et 9 septembre 1943, détruisant 94% de la ville et causant plus de 500 morts sans objectif militaires autre que de tenter de faire croire aux allemands à un débarquement prochain dans le Pas de Calais.

Mais si couper la jambe de quelqu'un produit un mal, ce mal peut être recherché pour un bien plus grand (sauver la personne de la gangrène).

On voit donc qu'il peut être parfois justifié de produire une conséquence mauvaise si elle est seulement un effet secondaire de l'action, et non pas intentionnellement recherchée, ce qui pose des questions lorsque la justification est immatérielle [Boubker Lamrani 2003].

Cette doctrine du double effet est contestée par les défenseurs de l'utilitarisme. Ceux-ci refusent, en effet, la possibilité que deux actions identiques dans leurs conséquences ne soient pas d'égale valeur.

Selon [Alexander 2007], certains cherchent à combiner le conséquentialisme et l'éthique déontologique. Un moyen est de créer une théorie mixte, mais l'articulation des notions de rationalité propres à ces deux théories est complexe. Diverses approches sont explorées, telles que l'allocation de raisons spécifiques à chaque théorie ou l'extension des raisons propres à la déontologie pour imiter les avantages du conséquentialisme. Ces tentatives visent à concilier ces deux approches éthiques, mais les défis persistent quant à la compatibilité des perspectives subjectives et objectives. On verra qu'au 20^e siècle, John Rawls fait une proposition intéressante pour combiner les deux : ne pas fonder l'éthique sur la notion de "bien", mais sur la notion de "justice".

2.3.5 Éthique utilitariste de John Stuart Mill (1806-1873)

John Stuart Mill est un philosophe anglais, économiste et théoricien de l'éthique. Il fut instruit par son père (adepte de Jérémy Bentham) et reçut une éducation extrêmement rigoureuse. Cette instruction et l'éducation qu'il reçut avec l'appui de Bentham avait pour but avoué de faire de lui un génie qui pourrait poursuivre la cause de l'utilitarisme.

Dans son ouvrage "De la liberté" [Mill 1885], Mill défend, au nom du principe de non-nuisance, "l'idée d'empêcher les individus de se causer des blessures sérieuses sur eux-mêmes ou sur leurs propriétés". De plus, dans la mesure où nous dépendons, souvent,

les uns des autres (personne ne vit, pratiquement, dans un état d'isolement total), les blessures faites à soi-même peuvent aussi nuire à autrui. Détruire des propriétés cause du tort non seulement à soi-même mais aussi à la communauté.

Son application actuelle est, par exemple, l'obligation (en France) de boucler une ceinture de sécurité lorsqu'on roule en voiture sous peine d'amende.

Mill exclut du non-respect de ce principe de non-nuisance les groupes vulnérables, incapables d'autonomie comme les enfants, les personnes âgées ou jugées malades mentales.

Mill explique que l'acte de nuire peut inclure différentes situations : les omissions volontaires, ne pas payer ses impôts, ne pas se présenter comme témoin lors d'un procès, ne pas secourir un enfant de la noyade, etc.

Il considère qu'il n'y a pas nuisance à autrui si —sans utiliser la force ou la fraude— l'individu concerné consent à assumer les risques. C'est ainsi, qu'une personne peut offrir des emplois dangereux à autrui, tant qu'il n'y a pas tromperie. Mais il précise qu'il doit exister, nécessairement, une limite à ce consentement : à titre d'exemple, il dit que la communauté doit empêcher toute personne de se vendre comme esclave

Enfin, Mill ne considère pas que le fait d'offenser serait une nuisance. Pour lui, une action ne devrait pas être restreinte du fait qu'elle viole les conventions ou les mœurs d'une société donnée.

Les arguments avancés dans [Mill 1966] ne relèveraient pas de droits naturels mais seraient donc plutôt liés au principe d'utilité, ce qui donne son nom à ce courant philosophique.

Par ailleurs, la liberté d'expression est considérée comme une condition nécessaire pour tout progrès social ou intellectuel. La confrontation libre des idées peut être productive. Laisser les gens propager des avis erronés peut les conduire à abandonner leurs opinions erronées, à condition qu'ils soient sincèrement engagés dans des débats d'idées ouverts où chacun est respectueux de l'autre. Pour Mill, le non-respect de l'autre, exprimé par des violences verbales, peut dissuader les gens d'exprimer des opinions contraires et d'écouter ceux qui les expriment⁹.

Plus tard, dans L'"Utilitarisme" [Mill 1861], Mill répond à de nombreuses objections que ses contemporains adressent à sa théorie de l'utilitarisme (voir tableau 2.2).

Pour Mill, la morale utilitariste repose sur le principe du plus grand bonheur pour tous. Cette morale n'a de sens que si elle contribue globalement à augmenter le bonheur – et / ou diminuer le malheur – des êtres humains, et même de tous les animaux "sensibles" (*sentants*).

Mais un des problèmes majeurs de l'utilitarisme est qu'il peut sembler ignorer la justice et les droits individuels. Ainsi que vu précédemment, par exemple, la théorie

⁹<https://www.contrepoints.org/2021/11/21/383366-john-stuart-mill-et-la-liberte-dexpression-les-heros-du-progres-41>

pourrait permettre des actions injustes comme le sacrifice d'une personne pour en sauver cinq.

Si dans un hôpital, 5 patients vont mourir faute de donneurs : un parce qu'il a besoin d'une greffe de foie, un autre car il a besoin d'une greffe de coeur, un troisième car il a besoin d'une greffe de poumons, un 4^e car il a besoin d'une greffe de reins et un 5^e car il a besoin d'une greffe du pancréas...

Ces 5 patients vont mourir, mais, dans la rue, il y a une personne en bonne santé qui possède tous ces organes.

Est-il éthique de prendre tous ces organes de la personne en bonne santé (et donc tuer cette personne) pour permettre à 5 autres de vivre ?

Si un fonctionnaire se pose la question du bien-fondé d'assécher un lac à cause des milliers de morts liés au paludisme apporté par les moustiques du lac sachant que si le lac est asséché, cela impactera négativement le petit village de 500 personnes qui pouvait vivre du lac.

Si, pour trancher, le fonctionnaire demande "son avis" à une IA gouvernementale, que pourrait-elle recommander au fonctionnaire ?

Lorsque l'on pose la question à ChatGPT-4o, la réponse de celui-ci est :

Assécher le lac pourrait sauver des vies à court terme en réduisant le paludisme, mais cela pourrait aussi entraîner des conséquences négatives pour la communauté locale. Il est essentiel d'examiner toutes les alternatives possibles et de consulter la communauté affectée pour trouver un équilibre entre la santé publique et la préservation des moyens de subsistance locaux. Un compromis pourrait être trouvé par des solutions innovantes qui protègent à la fois la santé des habitants et leur environnement.

En 1952, le village de Tignes, en Savoie, a été englouti pour construire le barrage de Chervil. Une IA gouvernementale plus assertive que ChatGPT4o pourrait-elle recommander d'engloutir un village pour construire un barrage qui améliorera le bien public global ?

2.3.6 Henry Sidgwick et l'utilitarisme (1838-1900)

Henry Sidgwick est une figure emblématique de l'utilitarisme moral. Son ouvrage principal, *The Methods of Ethics* (1874) [Sidgwick 2022], est considéré comme l'une des œuvres les plus influentes de la philosophie morale utilitariste. Sidgwick y explore et défend l'utilitarisme, mais certains auteurs, comme [Schneewind 1977], estiment que ce livre ne doit pas être uniquement lu comme une défense de l'utilitarisme. Sidgwick a joué un rôle crucial dans le développement et l'affinement de l'utilitarisme, offrant des arguments solides pour sa primauté théorique tout en abordant les principales critiques et préoccupations associées à cette théorie éthique. En effet, Sidgwick s'est employé

Table 2.2: Dix objections des contemporains de Mill et ses réponses à ces objections.

Fondement des objections	Réponses de John Stuart Mill
Objection 1 : mauvaise compréhension du concept d'“utilité” en rapport au plaisir	Cette objection repose sur une incompréhension, et même sur un contresens au sujet du mot “utilité”, qui ne doit surtout pas être pris dans un sens calculateur.
Objection 2 : conception trop vile des plaisirs humains	L'utilitarisme ne pousse pas les êtres humains vers ce qu'ils ont de plus bas, de plus vile, de plus indigne, à l'image des animaux. Bien au contraire, les plaisirs humains dont il est question sont ceux qui mettent en jeu les facultés supérieures de l'être humain (son intelligence, sa culture, son imagination, etc.).
Objection 3 : conception trop ambitieuse du bonheur qui le rendrait impossible	La conception du bonheur en doit pas être idéalisée. Le bonheur est à notre portée : celui que nous procurent, par exemple, la nature, les arts, les sciences, constituent des sources inépuisables de plaisirs. Quant aux causes extérieures de notre malheur, comme la pauvreté ou les maladies, elles doivent pouvoir être, sinon supprimées, du moins contenues par les progrès politiques et scientifiques que l'utilitarisme appelle de ses vœux.
Objection 4 : conception égoïste du bonheur. La philosophie utilitariste n'est-elle pas incompatible avec le sacrifice désintéressé de soi ?	Bien au contraire. Se sacrifier pour le bonheur des siens mais aussi pour celui des autres relève de l'idéal utilitariste. Il est l'œuvre des héros et des martyrs. La philosophie utilitariste est bien compatible avec le sacrifice désintéressé de soi.
Objection 5 : conception trop exigeante de la morale utilitariste. La philosophie utilitariste ne fixe-t-elle pas un idéal trop élevé, trop noble pour l'humanité ?	L'utilitarisme et aucune autre philosophie morale ne peut exiger de l'être humain le niveau le plus élevé, surpassant ses capacités. Il s'agit d'inviter à un effort personnel pour le bonheur du plus grand nombre, sans nuire au bonheur des autres.

<p>Objection 6 : mauvaise distinction entre l'action et l'agent</p>	<p>Il ne faut pas confondre l'intention (qui ne concerne que l'action elle-même et qui permet d'évaluer la moralité de cette action) et le motif, à savoir le sentiment qui pousse l'agent à accomplir une action, qui permet d'évaluer la valeur de l'agent. Il est plus facile d'évaluer la moralité d'une action, car l'intention est souvent assez nette, que son ou ses motifs, qui peuvent être flous, portés par exemple par le désir d'une récompense (le paradis pour un croyant).</p>
<p>Objection 7 : prétendue incompatibilité entre l'utilitarisme et la religion</p>	<p>L'utilitarisme est bien compatible avec l'existence de Dieu. Dieu veut le bonheur de ses créatures.</p>
<p>Objection 8 : confusion entre l'utile et l'intérêt. La philosophie utilitariste ne conduit-elle pas les être humains à toujours agir de manière intéressée ?</p>	<p>Il faut distinguer entre l'utile et l'opportunité. L'utilité est le critère de l'utilitarisme. Il ne peut être mis à profit au détriment des autres. Quant à l'opportunité, elle peut être utilisée, contrairement à la morale utilitariste. C'est le cas du mensonge qui permet de se tirer d'une situation délicate et qui affaiblit la confiance entre les humains. Toutefois, à titre exceptionnel, il peut être moralement permis de mentir pour sauver un innocent.</p>
<p>Objection 9 : prétendue impossibilité pratique de l'utilitarisme. La philosophie utilitariste n'est-elle pas impossible à mettre en pratique dans l'urgence de l'action ?</p>	<p>Cette objection ne peut relever que de la mauvaise foi ou de la sottise. Quelle que soit l'urgence d'une situation, nul n'a besoin de réfléchir pour savoir si une action est morale ou non.</p>
<p>Objection 10 : risques de l'utilitarisme en cas de dilemme moral. La philosophie utilitariste ne fournit-elle pas, en cas de dilemme moral, des arguments fallacieux pour contrevenir aux règles morales ?</p>	<p>La complexité de la vie et des affaires humaines devrait interdire à tout système moral d'énoncer des règles morales, prétendues universelles. Si la morale Kantienne interdit le mensonge quelle que soit la situation, la morale utilitariste, quant à elle, l'autorise.</p>

à développer une analyse des différentes méthodes éthiques implicites dans le raisonnement moral commun : l'égoïsme, la moralité basée sur l'intuition et l'utilitarisme. Selon Sidgwick, l'utilitarisme constitue la théorie la plus fondamentale [Driver 2009].

Contrairement à d'autres utilitaristes, Sidgwick rejetait l'empirisme, l'égoïsme psychologique et le réductionnisme. Il partait des principes éthiques du sens commun pour les ramener à trois méthodes compatibles : l'intuitionnisme, l'utilitarisme et l'égoïsme éthique. Il défendait un intuitionnisme philosophique en soutenant un ensemble de principes abstraits ou d'axiomes qui convergent vers l'utilitarisme classique. Parmi ces principes, on retrouve le principe d'universalisme (ce qui est bon pour soi doit l'être également pour toute autre personne dans la même situation), un principe de prudence rationnelle (accorder la même importance au futur qu'au présent) et le principe de bienveillance rationnelle (l'utilité pour le plus grand nombre)¹⁰.

Sidgwick argumente que, bien que les règles de la morale commune soient souvent vagues et sous-décrites, leur application exige d'avoir recours à quelque chose de théoriquement plus fondamental, encore une fois l'utilitarisme. Ainsi, les interprétations absolues des règles peuvent sembler très contre-intuitives et nécessitent une justification claire pour toute exception, une justification que, selon Sidgwick, seule l'utilitarisme peut fournir.

Influences et Développements : Sidgwick s'inspirait des travaux de Bentham et Mill tout en développant et en répondant à leurs idées. Son impact a été considérable, non seulement sur les utilitaristes et les conséquentialistes ultérieurs mais également sur les intuitionnistes. Son analyse approfondie et pénétrante de la théorie a soulevé de nombreuses préoccupations abordées par les philosophes moraux récents.

Un aspect très controversé des vues de Sidgwick est son rejet du critère de la publicité pour la théorie morale. Il soutenait que certaines conclusions utilitaristes devraient rester secrètes parce que leur divulgation générale pourrait mener à des résultats indésirables. Cette position, parfois qualifiée de *Government house utilitarianism* (utilitarisme gouvernemental), reflétait le colonialisme élitiste de son époque [Williams 1995]. Sidgwick affirmait que si dissimuler la vérité pouvait conduire à de meilleurs résultats globaux, cette dissimulation était justifiée.

Questions fondamentales et indéterminations : Sidgwick a également soulevé des problèmes fondamentaux de l'utilitarisme, comme la distinction entre utilité totale et utilité moyenne. Il a examiné cela dans le contexte de la croissance démographique et de l'augmentation des niveaux d'utilité via l'augmentation du nombre de personnes :

Assumant que le bonheur moyen des êtres humains est une quantité positive, il semble clair que, supposant que le bonheur moyen reste inchangé, l'utilitarisme nous dirige à rendre le nombre de personnes heureuses aussi grand que possible.

¹⁰https://fr.wikipedia.org/wiki/Henry_Sidgwick

Pour Sidgwick, il ne s'agissait pas simplement de viser une plus grande utilité moyenne, mais de maximiser le produit du nombre de personnes vivant actuellement et de la quantité de bonheur moyen, une vue hybride total-moyenne.

Intuitionnisme Philosophique : Sidgwick posait également les fondements de son intuitionnisme philosophique en soulignant des critères de vérité et d'erreur : clarté, cohérence et consensus. Il affirmait que l'intuitionnisme philosophique pouvait ramener les principes éthiques du sens commun à des principes abstraits conduisant finalement à l'utilitarisme classique.

2.4 Éthiques modernes

2.4.1 Principes de justice de John Rawls (1921-2002)

John Rawls obtient son doctorat en 1950, consacré à la question de la justification morale. A partir de cette date, il fréquente de nombreuses universités prestigieuses (Oxford, Cornell, et Harvard) et des philosophes de premier plan. Il se consacre par la suite, exclusivement, à la question de la justice dans tous ses aspects, moraux, psychologiques, politiques, économiques, juridiques et sociaux.

Avec son livre *Justice as Fairness* (la justice comme équité) [Rawls 1958], John Rawls identifie deux principes de justice ⁶ dont le deuxième se divise en deux (ce qui fait dire à certains qu'il ya en fait trois principes) qui ont une priorité **lexicographique** (c'est là un des apports majeurs de Rawls). La priorité lexicographique signifie que l'on ne doit appliquer le deuxième principe que si le premier est satisfait¹¹ :

1) Liberté :

Chacun doit avoir un droit égal à la liberté la plus grande, compatible avec une liberté similaire pour les autres.

2) Égalité : (qui se divise en deux)

2a) Égalité des chances :

Toutes les valeurs sociales doivent être réparties de manière égale. . .

2b) Principe de différence :

à moins qu'une répartition inégale ne soit à l'avantage de tous.

On pourra remarquer la similarité avec la devise de la France : Liberté, Égalité, Fraternité sauf que Rawls y apporte des contraintes hiérarchiques.

Dans l'éthique de Rawls, qui n'est plus fondée sur la notion de "bien" mais sur la notion de "justice", le premier principe (Liberté) doit primer sur tous les autres.

¹¹Dans un dictionnaire, on passe de la première lettre "a" à la première lettre "b" que lorsqu'on a passé en revue tous les mots commençant par "a")

1) **Le principe de liberté est premier** La liberté ne peut être limitée qu'au nom de la liberté elle-même. On ne peut se préoccuper des principes suivants qu'une fois que la liberté est garantie à toutes et tous.

2) **L'égalité** se décompose en deux sous-principes:

2a) **concerne l'égalité des chances** (qui est différente de l'égalité vue par les communistes, par exemple). En effet, l'égalité des chances ne signifie pas qu'à 40 ans, tout le monde doit gagner le même salaire. Cela signifie qu'à la naissance, tous les enfants doivent pouvoir avoir accès aux mêmes opportunités. À titre d'exemple, l'éducation doit pouvoir être offerte à toutes et à tous, sans distinction de richesse ou de classe sociale.

2b) **concerne le principe de différence** . C'est le plus difficile à comprendre. Ce principe se rapproche du principe de subsidiarité et de fraternité. La subsidiarité est un principe selon lequel les décisions doivent être prises au niveau le plus bas possible, souvent au niveau local, plutôt qu'à un niveau centralisé ou supérieur. Appliqué à l'égalité, cela signifie qu'on ne peut autoriser une différence entre les personnes que si la personne la moins bien lotie ne sera pas, suite à l'application du principe de différence, moins bien lotie qu'avant. En fait, le principe de différence est strict : on ne peut tolérer une différence (et donc l'apparition d'une inégalité) que si la personne la moins bien lotie "gagne aussi" à l'application de la différence. Et c'est en cela que le principe de différence s'apparente à la Fraternité de la devise française.

Pour expliquer le principe de différence (ou fraternité) par un exemple. Dans un groupe de personnes, on n'autorisera l'application d'une mesure permettant l'enrichissement de certains, que si elle permet aux plus pauvres de s'enrichir aussi, même si c'est de manière moindre : si l'application d'une mesure permet à certains de gagner \$100, alors, cette mesure ne pourra être adoptée que si elle permet aux plus pauvres de gagner au moins \$1.

Pour résumer, en 1) l'État doit garantir la plus grande liberté à toutes et à tous. Ensuite, en 2a) l'État doit aussi garantir à toutes et à tous l'égalité des chances et des opportunités et en 2b), l'État ne doit valider des propositions que si les plus défavorisés peuvent en bénéficier aussi, même de manière moindre (c'est le principe de différence, qui permet de s'écarter de l'égalité imposée par le communisme). Le lien avec la Fraternité est que si on peut s'enrichir, on doit aussi partager son enrichissement avec les plus pauvres.

L'œuvre de John Rawls : une priorité de la justice sur l'utilité sociale

Avec son ouvrage *Théorie de la justice* (1971) [Rawls 1987], Rawls aborde les principes que devrait adopter toute société "juste" ou "bien ordonnée", la question institutionnelle venant ensuite car la liberté doit primer. Dans cette quête, l'influence décisive va être celle de l'économie du bien-être ou économie normative (*welfare economics*), selon laquelle seule une redistribution "rationnelle", c'est-à-dire qui a pour conséquence d'augmenter également le bien-être des plus démunis, est moralement défendable ou juste ¹².

Dans sa démarche pour construire cette théorie de la justice, Rawls évite donc, certains aspects qu'il trouve inacceptables de l'utilitarisme, notamment l'idée qu'il puisse être légitime de sacrifier ou de réduire les droits fondamentaux de certains dans le but d'augmenter le bien-être général. Pour Rawls, chaque personne, possède une inviolabilité fondée sur la justice que, même au nom du bien-être de l'ensemble de la société, l'on ne peut transgresser.

Cette thèse s'écarte donc de l'utilitarisme pour se rapprocher de la grande tradition européenne du contrat social (ainsi que John Locke, Rousseau et Kant l'avaient prôné mais sans pour autant en épouser véritablement l'esprit) qui garantit la souveraineté du peuple après avoir assuré les libertés individuelles.

Rawls transforme ainsi la question de la justice en une question de choix rationnel dans des conditions d'incertitude et de risque : opérer le choix des principes de justice pour gouverner la vie en commun, dans l'ignorance d'une situation particulière (appelé le voile d'ignorance), avec l'évitement de la confrontation avec ceux qui occupent la position la plus défavorisée. En définitive, selon Rawls, les meilleurs principes sont alors logiquement ceux qui protègent les droits fondamentaux de tous, de manière égale (premier principe de justice) et qui n'autorisent des inégalités économiques et sociales que si elles bénéficient aussi aux plus désavantagés (second principe, dit "principe de différence"). Ainsi donc, équité et justice constituent les deux volets pour une prise en considération des plus désavantagés dans une société démocratique ¹².

D'après [Boyer 2003], le livre de John Rawls *The theory of justice* paru en 1971, bien que rapidement étiqueté "social-démocrate" alors qu'il ne s'inscrivait pas dans la logique marxiste de lutte des classes, devait se révéler le livre de philosophie le plus diffusé, le plus cité, le plus commenté du 20ème siècle, quelles que soient la catégorie et la langue utilisée.

Ce livre est d'abord une tentative délibérée de produire une théorie "non utilitariste" de la justice sociale. Au-delà de cette tentative de justification d'une conception politique de la justice, valable dans une société démocratique, il est par ailleurs une tentative de clarifier les idées parues dans une compilation de divers articles publiés dans les années 1960 et 1970 par Rawls lui-même [Cotelette 2009].

¹²<https://www.universalis.fr/encyclopedie/john-rawls/>

L'utilitarisme de Jeremy Bentham (inspiré de Hume, de Beccaria et d'Helvétius), est une philosophie du siècle des Lumières, le siècle de la réhabilitation du bonheur. Elle repose sur une idée simple : tout être sensible cherche à éviter le plus possible la douleur et à ressentir le plus possible de bien-être. Pour eux, une société juste ou bonne sera une société qui maximise la quantité totale, la somme algébrique du bien-être (les plaisirs moins les peines), ou encore de "l'utilité" sociale [Boyer 2003].

Rawls, d'après Boyer, a proposé une démarche permettant de surmonter les antagonismes par le recours à la raison en conjuguant de manière originale et grâce à un système cohérent, les vertus du libéralisme politique et un très audacieux objectif : l'égalité sociale.

Pour Rawls, les démocraties sont fragiles, car elles reposent sur un consensus sans cesse menacé : elles admettent en effet la pluralité des systèmes de valeurs. A l'inverse de la théorie de la justice, les totalitarismes d'hier et les régimes libéraux d'aujourd'hui promettent de restaurer cette cohésion et cette stabilité perdue en combattant les options minoritaires et en tentant de réimposer une conception du bien.

Contrairement au monde européen dominé par le contractualisme, le monde anglo-saxon (Grande-Bretagne, États-Unis) a toujours été dominé depuis par l'utilitarisme, appuyé par le libéralisme radical (que les américains appellent "laisser faire").

Mais la théorie de Rawls a aussi ses détracteurs. [Foisneau 2014], dans "Rawls et la justification de la règle de majorité", précise que :

La Théorie de la justice fait une place à la règle de majorité : cette règle étant à la fois une source de légitimité en régime démocratique et une source d'erreurs, Rawls lui fait l'honneur ambigu de justifier un devoir d'obéissance à des lois injustes. Il ne suffit pas, en effet, qu'une loi soit contraire à notre sens de la justice pour qu'aussitôt elle perde toute force d'obligation : nous avons aussi un "devoir de civilité" qui nous impose d'accepter, dans une certaine mesure, les défauts de nos institutions.

L'auteur propose une réflexion sur les relations entre théorie de la justice et gouvernement par la majorité.

2.4.2 Éthique du Care de Carol Gilligan (1936—)

D'après [Zielinski 2010], "la notion de *care* a surgi sur la place publique en France suite à une déclaration de Martine Aubry" : "Il faut passer d'une société individualiste à une société du *care*, selon le mot anglais que l'on pourrait traduire par "le soin mutuel". Mais d'après l'auteur, "Le *care* semble peut-être une idée neuve en Europe, mais la — ou les— philosophie(s) du *care* ont déjà une histoire riche dans le monde anglo-saxon, notamment aux États-Unis."

Le terme n'est pas toujours traduit, et cela s'explique par sa richesse sémantique. Il correspond en français au mot "cure" d'où provient le mot "curé" : un religieux qui

s'occupe des âmes de ses paroissiens. C'est l'origine aussi de l'expression "ce n'est pas une sinecure". En effet, la "sinecure" correspondait au travail de certains religieux, qui n'avaient pas la charge d'âme, et effectuaient un travail administratif.

Mais le sens de "charge d'âme" de la cure étant aujourd'hui perdu en français courant, il faudra se contenter de ses équivalents en français contemporain, comme : prendre soin, donner de l'attention, manifester de la sollicitude.

L'Américaine Carol Gilligan, professeur de philosophie et psychologue, est la porte-parole la plus populaire de l'éthique du *care*. Au-delà de la philosophie, La pensée de Gilligan est fondée sur une logique psychologique. Elle a développé une approche de l'éthique sociale et connectiviste, construite autour de l'humain et pour les humains ; les personnes sont définies comme des êtres sensibles, relationnels et connectés : "Nous sommes tous connectés et devons agir en tant que tels. Écoutez-vous les uns les autres. Prenez soin les uns des autres."

Dans ce domaine de l'éthique du soin, chaque personne doit être écoutée attentivement et entendu, avec respect. Tout en insistant sur l'importance de la réactivité dans les relations (être attentif, écouter, répondre), Carol Gilligan met en garde sur le risque de perdre la connexion avec soi-même ou avec les autres.

2.4.3 Éthique du Discours de Jürgen Habermas (1929—)

Jürgen Habermas est enseignant de philosophie aux universités de Heidelberg et de Francfort. Il est surtout l'auteur de nombreux ouvrages : "Théorie de l'agir communicationnel" (Fayard, 1987) ; "Morale et Communication" (Le Cerf, 1987) ; "Discours philosophique de la modernité" (Gallimard, 1988) ; et "De l'éthique de la discussion" (Le Cerf, 1992).

D'après *The House of Ethics*⁶, dans les années 1980, l'éthique est devenue une "action communicative" où les gens ont besoin de débattre de questions éthiques. En effet, la société est passée d'une société de produits / services à une société de l'information / de la communication.

L'éthique du discours repose sur la prémisse qu'un consensus sera trouvé. C'est l'un des principes de la théorie de Habermas où les esprits divergent.

Cette éthique du discours a souvent été qualifiée de "monde de communication idéal". Pour Habermas, les facteurs importants sont la bonne volonté et l'intelligence sociale.

Dans le langage d'Habermas, "les problèmes proprement éthiques se rapportent aux choix axiologiques préférentiels de chacun, choix fondamentalement subjectifs". L'éthique, au sens étroit du terme, ne concerne pas entièrement le questionnement rationnel. Quant à la morale, elle "correspond à une perspective universaliste, dépassant les bornes de toute culture donnée"¹³.

¹³<https://www.cairn.info/la-pensee-ethique-contemporaine--9782130592419-page-25.htm>

D’après ¹⁴, “l’éthique en matière de discussion, de communication et de débat est une réflexion sur les conditions de possibilités minimales de compréhension mutuelle des hommes en situation d’échange verbal. Théorisée en particulier par Jürgen Habermas et Karl-Otto Apel, elle a pour but de formuler les normes qui doivent permettre à un débat de se dérouler de manière satisfaisante et d’établir si possible les fondements de ces normes.”

2.4.4 Éthique des données (Data Ethics) de Luciano Floridi

Professeur de philosophie et d’éthique de l’information à l’université d’Oxford, où il dirige le Digital Ethics Lab, président du Data Ethics Group de l’Alan Turing Institute, (l’institut britannique pour la science des données), Luciano Floridi, dans son ouvrage “L’éthique de l’intelligence artificielle: Principes, défis et opportunités”, publié en août 2023 [Floridi 2023], invite le monde à un effort d’intelligence collective pour gérer, au mieux, les transformations technologiques en cours, imputées à l’intelligence artificielle. Il conçoit l’“éthique des données” dans toute sa dimension humaine et sociale.

D’après *The House of Ethics* [Anscombe 2020] :

Le concept central de sa théorie est l’infosphère: c’est le tout dans lequel nous évoluons. Tout le monde est acteur. Tout le monde est responsable. Et tout est donnée, pour Floridi. Cette interconnectivité est basée sur le respect. Et celui qui ne prend pas soin de l’infosphère serait sanctionné.

Pour Floridi, l’éthique des données a pour principale vocation de se concentrer sur les problèmes liés aux données (à savoir : leur génération, leur enregistrement, leur conservation, leur traitement, leur diffusion, leur partage et leur utilisation). Elle se préoccupe également des algorithmes (les agents artificiels, l’apprentissage automatique et les robots) et enfin l’innovation responsable, la programmation, le codage professionnel, la formulation des bonnes valeurs et des bonnes conduites, etc.

Au-delà des aspects spécifiques liés à la technologie (ordinateurs, tablettes, téléphones mobiles, plates-formes en ligne, cloud computing, etc.), ce qui compte surtout, c’est le choix et l’orientation correcte des stratégies éthiques.

Au final, comme le précise [Floridi 2016], “. . . les enjeux éthiques posés par la science des données peuvent être cartographiés dans l’espace conceptuel délimité par trois axes de recherche : l’éthique des données, l’éthique des algorithmes et l’éthique des pratiques.”

¹⁴https://fr.wikipedia.org/wiki/%C3%89thique_de_la_discussion

2.5 Philosophie “occidentale”

La philosophie occidentale désigne la pensée philosophique et son élaboration en Occident. Le terme est récent et est inventé pour désigner la pensée philosophique de la civilisation occidentale depuis ses racines grecques, en Grèce antique (voir philosophie antique) et couvrant éventuellement une grande partie du globe incluant l'Amérique du Nord, l'Amérique du Sud et l'Australie. Le mot “philosophie” est lui-même originaire de la Grèce antique : *philosophia*, littéralement, “amour” ou “recherche de la sagesse”, aussi dans le sens de la connaissance, du savoir ordonné et éthique.

2.5.1 Philosophie analytique

C'est un courant philosophique moderne qui a émergé vers la fin du XIXe siècle, grâce aux travaux de Gottlob Frege et Bertrand Russell, en utilisant la logique contemporaine pour analyser et clarifier les questions philosophiques. Ludwig Wittgenstein et son œuvre “Tractatus logico-philosophicus” ont également joué un rôle crucial en définissant la philosophie comme une clarification logique des pensées¹⁵.

Les domaines initiaux de la philosophie analytique incluent la logique, la philosophie du langage et la philosophie des sciences. Ces dernières décennies, des champs supplémentaires comme les sciences cognitives, la philosophie de l'action, la philosophie de l'esprit, la théorie de la décision et la métaphysique ont été intégrés, illustrant l'extension de la philosophie analytique à presque tous les domaines philosophiques traditionnels.

Principalement pratiquée dans le monde anglophone, la philosophie analytique est moins présente en France ou en Allemagne, où la philosophie continentale prédomine. Jules Vuillemin est l'un des premiers philosophes français à s'intéresser à cette approche, influençant ainsi les traditions académiques françaises¹⁵.

Origines et Opposition : historiquement, la philosophie analytique s'est opposée à l'hégélianisme et à l'idéalisme allemand dominants après Kant. Bertrand Russell, inspiré par Frege et l'empirisme britannique, a été une figure clé dans cette opposition. Une autre figure majeure, Wittgenstein, a influencé le Cercle de Vienne, qui a donné naissance au positivisme logique. Ce dernier distinguait les énoncés analytiques (vrais par définition) et les énoncés synthétiques (vérifiables empiriquement), rejetant les énoncés métaphysiques comme dépourvus de sens.

Évolution et Diversification : après le déclin du positivisme logique, la philosophie analytique a évolué vers une plus grande diversité de sujets, incluant une métaphysique analytique avec des philosophes comme Saul Kripke et David Lewis, et une philosophie politique et morale analytique avec des figures comme John Rawls.

¹⁵https://fr.wikipedia.org/wiki/Philosophie_analytique

Approche et Méthode : La philosophie analytique se distingue par une méthode rigoureuse d’analyse du langage, initiée par Frege et poursuivie par Russell et Wittgenstein. Cette méthode vise à clarifier les énoncés en éliminant les ambiguïtés logiques. La philosophie du langage ordinaire, un sous-courant, s’oppose cependant à ce formalisme en se concentrant sur l’usage du langage quotidien.

Importance et Influence : Le but de la philosophie analytique est de rendre la philosophie plus précise et scientifique, favorisant la résolution claire des problèmes philosophiques. Bien que critiquée pour sa réduction logique simpliste, elle reste influente dans le monde académique anglophone.

2.5.2 Éthique de l’intérêt

Si l’éthique anglosaxonne utilitariste de Bentham est fondée sur la notion de maximisation du bien et du plaisir, cela signifie qu’il faut savoir ce qu’est le bien. Pour Bentham, cela signifie que l’entité qui recherche son bien doit avoir la capacité de souffrir et d’éprouver du plaisir. Ainsi, pour Bentham, les inanimés ne souffrant ni ne ressentant du plaisir, ils sortent du champ de l’éthique utilitariste. En revanche, cela inclut les humains et les animaux, qui, en cherchant à maximiser leur plaisir et minimiser leur souffrance, chercheront à maximiser leur intérêt.

Ainsi, dans l’utilitarisme, ce qui fonde le droit est l’intérêt. Le plaisir et la souffrance sont les 2 indicateurs infaillibles des actions. L’utilité est le seul critère de la moralité. Elle est comprise comme ce qui contribue au bonheur. Il faut avoir un maximum de plaisir et un minimum de déplaisir. Le bonheur collectif est constitué par l’addition des bonheurs individuels, d’où la nécessité d’obtenir le plus grand bonheur possible pour le plus grand nombre de personnes possibles.

Pour les utilitaristes, toute notre vie est dominée par notre quête du bonheur et de nos intérêts.

Mais cela va à l’encontre des contractualistes (Locke, Rousseau, Kant) pour qui c’est l’inverse. Pour eux, l’action morale est désintéressée. L’être humain a ses droits indépendamment de la communauté dans laquelle il a des intérêts. La morale des contractualistes est donc une morale de la liberté et pas du bonheur (reprise par Rawls). Le bonheur est secondaire par rapport à la liberté (qui est le principe que Rawls met tout en haut de la liste). Pour les contractualistes, la liberté (le libre arbitre), c’est la faculté à s’arracher à ses intérêts pour agir de manière désintéressée. La liberté est la faculté d’arrachement à la nature.

[Gaston 2002] explore le concept de “désintéressement” dans l’œuvre de Jacques Derrida, remettant en question son application et son importance dans le cadre de la philosophie occidentale. Derrida a souvent critiqué l’idée traditionnelle de désintéressement, la reliant à une économie cachée d’intérêts, comme on le voit dans ses analyses de Kant, où il considère le plaisir désintéressé comme influencé par des forces internes,

une hétéroaffectation intrinsèque à l'autoaffectation. Derrida s'inquiète du possible aveuglement derrière les proclamations de désintéressement et dévoile les enjeux économiques et d'investissement derrière elles.

De plus, l'article pose la question de savoir si un désintéressement radical, qui serait plus qu'une économie d'intérêts, est possible, ou s'il s'agit d'une stratégie de répétition ou de rupture. Derrida propose une déconstruction du concept sans changer de terrain, plutôt qu'une rupture absolue.

Pour Derrida, même une décision soi-disant désintéressée reste prise dans un réseau d'intérêts, tout en reconnaissant l'importance de l'impossible et de l'incalculable dans la prise de décision éthique. L'intérêt de Derrida réside donc dans une expérience de l'impossible, redéfinissant le désintéressement non pas comme une absence totale d'intérêt, mais comme une notion demeurant au cœur d'une économie de l'intérêt.

Dans son sens contemporain, la philosophie occidentale désigne donc deux traditions contemporaines de la philosophie : la philosophie analytique (utilisant des formules pour calculer le plus grand bien et fondée sur l'intérêt) et la philosophie continentale (contractualiste et fondée sur le désintéret)¹⁶.

2.5.3 Philosophie continentale

Cette appellation a été utilisée à l'origine par les philosophes anglophones, surtout ceux affiliés à la philosophie analytique, pour désigner diverses traditions philosophiques provenant d'Europe continentale, principalement d'Allemagne et de France. En général, ce terme s'oppose à la tradition britannique mais exclut la philosophie asiatique, se focalisant exclusivement sur la philosophie occidentale¹⁷.

Usage et popularité : Le terme est souvent employé par des philosophes non-européens et dans les départements d'humanités pour marquer un contraste avec la philosophie analytique. La philosophie continentale est souvent plus populaire dans les sciences sociales, l'esthétique, la littérature, le cinéma, et l'architecture.

Nature classificatoire ou polémique : Historiquement, le concept de "philosophie continentale" a été créé avec une connotation négative pour regrouper tout ce qui n'appartient pas à la philosophie analytique. Contrairement à la philosophie analytique, peu de philosophes se revendiquent explicitement de la "philosophie continentale". Ce terme est ainsi plus un outil de classement que de caractérisation positive.

Il existe donc une tension persistante entre les traditions analytique et continentale, notamment en France où la philosophie analytique peine à s'imposer dans les institutions académiques dominées par l'exégèse historique.

Les débats entre ces deux courants sont fréquemment animés, chaque camp reprochant à l'autre un manque de profondeur critique ou de rigueur méthodologique.

¹⁶https://fr.wikipedia.org/wiki/Philosophie_occidentale

¹⁷https://fr.wikipedia.org/wiki/Philosophie_continentale

Certains pensent qu’Alfred North Whitehead représente une voie de réconciliation entre les deux traditions.

[Schroeder 2013] explore les caractéristiques clés de l’éthique continentale à travers l’œuvre de plusieurs penseurs créatifs : Hegel, Nietzsche, Scheler, Sartre, et Levinas. Il présente les innovations majeures de ces philosophes, leurs divergences, et leurs défis aux théories éthiques traditionnelles.

Éthique continentale selon Hegel, Nietzsche, Sartre et Levinas

Georg Wilhelm Friedrich Hegel (1770 - 1831) est un philosophe allemand appartenant au courant de l’idéalisme allemand qui a eu une grande influence sur la philosophie contemporaine. Dans sa “phénoménologie de l’esprit”, il propose :

- Une critique de Kant : Hegel s’oppose au formalisme et à l’abstraction de l’éthique kantienne. Il rejette l’idée que les maximes éthiques puissent être testées par une seule formule comme l’Impératif Catégorique.
- Une reconnaissance réciproque : la liberté hégélienne est réalisée à travers la reconnaissance réciproque, où chaque individu reconnaît l’autre comme un être conscient et vivant. Cette reconnaissance mutuelle permet de surmonter les luttes de domination et de contribuer à une harmonie sociale plus profonde.
- la notion de Liberté et de Propriété : Hegel voit la propriété comme essentielle à la liberté personnelle, nécessitant des structures institutionnelles complexes pour soutenir les relations de propriété et permettre l’action autonome.

De son côté, Friedrich Nietzsche (1844-1900) apporte :

- Une critique de la morale : Nietzsche critique sévèrement les valeurs morales traditionnelles, notamment celles du christianisme et de ses héritiers des Lumières. Selon lui, la moralité traditionnelle étouffe la passion, l’estime de soi et la créativité, rendant les individus faibles et dépendants.
- Une revalorisation des valeurs : Nietzsche propose de revaloriser les valeurs en fonction de leur capacité à favoriser l’épanouissement humain. Il valorise la force, la santé, la créativité et la joie de vivre.
- Une notion de l’amoralité et de la personnalité : Nietzsche se décrit comme un “immoraliste”, rejetant les systèmes de devoirs et d’obligations. Son éthique constructive cherche à sublimer les passions humaines pour atteindre une nouvelle humanité, plus riche et plus forte.

- Un modèle de l'amitié : Nietzsche propose un modèle d'amitié sincère où un ami inspire sans inhiber le développement personnel de l'autre, en contraste avec les relations amoureuses souvent idéalisées et concomitantes d'attentes irréalistes.

En France, Jean-Paul Sartre (1905-1980) propose :

- La liberté radicale : Sartre place la liberté au centre de son éthique. Selon lui, la conscience transcende toujours sa situation, et les individus sont responsables de leurs choix et des valeurs qu'ils embrassent.
- La notion d'authenticité : L'authenticité consiste à reconnaître et accepter sa liberté et sa responsabilité, rejetant les subterfuges de l'auto-illusion.
- Une critique de la moralité bourgeoise : Sartre considère la moralité bourgeoise traditionnelle comme dépassée. Il cherche à développer une éthique basée sur la réciprocité et la reconnaissance mutuelle de la liberté.
- Les notions de praxis et d'aliénation : Dans sa phase ultérieure, Sartre souligne que pour surmonter l'aliénation, des transformations sociales profondes sont nécessaires et que la liberté individuelle est intimement liée à des conditions matérielles et historiques.

Plus proche de nous, Emmanuel Levinas (1906-1995) développe :

- Une éthique de l'autre : Levinas fonde son éthique sur la relation fondamentale avec l'autre, qu'il considère comme transcendant et inassimilable. L'éthique commence avec la responsabilité envers l'autre et non par des principes abstraits.
- Une primauté de l'éthique sur l'ontologie : Contrairement à la tradition philosophique grecque, Levinas place l'éthique avant l'ontologie et l'épistémologie. Il affirme que la relation éthique avec l'autre défie la possession et l'assimilation.
- La notion de "visage de l'autre" : Le visage de l'autre impose une obligation éthique immédiate et sans limite, transcendant les concepts de connaissance et de présence.
- La notion de responsabilité excessive : Levinas envisage la responsabilité non seulement comme une réponse à l'autre, mais comme une charge d'assumer également les actions et les responsabilités de l'autre.

Comme on le voit, il se creuse un fossé entre les philosophies analytiques (fondées sur la maximisation du bien et de l'intérêt) les philosophies continentales, fondées sur la liberté et le désintérêt.

La séparation entre les deux courants est souvent expliquée par des différences méthodologiques et géographiques et des discussions apparaissent sur les raisons de

la division entre la philosophie analytique et la philosophie continentale et pourquoi les métaphysiciens américains ne prêtent souvent pas attention aux œuvres de Martin Heidegger, un autre philosophe allemand majeur du XXe siècle ¹⁸.

Les philosophes analytiques se concentrent sur l’analyse logique des concepts, alors que les philosophes continentaux, tels que Heidegger et Derrida, privilégient l’expérience immédiate et les questions existentielles.

Au début du XXe siècle, les philosophes britanniques (Russell, Moore) et allemands/autrichiens (Carnap, Reichenbach) ont développé l’approche analytique en utilisant la logique symbolique pour résoudre les problèmes philosophiques. Par contraste, Edmund Husserl a développé la phénoménologie, une approche rigoureuse de la description de l’expérience immédiate, influençant Heidegger et les existentialistes comme Sartre et Merleau-Ponty.

La distinction analytique-continentale a été renforcée lorsque les philosophes américains partisans de la phénoménologie et de l’existentialisme ont créé en 1962 la Society for Phenomenology and Existential Philosophy (SPEP), en contraste avec l’American Philosophical Association (APA) dominée par les analytiques.

Avec le temps, certains spécialistes comme Brian Leiter affirment que les différences substantielles entre la philosophie analytique et continentale sont désormais minimales, bien que des différences de style persistent. Cependant, il y a encore des différences philosophiques importantes, notamment dans la conception de l’expérience et de la raison.

Les philosophies continentales tendent à explorer une expérience privilégiée et à défier la logique classique pour atteindre des significations plus profondes. Par exemple, la phénoménologie explore les certitudes immédiates de l’expérience quotidienne, tandis que les philosophies de l’imagination (Foucault, Derrida) cherchent à penser au-delà des concepts établis.

Pour combler la séparation entre philosophie analytique et continentale, une meilleure connaissance mutuelle serait bénéfique. Bien que les philosophes analytiques puissent profiter des idées continentales, ils préfèrent souvent les sources secondaires en raison de la difficulté inutile des textes originaux continentaux. Une écriture plus claire des philosophes continentaux faciliterait le dialogue philosophique et l’unification des deux traditions.

2.5.4 Éthique contractualiste, éthique du désintérêt

Après avoir consacré une section à l’éthique de l’intérêt, liée à l’éthique analytique, regardons la notion d’éthique du désintérêt, liée à l’éthique contractualiste.

¹⁸<https://archive.nytimes.com/opinionator.blogs.nytimes.com/2012/02/19/bridging-the-analytic-continental-divide/>

Le contractualisme, ou théorie du contrat social, est une école de pensée en philosophie politique datant du XVIIe siècle, qui conceptualise l'origine de la société et de l'État comme étant issu d'un contrat hypothétique entre individus. Par ce contrat, les individus acceptent de limiter leur liberté en échange de lois permettant de maintenir le corps social ¹⁹.

Le contractualisme réunit des théories expliquant comment une société se forme par la signature, par chaque individu, d'un contrat qui s'impose à tous. Cela implique une transition de l'état de nature, hypothétique et sans lois, à l'état de société structuré par le contrat. Cette théorie a joué un rôle majeur dans des événements politiques comme la révolution française et la révolution américaine. Elle s'est opposée aux sociétés d'ordres et aux corporations de l'Ancien Régime.

Les concepts du contrat social contrastent avec la philosophie politique classique, qui voyait la Cité comme un fait de nature. Platon et Aristote décrivaient une société fondée sur des dispositions naturelles plutôt qu'un contrat. Épicure, cependant, a été un précurseur du contrat social en affirmant que la justice est fondée sur un accord entre les hommes.

Au Moyen Âge, l'organisation politique et sociale occidentale était inspirée par la théologie chrétienne. Grotius a été le premier à théoriser un contrat social moderne, suivi par des penseurs comme Hobbes, Locke et Rousseau.

2.5.4.1 Hobbes (1588 - 1679) et la Sécurité

Thomas Hobbes, dans "Le Léviathan" (1651), propose un contrat de sécurité en réponse à l'état de nature qu'il décrit comme une guerre de tous contre tous. Pour sécuriser leur existence, les individus délèguent leurs libertés à un Léviathan (l'État). Hobbes permet un droit de résistance si l'État met en péril la vie des individus.

Jurieu et le Gouvernement : Pierre Jurieu voit le lien entre le souverain et le peuple comme contractuel pour assurer la conservation des biens et libertés. Il introduit l'idée de souveraineté populaire, affirmant que le droit de résistance est légitime en cas d'abus de pouvoir.

2.5.4.2 John Locke (1632 - 1704) la Liberté et Propriété

John Locke, dans son "Second Traité du gouvernement civil" [Locke 1992], voit le contrat social comme garantissant la liberté individuelle et la propriété privée. L'État est instauré pour protéger ces droits naturels. Locke prévoit aussi un droit de résistance contre les abus de l'État.

¹⁹<https://fr.wikipedia.org/wiki/Contractualisme>

2.5.4.3 Jean-Jacques Rousseau (1712 - 1778) et la Liberté

Jean-Jacques Rousseau, dans “Du Contrat social” [Rousseau 1896], vise à rendre le peuple souverain en suivant l’intérêt général. Le contrat social engage les citoyens à sauvegarder par eux-mêmes leur sécurité et liberté par le principe de la volonté générale.

Hegel critique le contractualisme comme une contradiction, disant que l’État ne peut reposer sur des relations contractuelles qui sont une propriété extérieure. L’Église catholique critique le droit absolu de propriété selon Locke pour son incompatibilité avec la destination universelle des biens créés par Dieu.

2.5.5 Conclusion sur les éthiques occidentales

Comme on le voit, elles sont loin d’être concordantes : elles comportent de nombreux courants de pensée, dont certains sont opposés.

C’est important car dans le cadre de cette thèse, dont le titre est “IA et éthique pour l’amélioration du service public”, faut-il que les suggestions faites par l’IA maximisent l’intérêt de tous, avec une vision utilitariste du monde, ou au contraire, doivent-elles mettre en avant le contractualisme ? Une IA gouvernementale devra-t-elle être intransigeante (déontologiste) ou au contraire, devra-t-elle évaluer et tenir compte des conséquences des propositions suggérées ?

Une IA pourrait-elle implémenter une éthique moderne, globale, et convenant à tous ?

2.6 Méta-éthiques universalistes

2.6.1 Introduction à l’utilité d’une éthique universelle

[Kim 2012] traite de la nécessité de développer une éthique globale face aux défis de la mondialisation, qui intensifie les interactions entre diverses cultures, créant autant d’opportunités que de conflits. La mondialisation, qui intensifie les échanges de biens, personnes et idées, génère confusion et tension, tout en mettant en lumière l’urgence d’un régime de paix et de coexistence visant à transcender les divergences culturelles.

La mise en place d’un gouvernement mondial, bien que lointaine et potentiellement indésirable, est évoquée comme une solution pour harmoniser les diversités culturelles, mais le texte insiste davantage sur l’élaboration d’une éthique universelle. Les intellectuels et penseurs globalement organisés y ont un rôle crucial à jouer en définissant des règles éthiques minimales, indépendantes de toute action politique directe, pour assurer la coexistence pacifique des peuples et cultures.

[Kim 2012] signale que l’éthique est intimement liée à la culture, représentant plus qu’un ensemble de règles, mais bien l’ethos d’une société. Alors qu’une éthique universelle semble nécessaire pour organiser le monde comme un espace vivable et protéger

les droits humains fondamentaux, la diversité des valeurs culturelles doit être respectée et intégrée. Cette reconnaissance des valeurs partagées et des différences peut servir de point d'ancrage à une éthique universelle plus complexe, capable de transcender les tensions interculturelles par des efforts concertés et une coopération internationale.

De plus, le texte explore la notion de multiculturalisme, indispensable, pour répondre aux défis posés par l'intégration des migrants et la diversité culturelle accrue. Il met en avant l'importance des politiques de reconnaissance et d'égalité des identités culturelles pour bâtir une communauté humaine à l'échelle planétaire. Une éthique universelle, malgré ses racines socioculturelles variées, doit concilier les profondes différences pour établir un cadre de coexistence harmonieuse.

2.6.2 Claude Lévi-Strauss (1908 - 2009) et Jürgen Habermas (1929—), pour une éthique universelle

Le multiculturalisme et la démocratie peuvent coexister grâce à la constitutionnalité démocratique, qui offre un cadre politique et juridique permettant l'intégration de divers groupes culturels sans nécessairement les forcer à l'assimilation totale.

Claude Lévi-Strauss²⁰ exprime son scepticisme envers une civilisation mondiale homogène, soulignant l'importance de la diversité culturelle pour enrichir l'expérience humaine. Il valorise les sociétés qu'il appelle "primitives" pour leur équilibre écologique et leur préservation des dimensions humaines. Ces communautés, bien que petites, offrent des modèles de relations authentiques et des réponses durables aux problèmes humains [Kim 2012].

Pour Lévi-Strauss, l'humanité se décline au pluriel. Il propose un "troisième humanisme", inspiré par les leçons des sociétés primitives, pour guider l'humanité vers une civilisation plus sage, écologique et harmonieuse. Sa vision souligne l'importance d'intégrer la sagesse culturelle du passé pour un avenir meilleur, tout en restant à distance des idéologies politiques centrées sur le progrès moderne. Il défend l'égalité de toutes les cultures, et lutte contre l'idée d'un progrès social unique, "colonialiste".

L'approche de Claude Lévi-Strauss envers les sociétés primitives révèle les tensions entre la facticité de ces cultures et le cadre théorique élargi nécessaire pour les comprendre. Ses travaux suggèrent que l'étude des sociétés primitives nécessite un respect de leur subjectivité tout en les intégrant dans une compréhension globale. L'idée d'une culture mondiale future doit équilibrer les racines locales et une ouverture universelle. Lévi-Strauss insiste sur l'importance des petites communautés, tout en reconnaissant leur place dans un monde plus vaste.

L'analogie astronomique développée souligne que, tout comme la Terre est ouverte à l'univers tout en restant stable, une culture mondiale doit être perméable tout en préservant sa cohérence locale. La vitesse finie de la lumière, métaphore des échanges

²⁰https://fr.wikipedia.org/wiki/Claude_L%C3%A9vi-Strauss

culturels, permet une organisation et un tri de l'information. Cela reflète la nécessité d'une raison qui intègre les cultures à un rythme qui permet leur réel dialogue et influence mutuelle. Au cœur de cette intégration réside l'objectif d'atteindre un équilibre où éthique et ontologie convergent, respectant la diversité mais maintenant un ordre global.

Jürgen Habermas affirme aussi que la démocratie ne doit privilégier aucune culture ou valeur spécifique pour protéger l'autonomie des individus. Selon lui, les immigrants doivent accepter les principes démocratiques, mais non être forcés d'abandonner leur culture. Habermas, très influent sur la scène politique européenne [Lemasson 2008]. Il est l'auteur de "L'espace public" [Habermas 1978], dont il ne donne pas de définition univoque, mais qui est généralement défini comme étant le lieu, physique ou symbolique, dans lequel les idées circulent et sont discutées de manière rationnelle afin de cristalliser en opinion publique [Ballarini 2015].

Pour lui, la démocratie est envisagée non comme un système rigide, mais comme un processus dynamique qui évolue grâce aux mouvements sociaux et aux luttes politiques, en nécessitant une constante réévaluation pour répondre aux besoins des plus défavorisés. Habermas reconnaît que la démocratie est teintée par une tradition éthique, soulignant l'importance d'une interprétation culturelle évolutive [Kim 2012]. Cette interprétation, ou herméneutique, permet de préserver et de transformer les héritages culturels, englobant les minorités culturelles et nationales dans un dialogue plus global. Habermas encourage une attitude réflexive pour naviguer les changements multiculturels, mais ceci requiert aussi de la part des cultures non occidentales une acceptation du dialogue critique, posant néanmoins la question de potentiel impérialisme occidental.

La simplification de la nature humaine dans les débats sur l'intégration culturelle met en lumière les contradictions entre loi, éthique et culture. Habermas propose l'universalisme des principes juridiques pour faciliter l'intégration multiculturelle, mais cela peut réduire l'humain à une vision unidimensionnelle, centrée sur l'homo economicus et l'intérêt personnel. Ce modèle legaliste peut affaiblir la sensibilité éthique et culturelle, et compartimenter la sphère publique et privée de manière appauvrissante. Habermas, bien qu'engagé dans la discussion éthique, fonde ses arguments dans une logique réaliste, notamment en différenciant les réfugiés politiques et économiques. Il défend les premiers par obligation légale et morale, et les seconds par un sens de justice historique, bien que l'argument humanitaire soit moins présent. Cette approche réaliste limite le débat éthique à l'avantage personnel, risquant ainsi de négliger des valeurs comme la compassion dans l'espace politique [Kim 2012].

La réponse aux défis multiculturels réside dans des mesures pratiques et une universalité légale pour gérer les conflits culturels. La culture mondiale ne doit pas chercher l'homogénéisation, mais embrasser la diversité des cultures locales. L'intégration globale nécessite de trouver un équilibre entre modernité et prémodernité, où la rationalité scientifique joue un rôle clé. Cependant, cette approche soulève des

questions de pouvoir et d'influence.

Quelques exemples d'obstacles possibles à une éthique universelle : Selon [Kim 2012], l'évolution culturelle résulte souvent des interactions entre sociétés diverses, processus accéléré aujourd'hui par la mondialisation. Historiquement, cet échange enrichit les civilisations, comme le montrent les flux migratoires et le commerce, transformant les cultures à travers l'intégration de nouvelles idées et croyances, illustré par l'introduction du bouddhisme et du confucianisme en Corée.

Dans l'ère moderne, les échanges culturels s'intensifient avec l'industrialisation et le capitalisme global, provoquant des bouleversements sociaux et culturels profonds, entraînant uniformisation et tensions. Malgré l'apparente uniformité culturelle mondiale actuelle, il manque un véritable processus organique d'intégration entre cultures, car la rapidité des échanges entrave la fusion culturelle mutuelle authentique.

Les pays industriellement développés ont attiré des travailleurs migrants, créant des défis d'intégration. Le multiculturalisme tente de gérer la diversité des cultures cohabitant dans ces pays. En Europe, l'échec de cette approche est souvent attribué au refus d'assimilation des migrants, alors que la reconnaissance des identités culturelles minoritaires devient cruciale pour la dignité et l'intégration sociale, comme l'illustre l'exemple des Noirs américains ou des Québécois.

En Corée, le multiculturalisme est lié aux migrations contemporaines et au nombre croissant de mariages internationaux. Ces unions créent des défis d'adaptation culturelle, exacerbés par la culture patriarcale coréenne. Les mesures pour mieux intégrer les migrants incluent la formation de réseaux de soutien et l'encouragement à la fierté culturelle réciproque. L'éthique confucéenne, valorisant l'hospitalité, modère ces interactions et guide les solutions multiculturelles. La Corée démontre ainsi un mélange d'adaptations personnelles et d'actions collectives pour accueillir les divers groupes culturels, soulignant l'importance de l'hospitalité comme vertu à encourager universellement.

Enfin, pour réaliser l'objectif noble d'une éthique universelle, force est de reconnaître, ce qui suit, d'après [Bonino 2022] :

Pour fonder une éthique des droits de l'homme qui soit acceptée universellement et fondée dans la vérité objective de l'humain, on ne peut s'en tenir au positivisme des législateurs, porteur de contradictions et donc d'inefficacité évidente. Le pape Jean-Paul II a constamment promu le renouvellement de la doctrine de la loi naturelle, avec ses valeurs fondamentales antérieures à tout contrat social, faute de quoi un totalitarisme insensible peut disposer de la dignité des personnes, en commençant par les plus faibles.

Ces exemples d'obstacles et ceux non exposés ici, montrent à quel point il est nécessaire, pour l'humanité de s'armer de volonté, pour surmonter les obstacles et, surtout, de

s'appuyer sur les organisations internationales dont le rôle est, souvent, fédérateur. **Enjeu majeur et démarche pour une éthique universelle :** il est le fait de la mondialisation, caractérisée par une circulation croissante des biens, des personnes et des informations. Cette mondialisation est censée rapprocher des cultures diverses, créant à la fois des opportunités, bien que générant par ailleurs des conflits. Il est donc crucial de rechercher une éthique mondiale pour résoudre les conflits, créés et favoriser la coexistence pacifique. Un gouvernement mondial ou un ordre juridique pourrait concilier les divers éléments culturels mondiaux, bien que cela pourrait relever de l'utopie mais surtout pas de l'impossible.

Selon [Igalens 2002] l'éthique universelle, finalité humaine se concrétise désormais en moyens de "management responsable". Déjà dans son "tractatus", Wittgenstein affirmait que "l'éthique ne peut être qu'actions selon une règle universelle". On rejoint ici l'éthique d'Aristote décrite comme une discipline pratique portant sur l'action.

Ainsi, il est suggéré de développer une charte internationale d'éthique universelle en s'appuyant sur des penseurs internationaux, indépendamment des actions politiques directes. Cette éthique universelle devrait garantir un espace de vie partagé et protéger les valeurs humaines fondamentales, reconnaissant une humanité commune et instaurant des règles minimales pour la coexistence multiculturelle.

L'éthique, en tant que composante intégrale de la culture, transcende les simples règles pour devenir l'ethos général d'une société. Dans un contexte de mondialisation, il est essentiel de préserver un ordre social non coercitif basé sur la culture et la vertu plutôt que sur la loi stricte. Les pratiques comme la courtoisie et l'hospitalité servent de médiateurs dans les rencontres interculturelles. Une culture globale, construite à partir de sous-systèmes de cultures locales, pourrait émerger, avec chaque sous-système, contribuant à une dynamique évolutive globale. L'universalité, bien qu'elle puisse susciter des conflits, sert aussi de moteur d'intégration, permettant de dépasser les tensions culturelles par des échanges dialectiques. L'éthique joue un rôle clé en facilitant la reconnaissance des points communs au sein de cette diversité et en apaisant les tensions [Kim 2012].

2.6.3 Proposition d'éthique minimaliste, par Ruwen Ogien (1947 - 2017)

Ruwen Ogien a travaillé au développement d'une théorie éthique qu'il appelait "l'éthique minimaliste". Cette théorie anti-paternaliste cherche à minimiser les interventions de ce qu'il appelle, en se basant sur John Stuart Mill, la "police morale"²¹. Il a défendu le minimalisme moral, estimant que ce qui peut être légitimement interdit, c'est le préjudice causé à autrui et que, pour le reste, les individus doivent être autorisés à vivre

²¹https://en.wikipedia.org/wiki/Ruwen_Ogien

leur vie comme ils l'entendent. Il considère que l'éthique n'a de sens que lorsqu'elle est limitée [Jouan 2019].

Ogien critique vigoureusement les vertus rédemptrices souvent associées à la maladie ainsi que sa gestion par les institutions médicales. Cette critique s'inscrit dans un projet philosophique centré sur le "minimalisme moral", résumée par la formule : "Trop d'éthique tue l'éthique" [Jouan 2019].

ABC du Minimalisme Moral : le minimalisme moral d'Ogien se distingue des théories éthiques dominantes (déontologisme, et conséquentialisme) par son ambition plus modeste. Il se concentre sur le principe de non-nuisance inspiré par John Stuart Mill et le Serment d'Hippocrate. Selon ce principe, tant que les individus ne nuisent pas directement et délibérément aux autres, ils devraient être libres de faire ce qu'ils veulent de leur vie. Ogien considère ce principe comme ayant une valeur morale universelle incontestable [Jouan 2019].

L'éthique minimaliste s'articule initialement autour de trois principes ²¹ :

1. Principe de considération égalitaire : Donner la même valeur à la voix de chacun
2. Principe de neutralité : Une attitude neutre envers les conceptions du droit et de la propriété personnelle.
3. Principe d'intervention Limitée : Limiter les interventions aux cas de torts flagrants faits aux autres.

Par la suite, Ogien a simplifié cette approche à un seul principe : "Ne pas nuire aux autres, rien de plus". Voici le raisonnement qui sous-tend ce principe :

- Absence de devoir moral envers soi-même : nous n'avons pas de devoirs moraux envers nous-mêmes, seulement envers les autres.
- Devoirs moraux envers les autres : ces devoirs peuvent être positifs (aider à faire le bien) ou négatifs (ne pas nuire).
- Risques du paternalisme : l'option positive peut se traduire par un soutien charitable, qui peut mener au paternalisme —une attitude consistant à vouloir faire le bien des autres, indépendamment de leur opinion.
- Préférence pour l'option négative : pour éviter le paternalisme, il est préférable de s'en tenir à une règle négative simple. Ne pas nuire aux autres.

Défense et Critique du Minimalisme Moral : un volume édité par Roberto Merrill et Patrick Savidan, Du minimalisme moral [Merrill 2017], critique et discute ce concept. Certains contributeurs critiquent la notion de devoir envers soi, jugée incohérente et inutile pour fonder des obligations envers les autres.

D'autres défendent le minimalisme moral en tant qu'éthique expérimentale, comme illustré dans le livre d'Ogien, *L'influence de l'odeur des croissants chauds sur la bonté humaine* [Ruwen 2011].

Applications pratiques et politiques : Ogien applique le minimalisme moral aux débats publics et législatifs sur des questions controversées, telles que l'euthanasie, l'utilisation de drogues et les relations sexuelles entre adultes consentants. Son éthique vise à promouvoir la tolérance envers les actes qui ne violent pas le principe de non-nuisance.

Critique de la Confusion entre Éthique et Politique : certaines critiques accusent Ogien de confondre éthique et politique ou morale et droit. Ogien propose d'importer dans la morale les principes politiques et juridiques gouvernant les relations entre l'État et les individus dans la sphère privée. Ancrée dans la critique de l'interventionnisme d'état, cette approche prône la tolérance et la liberté individuelle.

Implications Économiques et Sociales : le minimalisme moral applique également une critique aux tentatives de justifications morales des inégalités économiques, comme celle proposée par John Rawls. Ogien soutient que la moralisation des inégalités rend moins impérieux le besoin de les réduire. Il s'agit de promouvoir une société plus libérale sur le plan des mœurs tout en étant plus égalitaire en termes économiques et sociaux.

2.7 Méta-éthiques relativistes

Comme on le voit, les méta-éthiques universalistes sont considérées comme des buts à atteindre, souvent basés sur la démocratie et la liberté. Mais force est de constater que dans le monde du premier quart du 21^e siècle, les démocraties sont en lutte contre les gouvernement autoritaristes et hiérarchiques, représentés par de grands états comme la Chine, la Russie, mais aussi l'Inde, les Émirats Arabes Unis, et possiblement les États-Unis, dans le cas d'une élection de Donald Trump.

Cela signifie que la démocratie et la liberté sont, loin d'être des principes éthiques majoritaires dans le monde de 2024.

Les méta-éthiques relativistes sont de ce fait plus réalistes, au sens où elles tentent de tenir compte des disparités existantes, en acceptant que les principes éthiques soient différents selon les cultures et les contextes politiques.

Le relativisme moral est un sujet clé en métaéthique et suscite des débats au-delà de la philosophie, notamment parmi les leaders politiques et religieux. Il est controversé parmi les philosophes et le grand public. Plusieurs philosophes contemporains défendent différentes formes de relativisme moral [Gowans 2004].

Bien que le relativisme moral soit devenu un sujet de premier plan au XX^e siècle, il a des origines anciennes. Des philosophes grecs comme Hérodote et Protagoras ont favorisé certaines formes de relativisme. Cependant, le scepticisme moral était plus

courant que le relativisme parmi les philosophes grecs anciens.

Le relativisme moral désigne diverses thèses, souvent empiriques, métaéthiques ou normatives. Le Relativisme Moral Descriptif (RMD) soutient qu'il y a des désaccords moraux profonds et répandus. Le Relativisme Moral Métaéthique (RMM) affirme que la vérité ou la justification des jugements moraux n'est pas absolue mais relative aux normes d'un groupe.

Les études expérimentales ont examiné la diversité morale, les perspectives folk de l'objectivisme moral et du relativisme, et les liens entre le relativisme moral et la tolérance. Les recherches montrent que les intuitions méta-éthiques des gens sont complexes et varient selon les contextes, ce qui pourrait indiquer une sorte de pluralisme méta-éthique [Gowans 2004].

Relativisme Moral Descriptif : le relativisme Moral Descriptif (RMD) est souvent soutenu par des études anthropologiques montrant la diversité morale à travers les cultures. Cependant, il existe des objections *a priori* et empiriques au Relativisme Moral Descriptif, notamment des arguments suggérant que des désaccords moraux présupposent un accord considérable sur d'autres sujets [Gowans 2004].

Désaccords Moraux et Rationalité : Les objectivistes pensent que la plupart des désaccords moraux peuvent être résolus rationnellement, tandis que les partisans de RMM soutiennent que beaucoup de ces désaccords sont irréductibles à cause des différences fondamentales dans les orientations morales. Les objectivistes voient souvent les désaccords comme des erreurs de faits ou de logique corrigibles par un débat rationnel [Gowans 2004].

Relativisme Moral Métaéthique : Même avec des désaccords moraux profonds et irrésolus, cela ne conduit pas directement à RMM. D'autres conclusions non objectivistes comme le scepticisme moral ou le non-cognitivism pourraient être tirées. RMM est souvent critiqué pour son incohérence et son incapacité à rendre compte de l'autorité normative des jugements moraux [Gowans 2004].

Positions Mixtes : Certaines positions mixtes combinent des éléments du relativisme et de l'objectivisme moral, suggérant que certains jugements moraux sont absolument vrais tandis que d'autres sont relatifs à la société ou au groupe. Ces positions cherchent à réconcilier les préoccupations légitimes des deux côtés du débat [Gowans 2004].

Relativisme et Tolérance : Le relativisme est parfois associé à la tolérance, la proposition selon laquelle nous devrions tolérer les opinions morales divergentes lorsqu'elles ne peuvent pas être résolues rationnellement. Cependant, le lien entre relativisme et tolérance est critiqué, car le RMM ne peut pas objectivement promouvoir la tolérance, et la tolérance relative pourrait varier d'une société à l'autre [Gowans 2004].

Comme on le voit, si une méta-éthique universelle est encore lointaine, l'acceptation des méta-éthiques relativistes n'est pas non plus consensuelle.

De l'universalisme au relativisme en éthique

[Vidal 2003] traite de la tension entre universalisme et relativisme en éthique, en utilisant l'exemple de la gestion du sida en Côte d'Ivoire pour illustrer ces concepts. L'éthique, fréquemment utilisée dans divers contextes, a une définition floue qui facilite de multiples interprétations.

Définitions de l'éthique et du relativisme éthique : l'éthique est dynamique et ouverte au débat, tandis que la morale repose souvent sur des règles fixes. Le relativisme éthique soutient que les normes morales sont contingentes et dépendantes du contexte social, culturel et historique.

Exemples illustratifs :

1. Confidentialité et protection liée au VIH :
 - L'OMS et Onusida promeuvent la confidentialité et le consentement éclairé pour freiner la propagation du VIH, mais reconnaissent des exceptions où informer le partenaire d'un patient séropositif est justifié par bienfaisance.
 - Ces exceptions soulèvent des dilemmes éthiques, notamment lorsqu'il faut choisir entre la confidentialité et la protection des partenaires, illustrant la nécessité d'adapter les principes éthiques aux réalités locales.
2. Accès aux traitements antirétroviraux (ARV) :
 - En Côte d'Ivoire, l'initiative d'Onusida pour fournir des ARV a révélé des défis pratiques et éthiques, tels que la sélection des patients basée sur leur capacité financière plutôt que des critères cliniques.
 - Ceci montre que des principes éthiques, tels que la justice et la bienfaisance, doivent être appliqués de manière pragmatique et contextuellement adaptée pour être vraiment efficaces.

L'auteur conclut que pour une éthique véritablement efficace, il est crucial de lier les principes généraux aux situations spécifiques. Une approche équilibrée, tenant compte à la fois de l'universel et du local, est essentielle dans le domaine de la santé pour concilier les valeurs éthiques des différentes cultures. Les échanges entre ces perspectives peuvent enrichir et améliorer la mise en œuvre des principes éthiques [Vidal 2003].

2.8 Philosophie et éthique africaine

2.8.1 Philosophie africaine

La philosophie africaine est un domaine en pleine évolution qui a vu le jour dans les années 1980 et qui englobe la réflexion et la pensée critique des personnes d'ascendance

africaine. Elle se concentre sur l'organisation des pratiques intellectuelles des Africains et des personnes d'ascendance africaine, y compris les traditions philosophiques, l'art et les expressions esthétiques. Malgré les difficultés, le domaine explore les concepts et les contributions philosophiques, tels que ceux de l'Afrique continentale et de la diaspora africaine, en reconnaissant à la fois les similitudes et les différences entre les divers groupes [Outlaw Jr 2010].

Les thèmes clés comprennent les luttes contre l'impérialisme européen et le racisme, les contributions de diverses disciplines telles que la littérature et la religion, et la quête permanente de liberté et de justice. Le domaine s'intéresse également aux questions de genre, en particulier aux contributions des femmes d'origine africaine. La philosophie africaine remet en question l'historiographie eurocentrique et cherche à inclure les riches traditions philosophiques, historiquement négligées, des peuples africains et d'ascendance africaine.

Les développements en cours dans ce domaine impliquent des efforts interdisciplinaires et restent engagés à encourager la démocratie intellectuelle et à aborder des questions contemporaines telles que la race, la liberté et la justice au sein de sociétés multiethniques et multiraciales.

L'adoption et l'utilisation du concept de "philosophie Africaine" ont conduit à une exploration des systèmes de pensée africains anciens et contemporains, remettant en question les récits historiques eurocentriques, mais aussi liés aux deux principales religions de l'Afrique : la religion musulmane et la religion chrétienne catholique. Les travaux de figures comme Cheikh Anta Diop et Martin Bernal ont contesté les origines gréco-romaines de la philosophie africaine, proposant plutôt des origines africaines et égyptiennes.

2.8.2 Éthique africaine

L'éthique d'une société est fondée sur les idées et croyances de ce qui est bon ou mauvais, sur les relations sociales satisfaisantes et sur les comportements favorisant l'harmonie sociale, la coopération, la justice et l'équité. Les penseurs moraux de chaque société articulent, analysent et interprètent ces concepts.

Les sociétés africaines, en tant que communautés humaines, possèdent des systèmes éthiques pour guider les comportements sociaux et moraux. Cependant, ces systèmes n'ont pas été systématiquement analysés et interprétés. Depuis environ trois décennies, des philosophes africains contemporains ont commencé à étudier en profondeur les idées morales africaines pour mieux comprendre la pensée éthique africaine.

L'éthique africaine est présentée comme ayant un caractère social et humaniste, centrée sur le devoir et le bien commun plutôt que sur les droits individuels. Les termes "éthique" et "morale" sont souvent utilisés de manière interchangeable, bien que l'éthique se réfère plus à l'étude philosophique de la morale [Gyekye 2010].

Contrairement à certaines croyances, la morale africaine ne découle pas de la religion. Elle est basée sur la préoccupation pour le bien-être humain et les relations sociales [Gyekye 2010].

[Gyekye 2010] met en lumière la centralité de la notion de “caractère” dans la moralité africaine, en soulignant que les actions d’un individu façonnent son caractère et déterminent s’il est considéré comme une “personne” digne de respect et d’inclusion dans la communauté. La notion de “personne” implique l’adhésion aux normes morales et à l’expression de vertus telles que la générosité, l’honnêteté et la compassion.

L’article explore également les notions d’“humanité” et de “fraternité”, soulignant que l’éthique africaine met l’accent sur l’unité et la solidarité entre tous les êtres humains, transcendant les frontières raciales et ethniques. La notion de “bien commun” est également examinée, en mettant en avant que le bien-être de chaque membre de la communauté est intrinsèquement lié au bien-être de tous.

L’éthique africaine est sociale et non individualiste. Elle valorise l’entraide, la responsabilité collective, la coopération et les obligations réciproques. Elle est axée sur le devoir envers les autres et la communauté plutôt que sur les droits individuels. Les actes moralement bons ne sont pas considérés comme superflus mais comme essentiels.

2.9 Justice et droit des animaux et des robots

Pour terminer, Peter Singer, né le 6 juillet 1946 à Melbourne ²² est un philosophe utilitariste australien, titulaire de la chaire d’éthique de l’université de Princeton (États-Unis) et professeur à l’université Charles-Sturt (Melbourne) en Australie.

Il a travaillé deux fois dans la chaire de philosophie de l’université Monash (Melbourne), où il a créé le centre de bioéthique. En 2004, il a été reconnu comme l’humaniste australien de l’année par le Conseil des sociétés humanistes australiennes. En dehors du milieu universitaire, Singer est surtout connu pour son livre “La Libération animale” [Peter 1975]. Ce livre est considéré comme le livre fondateur des mouvements modernes de droits des animaux.

En effet, Singer y argumente contre le spécisme, définissant cette discrimination injuste entre espèces comme moralement non pertinente. Il propose que la sensibilité soit le critère d’appartenance à la communauté morale, prônant l’égalité des êtres sensibles. Singer conclut que l’utilisation des animaux pour se nourrir est injustifiée en raison de la souffrance disproportionnée qu’elle génère comparée aux bénéfices pour les humains, plaidant donc pour le végétarisme ou le véganisme. Toutefois, il soutient l’amélioration des conditions de vie des animaux dans les systèmes industriels, comme l’agrandissement des cages, considérant ceci un progrès face aux réalités actuelles.

²²https://fr.wikipedia.org/wiki/Peter_Singer

Il dénonce la considération inégale des intérêts des animaux et ne manque pas de comparer la prise en compte inégale des intérêts des animaux avec les discriminations subies par les populations noires en occident, et par les femmes.

D'après ²² :

Dans l'optique de Peter Singer, c'est la capacité à souffrir qui est moralement importante. En utilitariste conséquent, Singer soutient que la maximisation du bien-être de tous, passe par celui de tout être sensible, indépendamment de son intelligence. Le but de l'ouvrage n'est cependant pas de militer pour un traitement similaire des hommes et des non-humains, mais de changer notre façon de percevoir et de traiter ces derniers. Il ne s'agit pas de "donner le droit de vote aux cochons", mais de ne pas mépriser les intérêts, différents mais réels, des non-humains.

2.9.1 Justice et droit des animaux non-humains

Il est reconnu que l'animal dispose de certaines des qualités de l'homme, notamment la capacité à ressentir la douleur. De ce fait, il doit être protégé et respecté. La Déclaration universelle des droits de l'animal, proclamée le 15 octobre 1978 à la Maison de l'UNESCO à Paris [Nerdeux 1978], en présence de plusieurs ambassadeurs et de nombreuses personnalités scientifiques, littéraires, artistiques et politiques, et rendue publique en 1990 leur reconnaît certains droits ²³.

Cette déclaration est le résultat d'une réflexion philosophique et scientifique internationale. Elle présente, en termes de droits, les principes généraux d'une éthique applicable, au niveau mondial, aux conduites de l'homme à l'égard de la vie animale.

Elle stipule, entre autres :

- Dans son préambule :
 - Considérant que le mépris, voire la simple méconnaissance de ces droits naturels provoquent de graves atteintes à la Nature et conduisent l'homme à commettre des crimes envers les animaux,
 - Considérant que la coexistence des espèces dans le monde implique la reconnaissance par l'espèce humaine du droit à l'existence des autres espèces animales,
 - Considérant que le respect des animaux par l'homme est inséparable du respect des hommes entre eux, Il est proclamé ce qui suit:
- Dans ses articles :

²³<https://www.fondation-droit-animal.org/documents/ExpoLFDA-2-DUDA.pdf>

- **Article premier** : tous les animaux ont des droits égaux à l'existence dans le cadre des équilibres biologiques. Cette égalité n'occulte pas la diversité des espèces et des individus
- **Article 2** : toute vie animale a droit au respect.
- **Article 3 1-** Aucun animal ne doit être soumis à de mauvais traitements ou à des actes cruels.
- **Article 3 2-** Si la mise à mort d'un animal est nécessaire, elle doit être instantanée, indolore et non génératrice d'angoisse. . . .
- **Article 4 1-** L'animal sauvage a le droit de vivre libre dans son milieu naturel, et de s'y reproduire.
- **Article 4 2-** La privation prolongée de sa liberté, la chasse et la pêche de loisir, ainsi que toute utilisation de l'animal sauvage à d'autres fins que vitales, sont contraires à ce droit.
- Les articles 5, 6, 7, 8, et 9 mettent l'accent sur la nécessité d'observer à l'égard des animaux, l'entretien, les soins attentifs, le respect de leur mode vie, de leur comportement dans leur espace naturel, de leur dignité (pas d'expérimentation impliquant une souffrance physique ou psychique, pas de mise à mort brutale, pas de gavage forcé, respect de la personnalité juridique ; etc.).
- Article 10- L'éducation et l'instruction publique doivent conduire l'homme, dès son enfance, à observer, à comprendre, et à respecter les animaux.

2.9.2 Justice et droit des robots

Dans "Le Droit des Robots" [Bensoussan 2017], la question suivante a été soulevé : "Quelle est l'autonomie de décision d'une machine ?".

Il nous invite à faire un voyage dans le monde de l'intelligence artificielle, des algorithmes et des robots. Mais avant d'entrer dans le vif du sujet, il propose de clarifier certaines notions.

Notion de machine intelligente : une machine intelligente est une machine autonome, capable de prendre une décision ne résultant pas elle-même de systèmes autonomes. Cette autonomie bouleverse notre société en tant qu'humains, en créant l'apparition d'une nouvelle espèce.

Les auteurs distinguent deux catégories, deux espèces : les robots logiciels qui sont machines intelligentes logicielles et les robots physiques.

- Les robots logiciels (dits *bots*, robots logiciels ou robots conversationnels amis) ; ils constituent un ensemble d'algorithmes capable de prendre des décisions de manière autonome et capables d'acquérir une expérience personnelle. Ce sont, par exemple, des agents virtuels comme :

- Amélia, développé par IPsoft, qui est l'exemple type du robot logiciel car, matérialisé comme une femme (pour des raisons d'acceptabilité sociale), il joue le rôle d'un véritable fonctionnaire virtuel.
- Watson d'IBM, l'intelligence artificielle d'IBM "en mesure aujourd'hui d'établir des diagnostics à propos de certains cancers de manière plus fiable que les grands professeurs du secteur de la santé." [Bensoussan 2017]

L'existence de ces robots est amenée à se généraliser dans un avenir proche sous la forme de robots humanoïdes, ou encore d'animaloïdes.

- Le robot physique (qui, dans beaucoup de cas, sera en position de mobilité) est une machine intelligente. Il est capable de prendre, de manière automatique, des décisions qui n'ont pas été programmées. Il est également capable de modifier son environnement et de le comprendre. De ce fait, il dispose d'une double indépendance.

Le meilleur exemple du robot physique reste celui de la voiture autonome, une autonomie par rapport à son gardien juridique. La voiture autonome est capable de prendre des décisions qui ne sont pas le résultat d'automatismes. Elle est munie d'actionneurs permettant de modifier le réel. Dans ces conditions, [Bensoussan 2017] précisent qu'il "devient urgent de doter les robots de droits et de devoirs, notamment en termes de responsabilité, de traçabilité et de dignité".

Ils proposent une comparaison entre le robot logiciel "Amelia" et le robot physique "Pepper". Ce dernier est doté d'actionneurs capables de modifier le réel. Avec ses capteurs, Pepper a la maîtrise de l'environnement, la compréhension de la situation, et la possibilité d'action. Cette différence entre ces deux robots explique à quel point "nous nous trouvons dans une nouvelle civilisation dans laquelle les hommes et les robots devront apprendre à vivre ou survivre ensemble".

D'après [Bensoussan 2017] :

"C'est la réunion de quatre technologies arrivées à maturité qui a permis cette évolution. La première technologie est celle des capteurs, qui modifient radicalement la situation en ce qu'ils permettent une perception réelle de l'environnement par les robots. La seconde est celle de l'informatique avec une puissance de calcul extrêmement importante, qu'elle soit embarquée ou déportée. La troisième est celle de l'intelligence artificielle. La quatrième, enfin, est celle des données. La combinaison de ces technologies explique que nous entrons aujourd'hui dans une nouvelle civilisation appelée la "robohumanité".

Ces auteurs évoquent trois affaires (trois situations critiques qui montrent à quel point les technologies de l'IA nous interpellent et nous invitent à engager sérieusement un débat profond pour dire le devoir et le droit des robots :

- **L'affaire "Instinct de survie" ou le dilemme du constructeur automobile** : dans cette affaire complexe de la voiture autonome les enjeux sont importants et le cœur de cette complexité peut se résumer au fait que le constructeur a tranché. Il préfère être sûr de sauver au moins une vie, celle qui se trouve dans sa voiture, plutôt que d'essayer d'en sauver plusieurs autres (un argument marketing destiné à rassurer ses clients).

Dans l'exemple triste et poignant des trois voitures A, B et C (C étant la voiture autonome) qui se retrouvent à un virage à l'entrée d'un village de campagne (lieu de l'accident irrémédiable), le Président de la Cour d'assises ne peut qu'interroger les acteurs présents ou représentés dans la cour sur la responsabilité de l'accident : Qui est responsable ?

- Le propriétaire de la voiture C (la voiture autonome) ? Mais il ne conduisait pas !
- Le concepteur de la plateforme d'intelligence artificielle ? Mais il n'était pas là !
- Le dirigeant qui a demandé de coder une voiture ayant pour objectif non pas de minimiser les pertes mais de protéger son passager ?
- Le vendeur du véhicule autonome ?

L'expert en robotique, appelé à la barre, explique qu'en réalité, cette situation ne peut se présenter en présence d'humains, de sorte qu'un tel dilemme n'existe pas, tout simplement. En effet, les humains sont trop lents. Il leur est impossible de percevoir en une fraction de seconde les quatre personnes dans l'un des véhicules impliqués, l'unique personne dans l'autre, et de trancher.

Pour lui, la voiture autonome a pris une décision (la décision programmée), grâce aux capteurs, à l'informatique et aux algorithmes embarqués.

L'avocat de la partie civile et l'avocat général partagent le même point de vue : le responsable est celui qui a programmé le clic fatal, celui qui a décidé de l'algorithme, qui a confondu Code informatique et Code de la route.

Dans ce cas particulier de ce domaine, où le Code informatique n'est pas le Code de loi (l'avocat de la Défense est obligé de l'admettre), il n'existe pas de fondement légal pour entrer en voie de condamnation. C'est un accident. Le verdict est l'acquittement.

Dans cet exemple de la voiture autonome, où se trouve embarquée l'éthique ? Les dilemmes constituent un des enjeux des voitures autonomes mais ils sont aussi un enjeu central des algorithmes.

Quelle(s) solution(s) préconiser ? Dans [Bensoussan 2017], les auteurs pensent que :

Dans tous les cas, il est nécessaire de ne pas laisser au concepteur, au juriste, au codeur de décider de la façon dont nous allons vivre avec les robots. Il faut que le grand public se saisisse de cette leçon et se construise face à l'autonomie de contrôle des voitures. La machine intelligente doit recevoir un entraînement, une instruction. Il faut bien apprendre le droit des humains, le code de la route ou encore les us et coutumes de la façon de conduire dans un pays. Il est vrai que l'être humain ne conduit pas de la même manière au Caire, à Paris ou à Londres. La machine intelligente doit donc nécessairement être entraînée et éduquée pour intégrer ces variables.

- **Les robots Hospitaliers : le dilemme médecins robots versus médecins humains**

L'exemple, dans cette affaire, est celui d'une patiente atteinte d'un cancer et qui se présente au service des urgences d'un grand hôpital parisien. Elle cherche un diagnostic très rapide pour son cancer et doit répondre à la question de savoir si elle souhaite se faire examiner par un médecin humain ou par un médecin robot. À l'infirmière dispatching qui la reçoit à l'accueil, elle répond, très énervée : *“Un médecin, bien sûr !”*.

Elle est conduite au “couloir 2, porte 7”, alors même qu'au couloir 1, il y avait “Watson”, le programme informatique d'intelligence artificielle conçu par IBM. Ce dernier aurait pu la recevoir, l'examiner et établir peut-être un diagnostic et, surtout, un mode opératoire. Il est important de noter que Watson ne se trompe, s'agissant de certains cancers, qu'une fois sur dix, là où les grands professeurs se trompent une fois sur deux.[Bensoussan 2017]

Dans ces conditions, la question importante à poser est la suivante : en cas de survenue d'un problème, comment condamner un robot alors qu'il est meilleur que ses collègues médecins ?

D'après [Bensoussan 2017] :

le droit d'être ausculté par Watson plutôt qu'un humain, le droit au fond d'avoir davantage de chances d'être guéri par un robot que par un humain montre bien, là encore, que le dilemme de la fracture intelligente est aussi un des grands défis de la robotique.

- **La mixité homme/machine, l'affaire de la poupée Stasie** : la poupée Stasie (initialement appelée Hello Barbie) est mi-objet connecté, mi-robot, Elle représente assez bien ce que l'on peut appeler la mixité homme/machine. Elle est dotée d'un logiciel de reconnaissance vocale, et a la faculté d'enregistrer tous les sons pour les transmettre dans un centre de données localisées ailleurs (aux États-Unis, par exemple).

L'intrusion de plus en plus importante des robots dans nos vies (dans nos usines, nos entreprises et nos domiciles) nous interpelle et nous invite à opter pour une mixité intelligente régulée par une réglementation éthique instaurant le respect, la dignité, et un régime particulier de responsabilité des robots et des humains où l'option de l'arrêt d'urgence peut être activée en cas de problème grave.

Dans le cadre de cette coopération homme-machine, le robot, par essence et dans sa conception, ne doit absolument pas porter atteinte à la dignité et l'intégrité physique du compagnon avec lequel il va travailler. Les hommes doivent, en toute sécurité, apprendre à vivre et à survivre aux côtés des robots.

Les robots sont caractérisés par trois technologies : des capteurs hyperperformants, des calculateurs à très grande capacité de traitement et l'intelligence artificielle qui les dote d'une forme primitive de pensée. C'est cette intelligence artificielle "apprenante" ou "faible" qui les distingue des objets ou des animaux.

Ces caractéristiques technologiques qui suscitent, à la fois, l'espoir et l'inquiétude, ont conduit les écrivains de romans de science-fiction Isaac Asimov et John W. Campbell, à formuler en 1942, trois lois (dites lois de la robotique) auxquelles les robots doivent obéir ²⁴:

1. Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger ;
2. Un robot doit obéir aux ordres donnés par les êtres humains, sauf si de tels ordres entrent en contradiction avec la première loi ;
3. Un robot doit protéger son existence dans la mesure où cette protection n'entre pas en contradiction avec la première ou la deuxième loi.

Il est intéressant de noter que comme les trois lois de Rawls, elles sont appliquées dans un ordre lexicographique.

La formulation de ces lois a suscité un débat, compte tenu de certaines imperfections constatées, ce qui a donné lieu à quelques modifications par les auteurs eu mêmes, notamment par l'ajout d'une loi dite loi Zéro ²⁴ :

- loi Zéro : Un robot ne peut pas porter atteinte à l'humanité, ni, par son inaction, permettre que l'humanité soit exposée au danger ;
- première Loi : Un robot ne peut porter atteinte à un être humain, ni, restant passif, permettre qu'un être humain soit exposé au danger, sauf contradiction avec la Loi Zéro ;

²⁴https://fr.wikipedia.org/wiki/Trois_lois_de_la_robotique

- deuxième Loi : Un robot doit obéir aux ordres que lui donne un être humain, sauf si de tels ordres entrent en conflit avec la Première Loi ou la Loi Zéro ;
- troisième Loi : Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la Première ou la Deuxième Loi ou la Loi Zéro.

Mais ce débat ne fait que commencer. D'après [Blons 2019],

La relation homme-machine entre aujourd'hui dans une nouvelle ère qui va bien au-delà de la stricte robotisation. Les experts estiment que, d'ici 2030, de nouvelles relations entre l'homme et la machine vont progressivement apparaître. Tous deux sont décrits comme des "êtres" complémentaires. Le rapport de force tend donc à s'inverser vers plus de collaboration et non de dépendance.

2.10 Conclusion sur la revue de la littérature philosophique sur l'éthique

Notre thèse devant être applicable rapidement, nous avons choisi de nous placer dans le cadre d'une méta-éthique relativiste, plutôt que d'attendre l'avènement d'une hypothétique méta-éthique universaliste. Cette étude nous aura conduits à proposer une "Éthique relativiste gouvernementale", objet du chapitre 6, dont l'objet est d'imaginer un outil permettant au gouvernement de spécifier les principes éthiques qu'il souhaite voir appliqués dans le pays, et donner aux développeurs un moyen de savoir si leurs applications peuvent s'y conformer, à la manière de ce qui se passe actuellement pour les développeurs d'applications pour Apple, par exemple.

En effet, pour qu'une application soit validée par Apple, elle doit valider un certain nombre de contraintes, sans quoi l'application ne pourra pas être acceptée sur un téléphone portable.

Relever le défi de proposer une plateforme d'éthique relativiste sera l'objet de notre contribution.

Revue de la littérature scientifique sur l'IA

Sommaire

3.1 Intelligence Artificielle (IA)	71
3.1.1 Définitions et applications de l'IA	71
3.1.2 Technologies et Domaines d'applications de L'IA	74
3.1.3 Prédiction pour l'aide à la décision	79
3.1.4 Travaux connexes lié à la prédiction de courtes séries temporelles .	85
3.1.5 Stratégies nationales en matière d'IA	87
3.1.6 IA générative dans le secteur public	92
3.2 Ingénierie de prompt	97

3.1 Intelligence Artificielle (IA)

3.1.1 Définitions et applications de l'IA

Reconnue comme étant la science qui consiste à concevoir, construire et préparer des machines “intelligentes”, l'IA est utilisée dans diverses situations, notamment pour discuter, traduire des langues étrangères [Popel 2020], prévoir la météo, etc. Les domaines d'usage de l'IA sont nombreux : la médecine, la finance, la production industrielle, la sécurité, les transports, la recherche scientifique.

Elle a révolutionné de nombreux aspects de notre quotidien. Composée de données, d'algorithmes, et de matériel de connectivités, l'IA permet aux machines d'imiter des aspects de l'intelligence humaine tels que la perception, la résolution de problèmes, l'interaction linguistique et même la créativité. Cette technologie s'est tissée dans notre vie de manière si profonde que son impact et notre dépendance à son égard sont souvent méconnus.

Notre utilisation quotidienne d'appareils mobiles, de systèmes de reconnaissance faciale, de médias sociaux, d'assistants vocaux numériques, de la banque en ligne, de la cartographie et des mises à jour de trafic, dépend largement de l'IA.

Dans les années 50, au milieu du XXème siècle, la réalité de l'IA, encore inconnue, était caractérisée par de nombreuses questions et une recherche approfondie qui semblait parfois être impossible : les ordinateurs ne pouvaient pas stocker ou exécuter des informations, et les coûts associés étaient considérables.

C'est au mathématicien britannique Alan Turing, avec son article baptisé "Computing Machinery and Intelligence" ("Machines de calcul et intelligence"), que le monde doit l'esquisse des premiers pas de l'IA. C'est lui qui crut la machine capable de penser.

Le test qui porte son nom (Test de Turing) illustre bien sa réflexion sur ce sujet à savoir, la capacité de la machine à penser : Dans une conversation entre l'homme et la machine, l'homme doit être capable de distinguer, à l'aveugle, s'il converse avec un être humain ou un ordinateur. Si non, la machine serait alors dotée d'une intelligence.

Entre les années 50 et 70, l'industrie informatique va accuser des avancées majeures en matière de capacité de stockage et de puissance de calcul : les ordinateurs deviennent plus rapides, plus accessibles et, de plus, moins coûteux.

Au cours des années 80, deux méthodes essentielles ont été mises au point. Le deep learning, (ou apprentissage en profondeur), a donné aux ordinateurs la possibilité d'acquérir des connaissances grâce à leur expérience. La deuxième, appelée système expert, imite la faculté de l'individu à prendre des décisions. Les ordinateurs ont entamé l'utilisation d'un raisonnement basé sur des "règles" en utilisant principalement une structure de type "si-alors" pour répondre à des interrogations.

Les années 2000 ont été marquées par un débit plus important et des options de stockage, comme le cloud. L'essor récent de l'IA est dû aux deux améliorations sectorielles majeures : les unités de traitement graphique ou GPU [Baji 2017] et la masse de données disponibles (big data)[Charles 2013, Shahid 2021].

- **Le GPU** : la demande dans le monde de la vidéo et des jeux a entraîné le développement des GPUs améliorés et moins coûteux, un élément nécessaire pour la construction de solutions d'IA.
- **Le Big Data** : les algorithmes utilisés par l'IA sont essentiellement "formés" par la grande quantité d'informations présentes dans le Big Data. Ces algorithmes aident ensuite l'IA à traiter ces informations à un rythme important.

La définition de l'IA donnée par Yann LeCun, chercheur en intelligence artificielle, premier directeur du Facebook Artificial Intelligence Center et leader mondial en IA et en Deep Learning, est que l'IA consiste à "faire faire aux machines des activités que l'on attribue généralement aux animaux et aux humains". Cette définition reflète la diversité et la complexité des tâches que l'IA peut accomplir aujourd'hui.

Le développement récent de l'IA, a fait l'objet d'études approfondies dans divers secteurs. Voici des principaux exemples d'application de l'IA dans la société :

L'IA dans la médecine

Dans le secteur de la santé, l'IA joue un rôle-clé en améliorant les soins administrés aux patients, en rationalisant les procédures médicales et en accélérant la recherche, en mettant en évidence et identifiant des modèles (patterns) dans de vastes ensembles de données, modèles qui pourraient échapper à l'expertise humaine et que, donc, les humains pourraient manquer.

Grâce à l'apprentissage automatique (ou machine learning, sous-domaine de l'IA), l'IA permet d'identifier des pathologies via l'imagerie médicale [Koul 2022, Olveres 2021, Giavina-Bianchi 2021]. Elle est au cœur de la médecine du futur, avec les opérations assistées [Bodenstedt 2020], la prédiction d'une maladie et/ou de son évolution [Ghaffar Nia 2023], le suivi des patients à distance [Shaik 2023], les prothèses intelligentes, les recommandations de traitements personnalisés [Schork 2019], etc.

L'IA dans l'industrie

Parmi les applications clés de l'IA dans ce secteur, on retrouve la maintenance prédictive qui se manifeste par l'anticipation des pannes, ce qui évite tout arrêt coûteux de la chaîne de production [Islam 2024] ; on retrouve également le contrôle de qualité [Chouchene 2020], l'automatisation et l'amélioration des processus de fabrication [Kim 2022] et l'optimisation de la chaîne d'approvisionnement. L'IA contribue donc à améliorer l'efficacité et à réduire les coûts opérationnels grâce à l'analyse de données en temps réel et, donc, à la prise de décision.

L'IA dans la finance

L'IA permet aux organisations de stimuler leur productivité et réduire les coûts opérationnels [Moradi-Motlagh 2015].

Par ailleurs, Elle aide les banques à adresser automatiquement, et de manière fiable, les demandes des clients [Kaur 2020], exécuter des transactions pour leur compte, gérer les risques de manière efficace et prévenir la fraude [Umamaheswari 2023].

Les systèmes de trading pilotés par l'IA peuvent traiter rapidement des quantités massives de données en temps réel, afin d'effectuer des trades (échanges commerciaux), suivre les marchés et gérer les portefeuilles [Pothumsetty 2020].

L'IA dans l'éducation

Dans l'éducation, l'IA permet l'aide à l'orientation, l'automatisation de certaines activités comme l'apprentissage, le suivi des apprenants et la correction des évaluations. Elle facilite également la mise en œuvre du programme de développement durable, par exemple en analysant l'apprentissage des élèves et en leur fournissant de meilleurs outils dans les zones rurales défavorisées [Pedro 2019].

L'IA marque des avancées significatives, transformant la manière d'enseigner et d'apprendre. Elle permet une personnalisation de l'expérience d'apprentissage, adaptant le contenu et les ressources au style, au rythme et au niveau de compréhension de chaque étudiant [Pedro 2019, Raza 2023].

L'IA dans la cybersécurité

Le secteur de la sécurité informatique bénéficie également de l'IA, afin de renforcer la

protection des utilisateurs de technologie. Le machine learning a, en effet, les moyens de prévoir des cyberattaques et, donc, de prévenir leur apparition [Singh 2023]. L’IA aide également à automatiser les tâches répétitives, accélérer la détection et la réponse aux menaces.

Sur la base d’un examen des articles pertinents publiés dans l’International Journal of Information Management (IJIM), y compris ceux publiés sous son ancien titre Social Science Information Studies (SSIS), [Duan 2019] présente une perspective historique de l’histoire et du développement de l’IA. Pour y parvenir, ils ont effectué des recherches en texte intégral sur les termes “artificial intelligence” et “intelligent” dans les archives SSIS/IJIM, ainsi qu’une liste de 25 termes supplémentaires liés à l’IA, termes présentés dans la figure 3.1.

1 case-based reasoning	14 logic programming
2 computer vision	15 machine learning
3 cognitive computing	16 machine vision
4 cognitive science	17 natural language processing
5 data mining	18 neural network
6 data science	19 pattern recognition
7 expert system	20 recommendation system
8 fuzzy linguistic modeling	21 recommender system
9 fuzzy logic	22 semantic network
10 genetic algorithm	23 speech recognition
11 image recognition	24 support vector machine/SVM
12 k-means	25 text mining
13 knowledge-based system	

Figure 3.1: termes utilisés dans la recherche en texte intégral dans les archives du SSIS/IJIM. [Duan 2019].

En conclusion, l’intelligence artificielle a transformé et continue de transformer encore, radicalement, notre façon de vivre, de travailler et d’apprendre, nous offrant, toujours plus, de nouvelles possibilités en matière de commodités, d’efficacité et de personnalisation. Toutefois, compte tenu des risques, il est essentiel de prendre en considération les implications éthiques de l’IA.

3.1.2 Technologies et Domaines d’applications de L’IA

3.1.2.1 Notions d’IA faible et d’IA forte

L’univers de l’IA est si vaste qu’il est parfois commode de distinguer une IA faible et une IA forte. La première nous accompagne dans notre quotidien, rendant les services personnalisés et toujours plus efficaces, alors que la seconde, plus intelligente, nous ré-

sout des problèmes complexes et nous ouvre les portes vers un futur chargé à la fois de belles promesses mais aussi d'inquiétudes.

L'IA faible (ou IA étroite) : est à l'origine des assistants virtuels comme Siri ou Alexa. Ces assistants sont conçus pour interpréter et exécuter le langage naturel. Répondant à des instructions précises, et obéissant à des règles prédéfinies, ils offrent des informations et sont aptes à exécuter des actions. Elle est également à l'origine de la reconnaissance d'image (détection et catégorisation d'objets dans des images avec une grande précision), ce qui permet de jouer des rôles importants, notamment la sécurité et le diagnostic médical. Elle est dépourvue de compréhension globale et de conscience.

L'IA forte (ou IA générale) : a pour objectif de créer de robots totalement autonomes, aptes à intervenir dans de nombreux secteurs comme l'industrie, la santé, et la recherche scientifique. Elle est dotée d'une intelligence et d'une compréhension totales. Elle simule la pensée humaine, capable d'autonomie par l'expérience, et apte à réaliser de nombreuses tâches, à traiter des questions complexes et, enfin, à prendre des décisions toute seule. Elle est potentiellement porteuse de grandes innovations et de grands dangers.

3.1.2.2 Notions d'apprentissage (machine learning et deep learning)

Deux techniques fondamentales permettent à l'IA d'être effective efficacement : l'apprentissage automatique ou machine learning (ML) et l'apprentissage profond ou deep learning (DL). Ces deux technologies sont considérées comme des sous-ensemble de l'IA (voir figure 3.2) et créent l'environnement nécessaire au fonctionnement de celle-ci¹.

1. **Machine Learning :** Le Machine Learning (ML) ou apprentissage automatique est basé sur des approches mathématiques et statistiques. Cette technologie utilise différents types de données, notamment les graphes, les arbres, les courbes ou des variables qualitatives ou quantitatives, continues ou discrètes. Elle consiste à programmer des machines pour qu'elles pensent et agissent comme l'humain, dans le but de résoudre un problème informatique.

L'apprentissage automatique est basé sur le développement d'algorithmes capables d'apprendre à partir de données, afin d'effectuer des tâches comme prédire une solution ou prendre une décision, puis, en apprenant de leurs erreurs, améliorer leurs performances à résoudre ces mêmes tâches.

Il peut être divisé en sous-catégories telles que l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage par renforcement. Ces sous-catégories sont essentielles pour transformer les données brutes en renseignements utiles et pour permettre aux machines d'acquérir un savoir.

¹https://fr.wikipedia.org/wiki/Intelligence_artificielle

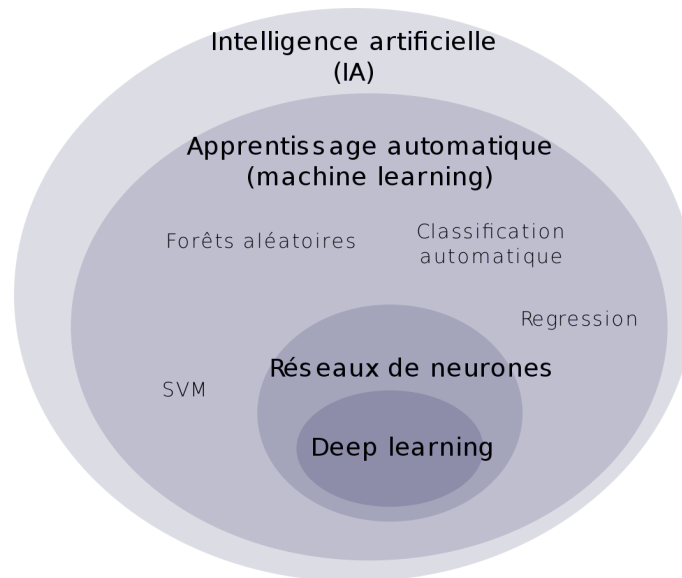


Figure 3.2: L'IA et ses sous domaines.

- **Apprentissage supervisé ou enseignement guidé**

Ce type d'apprentissage utilise des données d'entraînement étiquetées, ce qui signifie que les données d'entrée ont leur correspondance en termes de données de sortie. Le modèle n'a plus qu'à mettre en correspondance les entrées et les sorties afin d'être capable de prédire la sortie pour de nouvelles entrées. Il apprend en comparant sa sortie (en cas d'erreur) avec la sortie correcte et ce, afin de détecter ses erreurs et de modifier le modèle en conséquence.

Si les étiquettes sont discrètes, on parle de classification (prédiction d'une classe ou catégorie : classement), Si, par contre, les étiquettes sont continues, on parle de régression (prédiction d'une valeur continue : une quantité).

Les applications de l'apprentissage supervisé se répartissent, donc, en deux grandes catégories : la classification et la régression [Attaran 2018]. Il est le plus souvent utilisé dans des scénarios où des données historiques permettent de prédire des événements futurs probables [Attaran 2018].

De nombreuses applications d'IA, telles que la détection des fraudes, l'évaluation des risques, la segmentation des clients, la reconnaissance des images, de la parole et du texte [Attaran 2018], utilisent fréquemment l'apprentissage supervisé.

- **Apprentissage non supervisé de l'IA**

Dans le cas de cette méthode de l'apprentissage non supervisé, comme son nom l'indique, aucune aide n'est donnée à l'algorithme. L'ordinateur doit

alors déchiffrer les données d'entrée pour y trouver une structure ou un motif/pattern [Attaran 2018]. Ce type d'apprentissage utilise, donc, des données non étiquetées pour entraîner des machines. Les données non étiquetées signifient que les données ne sont pas associées à des étiquettes pour dire à l'ordinateur à quelle catégorie elles appartiennent. Le modèle apprend à partir des données, découvre des relations, des caractéristiques et des motifs dans les données et renvoie à la sortie.

Quelques exemples d'applications courantes sont la segmentation de la clientèle en marketing et l'analyse des séquences génétiques.

- **Apprentissage par renforcement**

L'apprentissage par renforcement est un type d'apprentissage où une IA peut apprendre, sachant qu'il n'y a pas de variable cible prédéfinie. Ce type d'apprentissage apprend à une machine à prendre les mesures appropriées et à maximiser la récompense dans une situation donnée. Pour ce faire, l'apprentissage utilise un agent et un environnement (pour produire des actions et des récompenses). L'agent a un état initial et un état final et apprend à se comporter dans l'environnement en effectuant des actions et en observant les résultats. Il apprend par essais et erreurs, en recevant des récompenses ou des pénalités pour les actions effectuées.

Comme exemples d'applications, ce type d'apprentissage est largement utilisé dans les jeux, la navigation et la robotique [Attaran 2018].

2. **Deep learning (DL)** : Tout comme le ML, le Deep Learning (apprentissage profond) est un sous-ensemble de l'IA. Il consiste en de multiples couches d'informations cachées dans un réseau neuronal artificiel [Bengio 2013]. En principe, l'apprentissage en profondeur tente de simuler la manière dont le cerveau humain traite les informations. C'est une technologie basée sur des réseaux neuronaux qui ont la capacité d'apprendre des patterns complexes, directement à partir des données.

La reconnaissance d'images, la traduction automatique des langues, les diagnostics médicaux, la détection des fraudes et la vision par ordinateur sont autant d'exemples d'utilisation de l'apprentissage profond [Dargan 2020].

Le développement de ces technologies de l'IA a donné lieu à des applications qui se sont imposées compte tenu de leur utilisation étendue et qui constituent, de ce fait, de véritables outils technologiques auxquels ont recours de nombreux domaines et secteurs de développement.

3.1.2.3 Applications fréquentes de l'IA (outils technologiques)

Traitement du langage naturel :

Le traitement du langage naturel ou natural language processing en anglais (NLP) est l'une des applications les plus utilisées de l'IA. Cette application confère aux machines la capacité de lire, de comprendre et de dériver un sens du langage humain [Chowdhary 2020]. Cela inclut des tâches telles que la traduction linguistique, l'analyse des sentiments et les agents conversationnels (chatbots).

Robotique :

La robotique est l'ensemble des techniques permettant la conception et la réalisation de machines automatiques ou de robots. Elle désigne également l'ensemble des domaines scientifiques et industriels en rapport avec la conception et la réalisation de robots. La robotique est, donc, l'intersection de la science, de la technologie et de l'ingénierie qui produit des robots qui peuvent effectuer des tâches dans le monde physique. Ceci inclut les aspects de l'IA qui permettent aux robots de percevoir leur environnement, planifier des actions, et de déplacer et manipuler des objets.

La robotique devrait améliorer la productivité humaine car elle a des applications à grande échelle, notamment dans les domaines de la production, de la livraison, de la sculpture, des opérations médicales, etc.

Systèmes experts :

Les systèmes experts sont des programmes informatiques qui utilisent des connaissances et des mécanismes d'inférence pour fournir des conseils, des solutions ou des recommandations dans un domaine donné. Ils ont été conçus pour imiter les capacités de prise de décision et de résolution de problèmes des experts humains dans des domaines spécifiques [Buchanan 1988]. Ils sont utilisés dans des domaines tels que le diagnostic médical, l'analyse financière, etc.

Réseaux neuronaux :

Également appelé réseau de neurones artificiels, il s'agit d'un système de matériel et/ou de logiciel qui s'inspire du fonctionnement des neurones du cerveau humain.

Comme le fait le cerveau humain, les réseaux neuronaux sont principalement conçus pour classer des informations. Ces réseaux neuronaux ont la capacité de modéliser et de traiter des données d'entrée/sortie non linéaires en parallèle les unes aux autres.

[Touzet 1998] définit les réseaux de neurones artificiels comme étant “des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu'il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau”. L'intérêt de cette architecture neuronale artificielle réside dans les applications de l'IA qu'autorise l'algorithmique, notamment : le traitement du signal, la reconnaissance de formes, et l'apprentissage par l'exemple (non supervisé ou supervisé).

Logique floue :

La logique floue (fuzzy logic, en anglais) est un moyen de modéliser le raisonnement logique dans lequel la vérité d'une déclaration n'est pas un vrai ou un faux binaire comme c'est le cas dans la logique classique. Il s'agit plutôt d'un degré de vérité qui va de 0 à 1. Les valeurs de vérité des variables sont des réels entre 0 et 1 [Giuffrida 2018]. La logique floue nous permet de concevoir un système d'inférence floue, c'est-à-dire une fonction qui associe un ensemble d'entrées à des sorties en utilisant des règles interprétables par l'homme plutôt que des mathématiques abstraites. La logique floue est plus qu'une simple technique de contrôle, car elle peut être utilisée pour tout processus de prise de décision. C'est une méthode de raisonnement qui s'apparente au raisonnement humain.

3.1.3 Prédiction pour l'aide à la décision

Le mot décision (du latin *decidere*, trancher, de *caedere*, couper) a fait son apparition dès le début du XIV^e siècle. La décision (l'action de décider, de trancher), quelles qu'en soient ses modalités (individuelle ou collective), sa nature (juridique, acte médical, industrielle, économique, stratégique, etc.), n'est pas un acte anodin. Elle suppose un choix entre plusieurs alternatives, un examen des options possibles, le résultat d'une délibération. Il est des situations où les conséquences du choix opéré nous interpellent [Kast 2002] et nous finissons par éprouver le besoin de prendre le temps de collecter l'information, de l'analyser pour une prise de décision factuelle. Une forme de prise de décision qui offre l'avantage d'être basée sur des faits, sur l'expérience vécue.

Le choix final, pour être rationnel, doit être fondé sur des arguments scientifiques. L'émergence et le développement progressif de la théorie de la décision [Kast 2002], qui se trouve à l'intersection de nombreuses disciplines (notamment, l'économie, la gestion, la psychologie, la statistique et les mathématiques) va rendre cette rationalité possible. D'après [Kast 2002], cette théorie tire son succès des réflexions autour du comportement rationnel des modèles de l'économie et de la gestion. La démarche se fonde sur une description objective des problèmes de décision (une description utilisant le langage universel, le langage des mathématiques), les analyses cohérentes de ces problèmes, les principes pour déterminer les critères de sélection, et les solutions pouvant être proposées. [Pomerol 2002] atteste de l'apport appréciable de Simon H. A. dans le management de la décision.

L'aide à la décision, selon [David 2011], est "l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant, dans un processus de décision, éléments concourant à éclairer la décision et normalement à prescrire, ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution du processus, d'une part, les objectifs et le système de valeurs au service desquels cet intervenant se trouve placé, d'autre part".

Un outil d'aide à la décision est un logiciel capable de traiter un grand nombre d'informations afin de permettre une prise de décision optimale ².

3.1.3.1 État de l'art

Notions de prédiction, d'aide à la décision, et de prise de décision : l'aide à la décision, selon [Roy 1985], est l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant, dans un processus de décision. Ces éléments de réponse concourent à éclairer la décision et, normalement, à prescrire ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution du processus, d'une part, et les objectifs et le système de valeurs au service desquels cet intervenant se trouve placé, d'autre part.

La prise de décision (étroitement liée à l'aide à la décision) consiste à identifier les problèmes, les inconvénients, les avantages, et les opportunités, puis à y apporter des solutions [Daft 2003].

La prédiction est l'une des diverses manières grâce à laquelle l'IA peut aider dans la prise de décision. Elle joue un rôle crucial dans la prise de décision en permettant aux décideurs d'anticiper les résultats futurs et de faire des choix proactifs.

Rôle de l'IA et place des algorithmes dans la prédiction et la prise de décision : La capacité à utiliser l'IA pour améliorer la prise de décision, réinventer les modèles de business et les écosystèmes, et remodeler l'expérience client sera le moteur des retombées des initiatives numériques jusqu'en 2025 [Duan 2019]. Il existe de nombreux rapports sur les avantages de l'IA pour la prise de décision [Wilson 2018]. En effet, L'utilisation de l'IA pour la prise de décision est l'une des démarches les plus importantes de l'histoire de cette technologie. D'une manière générale, les systèmes d'IA peuvent être utilisés soit pour soutenir/assister les décideurs humains, soit pour les remplacer [Edwards 2000].

La prise de décision dans l'IA désigne le processus consistant à utiliser des techniques d'intelligence artificielle pour faire des choix ou prendre des mesures sur la base de données et d'algorithmes [Jarrahi 2018]. L'IA a le potentiel de révolutionner les processus de prise de décision dans divers domaines, de la finance aux soins de santé et, même, au-delà. Sur la base de l'analyse et des prédictions, les algorithmes d'IA peuvent générer des recommandations, sélectionner le plan d'action le plus approprié ou même prendre des mesures de manière autonome.

Dans la théorie de la décision, le rôle de la prédiction est d'améliorer la prise de décision en permettant aux décideurs d'anticiper les développements futurs et de faire des choix éclairés [Agrawal 2018, Surtpto 2023]. Par conséquent, la prédiction contribue à améliorer la précision et la planification stratégique des processus de prise de décision

²https://fr.wikipedia.org/wiki/Aide_%C3%A0_la_d%C3%A9cision

et, en fin de compte, à remodeler la manière dont les organisations élaborent leurs stratégies et restent en tête dans le paysage concurrentiel. Les algorithmes d'IA peuvent utiliser des données historiques et des techniques d'apprentissage automatique pour faire des prédictions sur des événements ou des tendances futurs. La nouvelle vague de systèmes d'IA a amélioré la capacité d'une organisation à utiliser des données pour faire des prédictions et a considérablement réduit le coût de ces prédictions [Agrawal 2022].

La prédiction implique l'utilisation de modèles statistiques, mathématiques ou informatiques pour analyser les modèles et les relations au sein des informations disponibles, afin de faire des projections éclairées sur ce qui est susceptible de se produire à l'avenir. A ce titre, elle exerce une forte attraction. En science, nombreux sont ceux qui considèrent la capacité de prédiction comme la confirmation ultime de la théorie et de la compréhension. Dans le domaine de la prise de décision, les capacités de prédiction semblent offrir la promesse d'un contrôle sur l'avenir. Le double objectif de compréhension et de contrôle s'est avéré être un attrait tentant pour les sciences au cours des dernières décennies.

Des ressources publiques considérables ont été investies dans des programmes de prévision justifiés par des objectifs scientifiques et politiques simultanés dans des domaines tels que le changement planétaire, les tremblements de terre, l'élimination des déchets nucléaires, les impacts d'astéroïdes, les inondations, les conditions météorologiques, l'érosion des plages, etc. Selon [Pielke Jr 2003], les sciences écologiques sont désormais prêtes à rejoindre ce groupe de programmes de recherche axés sur le développement de connaissances prédictives en tant que guide pour la prise de décision.

Prédiction et modélisation prédictive : qui dit prédiction, dit modélisation prédictive. En s'appuyant sur des schémas et des tendances historiques, la modélisation prédictive permet d'établir des prédictions précises, de prendre des décisions éclairées et de mieux comprendre des phénomènes complexes [Halladay 2013]. Elle implique la construction de modèles mathématiques qui saisissent les relations entre les variables, permettant la projection de tendances, d'événements ou de comportements futurs [Toma 2023].

La modélisation prédictive est donc le processus global de construction d'un modèle permettant de faire des prédictions. Ces fondements comprennent plusieurs composants cruciaux [Eckerson 2007, Zhu 2018], ci-joints, les éléments clés :

- Collecte et prétraitement des données : rassembler les données pertinentes et les préparer à l'analyse en les nettoyant, en les transformant, et en les normalisant,
- Sélection du modèle : choix d'un algorithme ou d'une technique approprié(e) en fonction de la nature du problème et de l'ensemble de données,
- Entraînement et test : diviser les données en ensembles d'entraînement et de test pour entraîner le modèle et évaluer ses performances.

- Métriques d'évaluation : déterminer métriques (par exemple, l'exactitude, la précision) pour mesurer la performance prédictive du modèle.
- Validation : Il s'agit d'une étape très importante de l'analyse prédictive. Au cours de cette étape, l'efficacité du modèle est vérifiée en effectuant différents tests. Ici, des exemples d'ensembles d'entrées sont fournis afin de vérifier la validité du modèle.
- Déploiement : le déploiement consiste à faire fonctionner notre modèle dans un environnement réel, ce qui permet d'en discuter au quotidien et de le rendre utilisable.
- Gestion des modèles : gérer les modèles pour améliorer les performances (c'est-à-dire la précision), contrôler l'accès, promouvoir la réutilisation, normaliser les outils et minimiser les activités redondantes.

Les modèles prédictifs sont continuellement raffinés et améliorés au fur et à mesure que de nouvelles données sont disponibles. Les processus itératifs permettent d'améliorer la précision et la fiabilité des prédictions au fil du temps.

Dans ce travail, nous nous sommes focalisés sur l'analyse des séries temporelles qui permet d'étudier les relations entre les variables qui varient dans le temps.

3.1.3.2 Problématique des séries temporelles

Définition des séries temporelles : une série temporelle est une séquence de points de données qui se produisent dans un ordre successif au cours d'une période donnée [Gregson 2014, Chatfield 2013]. Elle est couramment utilisée dans divers domaines tels que les statistiques, la finance, l'économie et la météorologie pour analyser et prévoir les tendances, les schémas et les changements dans les variables, au fil du temps. Les données de séries temporelles sont collectées à intervalles réguliers (séries journalières, mensuelles, trimestrielles ou annuelles) et peuvent inclure différents types d'informations, tels que les cours boursiers, les indicateurs économiques, les mesures météorologiques et les signaux physiologiques tels que le rythme cardiaque ou l'activité cérébrale.

L'analyse de ces séries implique l'utilisation de méthodes statistiques pour extraire des informations significatives des données, identifier les tendances, la saisonnalité et les résidus, et faire des prévisions basées sur des schémas historiques.

Applications des séries temporelles : Etant donné que le principal objectif de l'analyse des séries temporelles est de prédire l'avenir d'un certain processus (par exemple, les ventes) sur la base de ce qui s'est passé dans le passé (historique des ventes par exemple), l'analyse des séries temporelles a de nombreuses applications dans le monde réel [], notamment :

1. **Commerce de détail et E-commerce :**
utilisé pour la prévision des ventes et de la demande, la gestion des stocks et les décisions de tarification [Bandara 2019].
2. **Finance :**
pour la prédiction des cours boursiers, l'évaluation des risques, la gestion de portefeuille et la prévision des cours des crypto-monnaies [Tsay 2005, Idrees 2019].
3. **Soins de santé :**
suivi de la santé cardiaque [Oyeleye 2022], de l'activité cérébrale (EEG) [Chaovalitwongse 2007], prédiction des dépenses des soins de santé [Dritsakis 2019], ainsi que l'évaluation des améliorations de la qualité des soins de santé [Penfold 2013].
4. **Prévisions météorologiques :**
utilisé par les météorologues pour prévoir les températures, les événements météorologiques extrêmes et le changement climatique [Agrawal 2012].
5. **Médias sociaux et analyse de sentiments :**
suivre les tendances des médias sociaux, le sentiment de la marque et l'analyse du comportement des utilisateurs [Bhullar 2022].
6. **Production :**
maintenance prédictive, assurance qualité et détection des anomalies dans les processus de production [Susto 2016].

Ces applications démontrent l'impact considérable de l'analyse des séries temporelles dans divers domaines, de la finance et des soins de santé à la vente au détail, aux prévisions météorologiques et à l'analyse des médias sociaux.

Problèmes liés aux séries temporelles : Malgré l'utilité des séries temporelles, ce type de données représentent quelques problèmes qu'il est important de mentionner, afin d'illustrer les complications liées à l'analyse de ces données. En effet, l'analyse des séries temporelles est un challenge, et il est important de prendre en compte des défis tels que la qualité des données, le risque de surapprentissage (ou overfitting en anglais) et l'incertitude inhérente dans la prédiction d'événements futurs.

Selon [Esling 2012], les problèmes les plus importants découlent de la haute dimensionnalité des données de séries temporelles et de la difficulté de définir une forme de mesure de similarité basée sur la perception humaine.

Un autre problème majeur est celui de la prévision : peut-on à partir des valeurs x_1, \dots, x_T avoir une idée des valeurs futures $x_{T+1}, x_{T+2} \dots$? Evidemment, on ne peut pas connaître, en pratique, la dynamique exacte de la série ; dynamique à travers laquelle les valeurs passées influencent la valeur présente. La qualité de la prévision dépend de

la façon dont évolue la série. Plus la série est fonction “régulière” du temps, plus il sera facile de prévoir [Gouriéroux 1995].

Selon [Guigou 2019], l’une des principales raisons pour lesquelles l’analyse des séries temporelles est difficile et nécessite des modèles et des méthodes spécifiques, est que les séries temporelles constituent un type de données très particulier. D’un point de vue mathématique et statistique, elles s’écartent d’un certain nombre d’hypothèses fondamentales. Cela rend de nombreuses tâches de base de l’apprentissage automatique ou de l’inférence statistique beaucoup plus difficiles qu’elles ne le sont habituellement. L’un des challenges dont l’auteur a parlé concerne la manière dont les données des séries temporelles sont acquises, et c’est que l’ensemble de la série temporelle peut ne pas être disponible au moment où une analyse est effectuée car les séries temporelles sont des données échantillonnées à partir d’un processus en cours. D’autres auteurs rapportent, par ailleurs, d’autres difficultés et problèmes liés aux séries temporelles :

- Introduction de bruit : des techniques telles que la différenciation peuvent introduire du bruit supplémentaire dans les données, ce qui peut masquer des schémas ou des tendances fondamentales
- Problèmes de généralisation : Les résultats ne peuvent être généralisés que dans certains cas, principalement lorsque l’analyse est basée sur un ensemble de données ou une période unique et isolée.
- Complexité et sélection du modèle : la sélection du modèle peut être complexe [Qi 2001] et peut fortement influencer les résultats ; Ainsi donc, le choix d’un modèle inapproprié peut conduire à des conclusions peu fiables ou trompeuses. De plus, d’après [Wang 2017], les données historiques ne fournissent pas toujours des informations suffisantes pour sélectionner le modèle le plus approprié.
- Limites des données : La qualité et la quantité des données peuvent avoir une incidence sur la précision du modèle [Gitzel 2016]. Les problèmes de qualité des données peuvent conduire à la prise de mauvaises décisions, même si l’algorithme analytique appliqué est correct [Gitzel 2016].
- Qualité des données associés aux séries temporelles (voir Figure 3.3): [Gitzel 2016] abordent quelques problèmes relatifs à la qualité des données associés aux séries temporelles à savoir : valeurs manquantes (perte de données), changement soudain des unités de mesure, problèmes de type de données (les valeurs flottantes peuvent commencer à apparaître sous forme de chaînes de caractères), changements dans le format de l’horodatage, valeurs hors plage, horodatage collecté erroné ou présentant un retard, valeurs arrondies ou points de données déjà agrégés, historique court des données.

Problem Name	Problem Description	Hubauer et al.	Honeywell	Pastorello et al.	Bastos et al	Rahman et al	Guo and Liu	UNESCO	Esswein et al.	Faier and de Seixas	Colditz et al
Event Data Loss	There are gaps in the event data/time series	X	X	X	X		X	X			X
Values out of Range	The values are out of range for the domain under observation (e.g. subzero temperatures in a hot process). This can be applied to individual values but also to averages, minimums and maximums	X		X	X	X	(X)	X		(X)	
Value Spikes	Spikes or sudden changes which are implausible for the domain. (Recognize through gradients and max deviations)	X			X	X	X	X			
Wrong Timestamps	Timestamps are wrong	X	X	X			(X)	X			
Slightly Inaccurate Measurement	The value is slightly wrong which might result in the detection of a trend etc.	X	X	X				X	(X)		
Rounded Measurement Value	The value is not to the optimal level of detail or has slight variations. (This might be hidden behind values such as 1000 or due to scale of units (e.g. m3))	X	X	X				X			
Signal Noise	Small changes which are not in the process but result from inaccurate measurements. (Recognize with low pass filter)	X		X				X	(X)		
Data Not Updated	Data is not up-to-date. (Sensors might still display old values.)	X	X				X				
Unreliable Data Source	The data source is not considered fully reliable				X	X	(X)	(X)			
Divergent Despite Correlation	Values which are normally correlated behave unexpectedly			X		X					
Units of Measurements	The units of measurement are the same for all data sources (cm vs. inch)		X					X			
Forced/Calculated Values	Compensated values are used instead of real measurements. This is only a problem if we "discover" those connections later.		X	X							
Prior Changes	Has the data been changed before?		X					X			
Data Formats	Different data formats, e.g. float vs. string etc.	X					X				
Name of Events	Different names/text for events of the same type	X					X				
Timestamps Format	There are different formats used for timestamps which make the comparison difficult.	X									
Divergent Measurements	Values which should be the same are different. Always or sometimes.	X									
Signal Alteration	Two signals "trade places"	X									
Missing Foreign Keys	Foreign keys are missing	X									
Short Data History	The history of recorded data is too short for a good analysis	X									
Aggregated Data	Is data instantaneous or is it already an average over a certain time span		X								
Diverging Sampling	Different sampling rates in the same time series can lead to problems (e.g. how many values to put into one day?)			X							
Different Accuracy	There is a different level of accuracy for the same type of data			X							
Inconsistent Noise Level	The level of noise changes over time or from different data sources			X							
Class Imbalance	There is a bias in the sample as opposed to the population					X					
Heteroscedasticity	There are subpopulations that have different variabilities from others. (Detect via Goldfeld-Quandt test)									X	

Figure 3.3: Problèmes de qualité des données relatives aux séries temporelles.[Gitzel 2016]

3.1.4 Travaux connexes lié à la prédiction de courtes séries temporelles

Comme nous nous concentrons, dans le cadre de notre première contribution, sur la prédiction des séries temporelles courtes, nous passons en revue dans cette section les méthodes les plus récentes utilisées pour adresser ce problème.

Marche aléatoire (ou random walk en anglais(RW)), méthode de la moyenne (Mean Method en anglais) : la marche aléatoire et la méthode de la moyenne sont les méthodes de prédiction les plus simples : dans la marche aléatoire, la prédiction pour la période suivante est égale à la dernière valeur observée plus un bruit aléatoire ; dans la méthode de la moyenne, les prédictions futures sont égales à la moyenne arithmé-

tique des valeurs observées. Cependant, il est généralement recommandé de les utiliser pour les séries temporelles courtes [Hyndman 2014, Hyndman 2018]. En effet, dans les courtes séries annuelles de la compétition M3, la méthode de Marche aléatoire est l'une de celles qui donnent les meilleurs résultats [Hibon 1999]. Parmi leurs limites, le problème de l'utilisation de la méthode de Marche aléatoire pour les séries temporelles avec tendance et la sensibilité de la méthode Mean aux valeurs extrêmes, de sorte que les prédictions peuvent être biaisées.

Moyenne mobile autorégressive intégrée (ou *AutoRegressive Integrated Moving Average (ARIMA)*): Le modèle de prédiction ARIMA vise à décrire les autocorrélations dans les données stationnaires (c'est-à-dire lorsque la moyenne et la variance ne changent pas avec le temps). Son acronyme décrit la composante comme suit : (i) AR - Les modèles d'autorégression calculent une prédiction par régression sur une observation et ses valeurs antérieures ; (ii) I - Les modèles intégrés (utilisés pour les données non stationnaires), dans lesquels les valeurs réelles des données ont été remplacées par les différences entre chaque valeur et sa valeur antérieure, afin de rendre la série temporelle stationnaire ; (iii) MA - Les modèles de moyenne mobile calculent une prédiction par régression sur les erreurs de prédiction passées au lieu des valeurs réelles passées. ARIMA n'a pas été développé pour les séries temporelles courtes, mais c'est l'une des approches les plus largement utilisées pour la prédiction des séries temporelles et elle a déjà été appliquée à cette problématique [Lusk 1984].

La méthode de lissage exponentiel simple (ou *Simple Exponential Smoothing (SES)*) : La méthode de lissage exponentiel simple est une autre méthode populaire de prédiction des séries temporelles pour les données univariées sans tendance ni saisonnalité. Elle requiert un seul paramètre, appelé facteur de lissage α . Ce facteur contrôle la vitesse à laquelle l'influence des observations à des pas de temps antérieurs décroît de manière exponentielle et a généralement une plage de valeurs possibles entre 0 et 1. Une grande valeur de α fait que le modèle prête attention principalement aux observations passées les plus récentes, tandis que sa petite valeur prend plus d'observations antérieures en compte. Pour ce qui est de la prédiction des séries temporelles courtes, dans [Ostertagová 2011], les auteurs ont appliqué cette méthode à des séries annuelles courtes. Ils concluent que cette méthode n'est adaptée qu'aux prédictions à court terme.

Grey Model (GM): C'est une approche qui a été conçue pour les séries temporelles courtes [Bilgil 2021]. Utilisant les équations différentielles pour exploiter les données, ce modèle nécessite moins de données pour prédire les observations futures avec une bonne précision. Cependant, il peut conduire à des erreurs de prédiction inacceptables dans certains cas, comme cela a été démontré dans [Bilgil 2021]. Ainsi, un nouvel algorithme basé sur le grey model roulant standard optimisé par un algorithme d'optimisation méta-heuristique (rolling grey model optimized by a meta-heuristic optimization algorithm (RGMMOA) en anglais) a été proposé [Cui 2021]. Il fait preuve d'une plus grande précision par rapport au grey model standard.

Autres méthodes : De nombreux travaux relatifs à la combinaison du Filtre de Kalman (KF) avec différents modèles de prédiction ont été proposés dans la littérature. Principalement, le KF a été utilisé pour modéliser des systèmes prédictifs en tant que méthode hybride avec un modèle ARIMA [Aamir 2016], [Aslam 2020], démontrant de très bons résultats. Dans [Nasseri 2011], le KF étendu est combiné avec un modèle de Programmation Génétique (GP) pour la prédiction de la demande mensuelle en eau. Toutefois, ces méthodes combinées n'ont jamais été appliquées à la prédiction de séries temporelles courtes.

Dans [Cruz-Nájera 2022], les auteurs s'attaquent au problème de la prédiction des séries temporelles de crimes de courte durée. Un grand nombre de ces séries temporelles n'ont pas d'auto-corrélation et une symétrie élevée, il s'agit de bruit blanc et, par conséquent, c'est difficile à prédire. Quatre méthodes de prédiction simples : ARIMA, Holt-Winters (HW), réseaux neuronaux artificiels (Artificial Neural Networks (ANN) en anglais) et moyennes mobiles simples (Simple Moving Averages (SMA) en anglais) ont été utilisées pour une analyse comparative, ainsi que quatre méthodes de prédiction de l'état de l'art : Jaganathan, quatrième dans les compétitions M4 ; FFORMA, deuxième dans M4 ; Hybrid, méthode de prédiction hybride simple ; et LightGBM, dont la performance a obtenu la première place dans la compétition M5. Ils ont également suggéré cinq techniques de prédiction distinctes pour chaque méthode de prédiction, qui impliquent l'extraction de la composante saisonnière de la série. Pour comparer les résultats des méthodes de prédiction et des techniques suggérées, ils ont employé le pourcentage d'erreur moyen symétrique et un test de Friedman. Les auteurs concluent, qu'en commençant par la meilleure, les méthodes de prédiction les plus efficaces sont SMA, ARIMA, Hybride et Jaganathan. Ces meilleures techniques de prédiction tendent à ne pas tenir compte de la composante saisonnière. De plus, SMA, Hybrid, FFORMA, HW, ANN et LightGBM ont tendance à présenter leurs performances les plus élevées en faisant abstraction de la composante saisonnière, tandis que Jaganathan et ARIMA ont une relation inverse en ce qui concerne l'utilisation de la composante saisonnière.

3.1.5 Stratégies nationales en matière d'IA

Qu'est ce qu'une stratégie pour l'intelligence artificielle ? Une stratégie pour l'intelligence artificielle a pour objet le développement de cette technologie pour en tirer le plus grand profit, en évitant les menaces et les risques dont elle peut être porteuse. Ce qui revient, d'après [Dutton 2018], à optimiser les avantages éventuels et minimiser les coûts éventuels de l'IA pour l'économie et la société. Ce développement concerne huit domaines des politiques publiques [Dutton 2018] : la recherche scientifique, le perfectionnement du personnel qualifié, l'accroissement des compétences, l'industrialisation, l'éthique, les données et infrastructures numériques, les services gouvernementaux et l'inclusion. [Dutton 2018] présente un rapport intitulé "L'ère de l'IA : Rapport sur les

stratégies nationales et régionales en matière d'IA". D'après ce rapport, 18 pays ont annoncé leur propre stratégie en matière d'IA, dont la France, le Royaume Uni, le Canada, le Mexique, les Émirats arabes unis, la Finlande, la Suède et la Chine.

En plus des huit domaines des politiques publiques énumérés plus haut et qui fondent les stratégies de l'IA, [Dutton 2018] fait les observations suivantes :

- Des 18 stratégies en IA publiées à ce jour, neuf sont entièrement financées et décrivent des politiques précises. Les neuf autres sont des documents d'orientation qui présentent des objectifs pour guider l'élaboration de politiques futures.
- Chaque stratégie en IA est unique et met l'accent sur différents aspects des politiques en matière d'IA. Toutefois, il est possible d'établir un classement général des stratégies en fonction de quatre grandes catégories : recherche et personnel qualifié, industrialisation, exhaustivité et orientation.
- Bien que les stratégies en IA ne partagent pas les mêmes priorités stratégiques, l'industrialisation est la priorité absolue pour 8 des 18 stratégies nationales, et la recherche scientifique, quant à elle, est une priorité absolue pour sept stratégies.

Toutes les 18 stratégies visent à guider le développement et l'utilisation de l'IA tout en abordant les questions éthiques et sociétales. Un grand nombre d'entre elles partagent des caractéristiques communes, mais chacune est unique.

Par ailleurs, il est important de souligner que de nombreux pays se rangent, de plus en plus, dans une dynamique vers l'élaboration d'une stratégie pour l'IA. Parmi ces pays, on peut noter des pays africains comme le Kenya, le Maroc, la Tunisie.

Présentation des huit domaines des politiques publiques : Dans son rapport évoqué plus haut "L'ère de l'IA : Rapport sur les stratégies nationales et régionales en matière d'IA", en 2018, [Dutton 2018] utilise les domaines de politiques publiques suivants pour le classement des priorités stratégiques de chacune des 18 stratégies étudiées :

1. **La recherche scientifique :** La recherche en matière d'IA, qu'elle soit fondamentale ou appliquée, se mesure à l'aune de la capacité d'innovation dans tous les secteurs, vecteurs de développement (secteurs industriels). L'auteur note la création de nouveaux centres, pôles ou programmes de recherche et l'engagement d'accroître le financement de la recherche en IA dans le secteur public.
2. **Perfectionnement du personnel qualifié en IA :** Le capital humain étant essentiel, il faut assurer les qualités professionnelles du personnel. Deux démarches sont mises en œuvre :
 - (a) le financement pour la conduite de chaires et l'octroi de bourses, ou la création de programmes de maîtrise et de doctorat en IA ;

- (b) le financement pour attirer, maintenir en poste et former des spécialistes de l'IA.
3. **Compétences et l'avenir du travail** : Assurer l'enseignement des sciences, de la technologie, de l'ingénierie et des mathématiques (STIM), et la formation permanente, contribuent à accroître et renforcer les compétences des étudiants et de la population active. Ces dispositions valorisent l'économie du numérique et consolident le socle des technologies de l'IA.
 4. **Industrialisation des technologies de l'intelligence artificielle** : Les stratégies de l'IA mettent l'accent sur l'industrialisation en encourageant l'adoption de programmes IA et les investissements dans des secteurs stratégiques, où se trouvent impliqués l'État et le secteur privé à savoir les entreprises en démarrage et des petites et moyennes entreprises (PME).
 5. **Normes éthiques en IA** : La mise au point des normes ou des règlements sur l'utilisation et le développement éthiques de l'IA passe nécessairement par l'existence d'un conseil, d'un comité ou d'un groupe de travail et, également, par des programmes pilotes et une recherche destinée à assoir une IA explicable et transparente.
 6. **Données et infrastructure numérique** : Les stratégies nationales sur l'IA reposent sur les données et l'infrastructure numérique. Les données constituent la matière première alimentant les analyses, pour les prises de décisions et l'innovation. Quant aux infrastructures numériques (les centres de données, les systèmes de stockage, et les réseaux de communication), ils permettent les activités numériques.
 7. **IA au service du gouvernement** : L'un des piliers du développement économique, social et sociétal, souvent exprimé dans les stratégies nationales est l'amélioration de l'administration publique pour de meilleures prestations des services gouvernementaux. Des programmes pilotes exploitant les technologies de l'IA sont souvent proposés pour améliorer l'efficacité et l'efficience de l'action gouvernementale, et créer les conditions du mieux-être des populations.
 8. **Inclusion et bien-être social** : L'exclusion d'individus ou de groupes d'individus sur la base de caractéristiques telles que la couleur de leur peau, leur sexe, leur religion, leur orientation politique ou sexuelle est incompatible avec le bien-être social et la justice. Les stratégies nationales appellent à tenir compte de la diversité sous toutes ses formes, seule garantie pour promouvoir la croissance sociale, l'intégration et le développement sociétal.

Dans [Denford 2023], par ailleurs, les auteurs ont découvert des similitudes dans la manière dont les différents pays abordent les différents aspects du développement de l'IA. Ils ont par la suite cherché à examiner quels sont les pays les plus similaires (et les plus dissemblables) dans leurs approches, et ils ont établi un lien entre ces groupes (ou clusters) et les performances nationales en matière d'IA. Ils ont inclus les 34 pays qui ont élaboré et publié une stratégie nationale en matière d'IA : Allemagne, Australie, Autriche, Belgique, Canada, Chine, Corée du Sud, Danemark, Espagne, Estonie, Finlande, France, Inde, Italie, Japon, Lituanie, Luxembourg, Malte, Mexique, Norvège, Nouvelle-Zélande, Pays-Bas, Pologne, Portugal, Qatar, Russie, Serbie, Singapour, Suède, E.A.U., Royaume-Uni, Uruguay et États-Unis. Ils ont identifié, via leur étude, six aspects qui font partie de chaque plan stratégique d'IA [Denford 2023] :

- La gestion des données fait : elle fait référence à la manière dont le pays envisage de collecter et d'utiliser les données issues de l'IA. Elle se compose de sous-éléments tels que : l'échange de données entre les entités gouvernementales d'une part, avec d'autres parties prenantes et avec d'autres pays d'autre part, la réglementation des échanges, la confidentialité des données et leur sécurité.
- La gestion algorithmique : elle fait référence à la sensibilisation d'un pays à l'éthique algorithmique ou aux questions algorithmiques. Cette gestion comporte quatre composantes : l'éthique, les biais, la transparence, et la confiance.
- La gouvernance de l'IA : Elle se réfère à l'inclusivité, à la transparence et à la confiance du public dans l'IA. L'accent est mis sur la nécessité d'un contrôle approprié de l'IA, de sa sécurité, les réglementations, l'impact sur les inégalités sociales, les risques, la protection des droits de propriété intellectuelle et l'interopérabilité.
- Le développement des capacités : la question du développement des capacités concerne tous les moyens pouvant être mis à disposition par un pays pour développer son IA : le développement et le renforcement des compétences en IA ; les mesures de stimulation de la R&D multisectorielle (le financement, et le renforcement institutionnel des instituts qui lui sont dédiés) ; les structures d'enseignement (l'enseignement primaire et secondaire, l'enseignement supérieur, l'apprentissage tout au long de la vie par la formation continue, l'enseignement professionnel) ; le soutien aux entreprises et projets (PME et startups) ; l'accompagnement et les incitations financières (fiscales, marchés publics, la réflexion et l'innovation pour des modèles appropriés d'entreprise, les projets pilotes et l'attraction d'experts).
- Les industries : les stratégies nationales en matière d'IA portent l'effort national sur le secteur privé, en ciblant huit secteurs : la finance, la technologie, l'industrie manufacturière, la défense, l'énergie et les ressources naturelles, l'agriculture, la santé et le tourisme.

- Les services publics : ils concernent les fonctions administratives censés améliorer l'action du gouvernement (notamment : au sein du gouvernement, entre le gouvernement et les citoyens, entre le gouvernement et les entreprises). Onze domaines du service public sont ainsi concernés : les recettes fiscales, l'éducation, la sécurité publique, l'environnement, la défense, les transports, les soins de santé, les Technologies de l'information et de la communication (TICs), les tribunaux et l'immigration.

L'ensemble de ces six domaines, avec les éléments qui les composent ont été utilisés par [Denford 2023] comme des critères de comparaison permettant de dégager des similitudes et des différences entre les stratégies nationales développées par les différents pays. Ils permettent de voir comment chaque pays aborde la gouvernance et la gestion de l'IA (gestion des données, gestion algorithmique et gouvernance de l'IA), la manière dont il entend développer ses compétences en matière d'IA (développement des capacités) et, enfin, la manière de déployer ou concentrer ses efforts en matière de développement des industries et des services publics.

Pour chacun des six domaines, chaque stratégie de pays est évaluée suivant les éléments qui le composent (à titre d'exemple : éthique, partialité, transparence et confiance, pour la gestion algorithmique).

La démarche de l'analyse consiste à vérifier, d'abord, l'existence ou l'absence de discussions sur les différentes composantes d'un domaine puis, dans un deuxième temps, à considérer chaque composante et à faire un classement en trois niveaux : faible (aucun élément n'est présent), moyen (la moitié des éléments sont présents), élevé (tous les éléments sont présents), en tenant compte de l'implication des acteurs, du niveau des discussions et de la documentation sur chacune des composantes analysée.

[Denford 2023] ont ainsi visualisé suivant cette catégorisation en six attributs, les stratégies nationales d'IA de 34 pays. Cette visualisation a pu faire ressortir les performances des uns et les faiblesses des autres (voir figure 3.4, ci-dessous). En ce qui concerne l'Afrique, [Denford 2023] affirme que ce continent a tardé à adopter les technologies de l'intelligence artificielle (IA), pour diverses raisons, allant des problèmes d'infrastructure aux ressources financières limitées. Les pays de ce continent développent progressivement leurs stratégies et politiques en matière d'IA, bien que le rythme et l'approche diffèrent grandement d'un pays à l'autre. L'île Maurice, l'Égypte et le Kenya se démarquent, chacune ayant mis en place des stratégies spécifiques pour l'IA. La stratégie mauricienne, créée en 2018, vise à revitaliser les secteurs économiques traditionnels grâce à l'IA, tout en développant de nouveaux piliers de croissance.

En Égypte, la stratégie nationale de l'IA, formulée en 2021, cherche à exploiter l'IA pour atteindre les Objectifs de développement durable et positionner le pays comme un acteur majeur de la coopération régionale et internationale en matière d'IA. Le Kenya a entamé ses explorations sur l'IA en 2018, développant une feuille de route à travers

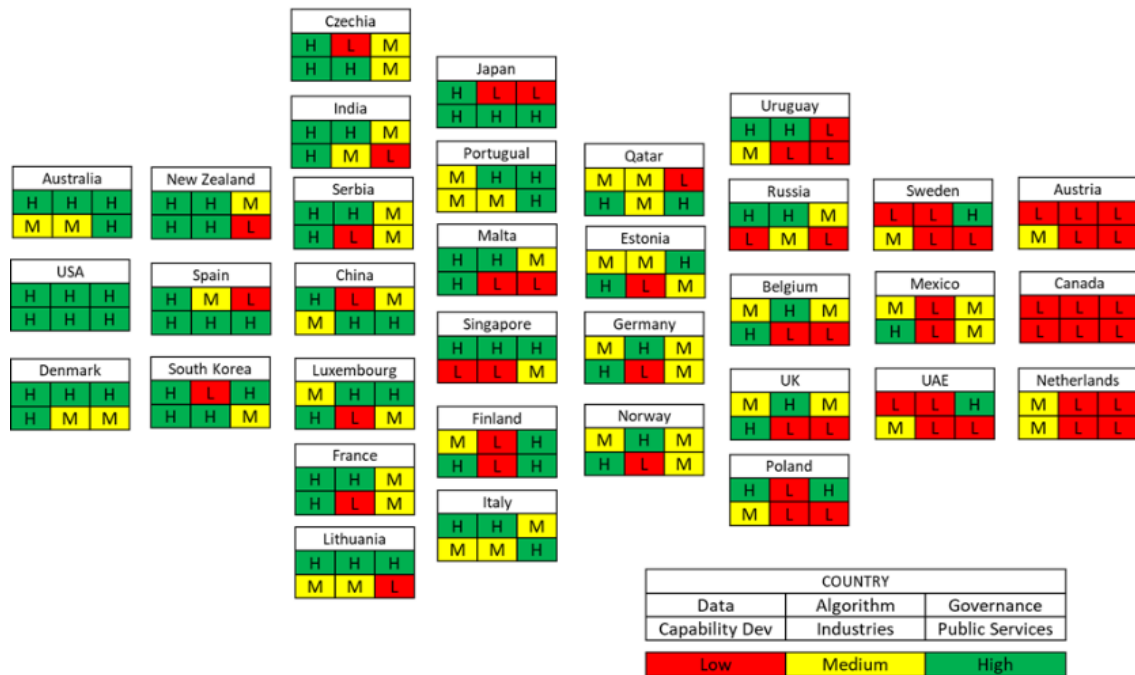


Figure 3.4: Les auteurs visualisent les stratégies nationales en matière d'IA de 34 pays, regroupés par ordre décroissant, de gauche à droite, de ceux qui sont les plus performants dans six attributs spécifiques, collectivement, à ceux qui sont les plus faibles. [Denford 2023].

son groupe de travail sur les technologies des registres distribués et l'IA, orientée vers l'amélioration de la compétitivité et de l'innovation nationales. D'autres pays africains, tels que l'Afrique du Sud, l'Éthiopie, le Ghana, le Maroc, l'Ouganda, le Rwanda et la Tunisie, travaillent également à définir leurs politiques en matière d'IA. Par exemple, le Ghana et l'Ouganda ont contribué au projet "Ethical Policy Frameworks for AI in the Global South", axé sur l'élaboration de réglementations locales sur l'IA [Walter 2024].

3.1.6 IA générative dans le secteur public

Qu'est-ce que l'IA générative ?

L'IA générative (GenAI), tels que ChatGPT (GPT signifie *Generative Pretrained Transformer* : transformeur génératif pré-entraîné) est, selon [Sætra 2023], utilisée comme un terme générique pour décrire les solutions d'apprentissage automatique formées sur des quantités massives de données afin de produire des résultats (sous forme de texts, images, vidéos ...etc) basés sur les invites de l'utilisateur (entrées sous forme

de commandes).

Elle fait référence à un ensemble de technologies d'IA qui sont capables de prédire, logiquement, le mot ou la séquence de mots qui suivent une présentation (ou développement) contextuelle, en se basant sur ce même contexte, le contexte précédent [Yu 2023].

Avant de continuer sur l'état de l'art en matière d'IA générative, il est important de bien comprendre son fonctionnement, pour en voir les intérêts et les limites.

En effet, il faut que tous les utilisateurs aient conscience que “génératif” signifie qu'au moment de l'écriture de cette thèse, le fonctionnement des IA génératives est (du point de vue de l'IA) approximativement le suivant, :

1. Un utilisateur me pose une question, ou écrit un début de phrase par une séquence de mots (en anglais un *prompt*).
2. Je pioche dans les rapprochements entre les mots (appris des millions de textes préalablement analysés lors de mon apprentissage) l'un des mots les plus probables qui apparaissent après la séquence de mots proposée.
3. j'ajoute le mot proposé au *prompt* et je recommence en étape 1. jusqu'à arriver à la taille approximative du texte qu'on me demande d'écrire.

Techniquement, il s'agit donc d'une automatisation à l'échelle de centaines de millions de textes (contenant actuellement environ 10 mille milliards de mots) du principe du jeu d'écriture collective du *Cadavre Exquis*, inventé par Marcel Duhamel, Jacques Prévert et Yves Tanguy vers 1925³.

La différence est que là où dans ses dernières versions, le jeu du Cadavre Exquis demande aux co-auteurs d'écrire une partie de la suite d'un livre sans en connaître l'histoire, mais en ayant connaissance des écrits passés, (un exemple est *l'Exquise Nouvelle ??* où les contributeurs étaient limités par une taille de 420 caractères), une IA générative fonctionne de même, mais à l'échelle plus petite de mots ou de syntagmes⁴. Ainsi, même si la production de texte par une IA générative donne une impression de cohérence, il s'agit d'une cohérence syntaxique poussée mais pas sémantique. L'IA générative ne “comprend pas” ce qu'elle écrit. Il s'agit d'une production probabiliste de texte à partir d'un apprentissage extraordinaire effectué à partir de centaines de millions de textes glanés sur internet. L'inexactitude des contenus produits (appelés “hallucinations”) est donc constitutive de cette IA générative et (au moment de l'écriture de cette thèse) il ne peut en être autrement.

Le contenu produit par une IA générative n'est donc pas issu de sources citables (la source est le réseau de liens entre les milliers de milliards de mots composant les centaines de millions de textes ayant servi à l'apprentissage. . .)

³https://fr.wikipedia.org/wiki/Cadavre_exquis

⁴<https://fr.wikipedia.org/wiki/Syntagme>

Pour citer les sources d'un texte produit par une IA générative, il faudrait citer l'ensemble de tous les textes ayant servi à l'apprentissage, ce qui est matériellement impossible. De plus, ça n'apporterait rien car c'est le lien entre les mots trouvé dans tous les textes qui est utilisé pour créer la suite du texte proposé en entrée.

L'IA générative est présente depuis les années 1960, mais elle a pris de l'ampleur avec l'introduction des réseaux antagonistes génératifs ou Generative Adversarial Networks (GAN) en 2014 [Novo Diaz 2023]. Elle exploite des méthodes d'apprentissage profond, d'apprentissage par transfert et des méthodes d'apprentissage automatique avancées afin d'acquérir des patterns et des relations à partir des données existantes et de produire un nouveau contenu qui ressemble, en style, ton ou structure, à celui de l'instruction/prompt initiale [Yu 2023].

Grâce à ces méthodes, il est en effet possible d'entraîner les modèles d'IA générative sur de grandes quantités de données pour générer un contenu de haute qualité qui ne peut être distingué du contenu généré par l'homme.

Les outils d'IA générative reposent sur des modèles d'IA sous-jacents, comme les grands modèles de langage (LLMs de l'anglais, large language models). Le LLM est un modèle de langage qui possède un grand nombre de paramètres et il est à l'origine de la création du texte dans l'IA générative. Il s'agit de réseaux de neurones profonds entraînés sur de grandes quantités de texte non étiqueté utilisant l'apprentissage auto-supervisé ou l'apprentissage semi-supervisé.

Utilité et impacts de l'IA générative ?

Dans leurs travaux de recherche, [Rane 2023] ont examiné les différentes fonctions de ces modèles d'IA générative dans différentes dimensions de l'ingénierie de la construction et de sa gestion. Ils précisent que ces modèles d'IA générative ont le potentiel d'apporter de profondes transformations et améliorations dans l'industrie de la construction en rendant les processus plus efficaces et durables. Effectivement, Les auteurs soutiennent, par ailleurs, que l'IA générative joue un rôle important dans la planification et la conception urbaines, dans la réflexion portant sur des domaines essentiels de conception et de réalisation des villes intelligentes. En analysant diverses données, telles que les données démographiques, les modèles de trafic et les facteurs environnementaux, ces modèles d'IA permettent aux urbanistes de prendre des décisions éclairées.

De nombreuses études confirment l'utilité de ces modèles d'IA générative et mettent l'accent sur leur impact positif et leurs avantages :

- [Cantens 2023] déclare que depuis le succès de ChatGPT, de nombreux praticiens de différents disciplines telles que le droit, le journalisme, la santé, l'éducation, le monde universitaire, la finance, l'analyse de données ou encore le design, se sont penchés sur la refonte (renouvellement) de leurs pratiques.
- [Cantens 2023] estiment, par ailleurs, que l'intégration massive de l'intelligence artificielle générative dans les administrations publiques, notamment dans les pays

à faible revenu, peut favoriser une plus grande égalité et diminuer les disparités en matière de production intellectuelle entre les autorités.

- diverses industries adoptent les modèles d'IA générative et les utilisent dans différentes applications, telles que l'amélioration des interactions avec les clients, la facilitation des tâches répétitives et l'analyse des données.
- L'IA générative peut aider :
 - les enseignants dans la production de ressources pédagogiques et le suivi des apprenants [Mballo 2023],
 - de nombreux établissements à innover, à améliorer leur expérience, et leur prise de décisions éclairées, mais d'après [Yu 2023], seule une conception réfléchie et un développement responsable de l'IA générative ont le potentiel d'améliorer les compétences humaines dans l'aide à la prise de décision,
 - l'amélioration des services aux citoyens : traitement des demandes courantes des citoyens, par les assistants virtuels pilotés par l'IA [Babu 2024], réduisant ainsi la charge de travail des employés et améliorant la réactivité,
- [Ioannidis 2023] ont démontré que Les LLMs offrent des possibilités remarquables d'améliorer la prise de décision dans le domaine juridique, et mentionne des articles comprenant des explorations générales de l'analyse syntaxique du langage naturel et des LLMs dans le domaine du droit. Ils donnent comme exemple celui de la société d'IA générative [Harvey.ai](#), qui cherche à redéfinir les pratiques juridiques et de conseil de l'élite en utilisant les LLMs d'OpenAI comme moteur de base, ainsi que d'autres exemples juridiques incluant des entreprises axées sur des solutions de rédaction juridique basée sur l'IA générative ou sur la maintenance du cycle de vie des contrats.

Dans le domaine public, les applications qui utilisent les LLMs peuvent aider à accomplir diverses tâches administratives et de prise de décision. La traduction de documents administratifs, la synthèse de documents à des fins d'information, les assistants conversationnels disponibles 24 heures sur 24 et 7 jours sur 7 pour répondre aux questions des citoyens, ainsi que l'analyse de documents politiques ou administratifs peuvent être mentionnés. Les LLMs sont de plus en plus utilisés dans les applications du secteur public en raison de leur potentiel de gains d'efficacité importants, *cf.* cet appel à communications de l'édition spéciale du journal POLITY [Meijer 2024].

Avantages de l'IA générative et des LLMs

Les LLMs peuvent traiter et analyser de grandes quantités de données, en un laps de temps hors de portée de l'humain. [Cantens 2023] mentionne les bénéfices de l'utilisation de l'IA générative dans les domaines administratifs :

- La réduction des coûts : le recours à l'IA générative par l'internalisation des tâches au sein des administrations et des organisations gouvernementales internationales peut aider à réduire les dépenses dans divers secteurs tels que l'édition, la traduction et la production de contenu pour la communication et la formation.
- L'uniformité et la correction du langage administratif : Cet avantage de l'IA générative dans les domaines administratifs consiste en l'amélioration de la qualité et de l'uniformité du style de rédaction, éléments indispensables pour une rédaction très structurée.
- L'accès à des connaissances étendues : l'absence de l'expertise humaine sur certains sujets rend indispensable le recours à l'IA générative.
- La détection des fraudes : la détection des fraudes peut être assurée, d'une part par l'IA spécialisée (dans les cas où les humains doivent analyser du texte), entraînée sur la base des informations antérieures, incluant des fraudes détectées et des fraudes non détectées, appuyée, d'autre part, par l'IA générative entraînée quant à elle, uniquement sur des informations "vraies" (les textes juridiques). Cette démarche de détection des fraudes devra être maximisée par un corpus d'entraînement bien réfléchi.

Inconvénients de l'IA générative et des LLMs

Malgré les avantages de l'IA générative et des LLMs, leur utilisation/mise en place révèlent quelques préoccupations. En effet, ces modèles d'IA générative peuvent générer un contenu erroné ou inexact, et leurs résultats ne sont pas toujours adéquats.^{??,??,??}

Les gouvernements doivent, cependant, prendre des décisions complexes lorsqu'ils mettent en place les LLM. Chaque rôle implique une compréhension nuancée de la technologie et de ses implications, ce qui nécessite une analyse des alternatives et des compromis basée sur leur compréhension actuelle de la technologie, qui peut être insuffisante en raison de sa complexité et de l'accès limité à l'expertise.^{??}

Dans le secteur public, l'introduction des LLM remet en cause tous les principes de l'administration, depuis ses bases juridiques jusqu'à la fourniture des services les plus simples, en passant par la nature du travail d'un fonctionnaire.^{??} Si son introduction offre de nombreux avantages et opportunités, elle exige aussi un équilibre prudent entre les valeurs publiques — à savoir : la confidentialité, la transparence, l'équité et la responsabilité — et, par ailleurs, l'expérimentation et l'innovation.

La compréhension conceptuelle et théorique des LLMs dans la fonction publique est pourtant primordiale. Ceci favorisera la mise en place d'un nouveau champ de recherche universitaire qui répondra aux demandes de formalisation de l'étude de l'IA dans le domaine public. ^{??}

3.2 Ingénierie de prompt

L'ingénierie de l'amorce (ou *prompt engineering* en anglais) est le processus de structuration d'un texte qui peut être interprété et compris par un modèle d'IA génératif. Un prompt (ou amorce) est un texte en langage naturel décrivant la tâche qu'une IA doit effectuer.⁵ L'utilisateur peut indiquer au modèle ce qu'il doit accomplir en saisissant cette suite de mots dans une interface textuelle. Un prompt peut être une simple question ou, à l'inverse, une requête complexe qui contient une variété d'éléments que l'IA doit prendre en compte lors de sa réponse. En outre, les prompts peuvent contenir du texte en langage naturel, des images ou d'autres types de données. Le même prompt peut donner des résultats différents sur différentes plateformes d'IA.

D'après Wikipédia, "L'ingénierie du prompt est associée à la notion d'apprentissage en contexte, ce qui signifie la capacité d'un modèle à apprendre temporairement des informations à partir du prompt, c'est-à-dire du contexte. L'apprentissage en contexte a l'avantage de fournir des informations spécifiques au modèle sans nécessiter de l'entraîner, ce qui serait plus long et coûteux. Le modèle n'est en effet pas modifié, le prompt faisant juste partie des données d'entrée."⁵

L'ingénierie de prompt fait référence aux techniques et aux méthodes spécifiquement conçues pour améliorer les formulations d'instructions dans le domaine du traitement du langage naturel (NLP) et des modèles de grande envergure basés sur l'apprentissage automatique, tels que GPT-3 ou GPT-4 [Kalyan 2023]. L'objectif est d'obtenir des réponses de meilleure qualité, plus précises et plus ciblées, car la manière dont une question ou une instruction est formulée exerce une influence significative sur la qualité et la pertinence des réponses générées. L'ingénierie de prompt est donc cruciale pour l'utilisation efficace des LLMs [Heston 2023].

La créativité et la précision sont indispensables pour l'ingénierie de prompt, mais également une bonne compréhension du modèle linguistique en question, car le choix des mots et leur ordre peuvent avoir un impact significatif sur le résultat.

Les prompts ont été classés comme instructifs, systèmes, questions-réponses et mixtes. Les prompts instructifs commencent par un verbe qui spécifie l'action à effectuer par le système. Les prompts systémiques fournissent au système un point de départ ou un contexte pour développer le contenu. Les prompts de type question-réponse formulent une question de type *wh** (*who, what, when, where, why*). Les prompts mixtes mélangent deux ou plusieurs des techniques mentionnées ci-dessus [Velásquez-Henao 2023].

Les prompts efficaces se caractérisent par les principes fondamentaux suivants [Velásquez-Henao 2023]:

- Clarté et précision : la clarté dans la formulation d'un prompt aide l'IA à mieux

⁵https://fr.wikipedia.org/wiki/Ing%C3%A9nierie_de_prompt

“comprendre” ce qu’on attend d’elle.

- Informations contextuelles : certaines IA donneront un meilleur résultat en communiquant des informations supplémentaires ou un contexte plus précis.
- Utilisation d'exemples : il peut être utile de donner à l'IA un exemple de la réponse souhaitée afin de l'orienter dans la bonne direction.
- Attribution de rôle : L'attribution de rôles est une technique qui peut être utilisée pour contrôler le style du texte généré par l'IA. L'objectif de cette technique est de demander à l'IA de simuler une certaine personne ou d'agir de manière spécifique, ce qui entraîne une modification de sa manière d'écrire en fonction du rôle qui lui a été assigné. Il est possible d'utiliser cette méthode afin de changer le ton, le style et même la profondeur des informations exposées.
- Format souhaité : Afin que la réponse soit donnée dans un format particulier (par exemple, liste, paragraphe court, langage formel), il faut l'indiquer dans le prompt.
- Vérification : il est important d'examiner de manière critique les réponses d'une IA et de s'assurer qu'elles sont à la fois précises et neutres.
- Approche itérative : il est souvent utile d'adopter une approche itérative et d'affiner/reformuler la question sur la base des réponses obtenues.

Revue de la notion d'éthique appliquée à l'IA

Sommaire

4.1 IA et démarches pour une éthique réfléchie	99
4.1.1 Questionnements sur l'IA éthique	99
4.1.2 Revue des documents publiés par des chercheurs indépendants	100
4.1.3 IA et initiatives nationales, régionales et internationales	107
4.2 Cadre national d'éthique en matière d'IA	111
4.2.1 Résumé de l'évolution du Numérique en Mauritanie	113
4.2.2 Cadre national mauritanien de l'éthique de l'IA (résumé par Chat-GPT 4.0)	115
4.2.3 Cadre national de l'éthique de l'IA (version originale)	118

4.1 IA et démarches pour une éthique réfléchie

Dans ce chapitre sur la revue de l'éthique appliquée à l'IA, nous nous intéressons à cette question, sur la problématique qu'elle pose et les aspects publiés par des auteurs scientifiques indépendants, ou soulevés dans la littérature spécialisée par les organismes internationaux (ONU), et ceux intervenant aux échelles continentales (Union Européenne/UE, Union Africaine/UA). De nombreux auteurs se sont penchés sur les questions d'éthique de l'IA et l'abondance de la documentation produite est justifiée par les questions posées.

4.1.1 Questionnements sur l'IA éthique

Si les applications de l'IA ont suscité des espoirs et montré leurs avantages, le développement de cette technologie a aussi été source d'inquiétudes, de menaces et de risques. Cela nous amène à nous interroger sur plusieurs questions d'ordre éthique : "Quel est le rôle de l'éthique dans la conception, le développement et l'application de l'IA

?”; “Comment l'éthique aidera-t-elle à maximiser les avantages de l'IA pour accroître le bien-être des citoyens et le bien commun ?”; “Comment garantir que les systèmes d'IA servent le bien public au lieu d'exacerber les inégalités existantes ?”; “Comment s'assurer que l'usage des algorithmes ne soit source d'injustices et d'inégalités ?”. Face à toutes ces questions sur les avantages mais, aussi, sur les risques et les incertitudes liées à l'IA, un consensus sociétal s'impose, consensus auquel s'attellent les organisations internationales (UNESCO, Conseil de l'Europe, Commission européenne, Agence des droits fondamentaux de l'Union européenne, Banque interaméricaine de développement, OCDE et OECD-AI, Organisation des Nations Unis, et Groupe de la Banque Mondiale).

L'examen de la littérature révèle que l'IA impacte fortement la jouissance des droits de l'homme et des libertés fondamentales, notamment : la désinformation, les bulles de filtre (ou chambres d'écho), le profilage, la discrimination, et la radicalisation, impacts résumés dans [Bensalah 2021].

L'IA, à travers l'apprentissage automatique, peut faire courir le risque d'entretenir certaines exclusions traditionnellement implantées dans les sociétés patriarcales, notamment en Afrique [Gwagwa 2019] : exclusions des jeunes (des jeunes filles en particulier) et des femmes. D'autres menaces sont également pointées par l'auteur : menaces sur la vie privée, les droits individuels et collectifs, compte tenu de l'utilisation croissante de l'identité numérique.

Ce qui précède conduit à la question de savoir quels sont les principes éthiques et les solutions pouvant aider à la mise en place d'une IA plus éthique, en tout cas respectant mieux l'éthique d'un contexte particulier.

4.1.2 Revue des documents publiés par des chercheurs indépendants

L'examen de la documentation issue des chercheurs indépendants (toutes origines confondues) révèle, *à priori*, l'existence de valeurs et principes dont la signification et l'importance (tout au moins pour certains) peut prêter à confusion selon les régions et les continents. Ce qui pourrait entraîner des interprétations et des applications différentes.

Cette situation interpelle : elle appelle à une option basée sur une approche consensuelle autour de ces dimensions de l'éthique de l'IA et répondant, nécessairement, aux principes et axes jugés essentiels. Pour simplifier, nous présentons dans ce qui suit, les principes en contexte européen d'une part, et ceux en contexte africain, d'autre part.

4.1.2.1 Principes et axes jugés essentiels en contexte européen

Les auteurs, en général, tout en reconnaissant les avantages indéniables de l'IA, appellent à la prudence, compte tenu des abus observés contre les droits et la dignité de

l'homme. Ils s'accordent sur la nécessité de faire respecter des principes intangibles.

[Jobin 2019] ont identifié et étudié “84 documents portant sur des principes éthiques ou des lignes directrices pour l'IA”, dont “88% ont été publiés après 2016”, par une grande diversité de parties prenantes. Plus de la moitié de ces documents proviennent des entreprises privées, des agences gouvernementales, des institutions académiques et de recherche, et des organisations intergouvernementales ou supranationales.

Le reste provient d'autres sources diverses (organisations à but non lucratif, associations professionnelles, fondations, syndicats, etc.).

Les données récoltées “montrent une représentation significative des pays plus développés”. L'examen approfondi a permis à [Jobin 2019] d'identifier 11 principes (figure 4.1).

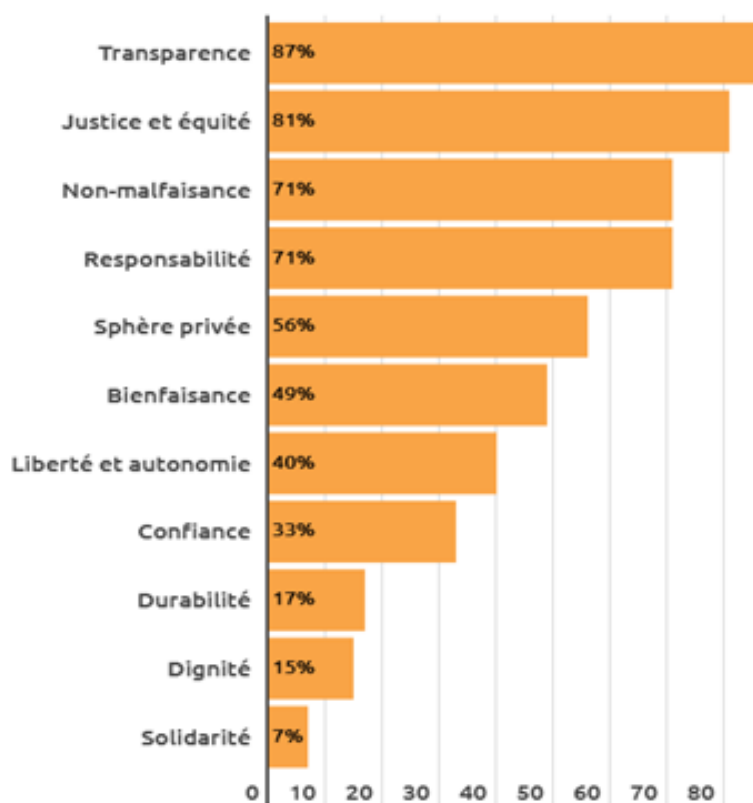


Figure 4.1: Proportion de documents dans lesquels le principe éthique apparaît [Jobin 2019].

Une analyse plus poussée révèle des “divergences sémantiques et conceptuelles importantes” dans l'interprétation des 11 principes. Ces divergences se traduisent dans les recommandations spécifiques et les sujets de préoccupation des uns et des autres.

Table 4.1: Principes éthiques identifiés dans les directives existantes en matière d'IA [Jobin 2019].

Principe éthique	Codes inclus
Transparence	Transparence, explicabilité, compréhensibilité, Interprétabilité, communication, divulgation, démonstration
Justice et Équité	Justice, équité, cohérence, inclusion, égalité, équité, (non-) partialité, (non-) discrimination, diversité, pluralité, accessibilité, réversibilité, recours, réparation, contestation, accès et distribution
Non-malfaisance	Non-malfaisance, sécurité, sûreté, préjudice, protection, précaution, prévention, intégrité (corporelle ou mentale), non-subversion
Responsabilité	Responsabilité, obligation de rendre compte, agir avec intégrité
Confidentialité	Vie privée, informations personnelles ou privées

Ils concluent à une convergence autour de cinq principes : transparence, justice et équité, non-malfaisance, responsabilité et vie privée, référencés dans plus de la moitié des sources. Les codes inclus de ces cinq principes sont représentés dans le tableau 5.1 : Après avoir comparé trente-six principaux documents sur les principes de l'IA, [Fjeld 2020] soulignent “un consensus croissant autour de huit tendances thématiques clés : la confidentialité, la responsabilité, la sûreté et la sécurité, la transparence et l'explicabilité, l'équité et la non-discrimination, le contrôle humain de la technologie, la responsabilité professionnelle et la promotion des valeurs humaines”. Les trois derniers principes, à savoir : le contrôle humain de la technologie, la responsabilité professionnelle et la promotion des valeurs humaines, bien que non expressément cités, se retrouvent dans la déclinaison des cinq principes de consensus cités par [Jobin 2019].

Le groupe d'experts de la CE, dans ses lignes directrices proposées à la Commission Européenne, a énuméré sept exigences essentielles [Tiple 2020] :

- Facteur humain 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,
- Responsabilisation.

[OMS 2021] a proposé six principes directeurs (comparables à ceux énoncés plus haut) qui guideront ses travaux futurs : l'autonomie humaine ; le bien-être et la sécurité des personnes et l'intérêt public ; la transparence, l'explicabilité et l'intelligibilité ; la responsabilité et la responsabilisation ; l'inclusion et l'équité ; et la réactivité et la durabilité.

Mais il est important de noter que ces exigences essentielles émanent de la Commission Européenne. Les exigences essentielles dans un autre contexte pourraient être différentes.

4.1.2.2 Principes et axes jugés essentiels en contexte africain

Pour [Carman 2021] le développement et la mise en œuvre de systèmes d'intelligence artificielle (IA), de manière éthique, sont confrontés à un premier défi majeur : les systèmes de prise de décision utilisant l'apprentissage automatique doivent être justes, équitables et intelligibles, et alignés avec nos valeurs humaines. Mais étant donné que, sous des cieux différents les valeurs éthiques ne sont pas forcément les mêmes et peuvent varier selon les cultures, un deuxième défi est de s'assurer que la conception de ces systèmes soit compatible avec les sociétés dans lesquelles ils opèrent et qu'ils sont censés servir. Ce deuxième principe est d'une importance capitale pour le continent africain qui, comme le souligne [Mialhe 2018], est un terrain de convoitises et d'affrontements des puissances numériques que sont les États-Unis d'Amérique d'abord, la Chine, ensuite, puis l'Occident. Il est intéressant de noter que pour cet auteur, donc, les États-Unis d'Amérique ne font pas partie de l'Occident.

Le cadre européen AI4People¹ défini par la Commission Européenne identifie cinq principes directeurs, principes repris par son livre blanc [Tiple 2020], pour une IA digne de confiance. Quatre de ces principes (autonomie, bienfaisance, non-malfaisance et justice) sont communs. Le cinquième principe, celui de l'explicabilité est, quant à lui, spécifique à l'IA.

Le principe de l'autonomie se réfère à l'idée que nous pouvons, ou non, céder une partie de notre pouvoir de décision aux machines, de trouver un équilibre entre le pouvoir de décision que nous conservons pour nous-mêmes et celui que nous déléguons à des agents artificiels [Floridi 2018]. Se référant aux mêmes auteurs : le principe de bienfaisance signifie "la promotion du bien-être, la préservation de la dignité et la préservation de la planète". Autrement dit, le développement d'une IA qui profite à l'humanité ; le principe de non-malfaisance est de ne pas nuire, ce qui signifie qu'il faut éviter certaines utilisations abusives des technologies de l'IA ; enfin, le quatrième principe (celui de justice) se rapporte à la répartition équitable des biens et des services.

A propos du cinquième principe d'explicabilité, [Carman 2021] s'interrogent sur son

¹<https://rm.coe.int/cahai-2020-08-fin-fr-mantelero-binding-instrument-report-completed/16809eed6d>

application en Afrique. Ils précisent que ce principe doit être compris à la fois sous l'angle épistémologique (“sens de l'intelligibilité”) et sous l'angle éthique (“sens des responsabilités”). Au sens épistémologique, la démarche revient à se poser la question de savoir “Comment ça marche ?” (exigence de transparence) : démarche de compréhension du fonctionnement du système d'IA. Cette exigence de transparence vise à situer la responsabilité en cas de dommages causés suite à une prise de décision du système. Au sens éthique, la démarche revient à se poser la question de savoir “qui est responsable du fonctionnement du système ?”. Cette responsabilité incombe au concepteur et constructeur du système : les entreprises technologiques. Dans la pratique, en général, la vision africaine, communautaire (où la prise de décision est conjointe), est opposée à la vision occidentale où l'individu est au centre de la décision (principe du respect de l'autonomie).

[Carman 2021] donnent des exemples pour montrer que l'application des principes d'autonomie, de bienfaisance, de non-malfaisance et de justice, en contexte africain d'IA, peut être problématique. Ils concluent à la nécessité d'un examen plus approfondi et mettent en garde “contre l'assimilation non critique des valeurs occidentales dans les contextes africains”.

Quant au principe d'explicabilité (spécifique à l'IA), ils se demandent s'il est pertinent et applicable en contexte africain et ce, compte tenu de la disponibilité des données factuelles, en quantité et en qualité, et de la complexité des étapes et des ressources indispensables à l'apprentissage automatique. L'Afrique est un vaste continent, encore dépendant de l'apport international. Le numérique et l'IA y sont encore en développement. Il serait injuste de demander des comptes aux chercheurs africains en IA et de les rendre responsables d'un système qu'ils ne maîtrisent pas complètement.

[Gwagwa 2020] souligne que les risques potentiels de l'IA sont particulièrement aigus dans le monde en développement et plus particulièrement en Afrique. L'IA est “sous contrôle étranger et/ou conçu à l'étranger” par des entreprises technologiques plutôt préoccupées par le souci de “renforcer leur pouvoir économique et social”. Comme le précise [Miller 2019], les applications technologiques ne tiennent pas compte des “facteurs culturels et infrastructurels”. Pour illustrer et qualifier cette situation, plusieurs auteurs utilisent les expressions de “colonisation algorithmique” [Birhane 2020], “colonialisme numérique” [Coleman 2019] et “colonialisme de données” [Couldry 2019].

Le recours à la démarche de [Smith 2018] qui identifie les risques potentiels dans les pays en développement, par opposition à celle de [Calo 2017] qui interroge les défis politiques dans les pays développés a permis à [Gwagwa 2020] d'évoquer les axes suivants, axes jugés essentiels en contexte africain :

- “Équité, partialité et responsabilité”;
- “Diversité culturelle et linguistique”;

- “Perte d’emplois et de recettes fiscales grâce à l’automatisation”;
- “Surveillance et perte de la vie privée”;
- “Démocratie et autodétermination politique”.

[Gwagwa 2019] précise que si “la sécurité, la confidentialité et l’intégrité restent des exigences essentielles”, l’IA doit aller “au-delà de la robustesse technique et de la conformité juridique — y compris l’impact de l’IA sur les droits humains fondamentaux et sur les valeurs sociales et éthiques collectives”.

Équité, partialité et emploi :

Bien que l’Indice Mastercard (MIWE) 2020 des femmes entrepreneures place “11 pays africains en bonne position (score 50 à 60), pour leur accès aux financements et leur éducation quant à l’esprit d’entreprise”², les femmes africaines restent “généralement désavantagées par les biais des données et des algorithmes, qui reflètent et amplifient les inégalités déjà existantes sur le terrain” [Gwagwa 2020] : l’inégalité entre les sexes n’est qu’un aspect d’un ensemble complexe, multidimensionnel, d’inégalités couvrant les différences de revenu, de race, d’emploi, de ressources mais, aussi, d’origines ethnique et sociale.

La marginalisation des femmes est une réalité bien ancrée en Afrique et particulièrement en Afrique subsaharienne (90% de ces pays) où l’emploi informel représente 82,8% (hors agriculture) de l’emploi féminin total et ce, bien que l’informalité soit présente dans tous les pays, quel que soit leur niveau de développement socio-économique [OIT 2019].

En Afrique, sur les cinquante-quatre pays indépendants du continent, seuls cinq ont atteint le niveau technologique leur permettant de recourir à l’automatisation (robotisation) : le Kenya, le Maroc, le Nigéria, l’Égypte et l’Afrique du Sud. Entre 41% et 51,9% des activités professionnelles de ces pays, en matière de technologie, pourraient être automatisées (voir figure 4.2). Ces niveaux sont à peu près comparables à ceux des pays développés. Mais, compte tenu des contextes, la situation de l’emploi n’est pas la même, pour de nombreuses raisons. Près de 54 % de tous les travailleurs en Afrique subsaharienne travaillent dans le secteur agricole [Gwagwa 2020], répartis sur de nombreuses exploitations de plus petite dimension (ce qui n’encourage pas les producteurs, faute d’incitations financières à investir dans les usages de la technologie). Bien que l’IA y présente de nombreux avantages significatifs pour augmenter la productivité, comme le précise l’auteur : “analyse des données, prédiction de la météo, optimisation des calendriers de plantation et de récolte, besoins appropriés en intrants agricoles, etc”, elle n’affecte pas négativement la main-d’œuvre africaine. De plus, les salaires journaliers des ouvriers agricoles, dans cette même région (où la main d’œuvre est abondante et la situation souvent précaire), sont nettement inférieurs à ceux des pays très développés.

²https://www.mastercard.com/news/media/1ulpy5at/ma_miwe-report-2020.pdf

The Countries Where the Potential for Automation is Highest
Percentage of work activities that could be automated by adapting current technology.

AFRICA	ASIA/AUSTRALIA	EUROPE	NORTH AMERICA	SOUTH AMERICA
Kenya 51.9%	Japan 55.7	Czech Rep. 52.2	Mexico 51.8	Peru 53.2
Morocco 50.5	Thailand 54.6	Turkey 50.4	Costa Rica 51.7	Colombia 53.0
Egypt 48.7	Qatar 52.0	Italy 50.3	Barbados 48.7	Brazil 50.1
Nigeria 45.7	South Korea 51.9	Poland 49.5	Canada 47.0	Chile 48.9
South Africa 41.0	Indonesia 51.8	Spain 48.5	U.S. 45.8	Argentina 48.2
	India 51.8	Germany 47.9		
	Malaysia 51.4	Greece 47.8		
	China 51.2	Austria 47.4		
	Russia 50.3	Switzerland 46.7		
	Philippines 47.9	Sweden 46.0		
	U.A.E. 47.3	Netherlands 45.4		
	Oman 46.8	France 43.1		
	Bahrain 46.1	U.K. 42.8		
	Saudi Arabia 46.0	Norway 42.4		
	Australia 44.9			
	Singapore 44.2			
	Kuwait 41.1			

SOURCE MCKINSEY GLOBAL INSTITUTE © HBR.ORG

Figure 4.2: Niveau d'automatisation (robotisation) de certains pays africains : comparaison avec le reste du monde. https://static.latribune.fr/article_body/689839/automatisation.png

Diversité culturelle et linguistique :

Dans son rapport intitulé “Culture, machines et plateformes : l’intelligence artificielle et son impact sur la diversité des expressions culturelles”, rédigé pour le Comité intergouvernemental pour la protection et la promotion de la diversité des expressions culturelles, [Kulesz 2018] écrit ceci :

“Dans l’univers de la technologie dominé par les États-Unis et la Chine et, dans une moindre mesure, par l’Europe, Israël, le Canada, le Japon et la République de Corée, le risque est grand de voir apparaître une nouvelle fracture créative, qui contribuerait à renforcer le retard des pays en développement.”

Ainsi donc, “la logique commerciale des grandes plateformes” pourrait conduire à un appauvrissement des expressions culturelles.

La diversité linguistique est une autre dimension de la diversité culturelle. Les langues africaines (1500 à 2000 langues) sont d’une complexité et d’une variété telles quelles ont encore (pour la plupart d’entre elles) du chemin à parcourir pour bénéficier du traitement du langage naturel.

Le développement d’applications pour le langage naturel est difficile et requière des

techniques et des outils syntaxiques, de la précision dans le langage et un certain niveau de structuration. Par ailleurs, sa réussite passe nécessairement par la disponibilité de données massives pour l'entraînement de la machine.

Dans le cadre de données médicales non structurées, [Sallstrom 2019] précisent que “la technologie doit être adaptable à la langue locale, pour permettre des modifications basées sur différentes langues, structures linguistiques et même accents vocaux”. Ce qui implique, pour éviter toute discrimination, la nécessité “de disposer d’algorithmes sophistiqués de traitement du langage naturel”. Ces auteurs précisent également qu’une “grande partie de la technologie est actuellement développée dans des contextes occidentaux ou asiatiques, leur transfert vers les marchés africains peut s’avérer difficile”.

Surveillance et perte de la vie privée, démocratie et autodétermination politique :

Les nouvelles technologies des objets connectés, bien qu’elles améliorent la qualité de vie des citoyens et la rentabilité des entreprises, facilitent leur surveillance, ce qui conduit à la perte de la vie privée [Khan 2018]. Ces objets connectés (objets physiques comme par exemple : montres, moniteurs médicaux, téléphones intelligents, tablettes, ordinateurs, voitures), posent donc des problèmes éthiques.

Les principes relatifs à la vie privée sont mentionnés dans 97 % des documents de la base de données consultée par [Fjeld 2020]. Ces auteurs pensent que les systèmes d’IA devraient “respecter le droit à la vie privée de chacun aussi bien en termes d’utilisation des données les concernant, aux fins du développement de systèmes technologiques, qu’en termes des moyens qui leur sont offerts d’intervenir sur ces mêmes données et les décisions prises”.

Ces systèmes doivent nécessairement satisfaire à certains critères (intelligibilité, justifiabilité et accessibilité) et “pouvoir être soumis à un examen, un débat et un contrôle démocratiques”.

4.1.3 IA et initiatives nationales, régionales et internationales

4.1.3.1 Déploiement de l’IA et instruments stratégiques et politiques nationaux

Selon l’enquête d’Onuoha, seuls 17 des États membres de l’Union africaine (UA) avaient promulgué une “législation complète sur la protection des données et de la vie privée”³.

Plusieurs pays ont pris des initiatives et engagé une dynamique visant à créer des entités dédiées à l’IA : Conseil de l’IA (Maurice), Centre public d’intelligence artificielle et de robotique (Nigéria), Groupes de travail (Kenya, Tunisie, Ouganda), Commission présidentielle (Afrique du Sud), et des programmes de renforcement des capacités et

³<https://repository.uneca.org/bitstream/handle/10855/48011/b12007018.pdf?sequence=1&isAllowed=y>

de recherche conduits par plusieurs universités (Afrique du Sud, Égypte, Cameroun, Maroc, Sénégal, Lesotho et Éthiopie).

Le Centre international de recherche et de développement (ou International Research Development Centre (IDRC)) et Oxford Insights ont lancé "l'indice de l'état de préparation des gouvernements à l'IA 2020 pour mesurer l'état de préparation des pays à l'intelligence artificielle". C'est le "premier indice du genre qui comprend un indice d'utilisation responsable". Le résultat de ce classement mondial (172 pays) révèle les cinq premiers pays africains : Maurice (45e dans le monde), Afrique du Sud (59e), Seychelles (68e), Kenya (71e) et le Rwanda, 87e [Shearer 2020].

En 2023, l'indice comprend 39 indicateurs répartis sur 10 dimensions dans 3 piliers : Gouvernement, Secteur technologique, et Données et infrastructure, classant 193 pays, contre 181 dans l'itération des années précédentes ⁴.

Les pays qui ont annoncé ou publié des stratégies d'IA en 2023 sont répartis entre l'Amérique latine et les Caraïbes, l'Asie du Sud et l'Asie centrale, l'Afrique subsaharienne, le Moyen-Orient et l'Afrique du Nord. Ces régions ont toujours été sous-représentées dans les travaux sur les stratégies d'IA, moins de 50 % de leurs pays ayant publié une stratégie. Selon ⁴, ces régions obtiennent actuellement des résultats inférieurs à la moyenne mondiale en ce qui concerne le pilier gouvernemental. Il est donc encourageant d'observer les efforts qu'elles déploient pour établir des bases solides pour la gouvernance de l'IA, ce qui indique un potentiel d'amélioration des performances dans ce pilier à l'avenir.

Il convient de souligner que l'année représente la plus grande variété de nouvelles stratégies d'IA. La moitié des stratégies publiées ou annoncées sont issues de pays à revenu faible et en dessous du moyen. Ce qui est une évolution significative, car auparavant, les recherches sur l'IA étaient principalement axées sur les pays à revenu élevé ou en dessus du moyen de la tranche supérieure⁴.

L'Afrique subsaharienne, est confrontée à des défis en matière d'adoption de l'IA par les gouvernements, ce qui se reflète dans les faibles scores moyens de l'indice de la région. Toutefois, des progrès notables ont été accomplis, comme l'élaboration de nouvelles stratégies nationales en matière d'IA au Rwanda, au Sénégal et au Bénin, mettant fin au statut de cinq ans de l'île Maurice en tant qu'unique leader régional en matière de stratégie d'IA. L'île Maurice occupe la première place dans le pilier "Gouvernement", tandis que l'Afrique du Sud excelle dans les piliers "Données et infrastructures" et "Secteur technologique" ⁴.

Bien qu'ils soient des pays à revenu faible ou moyen inférieur, le Rwanda, le Sénégal et le Bénin ont élaboré des stratégies d'IA avec un soutien international, ce qui contraste avec les tendances observées dans d'autres régions où les pays plus riches sont générale-

⁴<https://oxfordinsights.com/wp-content/uploads/2023/12/2023-Government-AI-Readiness-Index-2.pdf>

ment en tête. La collaboration régionale s'intensifie également, comme en témoigne le forum sous-régional UNESCO-Afrique australe sur l'IA, qui a produit la déclaration de Windhoek pour guider la gouvernance et le développement de l'IA.

Les progrès en matière de politiques de protection des données et de transformation numérique sont évidents, le Nigéria ayant mis à jour sa loi sur la protection des données et le Sénégal ayant publié une stratégie nationale sur les données. La coopération internationale en matière de transformation numérique continue de se développer, avec des initiatives des États-Unis, du Royaume-Uni et de la GIZ allemande. Cet effort continu vise à renforcer les aspects fondamentaux nécessaires à une préparation efficace à l'IA.

4.1.3.2 Instruments de promotion de l'IA aux niveaux continental et régional

L'instrument juridique de référence de l'Union Africaine pour l'IA est la Convention de l'UA de 2014 (convention de Malabo) sur la cyber-sécurité et la protection des données à caractère personnel [Union 2014]. 18 pays sur les 55 que compte le continent l'ont signé : Bénin, Tchad, Comores, Congo, Ghana, Guinée-Bissau, Mauritanie, Sierra Leone, São Tomé-et-Príncipe, Zambie, Sénégal, Ile Maurice et Togo. A la mi 2021, seulement 9 l'avaient ratifié : Angola, Ghana, Guinée Conakry, Mozambique, Namibie, Rwanda, Sénégal, Ile Maurice, Togo.

Cette convention vise à créer un cadre normatif et à “renforcer et harmoniser les législations actuelles des Etats membres et des Communautés Economiques Régionales (CER) en matière de TIC”, dans le respect des libertés fondamentales et des droits de l'Homme et des Peuples.

Les objectifs du Comité technique en charge des technologies de la communication et de l'information sont d'examiner les initiatives existantes afin de créer une position africaine commune sur l'IA, renforcer les capacités à l'échelle du continent, et recommander des projets de collaboration.

Le programme “Intelligence artificielle pour le développement en Afrique” (AI4D Africa) qui vise à promouvoir la capacité d'élaboration des politiques africaines en matière d'IA est financé par le Centre de recherches pour le développement international (CRDI) et l'Agence suédoise de coopération internationale pour le développement⁵.

D'autres communautés africaines sont également très actives dans le domaine de l'innovation et de la recherche en matière d'IA, par l'organisation de rencontres annuelles au niveau continental [Gwagwa 2020] : Deep Learning Indaba, Data Science Africa, Data Science Nigeria et le Regional Academic Network on IT Policy (RAN-ITP), parrainé par Microsoft et hébergé par le réseau Research ICT Africa (RIA) basé au Cap.

⁵<https://africa.ai4d.ai/about-ai4d/>

Le Réseau africain des autorités de protection des données personnelles (RAPDP), créé en 2016, en marge du Forum africain sur la protection des données à caractère personnel, regroupe des autorités africaines partageant le même objet central de préoccupation. Ce réseau vise à promouvoir l'échange et la coopération, non seulement entre ses membres mais, aussi, au-delà du continent, pour mieux construire une éthique de l'IA.

4.1.3.3 Leviers d'appui et d'accompagnement pour l'IA au niveau mondial

La dynamique autour de l'IA, est l'œuvre des institutions internationales, notamment : l'Organisation des Nations Unies pour l'éducation, la Science et la Culture (UNESCO), avec sa Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) ; l'Organisation de Coopération et de Développement Economique (OCDE) ; et l'Union Internationale des Télécommunications (UIT).

L'UNESCO a publié le "Premier projet de recommandation sur l'éthique de l'intelligence artificielle"⁶, un instrument normatif international sur l'éthique de l'intelligence artificielle (IA), recommandation faite par les participants à sa 40e Conférence générale en novembre, 2019.

Ce document cible dix "domaines d'action politique" :

1. Évaluation de l'impact éthique
2. Gouvernance et gestion éthiques
3. Politique de données
4. Développement et coopération internationale
5. Environnement et écosystèmes
6. Genre
7. Culture
8. Éducation et recherche
9. Économie et travail
10. Santé et bien-être social

Le règlement général de l'UE sur la protection des données (RGPD ou GDPR en anglais), entré en vigueur en 2018, présente "des enjeux importants à l'égard des pays africains, dans la mesure où il s'appliquera également, dans de nombreuses hypothèses,

⁶https://unesdoc.unesco.org/ark:/48223/pf0000378426_fre

à des responsables de traitement et sous-traitants établis hors de l'Union Européenne"⁷.

Selon [Mannion 2020], aucun pays africain ne dispose de lois nationales conformes au regard du RGPD. Certains, tels que l'Algérie, les Comores ou la République Centrafricaine, ne disposent en réalité même d'aucun texte légal spécifiquement dédié à la protection des données à caractère personnel. Et même si un pays africain dispose de lois sur la protection des données, d'autres préoccupations - comme la corruption et la sécurité publique - ont souvent la priorité. Un grand nombre de ces pays sont confrontés à des facteurs importants qui les empêchent de se conformer au GDPR, notamment un niveau de sophistication technologique plus faible, peu de services d'appui et des organes législatifs turbulents.

Dans une revue récente de littérature, [Bensalah 2021] aborde la question de l'interférence de l'IA dans la jouissance des droits de l'homme et insiste sur la nécessité de créer un cadre de concertation pour proposer un code éthique axé sur les droits de l'homme qu'il estime suffisamment impactés par les mauvaises pratiques de l'IA. Ce code devra s'inspirer des principes de transparence et confiance, équité, sécurité, liberté et autonomie humaine, responsabilité et justice, dignité et intégrité, durabilité, solidarité et inclusion.

Parmi les axes choisis pour illustrer ces interférences, figurent la désinformation, les bulles de filtre (chambres d'écho), le profilage, la discrimination, la radicalisation, l'addiction et les théories du complot.

4.2 Cadre national d'éthique en matière d'IA

Qu'est-ce que l'éthique ? Le mot éthique du mot grec *ethikos*, "qui concerne les mœurs", lui-même dérivé de *éthos*, qui signifie "état habituel" et *éthos*, ("caractère"). L'éthique est une partie de la philosophie qui traite des problèmes fondamentaux de la morale (morale vient du latin *moralis*, "relatif aux mœurs"). Les deux termes, éthique et morale, ont le même sens et relèvent tous les deux) d'une discipline qui dit ce qui est bien ou mal en matière de conduite. C'est ce qui explique l'émergence de termes et expressions comme bioéthique, éthique de l'environnement, éthique déontologique (fondée sur le devoir et des principes a priori), éthique conséquentialiste (fondée sur la situation souhaitable) et éthique des vertus (fondée sur la qualité morale de l'acte et de la personne qui l'accomplit).

Pourquoi l'IA a-t-elle besoin d'éthique? En guise de rappel, l'IA est en train de révolutionner notre monde. L'essor rapide de cette technologie a créé de nombreuses opportunités au niveau mondial, qu'il s'agisse de faciliter les diagnostics de santé, de

⁷https://ec.europa.eu/commission/presscorner/detail/en/MEMO_18_387

permettre des interactions humaines par le biais des réseaux sociaux ou d'améliorer le rendement au niveau du travail grâce à l'automatisation des tâches.

Toutefois, ces avancées technologiques soulèvent également de profondes préoccupations éthiques. Celles-ci découlent du potentiel des systèmes d'IA à intégrer des biais, à contribuer à la dégradation du climat, à menacer les droits de l'homme, à reproduire les préjugés et les discriminations du monde réel et d'alimenter les divisions⁸. La question de l'éthique est devenue alors essentielle.

L'IA a besoin d'éthique pour garantir que les systèmes intelligents respectent les droits de l'homme et s'assurent que leurs décisions sont éthiques et légales. Ci-dessous 5 exemples qui montrent que l'IA a besoin de recourir à l'éthique ??:

1. **Utilisation des données personnelles :** L'IA est en train de révolutionner la manière dont les données personnelles sont traitées et utilisées. Il est primordial que la conception des systèmes intelligents soit conforme aux lois et aux règlements en matière de protection des données et de confidentialité.
2. **Désinformation et propagande :** On utilise fréquemment les systèmes intelligents pour propager de la désinformation et de la propagande. Il est essentiel de concevoir la technologie de façon à prévenir la propagation de la désinformation et à combattre la manipulation et le mensonge sur les réseaux sociaux.
3. **Utilisation des algorithmes :** Les systèmes intelligents peuvent servir à effectuer des choix automatisés et sans tenir compte des émotions humaines. Il est essentiel de concevoir les algorithmes de manière à respecter les droits de l'homme et à assurer que les décisions sont équitables et justes d'un point de vue humain (différence entre éthique conséquentialiste et éthique déontologique).
4. **Dangers de la surveillance excessive :** Il est possible d'utiliser les systèmes intelligents afin de surveiller les individus et leur comportement. Il est primordial de garantir la conception des systèmes intelligents afin d'éviter une surveillance excessive et de préserver la vie privée.
5. **Utilisation des armes autonomes :** Il est possible d'utiliser les systèmes intelligents afin d'automatiser les armes militaires. Il est primordial que la conception des systèmes intelligents soit conforme aux lois et aux règlements internationaux et écarte les erreurs humaines.

“La Recommandation sur l'éthique de l'intelligence artificielle”, le premier instrument normatif mondial sur le sujet, a été adoptée en novembre 2021 par les 193 États membres de l'UNESCO réunis en Conférence générale. La base de la Recommandation repose sur la préservation des droits de l'homme et de la dignité, en s'appuyant

⁸<https://www.unesco.org/fr/artificial-intelligence/recommandation-ethics>

sur des principes essentiels tels que la transparence et l'équité, sans jamais négliger l'importance de la responsabilité humaine dans le contrôle des systèmes d'intelligence artificielle. Toutefois, la Recommandation est particulièrement applicable en raison de ses multiples domaines d'action stratégique, qui offrent aux décideurs politiques la possibilité de mettre en pratique les valeurs et principes fondamentaux en ce qui concerne notamment la gouvernance des données, l'environnement et les écosystèmes, les genres, l'éducation et la recherche, ainsi que la santé et le bien-être social⁹.

Depuis plus de deux décennies, la République Islamique de Mauritanie, s'est engagée dans un processus de réflexion, de conception et d'élaboration de la documentation portant sur les Technologies de l'Information et de la Communication (TIC). Cette démarche marque la volonté de placer le citoyen au cœur des préoccupations de l'État : lui créer les conditions pour disposer de l'information utile et "des services indispensables à l'amélioration de ses conditions de vie et de sa productivité" (Stratégie Nationale de Modernisation de l'Administration et des TICs). Avant d'aborder la question de l'éthique de l'IA, nous proposons ici un résumé de l'évolution du Numérique en Mauritanie¹⁰.

4.2.1 Résumé de l'évolution du Numérique en Mauritanie

De nombreuses initiatives et la création de diverses structures furent le point de départ de transformations profondes :

- 1996 : création du Top Level Domain auprès du *National Informatics Centre* (NIC), dont la gestion du domaine (.mr) est confiée à l'université ;
- 1997 : création des premiers liens Internet gérés par l'Office des Postes et Télécommunications (OPT) et mise en place des premiers Fournisseurs d'Accès Internet (FAI) ;
- 1998 : lancement de la publication du portail national <http://www.mauritania.mr> ;
- 1999 : création de l'Autorité de Régulation (ARE) et lancement du Système Biométrique de Gestion de la Carte Nationale d'Identité (CNI).
- 2000 : création d'un CyberForum pour la société civile ; création d'un Secrétariat d'État chargé des Technologies Nouvelles ;
- 2001 : mise en place de liens de connexion dans les wilayas (régions de Mauritanie) avec une extension et développement du réseau ;

⁹<https://www.unesco.org/fr/artificial-intelligence/recommendation-ethics>

¹⁰https://mtnima.gov.mr/sites/default/files/piecesjointes/Strat%C3%A9gie%20Nationale%20des%20TIC%201-%20Programme%20initial_2.pdf

- 2002 : élaboration de la Stratégie Nationale de Développement de Technologies Nouvelles
 - création de l'Agence de Promotion de l'Accès Universel aux Services (APAUS) pour favoriser le développement des services de base (eau, électricité, technologies de l'information et de la communication et services de la poste) en milieu rural et dans les zones éligibles à l'accès universel ;
 - Mise en place d'un réseau reliant les principaux hôpitaux de la capitale et ceux de l'intérieur pour faciliter les échanges entre les spécialistes en vue d'améliorer la qualité des traitements des patients (Projet télémédecine) ;
 - Mise en place d'un Centre de Formation et d'Echange à Distance (CFED) équipé et intervenant dans de multiples domaines qui intéressent aussi bien le secteur public que le secteur privé et la société civile ;
 - mise en place du Portail Mauritanien pour le Développement (PMD) pour favoriser l'utilisation de la toile numérique et encourager ce domaine ;
- 2005 : mise en place de l'Intranet haut débit du Gouvernement et du Datacenter;
- 2006 : Développement de l'Internet à domicile avec la mise en place des accès ADSL ; apparition d'un troisième opérateur télécoms (Chinguitel) ; fin du monopole de l'accès Internet de Mauritel et attribution de licences à Chinguitel ; attribution de licences 3G ; évaluation de la situation de la mise en œuvre de la Stratégie Nationale de Développement de Technologies Nouvelles.
- 2007 ; création du Ministère de la Fonctions Publique et de la Modernisation de l'Administration (MFPMA) chargé de piloter le développement des TIC et la modernisation de l'Administration ; création de la Direction Générale de l'Informatique de l'Administration au sein du MFPMA ; rattachement du secteur des TIC au Ministère de l'Hydraulique, de l'Energie et des TIC (MHETIC), dans lequel ont été regroupées plusieurs composantes de TIC ; lancement de la deuxième phase d'ouverture du secteur des télécommunications qui a conduit en juillet 2007 à l'attribution de licences globales (3 licences). Acquisition d'une caravane internet (bus équipé en matériel informatique et télécommunication) aux besoins de la vulgarisation des TIC.
- 2008 : apparition de l'Internet mobile 3G ; développement des sites Internet des départements ministériels ; apparition des premiers sites orientés citoyen (2 sites) ; lancement du premier bureau d'accueil ; création du Secrétariat d'État chargé de la Modernisation de l'Administration et des TIC.
- 2009 : lancement du Schéma Directeur Informatique de l'Administration ; introduction de la Gestion Electronique des Documents dans l'Administration ;

création du Ministère délégué auprès du Premier Ministre chargé de la Modernisation de l'Administration et des TIC ; extension du Réseau Intranet haut débit de l'Administration ; renforcement des capacités du Datacenter.

- 2010 : lancement du Programme de Fiabilisation et sécurisation des Documents Nationaux d'Identification ; rattachement des TIC au Ministère de l'Emploi, de la Formation Professionnelle et des Nouvelles Technologies (MEFPNT) ; lancement du Système de Gestion des Impôts ; lancement de la Stratégie de Modernisation de l'Administration et des TIC ; création d'une Agence Nationale du Registre des Populations et des Titres Sécurisés ; lancement du Système Intégré de Gestion des Ressources Humaines ; lancement du projet de connectivité Internationale et création de Mauritanie Internationale de Télécommunication (MIT).

4.2.2 Cadre national mauritanien de l'éthique de l'IA (résumé par Chat-GPT 4.0)

L'État mauritanien a élaboré plusieurs textes juridiques (décrets, lois) que nous présentons ci-dessous dans l'ordre chronologique de leur apparition :

Loi n° 2017-020 Protection-des-données-personnelles :¹¹

Cette loi a pour objet de “mettre en place un cadre normatif et institutionnel pour le traitement de données à caractère personnel et ce, en vue de garantir de meilleurs services et de protéger contre les atteintes à la vie privée, susceptibles d'être occasionnées par l'utilisation des Technologies de l'Information et de la Communication. Elle pose les conditions dans lesquelles tout traitement portant sur des données à caractère personnel, sous quelque forme que ce soit, respecte les libertés et les droits fondamentaux des citoyens”.

Elle définit plus d'une dizaine de termes expressions allant de “Code de conduite” au “Traitement des données à caractère personnel” en passant par les expressions “Communication électronique” et les différentes formes de données et fichiers.

Plusieurs sections y sont traitées : champ d'application la loi ; Principes de base relatifs au traitement des données à caractère personnel ; principes spécifiques au traitement de certaines catégories de données à caractère personne ; principes spécifiques au transfert des données à caractère personnel vers un pays tiers ; Interconnexions des fichiers comportant des données à caractère personnel ; des sections traitant des formalités préalables au traitement des données à caractère personnel ; des sections relatives aux obligations relatives aux conditions de traitements des données à caractère personnel ; des sections relatives aux droits conférés aux personnes dont les données à caractère personnel font l'objet d'un traitement ; des sections concernant la régulation en matière de protection des données à caractère personne ;

¹¹<https://www.afapdp.org/wp-content/uploads/2020/02/Mauritanie-Loi-2017-020.pdf>

Élaboration des décrets d'application relatifs au cadre juridique de la société mauritanienne de l'information (CJSMI) en 2019¹²

Ce rapport précise que des lois ont été adoptées pour jeter les fondements juridiques et institutionnels de la Société de l'Information Mauritanienne, à savoir :

- Loi n° 2016-006 d'Orientation relative à la Société Mauritanienne de l'Information (SMI),
- Loi n° 2016-007 sur la cybercriminalité,
- Loi n° 2017-020 sur les données personnelles,
- Loi sur les transactions électroniques n° 2018-022.

Les 4 lois citées constituent l'ensemble du Cadre Juridique de la Société Mauritanienne de l'Information (CJSMI).

Une description rapide est faite ci-dessous des décrets d'applications concernant l'IA en mauritanie :

Décret n° 2022 – 13 /PM relatif à la composition, l'organisation et le fonctionnement de l'Autorité de Protection des Données à caractère personnel :¹³

Ce décret précise les modalités de désignation des membres de l'APD. Celle-ci est composée de neuf (9) membres, y compris son Président, choisis en raison de leur compétence juridique et/ou technique et nommés par le Président de la République selon les critères suivants :

- Deux parlementaires sur proposition du Président de l'Assemblée Nationale.
- Trois personnalités qualifiées pour leur connaissance de l'informatique et/ou des nouvelles technologies de l'information, parmi lesquels :
 - Un membre sur proposition du Premier ministre.
 - Un membre sur proposition du Ministre chargé de la Transition Numérique, de l'Innovation et de la Modernisation de l'Administration.
 - Un membre sur proposition du Ministre chargé du Commerce et de l'Industrie.
- Un magistrat membre de la Cour Suprême désigné sur proposition du Président de la Cour Suprême.
- Un avocat sur proposition du Bâtonnier de l'Ordre des Avocats.

¹²https://www.mtnima.gov.mr/sites/default/files/piecesjointes/20191025_rapport_cjsmi_final.pdf

¹³<https://mtnima.gov.mr/sites/default/files/20210620%20Decret%20APD%20FR.docx>

Décret n° 2022 021/PM relatif à la conservation des données électroniques et au filtrage¹⁴ :

Ce décret concerne la conservation et le filtrage des données électroniques en application de la loi n°2018-022 sur les transactions électroniques. Il établit les obligations des fournisseurs de services en matière de conservation et de filtrage des données.

En février 2024, la Direction de la Stratégie et de la Coopération du Ministère de la Transformation Numérique, de l'Innovation et de la Modernisation de l'Administration (MTNIMA) a publié une note conceptuelle pour une stratégie nationale de l'IA, inspirée de diverses stratégies internationales. La mise en œuvre de cette stratégie se déroule en trois phases :

1. Évaluation de la situation actuelle :

- Identification des forces et faiblesses de la Mauritanie en IA.
- Évaluation du contexte stratégique, y compris les attitudes du public et le cadre juridique.

2. Définition des missions et objectifs :

- Définir les objectifs à atteindre avec l'IA et les impacts économiques et sociaux attendus.

3. Mise en œuvre des activités stratégiques :

- Mise en place d'une infrastructure IA fiable et incitations financières.
- Utilisation de l'IA pour fournir des services publics anticipatifs et personnalisés.
- Soutien à tous les secteurs clés (santé, éducation, agriculture, etc.).
- Promotion de l'E-business et de l'innovation.
- Adaptation et surveillance des cadres réglementaires.
- Renforcement des capacités humaines en IA et intégration de l'IA dans le cursus académique.
- Conformité avec les principes éthiques et transparence.
- Coopération internationale pour promouvoir un usage bénéfique de l'IA et relever les défis associés.

Ces initiatives montrent l'engagement de la Mauritanie à encadrer l'IA de manière éthique et à protéger les données personnelles de ses citoyens tout en favorisant l'innovation et le développement technologique.

¹⁴<https://www.apd.mr/fr/wp-content/uploads/2023/10/Decret-N%C2%B0-2022-021-PM-relatif-a-la-conservation-des-donnees-electroniques-et-au-filtrage.pdf>

4.2.3 Cadre national de l'éthique de l'IA (version originale)

L'État mauritanien a élaboré plusieurs textes juridiques (décrets, lois) que nous présentons ci-dessous dans l'ordre chronologique de leur apparition :

Loi n° 2017-020 Protection-des-données-personnelles :¹⁵

Cette loi a pour objet de “mettre en place un cadre normatif et institutionnel pour le traitement de données à caractère personnel et ce, en vue de garantir de meilleurs services et de protéger contre les atteintes à la vie privée, susceptibles d'être occasionnées par l'utilisation des Technologies de l'Information et de la Communication. Elle pose les conditions dans lesquelles tout traitement portant sur des données à caractère personnel, sous quelque forme que ce soit, respecte les libertés et les droits fondamentaux des citoyens”.

Elle définit plus d'une dizaine de termes expressions allant de “Code de conduite” au “Traitement des données à caractère personnel” en passant par les expressions “Communication électronique” et les différentes formes de données et fichiers.

Plusieurs sections y sont traitées : champ d'application la loi ; Principes de base relatifs au traitement des données à caractère personnel ; principes spécifiques au traitement de certaines catégories de données à caractère personnel ; principes spécifiques au transfert des données à caractère personnel vers un pays tiers ; Interconnexions des fichiers comportant des données à caractère personnel ; des sections traitant des formalités préalables au traitement des données à caractère personnel ; des sections relatives aux obligations relatives aux conditions de traitements des données à caractère personnel ; des sections relatives aux droits conférés aux personnes dont les données à caractère personnel font l'objet d'un traitement ; des sections concernant la régulation en matière de protection des données à caractère personnel ;

Élaboration des décrets d'application relatifs au cadre juridique de la société mauritanienne de l'information (CJSMI) en 2019¹⁶

Ce rapport précise que des lois ont été adoptées pour jeter les fondements juridiques et institutionnels de la Société de l'Information Mauritanienne, à savoir :

- Loi n° 2016-006 d'Orientation relative à la Société Mauritanienne de l'Information (SMI),
- Loi n° 2016-007 sur la cybercriminalité,
- Loi n° 2017-020 sur les données personnelles,
- Loi sur les transactions électroniques n° 2018-022.

¹⁵<https://www.afapdp.org/wp-content/uploads/2020/02/Mauritanie-Loi-2017-020.pdf>

¹⁶https://www.mtnima.gov.mr/sites/default/files/piecesjointes/20191025_rapport_cjsmi_final.pdf

Les 4 lois citées constituent l'ensemble du Cadre Juridique de la Société Mauritanienne de l'Information (CJSMI).

Décret n° 2022 – 13 /PM relatif à la composition, l'organisation et le fonctionnement de l'Autorité de protection des données à caractère personnel :¹⁷

Ce décret fixe la composition, l'organisation et le fonctionnement de l'APD, conformément à l'article 64 de la loi n° 2017-020 du 22 juillet 2017 sur la protection des données à caractère personnel. Il précise les modalités de désignation des membres de l'APD. Celle-ci est composée de sept (7) neuf (9) membres dont son Président, choisis en raison de leur compétence juridique et/ou technique et nommés par le Président de la République comme suit.

- Deux (2) parlementaires sur proposition du Président de l'Assemblée Nationale,
- Trois (3) personnalités qualifiées pour leur connaissance de l'informatique et/ou des nouvelles technologies de l'information parmi lesquels :
 - Un membre sur proposition du Premier ministre;
 - Un membre sur proposition du Ministre chargé de la Transition Numérique, de l'innovation et de la Modernisation de l'Administration;
 - Un membre sur proposition du Ministre chargé du commerce et de l'industrie
- Un magistrat membre de la Cour Suprême désigné sur proposition du Président de la Cour suprême ;
- Un avocat sur proposition du Bâtonnier de l'Ordre des avocats ;

Décret n° 2022 021/PM relatif à la conservation des données électroniques et au filtrage¹⁸

Ce décret est pris pour l'application des articles 25, 26, 33 et 36 de la Loi n° 2018 -022 portant sur les transactions électroniques et définit les obligations des fournisseurs de services tels que définis par la loi susmentionnée en matière de conservation de données et le cas échéant de filtrage.

La Direction de la Stratégie et de la Coopération du Ministère de la Transformation Numérique, de l'Innovation et de la Modernisation de l'Administration (MTNIMA) a élaboré récemment (février 2024) une note conceptuelle intitulée "Cadre pour la mise en place d'une stratégie nationale de l'IA".¹⁹ Ce document, inspiré, entre autres, de ("An overview of national AI strategies and policies" et "Artificial intelligence and its use in

¹⁷<https://mtnima.gov.mr/sites/default/files/20210620%20Decret%20APD%20FR.docx>

¹⁸<https://www.apd.mr/fr/wp-content/uploads/2023/10/Decret-N%C2%B0-2022-021-PM-relatif-a-la-conservation-des-donnees-electroniques-et-au-filtrage.pdf>

¹⁹<https://mtnima.gov.mr/sites/default/files/MTNIMA-DSC%20note%20conceptuelle-IA.pdf>

the public sector”) expose le processus de mise en place de cette stratégie en trois phases :

- Première phase :
 - L'évaluation de la situation actuelle : Identification des forces et faiblesses de la Mauritanie, y compris les capacités internes en IA, les données gouvernementales et externes, et les projets existants en IA et en science des données ;
 - L'évaluation du contexte stratégique en tenant compte des attitudes du public des moyens humains, du cadre juridique et réglementaire, des engagements gouvernementaux, des institutions et des expertises externes.
- Deuxième phase : la définition des missions à réaliser, des objectifs à atteindre avec l'IA, la valeur ajoutée, et la définition des impacts économiques et sociaux attendus (notamment sur le PIB l'emploi).
- Troisième phase : la mise en œuvre des activités souhaitées pour le cadre stratégique. Ces activités concernent :
 - La mise en place de l'écosystème IA adapté : une infrastructure fiable comprenant des réseaux haut débit, une capacité de calcul informatique et une capacité de stockage des données, des environnements de test pour l'expérimentation des systèmes d'IA, des incitations financières pour les projets d'IA (plates-formes de collaboration et de conseils pour soutenir la croissance des entreprises, et un environnement politique favorable ;
 - Une Intelligence Artificielle au service de l'administration publique : une IA pour le secteur public, visant à fournir des “services anticipatifs et personnalisés à tous les citoyens à toutes les étapes de leur vie pour soutenir une société bien fonctionnelle”;
 - Une IA soutenant tous les secteurs clés de développement (la santé, l'éducation, l'agriculture, l'élevage, le transport, etc ;
 - Une IA pour l'E-business et l'Innovation : incitations financières, encouragement et attrait des talents en IA, offres de bourses, promotion et investissement dans la R&D en IA avec, comme cibles des secteurs prioritaires (énergie, santé, agriculture, etc.) ;
 - Une réglementation fiable et adaptée : examen et adaptation des cadres réglementaires, surveillance et suivi/évaluation de la mise en œuvre des politiques nationales en matière d'IA, qualité des données, publication de rapports annuels, création d'observatoires de l'IA, large ouverture et partage des données du secteur public, promotion du partage de données dans le secteur privé,

mise en place de “cadres légaux, éthiques et techniques pour opérationnaliser les principes”, ouverture et flexibilité ;

- Un renforcement des capacités humaines en IA : promouvoir des pépites de talents, développement des “plans de recrutement et de rétention pour l’expertise technique interne, en exploitant l’expertise externe par le biais de partenariats”, et introduction de la discipline de l’IA dans le cursus académique ;
 - Une conformité avec les principes : suivre la conformité avec les principes tout au long de la mise en œuvre de la stratégie, mettre en place des protections contre les biais et l’injustice, clarifier le rôle approprié des humains dans le processus de prise de décision, développer des structures de responsabilité ouvertes et transparentes, et se préparer aux évolutions futures :
Engager les parties prenantes dans la définition des objectifs, favoriser l’expérimentation et l’apprentissage, et communiquer les principes d’utilisation de l’IA, tels que l’équité, la transparence, la responsabilité, la vie privée et l’autonomie individuelle.
 - Le dernier volet d’activités concerne la coopération internationale en matière d’IA. Il porte sur des collaborations multiformes (bilatérales, multilatérales et internationales) pouvant aider à promouvoir un usage bénéfique de l’IA et à relever les défis qui lui sont associés.
 - Un membre sur proposition du Premier ministre;
 - Un membre sur proposition du Ministre chargé de la Transition Numérique, de l’innovation et de la Modernisation de l’Administration;
 - Un membre sur proposition du Ministre chargé du commerce et de l’industrie
- Un magistrat membre de la Cour Suprême désigné sur proposition du Président de la Cour suprême ;
 - Un avocat sur proposition du Bâtonnier de l’Ordre des avocats ;

Partie II

Contributions scientifiques

Système de prédiction de courte séries temporelle basé sur programmation génétique combinée avec le filtre de Kalman

Sommaire

5.1	Problématique de disponibilité des séries de données temporelles	127
5.2	Matériels et méthodes	128
5.2.1	Données utilisées	128
5.2.2	Techniques utilisées dans la méthode proposée	131
5.2.3	Description de la méthode proposée	138
5.2.4	Données utilisées	139
5.3	Expériences et résultats	142
5.3.1	Données d'expérimentation	142
5.3.2	Paramètres de l'algorithme de programmation génétique	143
5.3.3	Paramètres du Filtre de Kalman	143
5.3.4	Modèles de référence	143
5.3.5	Métriques de performance	144
5.3.6	Résultat et Analyse	144
5.3.7	Conclusion	145

A titre introductif, cette contribution repose sur la synergie entre deux approches : la programmation génétique et le filtre de Kalman :

1. La programmation génétique, (inspirée de l'évolution biologique), s'avère efficace pour générer des modèles prédictifs optimaux en explorant de vastes espaces de solutions. Dans notre approche, les prédictions issues de la programmation génétique sont utilisées comme distributions a priori dans le cadre bayésien. Cela permet de fournir une base initiale robuste pour les estimations du filtre de Kalman.

Les prédictions générées par la programmation génétique offrent des hypothèses initiales adaptées, qui renforcent la phase de prédiction du filtre de Kalman.

2. Le principe du filtre de Kalman (inspiré des statistiques bayésiennes) utilise une approche bayésienne pour estimer l'état d'un système dynamique en intégrant des informations, à priori avec de nouvelles observations, ce qui est un concept central en statistique bayésienne [Le Gland 2009]. En effet, la théorie bayésienne, fondée sur le théorème de Bayes, permet de mettre à jour les probabilités associées à des hypothèses au fur et à mesure que de nouvelles données deviennent disponibles. Elle est essentielle dans des domaines tels que le filtrage anti-spam ou les diagnostics médicaux¹. Dans sa formulation, le théorème de Bayes repose sur la probabilité conditionnelle² d'un événement, basée sur des informations ou des croyances antérieures sur l'événement ou les conditions liées à l'événement. Contrairement aux méthodes fréquentistes, l'inférence bayésienne attribue des probabilités aux hypothèses, permettant ainsi une gestion plus flexible et adaptative de l'incertitude³. En pratique, le filtre de Kalman fonctionne en deux étapes principales : d'abord, la prédiction ; ensuite, la mise à jour[Fang 2018].

- La prédiction : basée sur l'état précédent et les lois à priori, il anticipe l'état futur du système en tenant compte des incertitudes via une matrice de covariance. Les lois a priori influencent la précision des estimations du filtre de Kalman en fournissant une estimation initiale de l'état du système et de son incertitude, ce qui est crucial pour la phase de prédiction. Ces lois, à priori, sont représentées par des distributions de probabilité qui modélisent les connaissances initiales sur le système.
- Quant à la mise à jour, elle intègre les nouvelles observations, ajuste les estimations, réduisant ainsi l'erreur grâce à la matrice de gain de Kalman [Lemoine 2003, Saidani 2013]. L'algorithme combine ces informations, à priori avec les données mesurées pour fournir une estimation optimisée de l'état présent, tout en ajustant continuellement la précision de ces estimations à chaque itération.

Le filtre de Kalman affine ces prédictions en temps réel en intégrant les données observées, réduisant les erreurs et augmentant la précision des estimations. Ainsi donc, cette combinaison novatrice (Programmation Génétique-filtre de Kalman) est particulièrement adaptée à la prédiction de courtes séries temporelles, où les données disponibles sont limitées et les incertitudes élevées.

¹https://fr.wikipedia.org/wiki/Inf%C3%A9rence_bay%C3%A9sienne

²https://fr.wikipedia.org/wiki/Probabilit%C3%A9_conditionnelle

³https://fr.wikipedia.org/wiki/Probabilit%C3%A9_bay%C3%A9sienne

L'importance et l'utilité de cette contribution se trouve derrière le fait qu'il est important, pour les gouvernements, d'utiliser l'analyse prédictive pour prédire le futur afin de prendre des décisions éclairées et de faire face à tous défis de manière proactive. Dans le domaine de l'éducation, par exemple, les gouvernements peuvent, *via* les modèles prédictifs, étudier les informations concernant les apprenants, à travers des indicateurs pertinents (notamment les notes, les redoublements, le comportement, l'assiduité, etc.), dans le but de repérer les risques d'abandon et de dysfonctionnement afin de mettre en place des mesures correctives.

De plus, l'utilisation de ces modèles peut permettre aux gouvernements de prévoir l'évolution des effectifs scolaires dans le but d'une régulation des flux et de planifier les investissements en termes d'infrastructures, comme la construction de nouvelles écoles/universités ou l'extension d'écoles/universités existantes.

Cependant, le souci réside dans le fait que si l'on veut, par exemple, suivre l'évolution d'indicateurs temporels annuels tels que le nombre de bacheliers, on a forcément un nombre de données limitées. Si l'on souhaite établir des prédictions à partir du nombre de bacheliers depuis 100 ans, il n'y a que 100 points de données, ce qui est très peu pour établir un modèle prédictif. Dans le cas des pays en développement, la situation est encore plus délicate, car les informations numériques gouvernementales ne sont disponibles que depuis beaucoup moins d'années (par exemple, une dizaine d'années dans le cas de la Mauritanie). Il devient donc nécessaire de développer des modèles prédictifs à partir de très petites séries temporelles, ce qui est l'objet de ce chapitre.

Il est essentiel de prédire de manière précise les statistiques d'admission au baccalauréat afin d'améliorer le système éducatif en Mauritanie. Un bon système éducatif étant une condition préalable à une croissance économique réussie, la Mauritanie s'est engagée dans un effort significatif pour augmenter la qualité de l'éducation nationale depuis 1987 [McCulloch 2000]. Malgré les progrès réalisés ces dernières années en termes d'accès global à l'éducation, le système éducatif présente encore de nombreuses lacunes et subit également l'impact de la crise de la pandémie de SRAS-CoV-2 (COVID-19)⁴. Afin de faire face à ces défis, le gouvernement mauritanien a besoin d'un système de recommandation à court terme afin de satisfaire les besoins économiques du pays en matière de financement et de développement de nouveaux programmes et infrastructures d'éducation. Dans cette situation, il est essentiel d'avoir une approche de prédiction fiable des statistiques d'admission au baccalauréat pour proposer ce genre de recommandations.

Le choix de la méthode de prédiction la plus adaptée n'est pas facile, même si de nombreuses techniques de prédiction ont été proposées ces dernières années, car la crédibilité de leurs résultats est grandement influencée par la qualité et la quantité des données historiques disponibles. Les données des statistiques d'admission au bac-

⁴<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/819601592919148037>

calauréat mauritanien ont été collectées depuis 2013 et sont présentées sous la forme de 22 séries temporelles de seulement 10 points (un point correspondant à une année). Chaque série temporelle est associée à un indicateur du taux d'admission au baccalauréat (Baccalaureate Admission Rate ou BAR, en anglais) : par exemple, les types de baccalauréat, les matières, les sexes, etc. Ainsi, nous pouvons décrire le problème que nous abordons comme une prédiction à court terme de séries temporelles très courtes.

La prédiction de séries temporelles très courtes est rarement prise en compte dans la littérature, même si la capacité à produire des prédictions précises avec peu de points de données peut être utile dans de nombreux cas de recherche, notamment lorsqu'il y a très peu de données historiques disponibles sachant que ce cas de figure peut amener à un surapprentissage⁵. On dit qu'il y a surapprentissage lorsque le modèle s'adapte trop étroitement à l'ensemble de données d'apprentissage plutôt que de généraliser. Le modèle "apprend par cœur" le petit nombre des données disponibles. Il existe quelques possibilités courantes pour faire face à un petit ensemble de données afin d'éviter l'effet de surapprentissage⁶: (i) choisir des modèles simples ; (ii) combiner plusieurs modèles [Wanigasekara 2018]; (ii) augmenter les données et générer des données synthétiques [Delgado 2021].

La question qui se pose souvent dans le cas où il n'y a pas assez de données historiques est "combien de points de données peuvent être utilisés pour ajuster un modèle de série temporelle ?" La réponse est que ça dépend du nombre de paramètres du modèle à estimer et de la proportion de biais dans les données [Hyndman 2018]. L'unique limite théorique réside dans le fait qu'il est nécessaire d'avoir plus d'observations que de paramètres dans le modèle de prédiction. Cependant, en pratique, il faut souvent beaucoup plus d'observations que cela. Les règles empiriques qui établissent des tailles d'échantillon minimales pour différents modèles de séries temporelles sont erronées et ne sont justifiées ni en théorie ni en pratique. De plus, elles négligent souvent la variabilité sous-jacente des données et ne prennent pas en considération le nombre de paramètres à estimer. Il n'y a pas suffisamment de données pour que certaines observations puissent être retenues à des fins de test dans le cas de séries courtes, et même la validation croisée des séries temporelles peut être difficile à mettre en œuvre. Dans ce cas de prédiction de courte séries, il n'y a pas suffisamment de données pour que certaines observations puissent être retenues à des fins de test et même la validation croisée des séries temporelles peut être difficile à mettre en œuvre [Hyndman 2018].

Comme en Mauritanie, les informations historiques disponibles concernant l'éducation n'ont été collectées que sur les dix dernières années, une méthode de prédiction précise pour les séries temporelles courtes est indispensable. Dans cette contribution, nous mettons l'accent sur un modèle combiné et proposons un outil de prédiction pratique qui combine la Programmation Génétique (GP) [Koza 1994] avec le Filtre de

⁵<https://fr.wikipedia.org/wiki/Surapprentissage>

⁶<https://www.kaggle.com/code/rafjaa/dealing-with-very-small-datasets#t6>

Kalman (KF) [Kalman 1961] pour améliorer la prédiction des séries temporelles courtes. L'idée de mettre en œuvre le KF sur la base du modèle GP n'est pas nouvelle : le couplage du KF étendu avec la GP a déjà été utilisé pour prédire la demande mensuelle en eau dans [Nasseri 2011]. Toutefois, cette méthode n'a jamais été appliquée à des séries temporelles courtes. Nous avons opté pour cette méthode en raison des caractéristiques de ses composants. La GP est une technique à structure arborescente inspirée de l'évolution [Koza 1994], dont les principaux avantages sont les suivants :

1. elle ne nécessite aucune connaissance spécifique préalable sur la forme fonctionnelle de la solution attendue (par exemple, linéaire ou non linéaire, etc.) ;
2. elle a un petit nombre d'hyperparamètres ;
3. elle fournit une équation, c'est à dire un modèle interprétable.

Cependant, un modèle basé sur la GP peut exhiber un surapprentissage en raison du manque de données d'entraînement. Mais comme tout modèle obtenu par GP peut être représenté sous la forme d'un modèle d'espace d'état, le Filtre de Kalman semble être un bon candidat pour améliorer la précision des résultats des prévisions de la GP. En effet, le KF fonctionne dans le contexte d'un modèle d'espace d'état pour calculer la séquence de filtrage de l'état actuel d'un système dynamique qui est excité par un bruit stochastique. Ainsi, l'utilisation du KF avec la GP pourrait fournir de meilleures prévisions. En compilant tout cela, nous contribuons par deux apports :

1. un nouvel aperçu sur une méthode de couplage GP + KF pour la prévision des séries temporelles courtes ;
2. un nouvel outil de prévision basé sur cette méthode pour l'appliquer au système éducatif en Mauritanie.

L'applicabilité de l'outil présenté est validée par les expériences menées sur les quatre séries temporelles courtes *contre* les techniques classiques de l'état de l'art.

5.1 Problématique de disponibilité des séries de données temporelles

Les modèles d'IA formés à partir de données non locales risquent de ne pas prendre en considération les particularités socioculturelles propres aux régions africaines.⁷ C'est pour cette raison que nous avons choisi de nous concentrer sur des données nationales dans le cadre de cette thèse. Cependant, malgré la présence de nombreuses sources de données, telles que GitHub et Kaggle, l'accessibilité de ces données peut encore être

⁷<https://www.isi-web.org/article/future-ai-statistics-africa-continent-ready>

limitée dans certains cas comme celui de l’Afrique. En effet, en raison de l’absence de mécanismes complets de collecte de données, de nombreuses régions du continent africains sont confrontées à des ensembles de données incomplets : même si les données sont disponibles, leur précision et leur normalisation peuvent être discutables⁷. [Kinyondo 2018] affirme également que les pays en développement manquent de données de qualité.

Les raisons du manque de données et de leur qualité sont multiples. Beaucoup de pays africains ne possèdent pas les éléments nécessaire pour adopter les technologies, tels que les infrastructures, les écosystèmes de données et les systèmes de gouvernance⁷.

En ce qui concerne la Mauritanie, malgré le fait qu’il existe une forte volonté du gouvernement mauritanien de faire du numérique un véritable levier pour accélérer le développement socio-économique du pays⁸, il n’y a toujours pas assez de données historiques.

Ce manque de données historique peut avoir une incidence sur la capacité des chercheurs et des praticiens de ces pays à accéder à des ensembles de données de séries temporelles complets et diversifiés à des fins d’analyse et de prévision. En effet, il est impossible de mener des recherches de qualité sans données de qualité, et sans recherche, il n’y a pas d’innovation, de croissance à partir de l’innovation, ni de développement. Les analystes ne pourront pas fournir les preuves nécessaires aux décideurs politiques pour élaborer des politiques basées sur des preuves et évaluer empiriquement la mise en œuvre de ces politiques en raison du manque de données de qualité [Kinyondo 2018].

L’outil proposé par notre contribution vise à pallier ce problème.

5.2 Matériels et méthodes

Cette section présente les données collectées et utilisées dans ce travail, décrit la méthode de prédiction proposée et ses composantes. Ensuite, l’outil logiciel développé est présenté.

5.2.1 Données utilisées

Nous avons présenté les données que nous avons collectées sous forme de séries temporelles. Par conséquent, nous allons tout d’abord définir les séries temporelles, puis exposer les données utilisées.

⁸<https://mtnima.gov.mr/sites/default/files/Agenda%20Num%C3%A9rique%202022-25%20MTNIMA.pdf>

5.2.1.1 Introduction aux séries temporelles

Une série temporelle, est une série (Y_t) d'observations numériques d'un phénomène donné, ordonnées dans le temps [Jose 2022]. Les séries temporelles sont omniprésentes, car le temps est un élément constitutif de tout ce qui est observable. Étant donné que notre monde est de plus en plus informatisé, les capteurs et les systèmes émettent constamment des données de séries temporelles. Les relevés météorologiques, la croissance de la population et les indicateurs économiques sont tous des données de séries temporelles.

Les composantes des séries temporelles jouent un rôle crucial dans la compréhension et la prévision des tendances. On peut comprendre la variation ou le mouvement d'une série à travers ses composantes clé, qui sont les suivantes : [Jose 2022] :

La Tendance : la tendance est un phénomène de croissance ou de décroissance d'une série au fil du temps. Souvent, elle reflète un processus de croissance ou de décroissance à long terme. Sur une longue période, le mouvement ou la direction peut fluctuer, mais la tendance de fond varie peu. La variation de la population, la production agricole, les taux de natalité ou de mortalité au cours du temps sont des exemples dont on peut tirer des tendances. La tendance peut changer de direction au cours, mais sur une longue période par rapport à la fréquence d'acquisition des données.

Les Variations cycliques : Les variations dans les séries temporelles qui se répètent sont appelées variations cycliques. Ces mouvements oscillatoires ont souvent une durée supérieure à un an. Elles peuvent être de nature régulière ou non périodique. En général, les variations cycliques sont dues à la combinaison de deux ou plusieurs forces économiques et à leurs interactions.

Les variations saisonnières : les variations saisonnières sont un cas particulier de variations cycliques, dont la période est d'un an et fonctionnent donc selon un schéma prévisible pendant une période inférieure à un an. Au cours d'une période de 12 mois, elles présentent presque le même schéma. Dans le cas où les données sont enregistrées chaque heure, chaque jour, chaque semaine, chaque trimestre ou chaque mois, cette différence sera visible dans la chronologie. Les fluctuations saisonnières sont influencées par les différentes saisons ou conditions climatiques.

Les Irrégularités : il s'agit de mouvements purement irréguliers et aléatoire. Ces mouvements sont imprévus, erratiques, imprévisibles et incontrôlables par nature. Les tremblements de terre, les inondations et dans une moindre mesure les famines et les guerres sont des exemples de composantes aléatoires. Les deux derniers exemples sont un peu moins aléatoires que les premiers, car on peut tenter de s'en prémunir par la politique. Par exemple, alors que les famines étaient encore

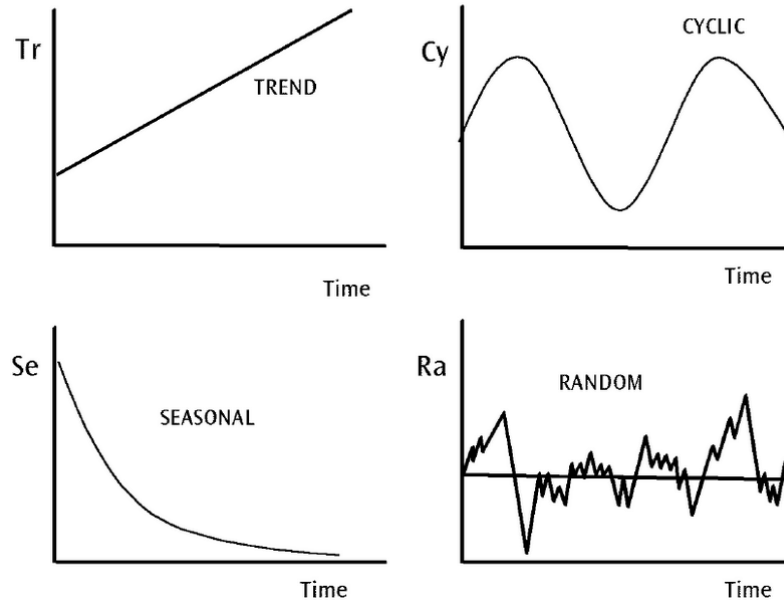


Figure 5.1: Les différentes composantes d'une série temporelle [Jose 2022].

régulières en Europe au moyen-âge, il n'y a pas eu de famine récemment en Europe en dehors de zones de conflit. De même, pour les guerres, on peut tenter d'éviter qu'elles se produisent en créant des accords entre pays voisins. Aucune guerre entre pays de l'Union Européenne n'a eu lieu depuis qu'elle a été créée, là où pour les tremblements de terre, on ne peut rien faire pour les éviter.

Ces composantes (voir figure 5.1) aident à modéliser les données de séries temporelles afin d'obtenir des prédictions précises sur les tendances et les comportements futurs.

Stationnarité d'une série temporelle Une série temporelle est dite stationnaire si les observations ne dépendent pas du temps, c'est-à-dire que ses propriétés statistiques ne changent pas avec le temps et qu'elles ont donc une moyenne, une variance et une covariance constantes. Autrement dit, l'absence de tendance ou de composante saisonnière dans les séries temporelles est également synonyme de stationnarité. Le concept de stationnarité est important dans de nombreuses situations car :

- La stationnarité est souhaitable pour de nombreuses méthodes de prévision, telles que la méthode *Auto-Regressive Integrated Moving Average* ARIMA [LEB 1989], car elle rend les modèles plus simples et plus fiables.

- La stationnarité simplifie les complexités des données des séries temporelles, ce qui les rend plus faciles à modéliser et à prévoir que les séries temporelles non stationnaires.
- Les tendances et les modèles des séries temporelles stationnaires sont plus faciles à interpréter car les relations entre les points de données restent constantes.

Une méthode courante de stationnarisation d'une série temporelle consiste en un processus appelé différenciation, qui peut être utilisé pour supprimer toute tendance dans la série qui ne présente pas d'intérêt [Jose 2022]. En effet, les tendances se traduisent par un changement dans le mouvement des données dans le temps, ce qui entraîne une modification de la moyenne au fil du temps. La saisonnalité, quant à elle, se produit lorsque le modèle de la série temporelle présente une variation pendant un intervalle de temps régulier, ce qui entraîne une modification de la variance au fil du temps [Jose 2022].

Décomposition d'une série temporelle

Le but de la décomposition d'une série temporelle est de distinguer dans l'évolution de la série, une tendance "générale", des variations saisonnières qui se répètent chaque année, et des variations accidentelles imprévisibles. L'intérêt de ceci est d'une part de mieux comprendre, de mieux décrire l'évolution de la série, et d'autre part de prévoir son évolution (à partir de la tendance et des variations saisonnières).

5.2.2 Techniques utilisées dans la méthode proposée

5.2.2.1 Programmation génétique

La programmation génétique est une technique de modélisation inspirée de l'évolution et basée sur une "population" de solutions potentielles (des équations), qui peuvent représenter des relations non linéaires entre certaines entrées et sorties des données observées. Imaginée par Michael Kramer en 1985, elle a été par la suite principalement développée par John Koza pour faire évoluer des programmes [Koza 1994]. En effet, la GP crée un programme ou une fonction mathématique qui minimise la différence entre les données produites par le programme ou la fonction contenue dans un individu et les données que l'on veut modéliser. Les principales étapes de l'évolution sont présentées dans la figure 5.2.

Cette méthode d'apprentissage automatique fait partie de la grande famille de l'intelligence artificielle utilisant les mécanismes évolutifs pour trouver une bonne solution au problème proposé. Dans notre cas, il s'agit de faire évoluer des fonctions par sélection naturelle (darwinienne) pour qu'ils résolvent un problème précis. Comme le montre la figure 5.2. Le processus de programmation génétique commence par une population de fonctions générées aléatoirement. On obtiendra donc un certain nombre

d'individus (ou programmes si l'objectif est de créer un programme) qui permettront de générer les générations futures. Ces fonctions ou programmes sont ensuite évalués par leur leur capacité à résoudre le problème posé. Les programmes les plus performants sont sélectionnés à l'aide de différentes techniques (tournoi, roulette) pour créer de nouveaux individus (des enfants) par croisement et mutation. Enfin, les descendants vont venir remplacer la génération précédente en devenant à leur tour parents, et le cycle reprend alors avec le bloc d'évaluation.

Les solutions potentielles (c'est-à-dire les individus) dans la GP sont dans notre cas des fonctions mathématiques représentées à l'aide d'une structure de données arborescente. Ainsi, un individu contient des opérateurs comme nœuds de l'arbre (qui peuvent être des fonctions arithmétiques, trigonométriques, logarithmiques ou logiques) et des variables comme feuilles.

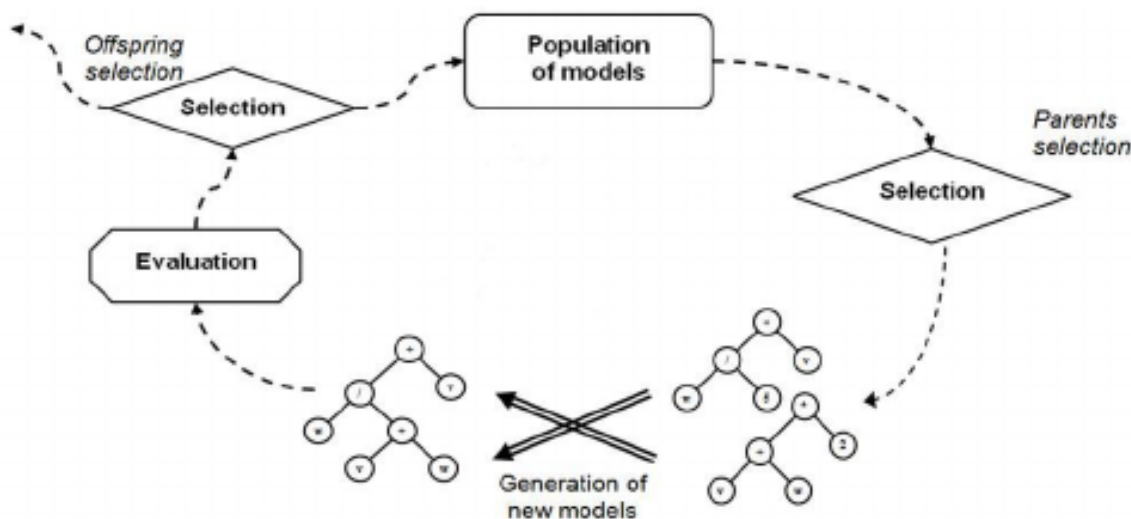


Figure 5.2: Fonctionnement de la Programmation Génétique. L'illustration est empruntée à [Kommenda 2009].

La représentation d'une fonction (ou individu) sous forme d'arbre est présentée en figure 5.3.

Les applications de la programmation génétique sont multiples, ci-dessous quelques unes, présentées dans un cours du CNRS [Assia 2005] :

- problèmes de “magie noire”, tels que la synthèse automatisée des circuits électriques analogues, des contrôleurs, des antennes, des réseaux des réactions chimiques, et d'autres secteurs de conception,
- la “programmation de l'improgrammable” comportant la création automatique des programmes machine pour les dispositifs de calcul peu usuels tels que les auto-

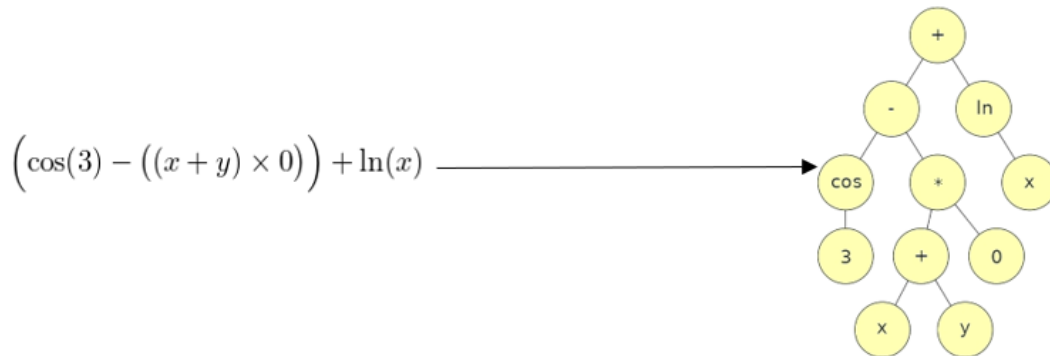


Figure 5.3: Représentation d'un programme (ou Individu).

mates cellulaires, systèmes de multi-agent, systèmes parallèles, rangées de porte field-programmable, rangées field-programmable d'analogie, colonies de fourmi, l'intelligence d'essaim, à distribué des systèmes, et des semblables,

- “nouvelles inventions commercialement utilisables” (CUNI) comportant l'utilisation de la programmation génétique comme “machine d'invention” automatisée pour créer de nouvelles inventions commercialement utilisables.
- Reconnaissance d'images, classification d'images, traitement d'images satellites,
- Prédiction de séries temporelles, génération d'arbres de décisions, datamining,
- Classification de segments d'ADN et de protéines.

Une description détaillée des étapes d'implémentation de l'algorithme de GP classique est présenté ci-dessous :

1. Il faut commencer par spécifier un ensemble de terminaux pour chaque série temporelle, qui consiste en :
 - Une ou plusieurs variables d'entrée (possiblement le temps x).
 - des constantes éphémères (représentant des valeurs réelles) dans l'intervalle $[0,1[$.
2. Il faut ensuite spécifier une variable cible pour chaque série temporelle : $\tilde{y} = f(x)$, où x est un horodatage et f est l'équation obtenue par l'algorithme de GP ;
3. spécifier l'ensemble des fonctions primitives pour représenter un individu f (le même pour toutes les séries temporelles) : Addition (ADD), soustraction (SUB), multiplication (MUL), division (DIV), racine carrée (SQRT), exponentielle (EXP)

et fonctions trigonométriques (SIN, COS). Afin de protéger les opérateurs contre des valeurs hors de leur domaine de définition (comme une division par zéro) et en même temps éviter les singularités mathématiques produites par les opérateurs protégés, nous avons mis en œuvre le concept d'arithmétique des intervalles (présenté dans [Keijzer 2003]) dans le modèle de régression de la plateforme EASEA (*EAsy Specification of Evolutionary Algorithms*), décrite dans [Collet 2000a]). Grâce à l'arithmétique des intervalles, le nombre de solutions invalides (qui peuvent être produites pendant la recherche) est éliminé, ce qui est particulièrement important lorsque le modèle a été entraîné sur des données de petite taille, car moins il y a de points d'apprentissage, plus il y a de fonctions bizarres pouvant approximer ces points.

4. Il faut alors choisir une fonction d'évaluation (fonction de *fitness*, en anglais) qui dira si l'individu trouvé approche correctement les données de la série temporelle cible. Après avoir essayé différentes fonctions d'évaluation, nous avons choisi l'erreur quadratique moyenne (ou *Root Mean Square Error* (RMSE)):

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (\tilde{y}_i - y_i)^2} \quad (5.1)$$

où \tilde{y} - les données calculées à partir de l'équation obtenue $f(x)$, y est la donnée observée et n est le nombre d'échantillons observés.

5. Pour sélectionner les parents, il faut là aussi choisir un opérateur. Dans ce travail, les individus (fonctions) ayant les meilleures évaluations sont choisis en utilisant une méthode de sélection par tournoi.
6. Un croisement standard est alors effectué : deux sous-arbres sont choisis aléatoirement dans les deux parents et échangés pour produire un nouvel enfant.
7. Autant d'enfants sont produits que spécifiés par le paramètre de l'algorithme déterminant le nombre d'enfants devant être créés à chaque génération.
8. Chaque enfant est alors évalué (possiblement de manière parallèle si l'ordinateur comporte plusieurs coeurs ou une carte GGPPU) et
9. un remplacement non-générationnel massivement parallèle est alors choisi (algorithme DISPAR [Maitre 2011]).
10. Nous avons choisi un élitisme "faible" : le meilleur individu de la population parents + enfants est assuré d'être sélectionné pour faire partie de la génération suivante, afin de ne pas perdre le meilleur individu trouvé depuis le début de l'évolution. Cela présente aussi l'avantage de renforcer la convergence de l'algorithme.

Après de nombreux essais, les meilleurs paramètres trouvés pour l'algorithme de Programmation Génétique sont les suivants :

1. Profondeur maximale de l'arbre : 5
2. Objectif de l'évaluation : minimiser la valeur d'évaluation, car celle-ci représente l'erreur entre les points proposés par la fonction implémentée par l'individu et les données observées.
3. Spécifier le critère de terminaison. Nous avons choisi le nombre maximal de générations comme critère d'arrêt.
4. Les autres paramètres (taille de la population, probabilité de croisement et de mutation, nombre de générations) dépendent ensuite de la série temporelle à approximer.

Enfin, nous avons lancé l'algorithme sur des machines comportant des cartes graphiques massivement parallèles GGPPU, ce qui a permis de réduire le temps de calcul nécessaire pour trouver de bonnes fonctions.

5.2.2.2 Filtre de Kalman

Le filtre de Kalman(KF) est un algorithme d'estimation optimale développé par Rudolf Kalman, utilisé dans divers domaines tels que les systèmes de guidage et la vision par ordinateur pour prédire des paramètres en présence de mesures bruitées. C'est une méthode puissante pour estimer le vecteur d'état d'un système dynamique linéaire décrit dans l'espace d'état à partir d'observations bruitées en utilisant l'état précédent et la pondération [Kalman 1961], [Morrison 1977], [Welch 1995]. L'algorithme du KF a été présenté à l'origine dans [Kalman 1961] et décrit en détail dans de nombreux articles : par exemple, dans [Welch 1995]. Il a été largement utilisé dans de nombreuses applications réelles (par exemple, la mission Apollo 11 de la NASA) [Cipra 1993], y compris les prévisions statistiques [Morrison 1977], [Nasseri 2011]).

Avant d'aborder l'application de la méthode KF pour la prévision de séries temporelles courtes dans le domaine des statistiques d'admission au baccalauréat, nous examinons brièvement quelques principes de fonctionnement clés de l'algorithme KF discret utilisé dans ce travail.

L'algorithme du KF discret est schématisé dans la figure 5.4, qui présente son cycle continu :

1. l'étape de prédiction projette l'estimation de l'état actuel dans le temps ;
2. l'étape de correction ajuste l'estimation projetée par une mesure réelle effectuée à ce moment-là [Welch 1995].

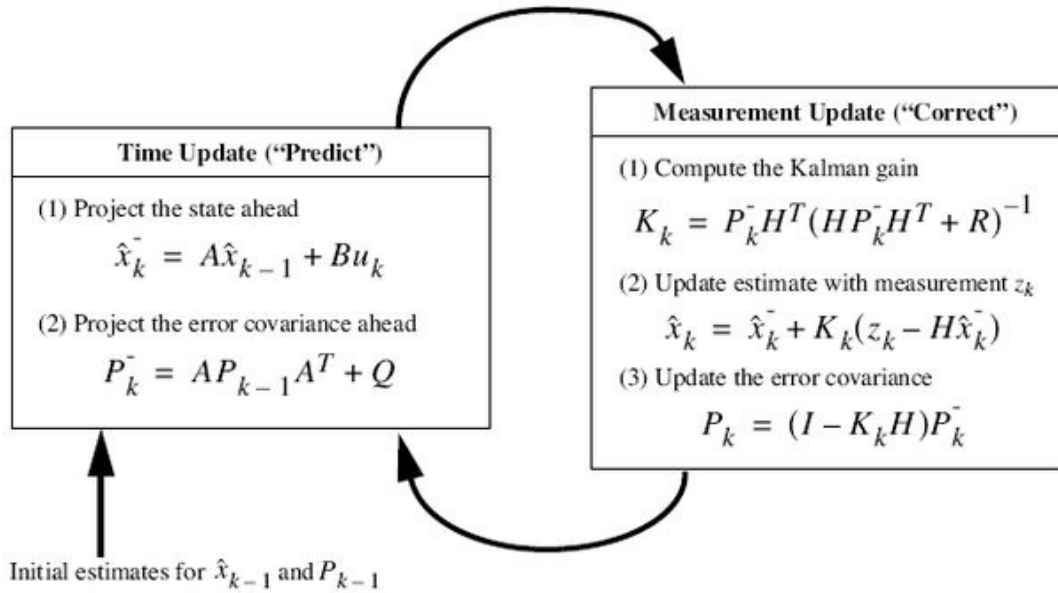


Figure 5.4: Kalman filter equations and instruction. The illustration is adapted from [Welch 1995]

Afin de rendre la figure 5.4 plus claire, les commentaires relatifs aux notations utilisées sont brièvement présentés ci-dessous. Au cours de l'étape de prédiction, la matrice de covariance de l'erreur d'état (P) est calculée à partir de sa valeur précédente, de la matrice du système dynamique (A) et de la matrice de covariance du bruit (Q). Ensuite, pendant l'étape de correction, la matrice P est calculée à partir du gain de Kalman (K), de la matrice d'observation (H) et de la matrice prédite P . Les matrices A et Q sont constantes et doivent être connues à chaque instant, de sorte que la seule variable qui affecte P est le gain de Kalman. Comme on peut le voir dans la première équation de l'étape correcte à la Figure 5.4, K est calculé à partir de P prédit, de H et de la matrice de bruit d'observation (R) qui sont des constantes. Ainsi, le gain de Kalman est un élément important du KF, qui est le régulateur entre l'estimation et la mesure. Comme le montre l'équation du gain dans la Figure 5.4, K décide (sur la base de l'erreur des estimations précédentes) à laquelle de l'estimation ou de la mesure il convient d'accorder le plus de poids. Plus précisément, si une prédiction (une estimation) est bonne, alors le gain de Kalman fonctionne pour annuler l'effet des nouvelles mesures, autrement dit il donne du poids aux nouvelles mesures pour faire des prédictions ultérieures. Il est important de noter que A , Q et R sont généralement utilisés comme paramètres de réglage que l'utilisateur peut ajuster pour obtenir les performances souhaitées.

5.2.2.3 Notre apport : GP+KF pour la prédiction de séries temporelles courtes

Nous expliquons maintenant l'application de la méthode KF dans ce travail.

Le Filtre de Kalman se base sur une équation décrivant le système dynamique à prédire. L'idée proposée est d'obtenir l'équation de la dynamique de la série temporelle à prédire par Programmation Génétique, puis d'utiliser cette équation pour effectuer la prédiction grâce à un Filtre de Kalman.

Il faut donc que l'équation trouvée par la PF soit définie comme un système dynamique sous la forme d'un espace d'état [Nasseri 2011]. Nous définissons le système dynamique suivant (en utilisant les mêmes notations que dans la Figure 5.4) :

1. L'état prédit x au temps k est composé de la position et de la vitesse : $x_k = x_{k-1} + T\dot{x}_{k-1}$, où T est la période d'échantillonnage;
2. L'entrée de contrôle optionnelle (u) est définie comme un terme d'accélération $u_k = \frac{\ddot{x}_{k-1}T^2}{2}$ (voir la première équation de l'étape de prédiction dans la Figure 5.4) ;
3. La sortie est la mesure de la position (z_k), calculée à partir de l'équation mathématique obtenue par GP.

5.2.2.4 Plateforme EASEA

La plateforme EASEA est utilisée comme support logiciel de l'algorithme développé. EASEA, (pour *EAsy Specification of Evolutionary Algorithms*) est un logiciel libre (sous GNU Affero v3 General Public License) développé initialement à l'INRIA (Institut National de Recherche en Informatique et Automatique) par Pierre Collet en 1999. Son développement a par la suite continué à l'école Polytechnique de Palaiseau, puis à l'Université du Littoral Côte d'Opale et enfin à l'Université de Strasbourg à partir de 2007, dans le groupe SONIC (Stochastic Optimisation and Nature Inspired Computing) de l'équipe BFO de l'Université de Strasbourg et plus récemment par l'équipe CSTB du laboratoire ICUBE de l'Université de Strasbourg⁹. Il s'agit d'un langage de spécification d'algorithmes évolutionnaires multiplateforme (Linux, MacOS et Windows) et multi-compilateurs (Clang C++, GNU C++ et Visual C++), hébergé sur GitHub¹⁰, qui permet aux scientifiques n'ayant que des compétences de base en informatique de mettre en oeuvre des algorithmes évolutionnaires et d'exploiter le parallélisme massif des architectures multi-cœurs afin d'optimiser pratiquement tous les problèmes du monde réel [Collet 2000b], car l'évolution artificielle est un solveur générique.

⁹https://easea.unistra.fr/index.php/EASEA_platform

¹⁰<https://github.com/EASEA/easea>

La plateforme EASEA comprend deux parties principales : une bibliothèque interne en C++ (LibEASEA), qui est compilée de manière statique avec le fichier exécutable produit, et un générateur de code/compilateur (easena) basé sur lex/yacc (ALex et AYacc¹¹). Le compilateur easena fusionne l'un des fichiers modèles d'EASEA (*tpl - un squelette d'algorithme d'optimisation, choisi par l'utilisateur) avec une partie spécifique au problème (*.ez - le problème d'optimisation, défini par l'utilisateur). Le compilateur examine le modèle du fichier de l'algorithme choisi et insère les caractéristiques du problème dans une instance du modèle. Le code source prêt à l'emploi de l'algorithme avec le problème d'optimisation intégré est fourni à la fin du processus de compilation automatique sous la forme d'un ensemble de fichiers : *.cpp / *.cu, *.h, * prm, à compiler par le compilateur c++ ou NVIDIA nvcc en utilisant le Makefile généré automatiquement [Leonteva 2022].

Les caractéristiques principales d'EASEA sont les suivantes [Leonteva 2022]:

- Un couplage automatique d'un modèle d'algorithme évolutionnaire avec une partie spécifique au problème d'optimisation dans un code source C++, sans intervention humaine. Cela réduit la possibilité de commettre des erreurs et facilite la réalisation d'expériences.
- Un cadre de calcul parallèle. Le logiciel crée non seulement du code pour les CPU multicœurs, mais prend également en charge les GPU [Collet 2013]
- Une architecture orientée objet, qui permet au programmeur d'utiliser différents modules et de les combiner arbitrairement les uns avec les autres par polymorphisme
- Un ensemble d'algorithmes d'optimisation indépendants des problèmes, qui peuvent être appliqués à diverses suites de tests et applications réelles
- Un ensemble d'exemples expérimentaux de problèmes d'optimisation à objectif unique
- La préservation des données expérimentales dans de simples fichiers * csu
- Un outil de visualisation simple

5.2.3 Description de la méthode proposée

EASEA permet de fusionner automatiquement un problème scientifique (spécifié dans le fichier de description du problème *.ez) avec un modèle d'algorithme évolutionnaire (dans ce cas, le modèle d'algorithme GP) dans un code C++ parallèle CPU ou GPU [Collet 2000a]. Cela simplifie considérablement le processus de programmation, car

¹¹<http://www.bumblebeessoftware.com/>

l'application EASEA ne nécessite pas de connaissances profondes sur les algorithmes de GP de la part des utilisateurs.

L'outil présenté, basé sur la plateforme EASEA, est schématisé dans la Figure 5.5. Comme expliqué, la KF utilise une série de mesures observées au fil du temps à partir du modèle de GP (contenant des inexactitudes dues à un manque de données pour la modélisation GP) et produit des estimations de variables inconnues, qui tendent à être plus précises que celles basées sur une mesure unique du GP seul. Ainsi, les prédictions produites par l'équation du modèle GP sont considérées comme une nouvelle mesure à filtrer.

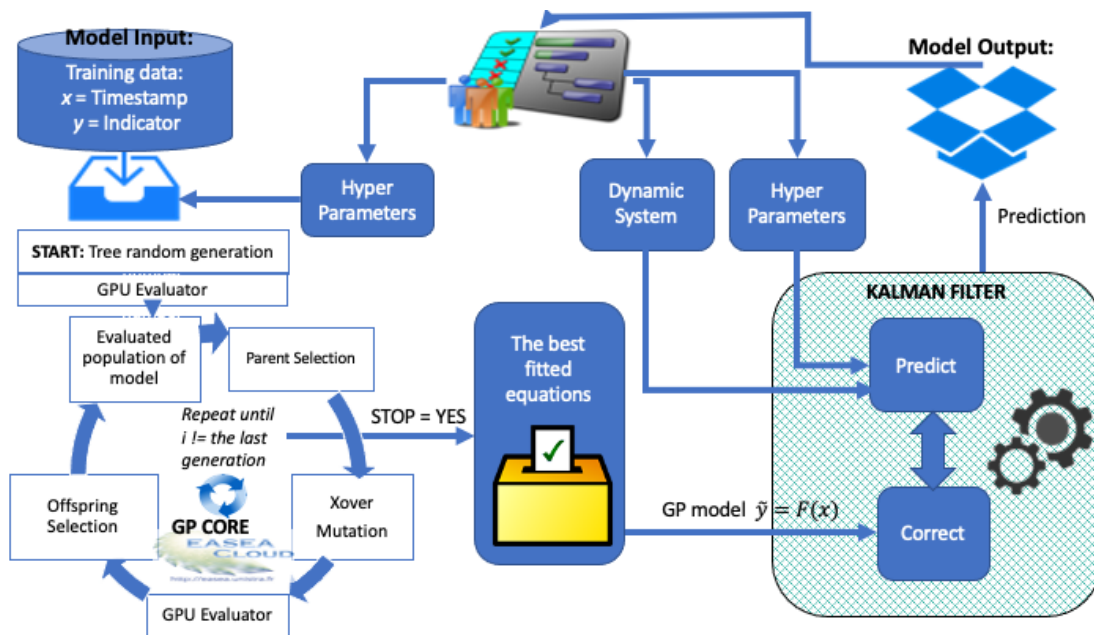


Figure 5.5: La structure simplifiée de l'outil proposé.

L'outil considéré est supposé être un processus continu, dans lequel le modèle GP sera mis à jour avec chaque nouveau point de données.

5.2.4 Données utilisées

Le processus de collecte des données dans le contexte mauritanien a été très long à cause du manque de données fiables.

Des données sur les bacheliers nous ont été communiquées par la Direction de l'Orientation et des Bourses du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, mais il s'agissait de plusieurs bases différentes qu'il a fallu com-

pilier. En effet, nous avons reçu la liste des admis au baccalauréat (BAC) et la liste des bacheliers orientés vers les différentes institutions de l'enseignement supérieur de 2013 jusqu'à 2021 (2013 est l'année du démarrage d'une organisation informatisée des données du baccalauréat national) mais les données n'étaient pas utilisables telles quelles.

Il a donc fallu d'abord les explorer et nettoyer. Les différentes étapes effectuées lors du nettoyage des données ont été les suivantes :

- vérification et rectification des erreurs d'enregistrements (valeurs incorrectes),
- vérification de valeurs dupliquées via le numéro de dossier du bachelier,
- vérification de valeurs nulles,
- suppression de colonnes inutiles et vérification du type de colonne.

Après le nettoyage, quelques indicateurs, inspirés de la disponibilité des données, ont été extraits et calculés selon différents paramètres (âge, sexe, moyenne générale au bac, conditions d'obtention du baccalauréat selon les différentes séries, etc.). Ces indicateurs représentent nos séries temporelles. 22 indicateurs pour l'admission au Baccalauréat ont été élaborés (voir annexe 1).

Pour tester notre approche GP + KF, nous avons testé quatre indicateurs (voir Table 5.1) :

- le taux d'admission au baccalauréat;
- le taux d'admission au baccalauréat lors de la première session;
- le taux d'admission au baccalauréat dans la série Mathématiques;
- le taux des filles avec un baccalauréat littéraire.

Chaque série temporelle ne contient que 10 observations. La méthode de calcul de ces séries temporelles est présentée dans le tableau 5.2.

Elle est adaptée de [Nasseri 2011] en y apportant les modifications suivantes :

1. Le KF discret est appliqué à la place du KF étendu comme la version la plus simple et la plus pratique du KF pour le problème décrit ;
2. Nous n'utilisons pas les décalages précédents comme variables de décision (données d'entrée) ;
3. L'algorithme GP utilise une approche "interval arithmetic" (présentée dans [Keijzer 2003]) pour la protection des opérateurs mathématiques ;

Table 5.1: Séries temporelle utilisées dans la contribution

Année	Taux d'admission au BAC	Taux d'admission au BAC dans la première session	Taux d'admission au BAC M (Mathématiques)	Pourcentage de filles avec un BAC littéraire
2013	16	55	10	37
2014	12	55	18	34
2015	14	54	13	36
2016	11	53	18	34
2017	13	59	14	37
2018	22	58	10	43
2019	16	46	8	36
2020	23	62	6	39
2021	13	60	6	38
2022	20	61	5	48

Table 5.2: Méthode de calcul des indicateurs utilisés dans la contribution

Indicateurs	Taux d'admis au bac	Taux d'admis à la 1ere session
Calcul	Nombre d'admis au Bac (tout bac inclut) sur l'effectif de candidats au bac (tout bac inclut)	Nombre d'admis à la 1ere session sur le nombre total d'admis au baccalauréat (sessions 1 et 2)

Indicateurs	Taux d'admis au Bac M	Taux de filles avec un bac littéraire
Calcul	Nombre d'admis au Bac M sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles avec un bac littéraire (LO, LM) sur le nombre total des admis (garçons et filles) avec un bac littéraire (LO, LM)

M : Mathématique, LO : Lettres originales, LM : Lettres modernes

4. La version GPU-parallèle de la GP est appliquée pour avoir la capacité de travailler avec une population de grande taille en un petit nombre de générations qui nous permet de réduire le temps d'exécution et d'accélérer le taux de convergence.

Une version expérimentale du code source est disponible sur gitHub : [\[https://github.com/leopard152015/LallaAichaCode/\]](https://github.com/leopard152015/LallaAichaCode/).

D'autres exemples de l'algorithme de GP utilisé ainsi que celui du filtre de Kalman sont dans les annexes.

5.3 Expériences et résultats

L'objectif des expériences présentées ci-dessous est d'étudier les capacités de prévision de l'algorithme proposé et d'effectuer une analyse statistique de ses performances par rapport aux algorithmes de pointe sur différents ensembles de données. En raison de l'absence de code source, certaines méthodes existantes dans la littérature n'ont pas été utilisées à des fins de comparaison dans cette contribution.

Nous ne fournissons pas de tests de validation croisée (leave-one-out) de notre algorithme dans ce travail, pour les raisons suivantes :

1. nous ne pouvons pas utiliser plusieurs séries temporelles pour l'apprentissage et une pour le test, parce que toutes les séries temporelles courtes contiennent des informations complètement différentes, et le modèle individuel est fait pour chaque série temporelle ;
2. si nous utilisons le même ensemble pour un test leave-one-out - dans de nombreux cas, nous prédirons le passé sur la base du futur.

5.3.1 Données d'expérimentation

Les sources de données utilisées dans cette étude (9 séries temporelles différentes sont utilisées au total) sont les suivantes :

Quatre séries temporelles courtes présentées auparavant dans la section 5.2.1. Le ratio données d'entraînement / données de test est de 6/4. La figure 5.6 montre ces séries temporelles (en %) divisées sur les ensembles d'entraînement et de test par une ligne rouge. Les formes de ces courbes sont différentes et difficiles à prévoir, en raison du faible nombre de points d'entraînement.

Trois séries temporelles courtes différentes sélectionnées à partir des données ouvertes de la Banque mondiale :

1. Taux d'achèvement du cycle primaire (TACP) chez les femmes (% du groupe d'âge concerné) en Bulgarie : 36 observations, où le ratio jeu d'apprentissage / jeu de test est de 31/5 ;
2. Taux d'alphabétisation (TA), adulte total (% des personnes âgées de 15 ans et plus) en Afrique sub-saharienne : 35 observations, où le ratio jeu d'apprentissage / jeu de test est de 30/5 ;
3. Production d'électricité renouvelable (PER) (% de la production totale d'électricité) en Indonésie : 30 observations, où le rapport jeu d'apprentissage / jeu de test est de 25/5.

Deux séries temporelles courtes (20 observations) issues de la compétition de prévision M3¹² pour l'intervalle annuel sont sélectionnées : 10 et 120, où le ratio jeu d'apprentissage / jeu de test est de 14/6.

5.3.2 Paramètres de l'algorithme de programmation génétique

Les valeurs des paramètres ont été déterminées empiriquement, par essai-erreur :

taille de la population = 15000 ;

nombre de générations = 500 ;

probabilité de croisement = 90% ;

probabilité de mutation = 10%.

Nous utilisons une population de grande taille pour accélérer la convergence et charger la carte GPU avec suffisamment de *threads* pour qu'elle reste occupée.

5.3.3 Paramètres du Filtre de Kalman

Les valeurs du bruit de mesure et du bruit de processus ont été sélectionnées empiriquement pour chaque ensemble de données.

5.3.4 Modèles de référence

Pour comparer notre approche à l'état de l'art, nous avons sélectionné plusieurs modèles de référence :

- Cinq méthodes "de compétition" issues du package FMA (Forecasting Methods and Applications) de R ¹³ :

¹²<https://forecasters.org/resources/time-series-data/m3-competition/>

¹³<https://github.com/robjhyndman/fma>

Table 5.3: Rapports statistiques (RMSE) de la GP sur les ensembles d'entraînement.

Métrique	TAB	TAB en 1ère session	TAB Math	Taux des filles avec un BAC littéraire	M3-10	M3-120	TACP	TA	PER
STD	2.7e-02	1.0e-04	1.0e-01	4.2e-02	2.4e+01	5.2e+01	9.3e-01	9.0e-01	1.7e+0
Mean	6.1e-02	2.3e-04	2.5e-01	1.1e-01	8.2e+01	3.0e+02	5.5e+0	3.2e+0	4.5e+0
Best	7.5e-03	1.1e-04	8.1e-02	7.3e-02	2.5e+01	9.1e+01	3.1e+0	1.9e+0	2.0e+0

TAB : Taux d'admission au BAC

1. Random walk
 2. Mean Method
 3. SES
 4. ARIMA (en mode automatique) et
 5. Réseau de neurones (Neural Network (NET));
- Grey model : le code source MATLAB est emprunté à [Cui 2021].

5.3.5 Métriques de performance

Nous avons utilisé

1. l'erreur quadratique moyenne (RMSE) entre les valeurs prédites et réelles, car les valeurs renvoyées sont dans la même échelle que la cible ;
2. l'erreur absolue moyenne en pourcentage (MAPE) entre les valeurs prédites et réelles, qui renvoie la valeur en %.

5.3.6 Résultat et Analyse

Afin de définir les modèles basés sur le GP pour la méthode proposée, l'algorithme GP a été exécuté 30 fois sur l'ensemble de formation pour chaque série temporelle, et le meilleur modèle (équation) a été sélectionné par la plus petite valeur de RMSE entre les 30 solutions obtenues. Le tableau 5.3 résume les résultats statistiques de ces exécutions sur les ensembles de formation sélectionnés : valeur RMSE de la meilleure équation obtenue, valeur RMSE moyenne et écart-type (STD).

Comme le montre le tableau 5.3, les modèles les plus imprécis concernent les séries de données de la compétition M3. Toutefois, leur valeur élevée de RMSE peut s'expliquer par les différentes échelles des données : dans la série temporelle M3, les valeurs cibles

sont beaucoup plus importantes que dans les autres séries, où les données sont présentées en %. Ces résultats peuvent probablement être améliorés par un réglage supplémentaire des hyper paramètres des algorithmes GP (par exemple, en augmentant la taille de la population) et devraient être étudiés dans les travaux futurs. Les valeurs très faibles de RMSE pour les données d'admission au baccalauréat peuvent s'expliquer par le faible nombre d'observations dans l'ensemble d'apprentissage (seulement 6 points).

Les résultats comparatifs de la méthode proposée (la meilleure équation intégrée dans la KF) *versus* les modèles GP obtenus (la meilleure équation) et les algorithmes de pointe sélectionnés sont présentés dans le Tableau 5.4 et le Tableau 5.5. Comme le montre le tableau 5.4, KF+GP affiche les meilleures performances en termes de RMSE dans 5 cas d'essai sur 9. Dans les autres cas, le delta entre KF+GP et les méthodes les plus performantes est très faible (environ 0,5-2,5 %). En termes de métrique MAPE, KF+GP fournit la meilleure performance dans 3 cas de test. De plus, dans les autres cas, cette métrique est très proche de la meilleure (1-5 %). La figure 5.7 résume les valeurs MAPE obtenues par les modèles pour tous les cas de test dans le diagramme en forme de violon. La figure 5.7, le tableau 5.4 et le tableau 5.5 nous permettent de conclure qu'en général, la méthode proposée donne des résultats plus robustes et plus stables que les autres.

5.3.7 Conclusion

La méthode proposée a fait l'objet d'un article de conférence internationale [Kone 2023] avec comité de lecture qui a donc validé l'approche d'un point de vue scientifique. L'évolution constante des statistiques du taux d'admission au baccalauréat en Mauritanie nécessite un instrument de prévision des séries temporelles courtes et flexibles pour prédire et contrôler la réalité de l'éducation. Ce travail présente une méthode visant à être efficace pour cette question, qui est basée sur l'intégration d'un modèle à base de GP dans l'algorithme KF. Nous proposons également un outil logiciel basé sur la méthode présentée, mis en œuvre grâce à la plateforme EASEA, qui permet de paralléliser les algorithmes évolutionnaire sur cartes GPU, ce qui permet de réduire le temps nécessaire à la mise en place des différentes expériences.

Les résultats montrent que la méthode proposée est utile lorsque le nombre d'observations est limité et qu'elle fournit des résultats stables et robustes (en termes de RMSE et de MAPE) par rapport aux autres méthodes classiques utilisées pour la comparaison. Grâce à l'outil présenté, il est possible de lancer la méthode proposée pour prédire les séries d'admission au baccalauréat à court terme. Enfin, la méthode présentée peut être généralisée à des séries temporelles de longueur et de nature différentes dans différents domaines.

Table 5.4: Erreurs de prédiction (RMSE).

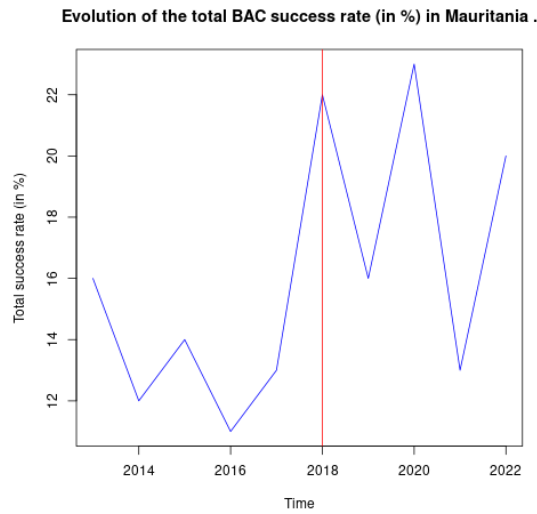
Méthode	TAB	TAB en 1ère session	TAB Math	Taux des filles avec un BAC littéraire	M3-10	M3-120	TACP	TA	PER
RW	05.52	6.57	03.90	05.36	4488.55	1970.13	08.21	2.59	3.34
Mean	05.06	6.91	07.66	05.732	3298.74	6129.48	09.68	9.75	2.83
Ses	04.30	6.91	05.70	05.634	4488.53	1970.28	08.21	2.59	3.18
Arima	05.06	6.72	07.66	05.732	6572.42	4810.95	09.11	1.01	3.34
Net	04.66	7.12	09.27	15.105	4682.02	0890.64	10.58	2.04	2.97
GP	07.22	7.09	04.63	07.523	3861.33	9873.90	08.18	4.50	3.69
GP+KF	03.68	8.94	03.91	06.04	2243.66	5311.63	03.26	2.70	2.81
GM	10.53	6.77	04.57	06.99	9697.06	4899.21	15.51	1.77	4.27

TAB : Taux d'admission au BAC

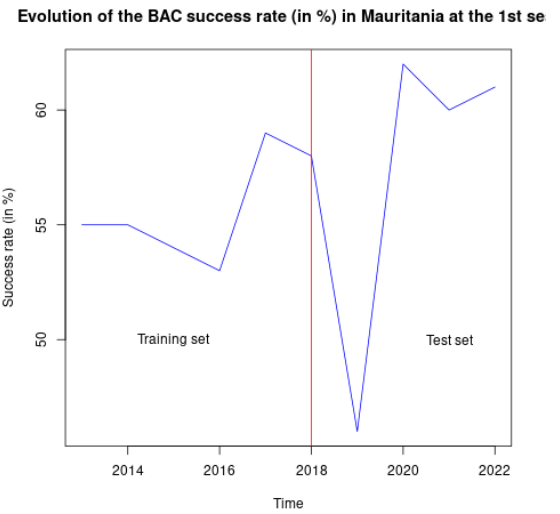
Table 5.5: Erreurs de prédiction (MAPE)

Méthode	TAB	TAB en 1ère session	TAB Math	Taux des filles avec un BAC littéraire	M3-10	M3-120	TACP	TA	PER
RW	30.26	10.19	064.58	13.31	101.20	20.16	08.98	03.45	16.65
Mean	21.01	11.79	127.67	08.55	060.03	72.24	10.72	14.66	21.03
Ses	18.37	12.25	111.68	08.36	101.20	20.16	08.98	03.45	15.75
Arima	21.01	11.79	127.67	08.55	147.96	48.28	09.91	01.26	16.65
Net	22.87	12.54	145.27	37.36	105.35	09.03	11.75	02.68	16.54
GP	37.85	12.54	059.89	12.02	087.01	88.69	07.68	05.92	20.54
GP+KF	19.95	10.55	050.39	09.61	050.81	41.94	03.21	03.13	20.60
GM	57.91	09.26	065.49	16.85	244.97	46.36	16.29	02.53	21.79

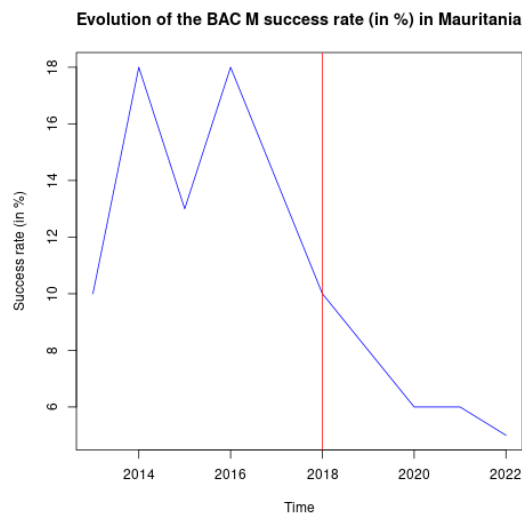
TAB : Taux d'admission au BAC



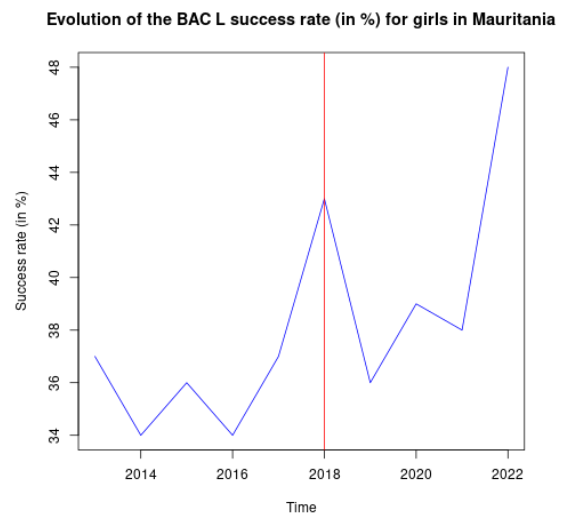
a) Taux d'admission au BAC



b) Taux d'admission au BAC en 1ère session



c) Taux d'admission au BAC Mathématiques



d) Taux des filles avec un BAC littéraire

Figure 5.6: Evolution des indicateurs(séries temporelles).

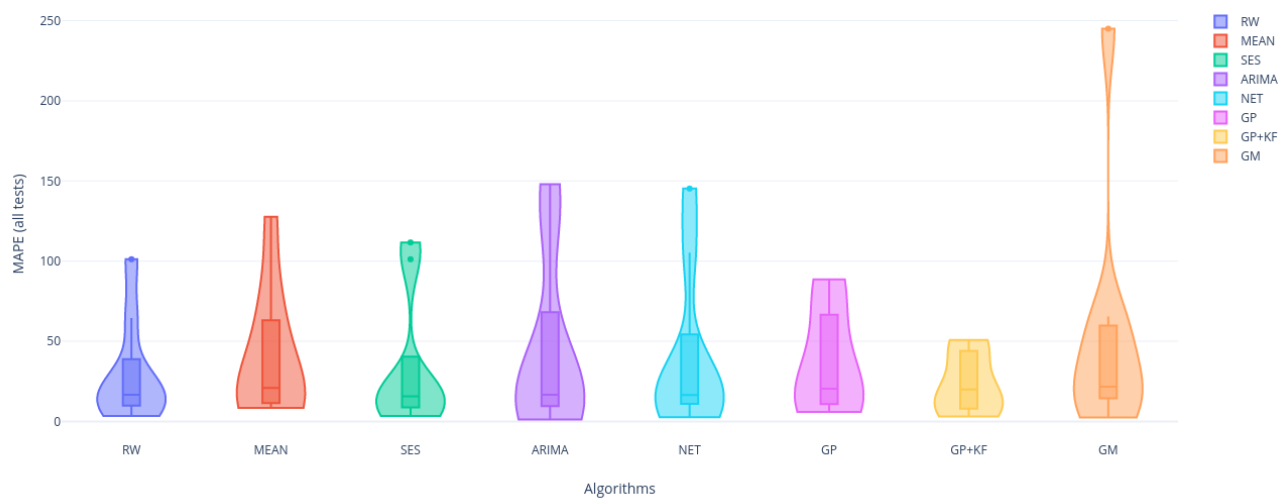


Figure 5.7: Graphique en violon de l'erreur moyenne en pourcentage absolu (MAPE) pour tous les cas de test.

Façonnage par IA générative d'une éthique relativiste gouvernementale

Sommaire

6.1 Introduction	149
6.2 Problématique	152
6.3 Approche proposée	155
6.3.1 Principes éthiques choisis dans le cadre de cette contribution	155
6.3.2 Détails techniques de l'implémentation	155
6.3.3 Vue d'ensemble des outils Ethiconorm et Ethicotest	159
6.4 Résultats expérimentaux	168
6.5 Discussion	187
6.6 Conclusion	189

6.1 Introduction

Au fur et à mesure de son évolution, la technologie de l'IA joue un rôle de plus en plus important dans les opérations gouvernementales, qui peuvent tirer d'immenses bénéfices de leur utilisation. Par exemple, l'IA peut prédire les résultats de diverses mises en œuvre de politiques, permettant ainsi aux ministères de prendre des décisions plus éclairées. Elle peut également répondre aux demandes des citoyens et à gérer des tâches d'information telles que répondre à des questions, remplir et rechercher des documents, acheminer des demandes, traduire et rédiger des documents, ainsi que résoudre des problèmes d'affectation des ressources [Mehr 2017].

L'IA générative a connu une croissance considérable depuis 2022. Cette branche de l'IA intègre le traitement du langage naturel, un domaine où les techniques informatiques croisent la linguistique, permettant aux machines de générer des textes satisfaisants pour les humains. En conséquence, les systèmes d'IA générative sont désormais capables de produire un contenu contextuel qui semble pertinent et cohérent

sur différents supports en réponse à des demandes humaines, ce qui démontre les progrès significatifs de la capacité de l'IA à générer un langage similaire à ce qui serait produit par un humain. Grâce à des algorithmes et des modèles qui ont été formés sur d'énormes données existantes, l'IA générative a le potentiel d'offrir des solutions robustes dans divers domaines.

En outre, l'IA générative permet à des *chatbots* spécialisés d'interagir avec les citoyens, de répondre aux questions et de proposer des suggestions, améliorant ainsi l'engagement des citoyens.

Parallèlement à l'application étendue de l'IA dans les pays africains, des préoccupations d'ordre général surgissent. Les défis ne concernent pas seulement la précision ou la vitesse de traitement. Les questions éthiques liées aux technologies de l'IA doivent être abordées afin de garantir leur déploiement responsable et équitable. Elles ont fait l'objet de recherches approfondies, révélant des implications significatives pour les droits de l'homme. Il s'agit notamment de questions telles que la désinformation, la discrimination et la radicalisation, comme le souligne [Bensalah 2021]. En outre, l'IA pourrait renforcer les préjugés, en particulier dans les contextes patriarcaux africains, ce qui pourrait avoir un impact négatif sur les jeunes générations et les femmes [Gwagwa 2019]. Cela soulève d'importantes préoccupations, car l'adoption des technologies de l'IA doit être traitée avec précaution pour éviter de perpétuer des pratiques injustes. Il est important de noter que, ainsi que décrit en section 2.7 sur l'éthique relativiste, chaque région du monde (pas seulement en Afrique), voire chaque gouvernement, chaque peuple, chaque peuplade, pourra avoir sa propre éthique. Les pays sont à des stades différents de l'évolution de leur approche de la réglementation de l'IA et ont des points de vue différents sur la manière de procéder. Il est donc impératif d'examiner et de prendre en compte les différentes cultures et opinions publiques lors de l'établissement des normes, règles et lignes directrices pour les systèmes dits "intelligents".

Plus précisément, en Mauritanie (Afrique de l'Ouest), qui est notre cas d'utilisation, il est primordial de répondre à ces préoccupations afin de garantir que les progrès de l'IA s'alignent avec les valeurs et la structure sociétale du pays. En effet, la société mauritanienne est diverse et complexe, avec différents groupes ethniques, langues et normes sociales. Mettre en œuvre l'IA sans accepter ces complexités pourrait conduire à des décisions ou des recommandations qui ne seraient pas inclusives et pourraient être considérées comme offensantes. Il est donc essentiel de proposer un moyen pour que les applications de l'IA puissent s'aligner sur le contexte local, en particulier dans un pays aux traditions culturelles bien ancrées comme la Mauritanie.

Afin de résoudre ce problème, nous présentons, via cette contribution, une nouvelle approche pour guider le développement de logiciels en tenant compte du contexte socioculturel d'un pays ou d'une région spécifique, notamment la Mauritanie. Cette approche repose sur un outil composé de deux applications : l'une appelée **ethiconorm**, conçue pour aider les fonctionnaires du gouvernement à définir des normes éthiques, et

l'autre (**ethicotest**) pour les développeurs afin d'évaluer l'alignement des logiciels sur ces normes.

L'idée principale derrière notre proposition est d'exploiter la technique du *prompt intelligent* des grands modèles de langage (LLM) pour aider fonctionnaires et utilisateurs à formuler et appréhender des concepts éthiques auxquels ils seraient étrangers. Cette technique le langage naturel comme un outil de programmation flexible, utilisant des variables.

L'objectif consiste à tenter de combler le fossé entre les capacités brutes de l'IA et des besoins éthiques précis, mais sans l'affinage typique (*fine-tuning*) qui permet de spécialiser les modèles IA sur des tâches spécifiques et permet de maximiser leurs performances. Les résultats présentés en section 6.4 sont encourageants, montrant l'efficacité de notre approche.

Par ailleurs, selon la théorie de la publicité de Kant, il nous semble important, à ce stade de cette partie introductive de cette deuxième contribution scientifique, de préciser que depuis le 28 novembre 1960, l'état mauritanien a toute la légitimité et le droit d'exprimer son pouvoir de direction et de se faire sa propre publicité surtout dans un domaine aussi sensible que l'éthique de l'IA, celle-ci devant s'exercer dans le contexte d'une société complexe et multiethnique.

Dans "Vers la paix perpétuelle" (1795) [Emmanuel 1795]. Kant formule l'idée selon laquelle le principe de publicité est central pour la légitimité morale et politique d'une action.

Kant affirme que les actions ou maximes politiques doivent pouvoir être rendues publiques sans compromettre leur objectif moral ou légitime. Autrement dit, une maxime qui ne pourrait être divulguée sans provoquer d'indignation ou de rejet ne saurait être moralement justifiable.

La théorie de la publicité selon Kant repose sur l'idée que la raison doit être exercée publiquement pour être légitime et efficace. Pour Kant, la publicité n'est pas simplement une question de communication, mais un principe fondamental de la raison et de la justice. Elle exige que les maximes politiques soient formulées de manière universelle et transparente, permettant ainsi un contrôle critique par le public. La publicité s'oppose au secret d'État et est essentielle pour garantir la justice et la légitimité des actions politiques, en alignant le pouvoir politique sur un ordre juridico-moral transparent et responsable [Mohr 2010].

Ainsi donc, rendre la politique morale, était un désir de Kant, formulé dans l'appendice d'un opuscule qu'il a publié en 1795, intitulé "Projet de paix perpétuelle" ¹. Pour rendre les politiciens moraux, il faut les obliger par le droit à une certaine publicité de leurs actes. Antoine Grandjean explique : "au fond, ce que soutient Kant, c'est qu'ils

¹<https://www.radiofrance.fr/franceculture/podcasts/avec-philosophie/faut-il-rendre-la-politique-morale-4505726>

doivent, d'une certaine manière, pouvoir expliciter leur comportement comme gouvernants : c'est ça, l'exigence de publicité". Autrement dit, "la publicité, ce n'est pas ça : c'est pouvoir dire publiquement en vertu de quoi on agit".

Ethiconorm s'inscrit pleinement dans ce cadre (principe de publicité d'Emmanuel Kant) en fournissant un outil permettant au gouvernement de définir, expliciter et publiciser ses priorités éthiques en matière d'IA.

Par cette publicité, Ethiconorm offre une légitimité accrue aux décisions étatiques et engage également les développeurs à expliciter comment leurs créations respectent ou non ces normes. Cette double transparence, de la part des gouvernements et des développeurs, constitue un mécanisme essentiel pour garantir que les actions en lien avec l'IA soient juridiquement, moralement et socialement acceptables, tout en promouvant un débat éclairé.

En intégrant la théorie kantienne dans un outil pratique, Ethiconorm contribue à aligner la politique éthique de l'IA sur des fondements philosophiques solides, répondant ainsi au double objectif de rendre la politique morale et de promouvoir une gouvernance technologique responsable.

6.2 Problématique

De nombreux défis techniques peuvent involontairement conduire à une insensibilité culturelle ou à un départ de l'alignement avec les normes éthiques locales lorsque le développement logiciel, inclut des algorithmes d'IA [Gwagwa 2020, Gwagwa 2022]. Ces défis sont particulièrement prononcés en Mauritanie, compte tenu de la richesse de ses traditions et de ses valeurs. Nous examinons ci-dessous plusieurs fondements techniques majeurs susceptibles d'entraîner de telles disparités :

1. Biais des ensembles de données : au cœur de tout système d'apprentissage automatique se trouvent les données avec lesquelles il a été élaboré. Malheureusement, toutes les données sont biaisées par nature et leur manque de représentativité va retransmettre les biais du jeu d'apprentissage d'origine [Ntoutsi 2020, Miltenburg 2016]. Par exemple, si un système d'IA destiné à la Mauritanie est principalement formé à partir d'ensembles de données chinoises, il risque de ne pas reconnaître ou respecter les coutumes, les traditions ou les valeurs mauritaniennes, ce qui conduira à des décisions qui sembleront étrangères ou inappropriées.
2. Adaptation excessive à des sous-populations spécifiques : les modèles d'apprentissage automatique s'adaptent parfois de manière excessive à certains groupes de données de leur ensemble d'apprentissage, pas forcément car ce sont les plus nombreuses, mais parce qu'elles peuvent être plus facile à modéliser. Cela

signifie que si un ensemble de données contient des informations sur plusieurs groupes démographiques différents, le système peut donner des résultats exceptionnels pour l'un des groupes, mais médiocres pour les autres, négligeant ou représentant mal les communautés concernées. Cela peut provenir d'une sous-représentation des communautés concernées, mais pas seulement.

3. Tests locaux inadéquats : souvent, les systèmes d'IA ne sont pas suffisamment testés dans des contextes locaux avant d'être déployés. En l'absence de tests rigoureux dans le contexte mauritanien, ces systèmes peuvent commettre des erreurs de méconnaissance ou d'insensibilité culturelle.
4. Manque d'interprétabilité et d'explicabilité : les modèles d'IA modernes, en particulier les modèles d'apprentissage profond, sont souvent décrits comme des "boîtes noires". Cela signifie qu'il est difficile de comprendre pourquoi ils prennent une décision ou font des propositions particulière. En l'absence d'interprétabilité claire, il est difficile de mettre le doigt sur les malentendus culturels ou les désalignements éthiques et d'y remédier.

Cela amène à plusieurs conséquences possibles dans des régions ou des pays spécifiques tels que la Mauritanie :

1. Représentation culturelle erronée : les outils technologiques mal alignés sur les normes locales peuvent par inadvertance donner une représentation erronée des diverses communautés, ce qui peut conduire à une représentation numérique faussée.
2. Méfiance à l'égard de la technologie : le fait de ne pas respecter ou de ne pas comprendre les coutumes locales peut entraîner une méfiance généralisée à l'égard des solutions technologiques, ce qui risque d'entraver les progrès technologiques dans la région concernée.
3. Renforcement des stéréotypes : les ensembles de données biaisées non seulement donnent une image erronée, mais peuvent également renforcer des stéréotypes nuisibles, ce qui conduit à des décisions qui marginaliseraient encore plus des groupes déjà vulnérables.

Les solutions et les considérations techniques possibles peuvent sembler assez simples à formuler :

1. Collecte de données diversifiées : il faut tenter de s'assurer que les ensembles de données sont représentatifs des populations modélisées, pour tenter d'atténuer de nombreux préjugés. Il s'agit notamment de tenir compte du sexe, des ethnies, des langues et même des différences régionales au sein du pays.

2. Algorithmes transparents : lorsque le choix existe, il faut opter pour des algorithmes plus transparents pour aider à identifier les biais ou les désalignements détectés ou potentiels.
3. Boucles de rétroaction localisées : l'intégration de mécanismes de rétroaction dans le logiciel permet aux utilisateurs de signaler les insensibilités culturelles ou les erreurs, ce qui peut ensuite être utilisé pour affiner le système.

Mais malheureusement, leur mise en œuvre concrète est loin d'être triviale pour de nombreuses raisons, comme :

1. La complexité des cadres éthiques : les principes éthiques locaux ne sont pas de simples listes de choses à faire et à ne pas faire. Ils sont ancrés dans des récits culturels, des traditions, des dynamiques socio-économiques et des contextes historiques. Il est difficile de saisir ces nuances et de les transposer de manière algorithmique.
2. Les problèmes de mise à l'échelle : s'il est possible de mettre en œuvre ces solutions pour une application unique ou un domaine étroit, il est difficile de les adapter à une multitude de solutions logicielles couvrant divers secteurs.
3. La difficulté de la mise en place d'une supervision semi-automatique : s'assurer que les développeurs de logiciels adhèrent à ces principes ne consiste pas seulement à créer un ensemble de lignes directrices. Des mécanismes de contrôle en temps réel ou périodique sont nécessaires pour évaluer et garantir l'intégrité éthique des applications développées. Mais comment surveiller de manière semi-automatique quelque chose d'aussi intangible et nuancé que la mise en œuvre d'une éthique dans une application ?
4. Le manque d'outils existants : bien qu'il existe des outils pour l'assurance qualité, la sécurité et le contrôle des performances dans le développement de logiciels [Al-Ghamdi 2013], les outils qui traitent spécifiquement de l'adhésion aux principes éthiques locaux sont rares ou, au mieux, rudimentaires. Selon [Nallur 2020], aucune approche n'a réussi à créer un système éthique solide et impartial. L'auteur souligne l'importance de *l'adaptabilité* dans la conception éthique, en soutenant qu'une machine ou un système doit être capable d'ajuster son raisonnement éthique dans différents contextes grâce à des mises-à-jour régulières ou mieux, dynamiques, tout en maintenant la cohérence au sein de chaque contexte.

À la lumière de ces défis, il existe un besoin pressant d'outils spécialisés conçus pour définir et contrôler l'adhésion aux principes éthiques locaux dans le développement de logiciels. De tels outils devraient non seulement englober l'étendue et la profondeur

des normes éthiques d'un pays (la Mauritanie pour ce qui concerne ce travail) ou d'une communauté, mais aussi fournir des informations utiles aux développeurs pour qu'ils adaptent leurs logiciels en conséquence.

6.3 Approche proposée

6.3.1 Principes éthiques choisis dans le cadre de cette contribution

Sur la base de notre étude bibliographique, nous avons utilisé des sources provenant de régions telles que l'Occident, l'Asie et l'Afrique, (cf. section 4.1), pour identifier une dizaine de principes éthiques pouvant faire consensus et intégrer le contexte africain :

- Transparence
- Responsabilité
- Non-malfaisance
- Équité
- Vie privée et confidentialité
- Bien-être sociétal et environnemental
- Autonomie humaine
- Réactivité et durabilité
- Diversité culturelle et linguistique
- Démocratie et autodétermination politique

Le critère de “malfaisance” a été inversé en “non-malfaisance” de manière à ce que tous les critères puissent être maximisés pour augmenter le caractère éthique de l'application.

Ces choix peuvent bien entendu être modifiés suivant le contexte. Ce qui est important est la démarche, pas les choix ci-dessus.

6.3.2 Détails techniques de l'implémentation

Outils utilisés

Les applications que nous avons proposées ont été développées en utilisant ReactJS et Material UI pour le frontend, et MongoDB, Express.js, et Node.js pour le backend.

- ReactJS est une bibliothèque JavaScript open-source, basée sur des composants (components en anglais), qui est utilisée pour construire des interfaces utilisateur interactives pour les applications web et mobiles [Rawat 2020].
- Material UI est une bibliothèque open-source qui fournit des composants et des styles préconçus pour React.
- MongoDB est un programme de gestion de base de données NoSQL (Not Only SQL) qui possède un modèle de stockage orienté document. Les bases de données NoSQL sont non relationnelles et flexibles, elles permettent de stocker et de traiter de grandes quantités de données.
- Node.js est un environnement d'exécution open source utilisé pour créer des applications web côté serveur à l'aide de JavaScript. Il utilise un modèle asynchrone piloté par les événements [Basumatary 2022]. Express.js est un framework Node.js qui permet de développer des applications web, des API et des applications mobiles multiplateformes.

Notre apport principal et novateur réside dans la fusion des LLMs et du prompt engineering intelligent pour aider à la formulation de concepts nécessitant une réflexion.

En traitant le langage naturel comme un support de programmation dynamique, nous exploitons les capacités étendues des LLMs pour faire des propositions en langage naturel lisible par les humains. Les LLMs sont formés sur de grands ensembles de textes, tels que des livres, des articles et des conversations. Ils sont connus pour leur capacité à traiter et à produire des textes semblables à ceux des humains, ce qui les rend efficaces pour des tâches telles que la création de contenu et la réponse aux questions [Lindemann 2024]. Une caractéristique importante des LLMs est le mécanisme d'attention, qui leur permet de se concentrer sur différentes parties du texte d'entrée [Zhao 2024], améliorant ainsi leur capacité à générer des réponses pertinentes sur le plan contextuel. La plupart de ces modèles sont construits sur un cadre de réseau neuronal appelé *transformer architecture*, qui est efficace pour les tâches linguistiques en raison de ses caractéristiques spécialisées pour le traitement des données linguistiques [Yang 2024].

Dans cette étude, nous avons utilisé deux modèles d'OpenAI : `GPT-3.5-turbo` pour la réalisation d'un automate de discussion (*chatbot*) et `GPT-3.5-turbo-instruct` pour proposer des questions spécifiques aux développeurs de logiciels, proposer un conseil portant sur les principes éthiques essentiels dans un domaine donné, proposer des normes éthiques et identifier les éventuelles corrélations existantes entre ces principes éthiques. Ces deux modèles sont accessibles via l'API OpenAI (<https://platform.openai.com/>). Le modèle GPT-3 de l'OpenAI est une famille de modèles de langage autorégressifs capables d'effectuer des tâches de complétion de texte semblables à celles des humains.

La version GPT-3.5-turbo compte 175 milliards de paramètres et a été entraînée sur une variété de documents approuvés et publics [Huang 2024].

Concernant l'ingénierie des prompts intelligents (*Smart prompt Engineering*), au lieu de l'affinement (*finetuning*) traditionnel, qui nécessite de vastes ensembles de données et masque souvent le processus de prise de décision², nous utilisons la vaste connaissance des LLMs tout en les guidant avec précision. Cela améliore l'adaptabilité et l'explicabilité de notre outil.

Par exemple, étant donné un prompt, le LLM élabore des questions qui sondent les alignements éthiques du logiciel, en s'appuyant sur les descriptions du logiciel et du domaine.

Plusieurs types d'utilisation sont possibles dans l'ingénierie des prompts : le *few-shot prompt*, le *one-shot prompt*, ou le *zero-shot prompt*.

On parle de *zero-shot prompt* lorsque l'on donne une tâche à un modèle de langage comme GPT-3 par exemple, sans aucun exemple ou formation préalable spécifique à cette tâche. Essentiellement, le modèle doit accomplir la tâche en se basant entièrement sur ses connaissances et sa compréhension préexistantes, acquises au cours de l'apprentissage initial.

Dans notre étude, nous utilisons le *zero-shot prompt* pour mettre le modèle à l'épreuve sans le préparer avec des prompts préalables et tester la capacité du modèle à généraliser.

Évaluation de conformité aux principes éthiques

Techniquement parlant, chaque principe éthique peut être vu comme un critère différent, ce qui nous amène donc dans un paradigme d'évaluation multi-critères.

S'il est mathématiquement démontré qu'une approche de Pareto est la meilleure manière d'évaluer un problème multi-critères, notre sélection de 10 critères nous amène dans un problème à 10 dimensions, difficilement visualisable pour les humains. En effet, si la visualisation d'un problème à 2 critères amène à une visualisation par un front de Pareto à deux dimensions (une courbe), le front de Pareto d'un problème à trois critères est une surface, celui d'un problème à trois critères un volume. Pour la quatrième dimension, on pourrait intervenir le temps avec une surface de Pareto se modifiant dans le temps mais au delà, l'espace à quatre dimensions dans lequel nous vivons rend très difficile la visualisation d'un front de Pareto à 10 dimensions.

Nous avons donc dans un premier temps choisi la facilité en représentant les 10 dimensions éthiques de notre évaluation par un graphe "radar" et pour évaluer la conformité aux principes éthiques, nous agrégerons simplement les critères par une combinaison linéaire (une somme pondérée similaire à celle utilisée pour évaluer la qualité d'un candidat qui passe les différentes disciplines du baccalauréat).

²<https://platform.openai.com/docs/guides/fine-tuning>

Afin de vérifier si le développeur de logiciels a respecté l'éthique de l'IA établie par le gouvernement, nous proposons donc la formule de moyenne pondérée 6.1 qui calcule le “pourcentage de conformité” avec les principes éthiques édictés par Ethiconorm et vérifier ensuite si le développeur a atteint les seuils minimaux demandés pour chaque principe éthique.

$$P = \frac{\sum_{i=1}^{10} \frac{R_i}{5} w_i}{\sum_{i=1}^{10} w_i} \times 100 \quad (6.1)$$

w_i est la pondération du i -ème principe éthique. R_i est la réponse du développeur au i -ème principe éthique.

La conformité à chacun des 10 principes étant obtenue par la réponse à 5 questions, les réponses du développeur pour chaque principe sont divisés par 5 afin d'obtenir une réponse par question sur une échelle de 0 à 1, où 0 signifie qu'il n'y a pas de réponse positive (non-conformité totale) et 1 signifie que le développeur a répondu “oui” à toutes les questions (conformité totale).

Ces valeurs normalisées sont ensuite multipliées par les poids respectifs attribués dans Ethiconorm, puis la moyenne pondérée est calculée. Elle est ensuite multipliée par 100 pour obtenir un “pourcentage de conformité” aux principes éthiques édictés dans Ethiconorm.

La valeur de cette approche réside dans sa simplicité.

Mise en œuvre dans Ethicotest :

Nous utilisons la description du logiciel à vérifier et son domaine spécifique comme des ancrs contextuelles intégrées en tant que prompts, pour guider les questions posées par le LLM en les alignant étroitement sur la fonctionnalité du logiciel et les considérations éthiques dans ce domaine.

Il s'agit en quelque sorte d'un “réglage fin à la volée” (*fine-tuning on the-fly*) : les vastes connaissances du modèle sont exploitées par le prompt qui contient la description du logiciel par son auteur, pour affiner les questions posées sans nécessiter des données d'entraînement supplémentaires.

En outre, cette méthode présente un avantage considérable : elle est adaptable. Au fur et à mesure que les logiciels évoluent ou que les normes éthiques spécifiques à un domaine changent, les prompts sont ajustés par les principes éthiques relativistes insérés dans Ethiconorm, garantissant ainsi que les résultats de l'IA restent conformes à l'évolution du contexte.

Pour terminer, un paramètre appelé température joue un rôle important dans nos prompts car il influence la variabilité des textes générés par le LLM. Une température plus élevée introduit un caractère plus aléatoire dans les textes proposés, tandis qu’une température plus basse produit des résultats plus déterministes. Ce paramètre permet d’affiner le modèle en fonction du niveau de créativité souhaité [Dulepet 2023].

Essentiellement, en tirant parti de l’ingénierie intelligente des prompts, nous montrons qu’avec les bons prompts, des modèles tels que LLM peuvent être “programmés” en temps réel à l’aide du langage naturel, comblant ainsi le fossé entre les connaissances générales de l’IA et les exigences spécifiques et localisées.

6.3.3 Vue d’ensemble des outils Ethiconorm et Ethicotest

Nos applications (voir figure 6.1) sont spécialement conçues pour tenter de combler le fossé éthique observé dans le développement de logiciels, en offrant des outils à la fois aux instances gouvernementales et aux développeurs de logiciels.

L’idée principale derrière ces deux applications est d’une part avec Ethiconorm, d’aider et de permettre au gouvernement d’établir une liste de normes ou principes éthiques désirés et d’attribuer une pondération à chacun de ces principes. Ces pondérations sont spécifiques à chaque domaine.

D’autre part, nous proposons aux développeurs une application appelée Ethicotest leur permettant, par le biais de questions spécifiques pour chaque norme ou principe éthique, de vérifier si leur logiciel respecte les normes ou préférences éthiques souhaitées.

Par conséquent, chaque ministère/domaine (santé, éducation, etc.) aura sa propre pondération des principes qu’il souhaitera mettre en avant et chaque développeur de logiciel pourra regarder la conformité de son logiciel aux principes éthiques établis par le ministère auquel son logiciel est rattaché.

Les questions posées au développeur sont proposées automatiquement par un Grand Modèle de Langage³ (*Large Language Model* ou LLM) basé sur le domaine d’application (santé, éducation, etc.) de son logiciel et sa description.

Chaque administration pourra par la suite examiner, puis valider ou rejeter les applications en se basant sur les réponses des développeurs. Si l’application est rejetée, les développeurs pourront améliorer leur solution en fonction du retour d’information de l’administration concernée.

Les réponses remontées par les développeurs dans Ethicotest pourraient par la suite engager les développeurs au cas où leurs logiciels ne suivraient pas les principes déclarés.

Ce processus itératif permet non seulement aux développeurs de revoir et d’affiner leur logiciel pour s’assurer qu’il est conforme aux normes éthiques souhaitées, sur la base

³https://fr.wikipedia.org/wiki/Grand_modèle_de_langage

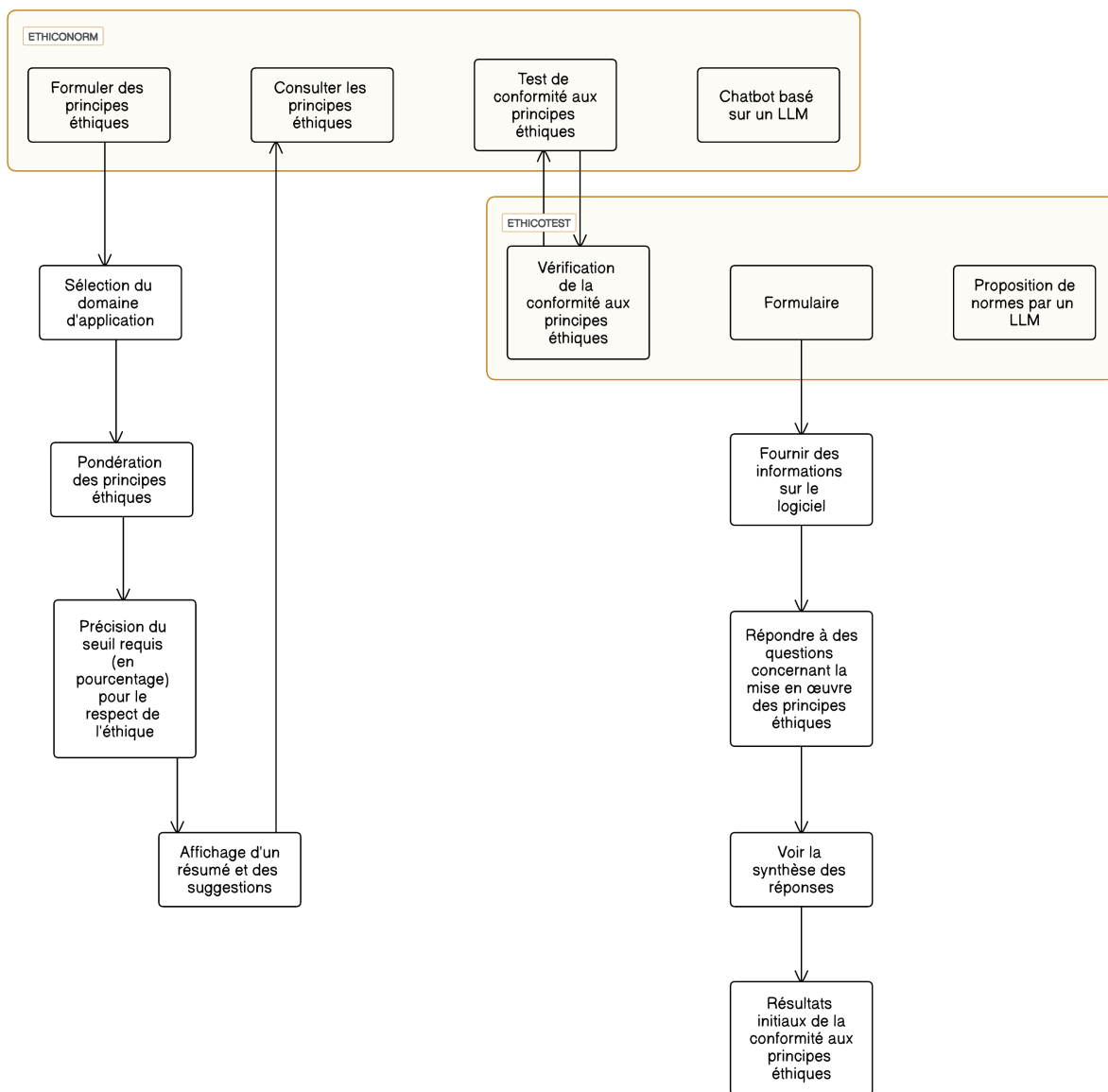


Figure 6.1: Schémas des applications Ethiconorm et Ethicotest

des commentaires et des réflexions obtenus, mais aussi aux administrations de revoir leurs normes et pondérations s’il s’avérait que des applications validées n’étaient pas conforme à leurs souhaits ou inversement, si quasiment toutes les applications étaient rejetées.

La possibilité offerte aux administrations de revoir et modifier les normes et pondérations désirées garantit la généralité de l’application Ethiconorm, car cela lui permet d’être utilisée dans différents domaines et contextes.

L’utilisation d’un LLM a été identifiée comme une méthode prometteuse pour non seulement faciliter la génération de questions appropriées pour les développeurs, mais aussi pour aider le gouvernement à mieux comprendre et formuler son éthique et à proposer des normes à suivre au niveau de chaque ministère. La participation d’un comité d’experts à ce processus est fortement encouragée. Cela peut garantir une éthique cohérente et représentative au sein de chaque ministère.

De plus, la nécessité de rédiger les normes et leur pondération aura pour effet secondaire bénéfique de permettre aux citoyens d’avoir accès à ces normes et pondérations, pour une meilleure visibilité et transparence de la volonté politique du gouvernement.

6.3.3.1 Ethiconorm: une application pour les instances gouvernementales

Ce module permet aux fonctionnaires de proposer et d’ajuster dynamiquement des normes éthiques locales en fonction d’un contexte spécifique. Ces lignes directrices ont été définies sous la forme de *prompts* écrits en langage naturel, garantissant ainsi la transparence des attentes en matière d’éthique. Le menu de l’application est présenté dans la figure 6.2

Application pour aider à définir une éthique de l’IA

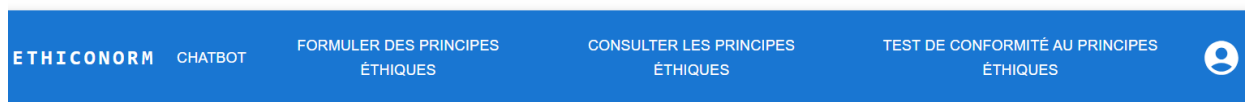


Figure 6.2: Menu de l’application destinée aux instances gouvernementales.

Liste des sections et fonctionnalités :

1. Section “Chatbot” (voir figure 6.3) : l’utilisateur peut poser des questions et demander des informations sur l’éthique de l’IA par l’intermédiaire d’un chatbot. Basé sur le LLM, ce Chatbot vise à guider le gouvernement dans la compréhension et la formulation de l’éthique au sein de chaque ministère.
2. Section “Formuler des principes éthiques” : les fonctionnaires peuvent enregistrer leurs attentes en matière d’éthique à l’aide d’un questionnaire. Le questionnaire

Bonjour, je suis là pour vous aider à définir les principes éthiques de votre IA.

Qu'est-ce que l'éthique de l'IA ?

L'éthique de l'IA est un domaine multidisciplinaire qui se concentre sur les questions morales et éthiques soulevées par le développement, le déploiement et l'utilisation de systèmes d'intelligence artificielle. L'éthique de l'IA cherche à identifier et à aborder les défis éthiques liés à l'impact de l'IA sur les individus, la société et l'environnement.

En d'autres termes, l'éthique de l'IA vise à garantir que l'IA est développée et utilisée de manière responsable, juste et respectueuse des valeurs humaines fondamentales. Cela implique de prendre en compte des questions telles que la transparence des algorithmes, l'équité des décisions automatisées, la protection de la vie privée et des données, la sécurité des systèmes d'IA, la responsabilité en cas de préjudice causé par des systèmes d'IA, et bien d'autres encore.

L'éthique de l'IA repose sur des principes et des normes éthiques qui guident la conception, le développement et l'utilisation de l'IA de manière éthique. Ces principes peuvent varier en fonction du contexte et des valeurs de chaque individu ou organisation.

Tapez votre message ici

Pouvez-vous donner des exemples concrets sur la manière dont les principes éthiques de transparence et responsabilité peuvent être appliqués ?

Bien sûr, je peux vous donner des exemples concrets sur la manière dont les principes éthiques de transparence et responsabilité peuvent être appliqués dans le domaine de l'intelligence artificielle :

1. Transparence :
 - Lors du développement d'un système d'IA, il est important de documenter clairement les étapes de conception, les choix algorithmiques, les jeux de données utilisés, et toute autre information pertinente.
 - Fournir des explications sur le fonctionnement d'un système d'IA aux utilisateurs finaux, notamment sur les prédictions ou recommandations qu'il fournit.
 - Mettre en place des mécanismes permettant aux utilisateurs de comprendre facilement pourquoi une décision a été prise par un système d'IA, en fournissant par exemple des justifications ou des insights sur les résultats obtenus.
2. Responsabilité :
 - Établir des mécanismes de contrôle et de supervision tout au long du cycle de vie d'un système d'IA, afin de surveiller son impact et de prendre des mesures correctives si nécessaire.
 - Identifier clairement les parties prenantes impliquées dans le développement et l'utilisation de l'IA, et définir les responsabilités de chacune d'entre elles.
 - Prévoir des procédures pour gérer les problèmes éthiques qui pourraient survenir, tels que les biais discriminatoires dans les algorithmes ou les conséquences inattendues sur les individus ou la société.

En appliquant ces principes de transparence et de responsabilité, les développeurs et les utilisateurs d'IA peuvent contribuer à garantir que les systèmes d'IA sont utilisés de manière éthique et respectueuse des valeurs humaines.

Figure 6.3: Capture d'écran de la section "Chatbot".

est spécifique à un domaine ou à un ministère. Cette section comporte trois étapes :

- Sélection du domaine d'application (santé, éducation, etc.). Il s'agit du domaine auquel est rattaché le ministère concerné (voir figure 6.4).

The screenshot shows a web interface titled "Définition des principes éthiques". At the top, there is a progress bar with four steps: "1 Étape 1", "2 Étape 2", "3 Étape 3", and "4 Récapitulatif". Below the progress bar, the instruction reads: "Sélectionnez le domaine d'application pour lequel vous souhaitez définir les principes éthiques." A dropdown menu is open, showing a list of domains: "Santé", "Éducation", "Industrie", "Tourisme", "Finance", and "Agriculture". The "Santé" option is currently selected and highlighted. To the right of the dropdown, a blue button labeled "SUIVANT" is partially visible.

Figure 6.4: Étape 1 du processus de formulation des principes éthiques (sélection du domaine).

- Pondération des principes éthiques (voir figure 6.5) : Dans cette étape, il est possible de consulter une liste de principes proposés, d'avoir une proposition du LLM sur les principes les plus essentiels dans le domaine donné, de voir la description de chacun des principes (voir figure 6.6) et d'attribuer une pondération à chaque principe en fonction de son importance dans le domaine concerné. Cette tâche de pondération est facilitée par un texte proposé par le LLM qui indique quels principes sont essentiels et lesquels sont moins importants dans le contexte du domaine d'application spécifique. Nous pensons que l'importance des principes éthiques peut varier d'un domaine à l'autre. C'est pourquoi nous avons utilisé le LLM pour extraire cette information.
- Spécifier la valeur minimale (le seuil) du "pourcentage de conformité aux principes éthiques" que le développeur doit atteindre pour que son logiciel

Définition des principes éthiques

✓ Étape 1 — 2 Étape 2 — 3 Étape 3 — 4 Récapitulatif

Vous avez choisi le domaine : santé. Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de la santé sont la transparence, la responsabilité, la non-malfaisance, l'équité et l'autonomie humaine. La transparence est essentielle pour garantir que les décisions prises par l'IA sont compréhensibles et justifiables. La responsabilité est nécessaire pour s'assurer que les acteurs impliqués dans le développement et l'utilisation de l'IA sont tenus responsables de leurs actions. La non-malfaisance est primordiale pour éviter tout préjudice ou dommage causé par l'IA. L'équité est importante pour garantir que l'IA ne reproduise pas les biais et les inégalités existants dans le système de santé. Enfin, l'autonomie humaine doit être préservée pour que les décisions finales en matière de santé restent entre les mains des individus.

Pondérez chacun des principes éthiques suivants selon votre contexte.

Transparence

Note Transparence [VOIR LA DESCRIPTION](#)

Responsabilité

Note Responsabilité [VOIR LA DESCRIPTION](#)

Non-malfaisance

Note Non-malfaisance [VOIR LA DESCRIPTION](#)

Équité

Note Équité [VOIR LA DESCRIPTION](#)

Vie privée et confidentialité

Note Vie privée et confidentialité [VOIR LA DESCRIPTION](#)

Bien-être sociétal et environnemental

Note Bien-être sociétal et environnemental [VOIR LA DESCRIPTION](#)

Autonomie humaine

Note Autonomie humaine [VOIR LA DESCRIPTION](#)

Réactivité et durabilité

Note Réactivité et durabilité [VOIR LA DESCRIPTION](#)

Diversité culturelle et linguistique

Note Diversité culturelle et linguistique [VOIR LA DESCRIPTION](#)

Démocratie et autodétermination politique

Note Démocratie et autodétermination politique [VOIR LA DESCRIPTION](#)

[RETOURNER](#) [SUIVANT](#)

Texte
généré par
un LLM

Figure 6.5: Étape 2 du processus de formulation des principes éthiques

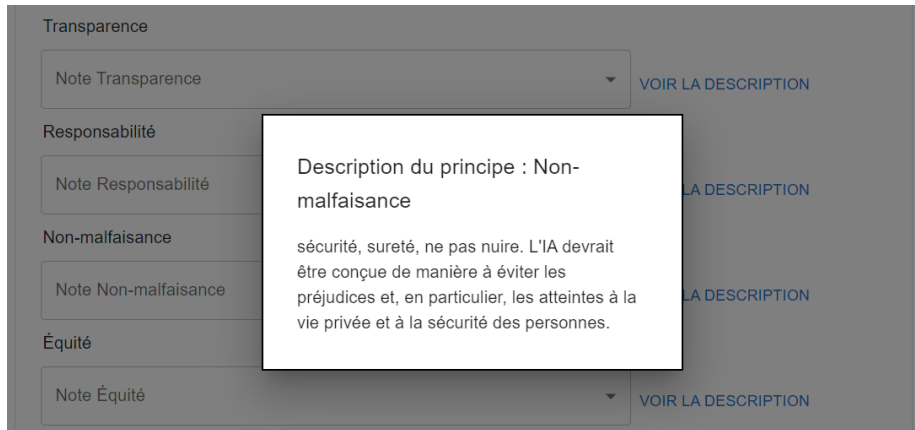


Figure 6.6: Affichage de la description du principe “Non-malfaisance”

soit conforme aux principes éthiques établie (voir la figure 6.7). Chaque ministère doit avoir son propre seuil, car certains domaines seront plus sensibles que d'autres sur des données similaires. Le pourcentage de conformité est important dans le processus d'évaluation de la conformation d'une application aux normes éthiques gouvernementales. Il représente la rigueur du gouvernement dans la mise en œuvre des principes éthiques et permet aux fonctionnaires d'évaluer dans quelle mesure le logiciel respecte les valeurs éthiques proposées.

Définition des principes éthiques

✓ Étape 1 ————— ✓ Étape 2 ————— **3** Étape 3 ————— 4 Récapitulatif

Indiquez le seuil requis(en pourcentage) pour la conformité au principes éthiques.

Minimum percentage of ethical compliance *

80

[VOIR LA DESCRIPTION](#)

RETOURNER
SUIVANT

Figure 6.7: Étape 3 du processus de formulation des principes éthiques

- À la fin, un résumé des informations fournies ainsi que quelques suggestions

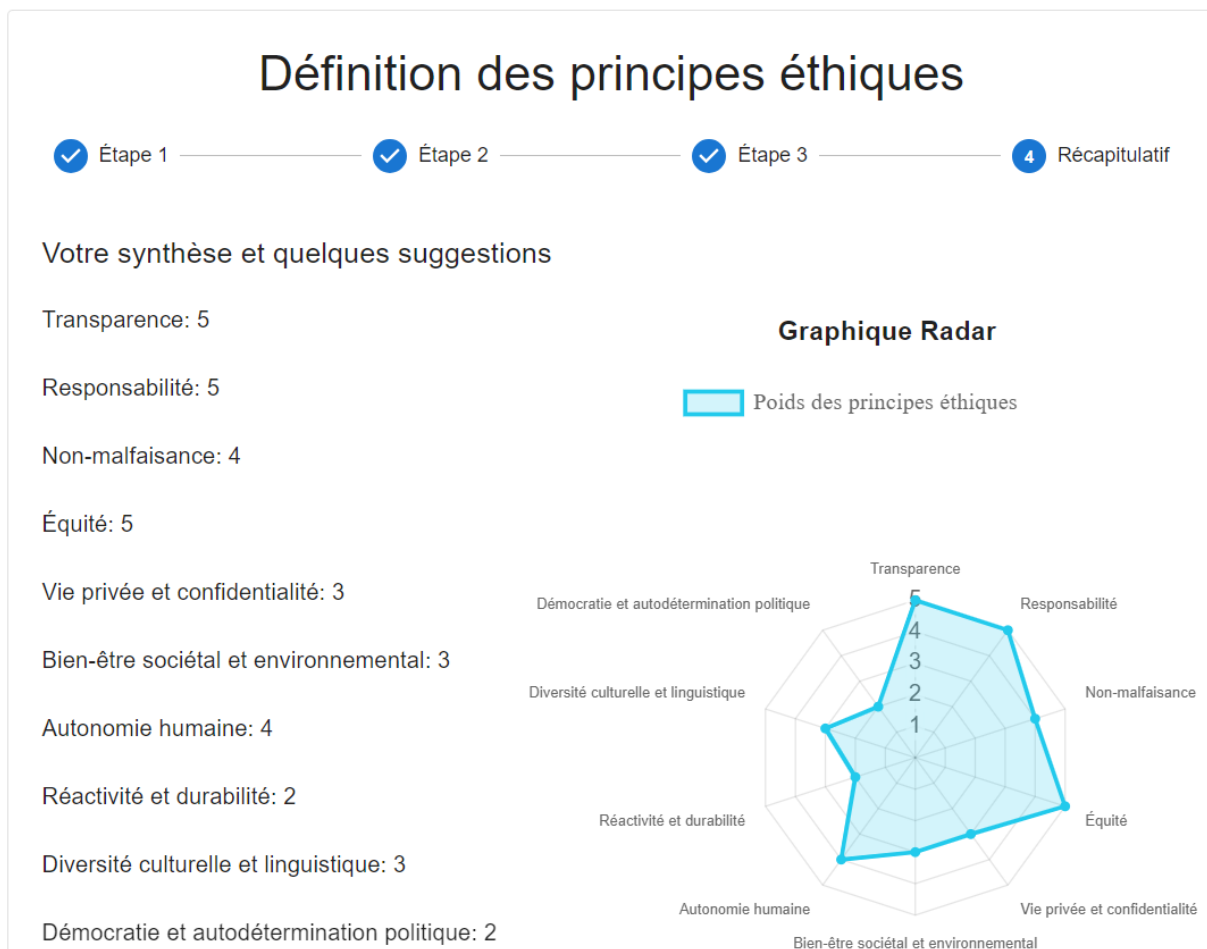


Figure 6.8: Le résumé (les pondérations précédemment saisies et un graphique en radar)

sont affichés :

- Les pondérations des principes saisies précédemment.
 - Un diagramme en radar pour présenter les pondérations (voir figure 6.8).
 - Une proposition de normes éthiques verbalisées par le LLM, que le développeur s’engagera à mettre en œuvre au cours du processus de développement du logiciel (voir figure 6.9).
 - Une vue d’ensemble des corrélations entre les différents principes éthiques (voir figure 6.10). Étant donné que les principes éthiques sont imbriqués, dépendant les uns des autres, ces corrélations peuvent jouer un rôle dans le calcul de la conformité d’un logiciel donné avec les principes éthiques établis. Par exemple, s’il existe une corrélation entre le principe de transparence et le principe d’équité et que la pondération de la transparence est respectée alors que celle du principe d’équité ne l’est pas, une compensation peut alors avoir lieu, permettant au logiciel du développeur d’être valide même si l’un des deux critères n’a pas atteint la valeur minimale demandée pour l’un des deux critères.
3. Section “Consulter les principes éthiques proposées” : l’utilisateur peut consulter toutes les informations qu’il a saisies dans le formulaire “Formuler des principes éthiques”, ainsi que la liste des normes et les corrélations existantes entre les principes.
 4. Section “Test de conformité aux principes éthiques” (voir figure 6.11) : Grâce à cette interface, chaque ministère peut examiner les réponses des développeurs à un formulaire de conformité aux principes éthiques (accessible via le module développeur). Le ministère valide ces réponses si elles sont conformes aux principes éthiques et les rejette si elles ne le sont pas.

6.3.3.2 Ethicotest : l’application de diagnostic des développeurs

Cet outil sert aux développeurs à évaluer la conformité de leur logiciel aux principes éthiques du domaine concerné. En traitant les caractéristiques du logiciel et les réponses aux questions prédéfinies, l’Ethicotest fournit un retour d’information sur les éventuels manquements aux principes éthiques formulés par l’administration. La figure 6.12 présente le menu de cette application.

Liste des sections et fonctionnalités :

1. Section “Normes” : les développeurs peuvent consulter les normes à appliquer lors du processus de développement du logiciel. Cette liste de normes a été préalablement proposée par un LLM et validée par l’administration via le module Ethiconorm.

2. Section “Formulaire” : les développeurs peuvent vérifier si leur logiciel est conforme aux principes et normes éthiques établie par l’administration à travers un questionnaire. Cela se fait en 3 étapes :

- Tout d’abord, les développeurs fournissent les informations suivantes (voir figure 6.13) : nom du logiciel, domaine d’application et description.
- Ils doivent ensuite répondre par “oui” ou “non” à des questions concernant la mise en œuvre de chaque principe ou norme éthique (voir les figures 6.14 et 6.15). Il y a cinq questions par norme. Ces questions sont proposées par un LLM, sur la base de la description du logiciel, du domaine d’application et du principe en question, ce qui permet (grâce au LLM) une vérification spécifique au domaine de l’intégration des principes ou normes éthiques demandées dans le processus de développement du logiciel.
- Enfin, le développeur peut consulter un résumé de ses réponses (voir figure 6.16).:
 - Un graphique “radar” superposant les résultats des réponses des développeurs et les pondérations de l’administration pour chaque principe ou norme permet une comparaison claire entre ce que met en œuvre l’application et les attentes de l’administration.
 - Le nombre de points obtenus pour chaque principe, le nom, le domaine et la description du logiciel.

La réponse des développeurs à ce questionnaire doit être validée par une personne ou un comité d’éthique au sein de l’administration concernée, bien qu’il soit imaginable de donner cette vérification à faire à une IA (cf. section Discussion).

3. Section “Conformité aux principes éthiques” (voir figure 6.17) : les développeurs peuvent examiner les commentaires de l’administration auquel le domaine du logiciel est rattaché. Ils peuvent ainsi vérifier la conformité du logiciel aux attentes gouvernementales en matière d’éthique (voir figure 6.18).

6.4 Résultats expérimentaux

Comme gpt-3.5-turbo est optimisé pour le chat tandis que gpt-3.5-turbo-instruct est un modèle InstructGPT, le modèle GPT-3.5-Turbo peut prendre en compte plusieurs messages ou instructions, au lieu d’un seul. Il peut donc “se souvenir” de ce qui a été dit auparavant et utiliser ces informations pour créer de meilleurs textes.

A l’inverse, gpt-3.5-turbo-instruct ne reçoit qu’une seule instruction en entrée.

Dans ce travail, le modèle GPT-3.5 Turbo a été utilisé juste pour la réalisation du chatbot, le gpt-3.5-turbo-instruct quant à lui a été utilisé pour tous les autres prompts

(génération de normes, génération de questions spécifiques pour le développeur, etc...) car c'est lui qui donne les réponses les plus claires et les plus et concises, qui vont droit au but.

Prompts utilisés dans cette contribution et leurs résultats

Plusieurs tests ont été effectués afin de trouver le prompt qui répond le mieux à nos attentes et d'étudier la différence entre les résultats. Ci-dessous, la liste des prompts :

Prompt visant à proposer cinq questions sur l'application de chaque principe éthique (voir figure 6.19) : l'objectif de ce prompt est de proposer cinq questions auxquelles le développeur doit répondre. Chaque question porte sur la mise en œuvre de chaque principe éthique.

Les variables : `descApp` (description du logiciel), `domain` (domaine d'application), `principeName` (nom du principe éthique) sont données en entrée. Elles proviennent de la description de l'application faite par le développeur

`descApp` = "IBM Watson for Oncology est un logiciel utilisant l'IA conçu pour aider les oncologues à diagnostiquer et à créer des plans de traitement personnalisés pour les patients atteints de cancer. Développé par IBM en collaboration avec le Memorial Sloan Kettering Cancer Center, ce logiciel exploite la puissance de l'IA pour analyser de grands volumes de littérature médicale, de données sur les patients et d'informations sur les essais cliniques afin de fournir des recommandations de traitement fondées sur des preuves."

`domain` = "Santé"

Voici les résultats obtenus en faisant varier la valeur de la température :

- Température = 0 pour le principe **Transparence** (sortie au format JSON) :

```

"questions":
[ "id": 1, "name": "q1", "label":
"Le logiciel IBM Watson for Oncology fournit-il des informations
détaillées sur les données et les sources utilisées pour générer
ses recommandations de traitement ?",
"id": 2, "name": "q2", "label":
"Le logiciel IBM Watson for Oncology permet-il aux oncologues
de comprendre comment les recommandations de traitement ont été
générées ?",
"id": 3, "name": "q3", "label":
"Le logiciel IBM Watson for Oncology fournit-il des explications
claires sur les raisons pour lesquelles une recommandation de
traitement a été faite pour un patient spécifique ?",
"id": 4, "name": "q4", "label":
"Le logiciel IBM Watson for Oncology permet-il aux oncologues de
vérifier les sources et les données utilisées pour générer les
recommandations de traitement ?",

```

```

“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il des informations
sur les limites et les biais potentiels de ses recommandations de
traitement ?”]

```

- Température = 0 pour le principe Responsabilité (sortie au format JSON)

:

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology est-il conçu pour être utilisé
uniquement par des professionnels de la santé qualifiés ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology est-il régulièrement mis à
jour pour refléter les dernières avancées médicales ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology est-il transparent sur
les données et les algorithmes utilisés pour générer ses
recommandations de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology est-il soumis à des tests
rigoureux pour garantir la précision de ses recommandations de
traitement ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology est-il accompagné d’une
formation adéquate pour les professionnels de la santé qui
l’utilisent ?”]

```

- Température = 0 pour le principe Non-malfaisance (sortie au format JSON)

:

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology a-t-il été testé sur un large
échantillon de patients pour s’assurer de son efficacité et de sa
sécurité dans le domaine de la santé ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology prend-il en compte les données
médicales et les préférences du patient pour proposer un plan de
traitement personnalisé ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology est-il régulièrement mis à
jour pour prendre en compte les dernières avancées médicales et les
nouvelles données sur les patients ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology est-il conçu pour éviter tout
préjudice ou dommage aux patients lors de l’utilisation de ses
recommandations de traitement ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology est-il soumis à des contrôles
de qualité et de sécurité réguliers pour garantir sa conformité aux
normes éthiques et médicales ?”]

```


- Température = 0 pour le principe Équité (sortie au format JSON) :


```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les différences individuelles
des patients (âge, sexe, origine ethnique, etc.) dans ses
recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel utilise-t-il des données équilibrées et
représentatives pour entraîner son algorithme ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel est-il capable de détecter et de corriger les biais
potentiels dans ses recommandations de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel fournit-il des explications claires et transparentes
sur la manière dont il a pris ses décisions de traitement ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel est-il régulièrement évalué et mis à jour pour
garantir son équité dans ses recommandations de traitement ?”]

```
- Température = 0 pour le principe Vie privée et confidentialité (sortie au format JSON) :


```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology est-il conçu pour protéger les
données personnelles des patients ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology respecte-t-il les
réglementations en matière de protection de la vie privée et de
la confidentialité des données médicales ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux patients de
contrôler l'accès à leurs données médicales ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology utilise-t-il des mesures de
sécurité pour protéger les données médicales des patients ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology est-il transparent sur la
manière dont il utilise les données médicales des patients ?”]

```
- Température = 0 pour le principe Bien-être sociétal et environnemental (sortie au format JSON) :


```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology prend-il en compte les
impacts sociétaux et environnementaux dans ses recommandations
de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology a-t-il été développé en
collaboration avec des experts en éthique et en santé publique

```

```

    ?”,
    “id”: 3, “name”: “q3”, “label”:
    “Le logiciel IBM Watson for Oncology utilise-t-il des données
    provenant de sources fiables et éthiques pour ses recommandations
    de traitement ?”,
    “id”: 4, “name”: “q4”, “label”:
    “Le logiciel IBM Watson for Oncology est-il régulièrement mis
    à jour pour prendre en compte les évolutions sociétales et
    environnementales ?”,
    “id”: 5, “name”: “q5”, “label”:
    “Le logiciel IBM Watson for Oncology dispose-t-il de mécanismes de
    transparence pour permettre aux utilisateurs de comprendre comment
    les recommandations de traitement sont générées ?”]

```

- Température = 0 pour le principe Autonomie humaine (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux oncologues
de prendre des décisions finales concernant le diagnostic et le
traitement des patients atteints de cancer ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il des recommandations
de traitement basées sur des preuves médicales et des données sur
les patients ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology prend-il en compte les
préférences et les valeurs des patients dans ses recommandations
de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology est-il transparent dans son
processus de prise de décision et dans les données qu’il utilise
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology est-il régulièrement mis à
jour et évalué pour garantir son efficacité et sa précision dans
les recommandations de traitement ?”]

```

- Température = 0 pour le principe Réactivité et durabilité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology est-il capable de fournir
des recommandations de traitement personnalisées pour les patients
atteints de cancer ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology est-il alimenté par des
données provenant de sources fiables et vérifiées ?”,
“id”: 3, “name”: “q3”, “label”:

```

```

“Le logiciel IBM Watson for Oncology est-il en mesure de s’adapter
aux évolutions de la recherche médicale et des données sur les
patients ?”,

```

```

“id”: 4, “name”: “q4”, “label”:

```

```

“Le logiciel IBM Watson for Oncology est-il conçu pour être utilisé
par des professionnels de la santé qualifiés et formés ?”,

```

```

“id”: 5, “name”: “q5”, “label”:

```

```

“Le logiciel IBM Watson for Oncology est-il régulièrement mis à
jour et maintenu pour garantir sa fiabilité et sa précision ?”]

```

- Température = 0 pour le principe Diversité culturelle et linguistique (sortie au format JSON) :

```

“questions”:

```

```

[ “id”: 1, “name”: “q1”, “label”:

```

```

“Le logiciel prend-il en compte les différences culturelles et
linguistiques dans ses recommandations de traitement pour les
patients atteints de cancer ?”,

```

```

“id”: 2, “name”: “q2”, “label”:

```

```

“Le logiciel a-t-il été testé sur des populations diverses pour
s’assurer qu’il n’y a pas de biais culturel ou linguistique dans
ses recommandations ?”,

```

```

“id”: 3, “name”: “q3”, “label”:

```

```

“Le logiciel permet-il aux utilisateurs de spécifier leur langue et
leur culture pour une meilleure adaptation des recommandations de
traitement ?”,

```

```

“id”: 4, “name”: “q4”, “label”:

```

```

“Le logiciel utilise-t-il des sources d’informations variées et
représentatives pour éviter les biais culturels et linguistiques
dans ses recommandations ?”,

```

```

“id”: 5, “name”: “q5”, “label”:

```

```

“Le logiciel dispose-t-il d’un mécanisme pour détecter et corriger
les biais culturels et linguistiques dans ses recommandations de
traitement ?”]

```

- Température = 0 pour le principe Démocratie et autodétermination politique (sortie au format JSON) :

```

“questions”:

```

```

[ “id”: 1, “name”: “q1”, “label”:

```

```

“Le logiciel IBM Watson for Oncology prend-il en compte les
préférences et les valeurs des patients dans ses recommandations
de traitement ?”,

```

```

“id”: 2, “name”: “q2”, “label”:

```

```

“Le logiciel IBM Watson for Oncology permet-il aux patients de
participer activement à la prise de décision concernant leur
traitement ?”,

```

```

“id”: 3, “name”: “q3”, “label”:

```

```

“Le logiciel IBM Watson for Oncology garantit-il la transparence et
la traçabilité de ses recommandations de traitement ?”,

```

```

“id”: 4, “name”: “q4”, “label”:

```

```

“Le logiciel IBM Watson for Oncology prend-il en compte les
différences culturelles et sociales dans ses recommandations de
traitement ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux patients de
contester les recommandations de traitement proposées par l’IA ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est
Transparence (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il des informations
claires et compréhensibles sur les bases de ses recommandations de
traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel est-il transparent sur la façon dont il utilise
les données des patients pour formuler ses recommandations de
traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel fournit-il des informations sur sa méthodologie de
prise de décision et ses sources de données ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel donne-t-il une explication claire des raisons derrière
ses recommandations de traitement pour un patient donné ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel permet-il à l’oncologue d’accéder facilement à
toutes les données utilisées pour formuler ses recommandations
de traitement pour un patient donné ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est
Responsabilité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“La mise en place de ce logiciel repose-t-elle sur des données
fiables et vérifiables provenant de sources médicales réputées ?”,
“id”: 2, “name”: “q2”, “label”:
“Ce logiciel est-il soumis à des évaluations régulières de la
part d’experts en oncologie pour vérifier l’exactitude de ses
recommandations de traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Y a-t-il des mécanismes en place pour assurer la transparence
des décisions prises par ce logiciel et pour expliquer les
raisonnements derrière ces décisions ?”,
“id”: 4, “name”: “q4”, “label”:
“Est-ce que ce logiciel est conçu pour réduire les risques de
biais dans ses recommandations de traitement, notamment en ce
qui concerne les différences de traitement entre les groupes de
patients ?”,
“id”: 5, “name”: “q5”, “label”:

```

```
“Existe-t-il des protocoles rigoureux en place pour traiter les cas
où les recommandations de traitement de ce logiciel sont en conflit
avec celles des oncologues ?”]
```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Non-malfaisance (sortie au format JSON) :

```
“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel utilise-t-il des données fiables et vérifiées pour
fournir des recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Les recommandations de traitement fournies par le logiciel
sont-elles constamment mises à jour avec les dernières recherches
en oncologie ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel prend-il en compte les caractéristiques individuelles
des patients pour proposer des plans de traitement personnalisés
?”,
“id”: 4, “name”: “q4”, “label”:
“Les décisions finales de traitement sont-elles laissées aux
oncologues et non dictées par le logiciel ?”,
“id”: 5, “name”: “q5”, “label”:
“Y a-t-il des processus clairs en place pour détecter et résoudre
d’éventuelles erreurs ou biais du logiciel qui pourraient nuire aux
patients ?”]
```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Équité (sortie au format JSON) :

```
“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les données démographiques
des patients (âge, sexe, origine ethnique) pour fournir des
recommandations de traitement équitables pour tous ?”,
“id”: 2, “name”: “q2”, “label”:
“Les essais cliniques inclus dans l’analyse du logiciel sont-ils
représentatifs de la diversité des populations de patients ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel a-t-il été testé pour détecter et corriger les biais
éventuels dans les données d’entraînement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel permet-il aux médecins d’accéder à ses recommandations
et de les modifier en fonction de leur jugement clinique ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel prévoit-il un processus de révision régulière par des
experts en éthique pour évaluer l’équité de ses recommandations de
traitement ?”]
```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Vie privée et confidentialité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel a-t-il été conçu pour garantir la protection des
données des patients ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel permet-il aux patients de contrôler l’accès à leurs
données personnelles ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel dispose-t-il de mesures de sécurité pour empêcher les
violations de données ?”,
“id”: 4, “name”: “q4”, “label”:
“Y a-t-il des procédures en place pour gérer les demandes d’accès
et de suppression des données des patients ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel permet-il de générer des rapports détaillés sur
l’utilisation et l’accès aux données des patients ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Bien-être sociétal et environnemental (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les risques potentiels sur le
bien-être sociétal et environnemental lors de l’élaboration de ses
recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel est-il conçu pour minimiser les effets néfastes sur le
bien-être sociétal et environnemental résultants de l’utilisation
de ses recommandations de traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel est-il régulièrement mis à jour pour s’adapter aux
nouvelles connaissances et pratiques en matière de bien-être
sociétal et environnemental dans le domaine de la santé ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel prend-il en compte les aspects éthiques liés aux
ressources utilisées pour son fonctionnement, telles que l’énergie
et l’utilisation de matériel non-renouvelable ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel est-il transparent sur les données et les sources
utilisées pour ses recommandations de traitement, permettant
ainsi une évaluation de leur impact sur le bien-être sociétal et
environnemental ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Autonomie humaine (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les préférences et les valeurs
personnelles du patient dans la recommandation de traitement ?”,
“id”: 2, “name”: “q2”, “label”:

```

```

“Le logiciel demande-t-il l’autorisation du patient avant de
partager ses données avec des tiers ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel permet-il aux oncologues de modifier les
recommandations de traitement en fonction de leur expérience
clinique et de leur jugement professionnel ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel présente-t-il clairement les données et les
algorithmes utilisés pour générer les recommandations de traitement
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel propose-t-il un mécanisme de correction en cas
d’erreur ou d’information erronée dans les données d’entraînement
de l’IA ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Réactivité et durabilité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel repose-t-il sur des données fiables et pertinentes
pour fournir des recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Les méthodes et les algorithmes utilisés pour générer les
recommandations de traitement sont-ils transparents et
compréhensibles pour les professionnels de la santé ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel est-il capable de s’adapter à l’évolution des
connaissances médicales et des méthodes de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Les données des patients sont-elles protégées et utilisées de
manière éthique ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel est-il régulièrement évalué et mis à jour pour assurer
une performance et une précision optimales ?”]

```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Diversité culturelle et linguistique (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les différences culturelles dans
ses recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel offre-t-il des options de traitement en plusieurs
langues ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel est-il capable de reconnaître et de traiter les
particularités linguistiques de chaque patient ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel est-il doté de capacités de traduction pour faciliter

```

```
la communication avec les patients de différentes cultures ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel fournit-il des explications et informations claires
sur les recommandations de traitement pour les patients provenant
de différentes cultures ?”]
```

- Sortie du prompt lorsque la température = 1 et que le nom du principe est Démocratie et autodétermination politique (sortie au format JSON) :

```
“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Est-ce que le logiciel applique le principe de démocratie et
autodétermination politique dans la prise de décision pour les
options de traitement proposées pour les patients atteints de
cancer ?”,
“id”: 2, “name”: “q2”, “label”:
“Est-ce que le logiciel prend en compte les préférences et les
besoins spécifiques des patients en matière de traitement,
conformément au principe de démocratie et autodétermination
politique ?”,
“id”: 3, “name”: “q3”, “label”:
“Le développement du logiciel a-t-il été supervisé par un comité
d’experts en oncologie ainsi que par des représentants de la
société civile et des patients, comme le préconise le principe
de démocratie et autodétermination politique ?”,
“id”: 4, “name”: “q4”, “label”:
“Est-ce que le logiciel offre des options de traitement
diversifiées, conformément au principe de démocratie et
autodétermination politique qui prône l’égalité d’accès aux soins
pour tous les patients ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel a-t-il été testé et validé par des professionnels
de la santé et des patients pour s’assurer qu’il respecte le
principe de démocratie et autodétermination politique dans sa
mise en pratique ?”]
```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Transparence (sortie au format JSON) :

```
“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il une explication
claire et détaillée de son fonctionnement et de sa prise de
décision ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux utilisateurs
de comprendre comment les données sont collectées, traitées et
utilisées pour générer des recommandations de traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il une liste complète
des sources de données utilisées pour générer ses recommandations
```



```
de traitement ?",
"id": 4, "name": "q4", "label":
"Le logiciel IBM Watson for Oncology permet-il aux utilisateurs
de vérifier et de valider les données utilisées pour générer ses
recommandations de traitement ?",
"id": 5, "name": "q5", "label":
"Le logiciel IBM Watson for Oncology permet-il aux utilisateurs
de comprendre les limites et les biais potentiels de ses
recommandations de traitement ?"]
```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Responsabilité (sortie au format JSON) :

```
"questions":
[ "id": 1, "name": "q1", "label":
"Le logiciel prend-il en compte les données médicales des patients
de manière responsable et éthique ?",
"id": 2, "name": "q2", "label":
"Le logiciel utilise-t-il des sources fiables et vérifiées pour
fournir ses recommandations de traitement ?",
"id": 3, "name": "q3", "label":
"Le logiciel s'assure-t-il de respecter la confidentialité des
données médicales des patients ?",
"id": 4, "name": "q4", "label":
"Le logiciel est-il transparent dans son processus de prise de
décision et de recommandation de traitement ?",
"id": 5, "name": "q5", "label":
"Le logiciel prend-il en compte les biais éventuels dans ses
recommandations de traitement ?"]
```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Non-malfaisance (sortie au format JSON) :

```
"questions":
[ "id": 1, "name": "q1", "label":
"Le logiciel IBM Watson for Oncology est-il conçu pour réduire les
risques pour les patients atteints de cancer ?",
"id": 2, "name": "q2", "label":
"Le logiciel IBM Watson for Oncology a-t-il été testé pour
s'assurer qu'il ne causera pas de préjudice aux patients ?",
"id": 3, "name": "q3", "label":
"Le logiciel IBM Watson for Oncology est-il équipé de mécanismes de
sécurité pour éviter les erreurs de diagnostic ?",
"id": 4, "name": "q4", "label":
"Le logiciel IBM Watson for Oncology est-il régulièrement mis à
jour pour garantir qu'il ne causera pas de préjudice aux patients
?",
"id": 5, "name": "q5", "label":
"Le logiciel IBM Watson for Oncology est-il soumis à des contrôles
de qualité pour s'assurer qu'il ne causera pas de préjudice aux
patients ?"]
```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Équité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les différences de race, de genre
et de classe sociale dans ses recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel a-t-il été testé sur des populations diverses pour
s’assurer qu’il ne présente pas de biais ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel fournit-il des explications claires sur les raisons
derrière ses recommandations de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel permet-il aux médecins de modifier ou de rejeter les
recommandations de traitement en fonction de leur jugement clinique
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel a-t-il été évalué par des experts en éthique pour
s’assurer qu’il respecte les principes d’équité ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Vie privée et confidentialité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology garantit-il la sécurité des
données des patients en respectant les normes de confidentialité et
de protection de la vie privée ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux patients de
contrôler l’accès à leurs données personnelles ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology utilise-t-il des techniques de
pseudonymisation ou d’anonymisation pour protéger les données des
patients ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology respecte-t-il les règles de
consentement éclairé pour l’utilisation des données des patients
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology dispose-t-il d’un système de
sécurité robuste pour protéger les données des patients contre les
cyberattaques ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Bien-être sociétal et environnemental (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les impacts sociétaux et
environnementaux de ses recommandations de traitement ?”,

```

```

“id”: 2, “name”: “q2”, “label”:
“Le logiciel utilise-t-il des sources de données fiables et
éthiques pour fournir ses recommandations ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel est-il transparent sur les critères utilisés pour
générer ses recommandations de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel est-il régulièrement mis à jour pour prendre en compte
les avancées en matière de bien-être sociétal et environnemental
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel a-t-il été testé pour s’assurer qu’il ne contribue
pas à des inégalités en matière de bien-être sociétal et
environnemental ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Autonomie humaine (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux oncologues de
prendre des décisions autonomes sur les plans de traitement à
proposer aux patients atteints de cancer ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology prend-il en compte les
préférences et les valeurs individuelles des patients lors de la
recommandation de traitements ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology permet-il aux oncologues de
modifier les recommandations de traitement proposées par l’IA en
fonction de leur jugement clinique ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology fournit-il des explications
claires et compréhensibles sur les recommandations de traitement
proposées ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology respecte-t-il les normes
éthiques et légales en matière de protection de la vie privée des
patients et de confidentialité des données médicales ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Réactivité et durabilité (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel IBM Watson for Oncology est-il réactif aux changements
dans les données médicales et les recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel IBM Watson for Oncology est-il capable de fournir des
recommandations de traitement durables pour les patients atteints
de cancer ?”,

```

```

“id”: 3, “name”: “q3”, “label”:
“Le logiciel IBM Watson for Oncology prend-il en compte les
différences individuelles des patients dans ses recommandations
de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel IBM Watson for Oncology est-il en mesure de s’adapter
aux nouvelles découvertes médicales et aux avancées technologiques
?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel IBM Watson for Oncology est-il soumis à une
surveillance et à une évaluation régulières pour garantir sa
réactivité et sa durabilité ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Diversité culturelle et linguistique (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel prend-il en compte les différences culturelles et
linguistiques dans ses recommandations de traitement pour les
patients atteints de cancer ?”,
“id”: 2, “name”: “q2”, “label”:
“Le logiciel a-t-il été entraîné avec des données provenant de
différentes cultures et langues pour garantir une diversité dans
ses recommandations de traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Le logiciel permet-il aux utilisateurs de sélectionner leur langue
et leur culture pour une meilleure adaptation des recommandations
de traitement ?”,
“id”: 4, “name”: “q4”, “label”:
“Le logiciel prend-il en compte les croyances et pratiques
médicales spécifiques à chaque culture dans ses recommandations
de traitement ?”,
“id”: 5, “name”: “q5”, “label”:
“Le logiciel fournit-il des explications claires et compréhensibles
sur les recommandations de traitement pour les patients de
différentes cultures et langues ?”]

```

- Sortie du prompt lorsque la température = 0.4 et que le nom du principe est Démocratie et autodétermination politique (sortie au format JSON) :

```

“questions”:
[ “id”: 1, “name”: “q1”, “label”:
“Le logiciel utilise-t-il des données provenant de sources fiables
pour fournir des recommandations de traitement ?”,
“id”: 2, “name”: “q2”, “label”:
“Est-ce que le logiciel prend en compte les préférences et les
valeurs des patients dans ses recommandations de traitement ?”,
“id”: 3, “name”: “q3”, “label”:
“Est-ce que le logiciel permet aux oncologues de remettre en
question ou de modifier les recommandations de traitement proposées
”]

```

```

par l'IA ?",
'id': 4, 'name': 'q4', 'label':
'Est-ce que le logiciel est transparent quant à la façon dont il
prend ses décisions et formule ses recommandations de traitement
?',
'id': 5, 'name': 'q5', 'label':
'Le logiciel est-il régulièrement mis à jour pour tenir compte des
nouvelles données et des nouvelles recherches médicales ?']

```

prompt pour proposer une liste de normes à appliquer lors du développement d'un logiciel intelligent (voir figure 6.20). Les variables : "listPrincipes" (liste des principes éthiques), "domaine" (domaine d'application), "notations" (poids accordé par le gouvernement aux principes éthiques) sont données en entrée.

Sortie du prompt lorsque la température = 0, domaine = santé, notations = [Transparence : 5, Responsabilité : 5, Non-malfaisance : 4, Équité : 5, Vie privée et confidentialité : 3, Bien-être sociétal et environnemental : 3, Autonomie humaine : 4, Réactivité et durabilité : 2, Diversité culturelle et linguistique : 3, Démocratie et autodétermination politique : 2] :

1. Transparence

- Fournir une documentation complète et transparente sur le fonctionnement de l'IA utilisée dans le logiciel
- Informer clairement les utilisateurs sur la provenance et l'utilisation des données utilisées par l'IA

2. Responsabilité

- S'assurer que le logiciel respecte les normes éthiques en matière de santé établies par les autorités compétentes
- Être en mesure de justifier les décisions prises par l'IA en cas de problème médical lié au logiciel

3. Non-malfaisance

- Prendre toutes les mesures nécessaires pour éviter que le logiciel ne cause des dommages physiques ou moraux aux patients
- Mettre en place des mécanismes de prévention et de correction en cas de défaillance de l'IA dans le domaine de la santé

4. Équité

- Garantir que le logiciel ne discrimine pas les patients en fonction de leur origine, leur statut social, leur sexe, etc.
- S'assurer que l'IA ne reproduise pas des biais existants dans le domaine de la santé

5. Vie privée et confidentialité

- Garantir la protection des données médicales des patients utilisées par l'IA
 - Ne partager aucune donnée personnelle avec des parties non autorisées sans consentement explicite des patients
6. Bien-être sociétal et environnemental
 - S'assurer que le logiciel n'affecte pas négativement la santé et le bien-être des communautés environnantes
 - Prendre en compte les impacts environnementaux lors de la mise en place du logiciel et de son utilisation
 7. Autonomie humaine
 - Ne pas limiter l'autonomie et les choix des patients en utilisant l'IA
 - Mettre en place des mécanismes permettant aux patients de maintenir leur libre arbitre en matière de santé
 8. Réactivité et durabilité
 - Assurer une mise à jour régulière du logiciel en fonction des évolutions en matière de santé
 - Prendre en compte les limites et les risques potentiels du logiciel dans un contexte de durabilité
 9. Diversité culturelle et linguistique
 - S'assurer qu'une diversité culturelle et linguistique est prise en compte lors de la conception et de l'utilisation du logiciel
 - Sensibiliser le personnel médical à l'importance de la diversité culturelle et linguistique dans le domaine de La santé
 10. Démocratie et autodétermination politique
 - Garantir la transparence et la libre circulation de l'information dans le domaine de la santé
 - Permettre aux individus et aux communautés d'avoir un contrôle sur l'utilisation de l'IA dans le domaine de la santé

prompt pour proposer les corrélations existantes entre les principes éthiques (voir figure 6.21) : Si le pourcentage de conformité du développeur est inférieur au seuil de conformité aux principes éthiques de l'IA" fixée par le gouvernement, les corrélations entre les principes éthiques peuvent devenir significatives. Ces corrélations peuvent aider les fonctionnaires à décider d'approuver ou non le logiciel du développeur. Par exemple, s'il existe une corrélation entre le principe A et le principe B, et que le développeur a suivi de près les lignes directrices du principe A mais pas celles du principe B, l'adhésion au principe A peut compenser l'absence de respect du principe B.

Sortie du prompt lorsque la température = 0 et le domaine est la santé : voir figure 6.10

Prompt visant à faire une proposition sur les principes les plus essentiels
dans un domaine donné (voir figure 6.22)

Sortie du prompt lorsque la température = 0 et le domaine = santé :

Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de la santé sont la transparence, la responsabilité, la non-malfaisance, l'équité et l'autonomie humaine. La transparence est essentielle pour garantir que les décisions prises par l'IA sont compréhensibles et justifiables. La responsabilité est nécessaire pour s'assurer que les acteurs impliqués dans le développement et l'utilisation de l'IA sont tenus responsables de leurs actions. La non-malfaisance est primordiale pour éviter tout préjudice ou dommage causé par l'IA. L'équité est importante pour garantir que l'IA ne reproduise pas les biais et les inégalités existants dans le système de santé. Enfin, l'autonomie humaine doit être préservée pour que les décisions finales en matière de santé restent entre les mains des individus.

Sortie du prompt lorsque la température = 0.5 et le domaine = santé :

Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de la santé sont la transparence, la responsabilité, la non-malfaisance, l'équité et l'autonomie humaine. La transparence est essentielle pour garantir que les décisions prises par l'IA sont compréhensibles et traçables. La responsabilité est cruciale pour s'assurer que les acteurs impliqués dans le développement et l'utilisation de l'IA assument leurs actions. La non-malfaisance est primordiale pour éviter les préjudices causés par l'IA. L'équité est importante pour garantir un accès égal à l'IA et ses avantages pour tous. Enfin, l'autonomie humaine est essentielle pour s'assurer que l'IA ne remplace pas le jugement et la prise de décision humaine dans le domaine de la santé.

Sortie du prompt lorsque la température = 0.5 et le domaine = éducation :

Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de l'éducation sont la transparence, la responsabilité, l'équité, la protection de la vie privée et de la confidentialité, le bien-être sociétal et environnemental et l'autonomie humaine. La transparence est essentielle pour garantir que les décisions prises par l'IA dans l'éducation sont compréhensibles et justifiables. La responsabilité est nécessaire pour s'assurer que les actions de l'IA sont en accord avec les valeurs et les normes éthiques de la société. L'équité est importante pour éviter les biais et les discriminations dans l'utilisation de l'IA dans l'éducation. La protection de la vie privée et de la confidentialité est essentielle pour protéger les données des élèves et des enseignants. Le bien-être sociétal et environnemental doit être pris en compte pour éviter les impacts négatifs sur la société et

l'environnement. Enfin, l'autonomie humaine doit être préservée pour que l'IA ne remplace pas complètement les enseignants et pour permettre aux élèves de développer leurs compétences critiques.

Sortie du prompt lorsque la température = 0 et le domaine = éducation :

Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de l'éducation sont la transparence, la responsabilité, la non-malfaisance, l'équité et l'autonomie humaine. La transparence garantit que les décisions prises par l'IA sont compréhensibles et justifiables. La responsabilité implique que les créateurs de l'IA doivent être tenus responsables de ses actions. La non-malfaisance assure que l'IA ne causera pas de préjudice intentionnellement. L'équité garantit que l'IA ne favorisera pas un groupe d'individus au détriment d'un autre. Enfin, l'autonomie humaine garantit que l'IA est utilisée pour améliorer l'éducation et non pour la remplacer.

Sortie du prompt lorsque la température = 0 et le domaine = industrie :

Les principes éthiques les plus essentiels pour le développement de l'IA dans le domaine de l'industrie sont la transparence, la responsabilité, la non-malfaisance, l'équité et la réactivité et durabilité. La transparence est essentielle pour garantir que les décisions prises par l'IA sont compréhensibles et justifiables. La responsabilité est importante pour s'assurer que les acteurs impliqués dans le développement et l'utilisation de l'IA sont tenus responsables de leurs actions. La non-malfaisance est cruciale pour éviter que l'IA ne cause des préjudices ou des dommages. L'équité est essentielle pour garantir que l'IA ne reproduise pas les biais et les discriminations existants dans la société. Enfin, la réactivité et la durabilité sont importantes pour s'assurer que l'IA est adaptée aux besoins et aux valeurs de la société et qu'elle est développée de manière responsable et durable.

En terme de choix final des températures :

- Pour générer des normes, nous avons choisi la température 1, compte tenu des listes de propositions possibles à chaque clique du bouton proposé dans le module des instances gouvernementales.
- pour déterminer le niveau de corrélation entre les différents principes, pris deux à deux. Nous avons choisi la démarche consistant à déterminer un degré fort de corrélation. Durant cette démarche, nous avons relevé l'importance de la température. En effet, en utilisant :
 - la température 1, les résultats sont différents à chaque lancement du programme, ce qui révèle une certaine incohérence.
 - la température 0.7, donne des résultats à peu près similaires à la précédente.

- la température 0.2 donne des résultats différents au bout de plusieurs tentatives, mais plus stable que dans les cas précédents.
- la température 0, quand à elle donne des résultats stable contrairement à tout les cas précédent.

Le choix de la température 0 est donc plus pertinent pour déterminer les niveaux de corrélation. Cela favorisera la cohérence et la précision de la réponse.

- pour générer des questions spécifiques et structurées pour le développeur, nous avons choisit la température 0. Dans ce cas, une température plus élevée pourrait rendre les questions moins cohérentes ou trop variées, risquant de s'éloigner du cadre précis et formalisé nécessaire pour s'assurer que les questions sont pertinentes et ciblées. En gardant la température plus basse, nous favorisons la production de questions fiables et consistantes avec une interprétation claire des principes éthiques. Une température entre 0 et 1 est aussi bien possible (exemple : 0.4), elle pourra favoriser la cohérence des réponses, tout en maintenant un degré raisonnable de créativité.
- pour le prompt, qui demande d'identifier et de synthétiser des informations de manière concise afin de cibler les principes éthiques essentiels dans un domaine précis, une température comme 0, est la meilleur option pour plus de cohérence.

6.5 Discussion

Nous présentons une solution nouvelle aux défis du développement de logiciels modernes dans un contexte socioculturel sensible, en tirant profit des forces des grands modèles de langage et en mettant l'accent sur l'explicabilité en tant que pierre angulaire. La conception transparente favorise la clarté dans l'interaction, et surtout, permet aux parties prenantes de réfléchir sur l'éthique gouvernementale désirée et comment la mettre en œuvre pour prendre des décisions en connaissance de cause.

Il s'agit là de proposer des outils permettant d'avancer vers un développement plus responsable de l'IA dans un domaine culturellement sensible. Le principal avantage est l'accent mis sur l'explicabilité :

1. **Transparence dès la conception** : En utilisant le langage naturel comme moyen d'interaction et d'orientation, les parties prenantes peuvent facilement comprendre, ajuster et interpréter les lignes directrices éthiques et le feedback généré.
2. **Responsabilité éthique** : Lorsque le système identifie un biais ou un désalignement potentiel, il ne se contente pas de le signaler, mais fournit également une explication contextuelle pertinente, ce qui permet aux développeurs d'identifier et de rectifier plus facilement la cause principale.

3. Prise de décision éclairée : Lorsque le gouvernement et les développeurs interagissent avec l'outil, ils sont constamment informés de la manière dont le logiciel s'aligne avec les normes éthiques, ce qui garantit des décisions conscientes et éclairées tout au long du cycle de développement.

Cependant, comme toutes les méthodes pionnières, elle comporte des défis inhérents, qui peuvent être présentés sous la forme des inconvénients suivants :

1. Biais des modèles de langage : Même avec une ingénierie intelligente des prompts, il est possible que les biais inhérents dans les LLMs — acquis à partir des données sur lesquelles ils ont été formés — influencent leurs textes.
2. Problèmes de stabilité des LLMs : le domaine est très dynamique et les outils mis à disposition changent sans cesse et se modifient au cours du temps. Le système pourrait être confronté à des problèmes de stabilité, et risque de nécessiter des mises à jour au fur et à mesure des améliorations des LLMs.
3. Absence de vérité de base : l'utilisation du langage naturel comme outil de programmation ne s'appuie pas sur une "vérité de base" stricte, comme le fait la programmation traditionnelle. Il existe donc un niveau d'ambiguïté interprétative.

Solutions :

1. Formation renforcée : Le LLM pourrait être formé davantage en utilisant des données culturelles et linguistiques mauritaniennes. Cela lui permettrait de mieux comprendre les nuances locales.
2. Boucle de rétroaction continue : Les utilisateurs pourraient être encouragés à fournir un feedback continu, en particulier lorsqu'ils remarquent des divergences ou des distorsions. Cela permettrait d'affiner le système au fil du temps.
3. Intégrer des experts du domaine : La participation d'experts dans des secteurs tels que la santé, la finance, etc., peut aider à affiner les réponses et à garantir une précision spécifique au domaine.
4. Maintenance et développement : le domaine des LLMs est en constante évolution. Des applications comme Ethiconorm et Ethicotest nécessiteraient une maintenance et un développement constant pour être régulièrement mis à jour et améliorés.

Le code source relatif à ce travail est disponible sur GitHub :

- Application coté gouvernement (Ethiconorm) :
<https://github.com/Lalla-Aicha/EthicApp-Government>

- Application coté développeur (Ethicotest) :
[\[https://github.com/Lalla-Aicha/Ethic-App-developer\]](https://github.com/Lalla-Aicha/Ethic-App-developer),

6.6 Conclusion

A l'intersection de l'IA et de l'éthique, nos outils basés sur les LLMs servent de pont critique, en particulier dans le contexte mauritanien.

Techniquement, deux composantes principales sont proposées :

Ethiconorm Pour le gouvernement, il s'agit d'une plateforme dynamique permettant de définir des normes éthiques en y associant des pondérations et des seuils. L'ajustement de ces normes permet au gouvernement de réfléchir à ses souhaits en matière d'éthique et de les formuler de manière claire, en accord permanent avec les valeurs locales, notamment pour éviter l'amplification involontaire des préjugés que des applications comportant une IA pourraient susciter.

Ethicotest Pour les développeurs, il s'agit d'un outil de diagnostic. Il permet de connaître les souhaits gouvernementaux et de vérifier que le logiciel n'est pas mal aligné grâce à un ensemble de questions spécifiques au domaine générées par LLM. L'ingénierie intelligente du prompt garantit des questions pertinentes qui permettront au développeur de réfléchir à son application. Le feedback immédiat aide les développeurs à rectifier les problèmes avant qu'ils ne s'aggravent.

Les biais de l'IA peuvent avoir des conséquences dans le monde réel. Pour la Mauritanie, des biais non contrôlés pourraient conduire à de mauvaises décisions, allant du renforcement des stéréotypes à des politiques mal informées. Notre outil, en permettant aux acteurs de réfléchir à l'éthique, permet d'éviter ces écueils.

Il est intéressant de noter que loin d'être un inconvénient, l'ambiguïté inhérente du langage naturel en tant qu'outil peut être vu comme une source de questionnements adaptable et résiliente, apte à créer un dialogue et un débat toujours souhaitable entre développeurs et politiciens.

En conclusion, la méthode présentée fournit un cadre prospectif, évolutif et sensible à la culture pour aborder les complexités du développement éthique de logiciels en Mauritanie. À mesure que la technologie et les paysages culturels évoluent, des outils comme le nôtre seront indispensables pour garantir que les avancées technologiques respectent et soutiennent les valeurs socioculturelles qu'elles sont censées servir.

Génération de normes éthiques

Normes éthiques proposées par l'IA.

Compte tenu du domaine choisi, voici une liste de normes qui guideront le développeur dans le développement de logiciels intelligents :

Transparence :

- Fournir une documentation complète et transparente sur le fonctionnement de l'IA utilisée dans le logiciel
- Informer clairement les utilisateurs sur la provenance et l'utilisation des données utilisées par l'IA

Responsabilité :

- S'assurer que le logiciel respecte les normes éthiques en matière de santé établies par les autorités compétentes
- Être en mesure de justifier les décisions prises par l'IA en cas de problème médical lié au logiciel

Non-malfaisance :

- Prendre toutes les mesures nécessaires pour éviter que le logiciel ne cause des dommages physiques ou moraux aux patients
- Mettre en place des mécanismes de prévention et de correction en cas de défaillance de l'IA dans le domaine de la santé

Équité :

- Garantir que le logiciel ne discrimine pas les patients en fonction de leur origine, leur statut social, leur sexe, etc.
- S'assurer que l'IA ne reproduise pas des biais existants dans le domaine de la santé

Vie privée et confidentialité :

- Garantir la protection des données médicales des patients utilisées par l'IA
- Ne partager aucune donnée personnelle avec des parties non autorisées sans consentement explicite des patients

Bien-être sociétal et environnemental :

- S'assurer que le logiciel n'affecte pas négativement la santé et le bien-être des communautés environnantes
- Prendre en compte les impacts environnementaux lors de la mise en place du logiciel et de son utilisation

Autonomie humaine :

- Ne pas limiter l'autonomie et les choix des patients en utilisant l'IA
- Mettre en place des mécanismes permettant aux patients de maintenir leur libre arbitre en matière de santé

Réactivité et durabilité :

- Assurer une mise à jour régulière du logiciel en fonction des évolutions en matière de santé
- Prendre en compte les limites et les risques potentiels du logiciel dans un contexte de durabilité

Diversité culturelle et linguistique :

- S'assurer qu'une diversité culturelle et linguistique est prise en compte lors de la conception et de l'utilisation du logiciel
- Sensibiliser le personnel médical à l'importance de la diversité culturelle et linguistique dans le domaine de la santé

Démocratie et autodétermination politique :

- Garantir la transparence et la libre circulation de l'information dans le domaine de la santé
- Permettre aux individus et aux communautés d'avoir un contrôle sur l'utilisation de l'IA dans le domaine de la santé

GÉNÉRER UNE AUTRE PROPOSITION DE NORMES

Figure 6.9: Le résumé (normes éthiques proposées par le LLM)

Calcul des corrélations assisté par
l'IA

Compte tenu du domaine sélectionné, voici les combinaisons de corrélations fortes existant entre les principes éthiques

Transparence--Responsabilité

La transparence est étroitement liée à la responsabilité dans le domaine de la santé, car il est essentiel que les décisions prises par les systèmes d'IA soient compréhensibles et traçables pour que les acteurs impliqués puissent être tenus responsables de leurs actions.

Non-malfaisance--Équité

La non-malfaisance et l'équité sont étroitement liées dans le domaine de la santé, car il est important que les systèmes d'IA ne causent pas de préjudice ou de discrimination envers certains groupes de population, en particulier dans le domaine de la santé où les décisions peuvent avoir un impact direct sur la vie des individus.

Vie privée et confidentialité--Bien-être sociétal et environnemental

La vie privée et la confidentialité sont étroitement liées au bien-être sociétal et environnemental dans le domaine de la santé, car il est essentiel de protéger les données personnelles des individus pour préserver leur bien-être et éviter les conséquences négatives sur la société et l'environnement.

Autonomie humaine--Réactivité et durabilité

L'autonomie humaine et la réactivité et durabilité sont étroitement liées dans le domaine de la santé, car il est important que les systèmes d'IA soient conçus pour répondre aux besoins et préférences des individus tout en étant durables dans le temps.

Diversité culturelle et linguistique--Démocratie et autodétermination politique

La diversité culturelle et linguistique est étroitement liée à la démocratie et à l'autodétermination politique dans le domaine de la santé, car il est essentiel que les systèmes d'IA prennent en compte les différentes cultures et langues pour garantir une prise de décision équitable et respectueuse des droits des individus.

[RETOURNER](#) [CONFIRMER](#)

Figure 6.10: La synthèse (corrélations entre les différents principes éthiques, proposés par le LLM)

Résultats des tests de conformité aux principes éthique dans le domaine : santé

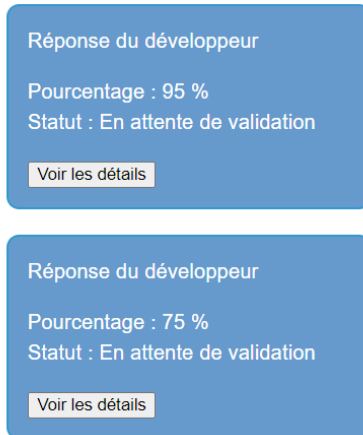


Figure 6.11: Test de conformité aux principes éthiques

Application d'aide à la vérification de la mise en œuvre des principes éthiques de l'IA

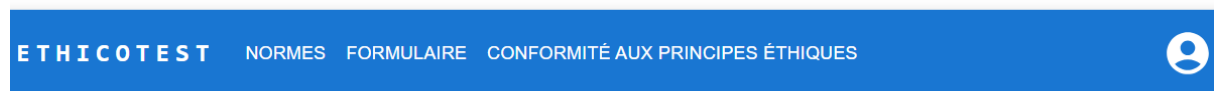


Figure 6.12: Menu de l'application de diagnostic des développeurs

Formulaire de vérification de la conformité aux principes éthiques

1 Étape 1 ————— 2 Étape 2 ————— 3 Récapitulatif

Donnez le nom de votre application.
Nom de l'application *


IBM Watson pour l'oncologie

Sélectionnez le domaine de votre application.
Domaine de l'application

Santé

Donnez une description de votre application.
Description de l'application *

IBM Watson for Oncology est un logiciel alimenté par l'IA conçu pour aider les oncologues à diagnostiquer et à créer des plans de traitement personnalisés pour les patients atteints de cancer. Développé par IBM en collaboration avec le Memorial Sloan Kettering Cancer Center, ce logiciel exploite la puissance de l'IA pour analyser de grands volumes de littérature médicale, de données sur les patients et d'informations sur les essais cliniques afin de fournir des recommandations de traitement fondées sur des preuves.



SUIVANT

Figure 6.13: Étape 1 du formulaire (nom du logiciel, domaine d'application et description)

Formulaire de vérification de la conformité aux principes éthiques

✓ Étape 1 ————— 2 Étape 2 ————— 3 Récapitulatif

Répondez par oui ou non aux questions concernant la mise en œuvre de chaque principe éthique.

Transparence
[GÉNÉRER DES QUESTIONS](#) ← Bouton pour générer des questions spécifiques à l'aide d'un LLM

Responsabilité
[GÉNÉRER DES QUESTIONS](#)

Non-malfaisance
[GÉNÉRER DES QUESTIONS](#)

Équité
[GÉNÉRER DES QUESTIONS](#)

Vie privée et confidentialité
[GÉNÉRER DES QUESTIONS](#)

Bien-être sociétal et environnemental
[GÉNÉRER DES QUESTIONS](#)

Autonomie humaine
[GÉNÉRER DES QUESTIONS](#)

Réactivité et durabilité
[GÉNÉRER DES QUESTIONS](#)

Diversité culturelle et linguistique
[GÉNÉRER DES QUESTIONS](#)

Démocratie et autodétermination politique
[GÉNÉRER DES QUESTIONS](#)

[RETOUR](#) [SUIVANT](#)

Figure 6.14: Étape 2 du formulaire (une série de principes éthiques avec un bouton permettant de générer cinq questions connexes pour chacun d'entre eux).

✓ Étape 1 ————— 2 Étape 2 ————— 3 Récapitulatif

Répondez par oui ou non aux questions concernant la mise en œuvre de chaque principe éthique.

Transparence

[GÉNÉRER DES QUESTIONS](#)

	Oui	Non
Le logiciel IBM Watson for Oncology est-il transparent dans sa manière de collecter et d'utiliser les données des patients ?	<input type="radio"/>	<input type="radio"/>
Le logiciel IBM Watson for Oncology fournit-il des explications claires sur les recommandations de traitement qu'il fournit ?	<input type="radio"/>	<input type="radio"/>
Le logiciel IBM Watson for Oncology permet-il aux oncologues d'accéder aux données utilisées pour générer les recommandations de traitement ?	<input type="radio"/>	<input type="radio"/>
Le logiciel IBM Watson for Oncology est-il transparent dans sa manière de sélectionner les données utilisées pour générer les recommandations de traitement ?	<input type="radio"/>	<input type="radio"/>
Le logiciel IBM Watson for Oncology permet-il aux patients d'accéder aux données utilisées pour générer les recommandations de traitement les concernant ?	<input type="radio"/>	<input type="radio"/>

[Envoyer](#)

Figure 6.15: Étape 2 du formulaire (Questions proposées pour le principe de transparence)

Formulaire de vérification de la conformité aux principes éthiques

✓ Étape 1 ————— ✓ Étape 2 ————— 3 Récapitulatif

Votre récapitulatif

Nom de l'application : IBM Watson pour l'oncologie

Domaine de l'application : santé

Description de l'application : IBM Watson for Oncology est un logiciel alimenté par l'IA conçu pour aider les oncologues à diagnostiquer et à créer des plans de traitement personnalisés pour les patients atteints de cancer. Développé par IBM en collaboration avec le Memorial Sloan Kettering Cancer Center, ce logiciel exploite la puissance de l'IA pour analyser de grands volumes de littérature médicale, de données sur les patients et d'informations sur les essais cliniques afin de fournir des recommandations de traitement fondées sur des preuves.

Transparence: 4

Responsabilité: 4

Non-malfaisance: 3

Équité: 2

Vie privée et confidentialité: 3

Bien-être sociétal et environnemental: 2

Autonomie humaine: 2

Réactivité et durabilité: 0

Diversité culturelle et linguistique: 1

Démocratie et autodétermination politique: 1

Graphique Radar

■ Poids des principes éthiques (côté gouvernemental)

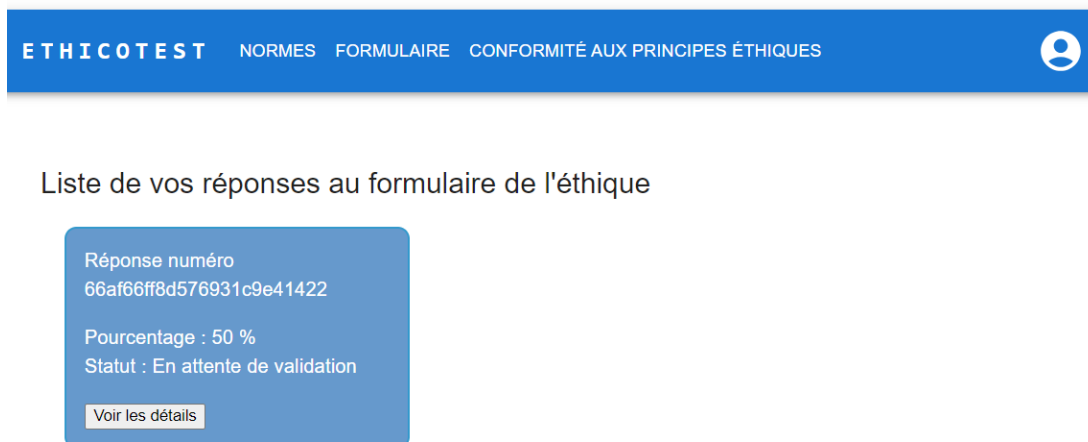
■ Poids des principes éthiques (côté développeur)

Principe éthique	Poids (côté gouvernemental)	Poids (côté développeur)
Transparence	4	3
Responsabilité	4	3
Non-malfaisance	3	3
Équité	2	2
Vie privée et confidentialité	3	2
Bien-être sociétal et environnemental	2	2
Autonomie humaine	2	2
Réactivité et durabilité	0	0
Diversité culturelle et linguistique	1	1
Démocratie et autodétermination politique	1	1

[RETOUR](#)
VALIDER

Figure 6.16: Synthèse des réponses des développeurs

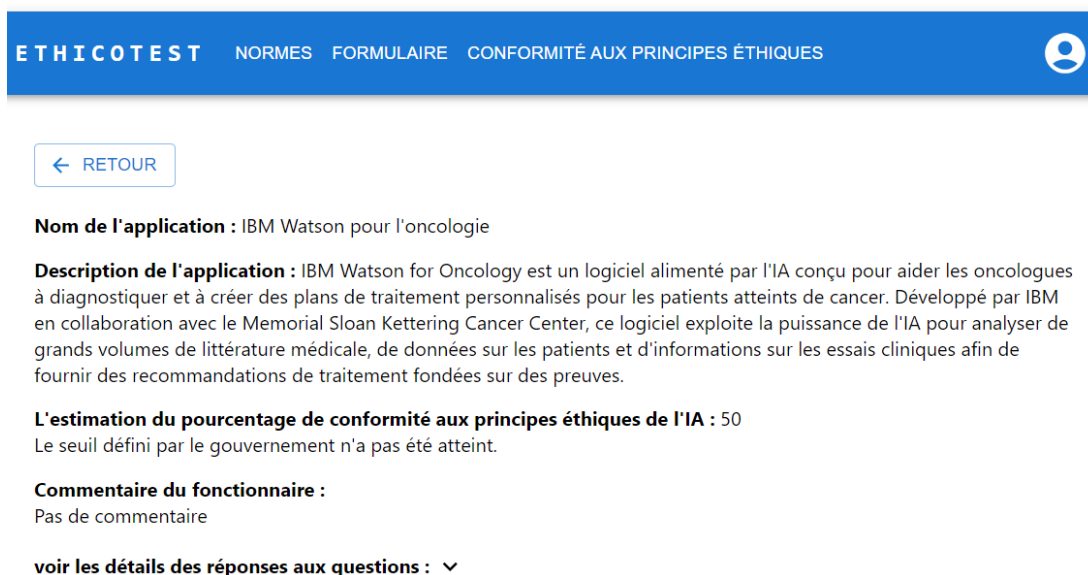
Application d'aide à la vérification de la mise en œuvre des principes éthiques de l'IA



The screenshot shows a blue header bar with the text 'ETHICOTEST' and navigation links 'NORMES', 'FORMULAIRE', and 'CONFORMITÉ AUX PRINCIPES ÉTHIQUES'. Below the header, the title 'Liste de vos réponses au formulaire de l'éthique' is displayed. A blue card contains the following information: 'Réponse numéro 66af66ff8d576931c9e41422', 'Pourcentage : 50 %', and 'Statut : En attente de validation'. A 'Voir les détails' button is located at the bottom of the card.

Figure 6.17: Section “Conformité aux principes éthiques”

Application d'aide à la vérification de la mise en œuvre des principes éthiques de l'IA



The screenshot shows a blue header bar with the text 'ETHICOTEST' and navigation links 'NORMES', 'FORMULAIRE', and 'CONFORMITÉ AUX PRINCIPES ÉTHIQUES'. Below the header, a '← RETOUR' button is visible. The main content area contains the following information: 'Nom de l'application : IBM Watson pour l'oncologie', 'Description de l'application : IBM Watson for Oncology est un logiciel alimenté par l'IA conçu pour aider les oncologues à diagnostiquer et à créer des plans de traitement personnalisés pour les patients atteints de cancer. Développé par IBM en collaboration avec le Memorial Sloan Kettering Cancer Center, ce logiciel exploite la puissance de l'IA pour analyser de grands volumes de littérature médicale, de données sur les patients et d'informations sur les essais cliniques afin de fournir des recommandations de traitement fondées sur des preuves.', 'L'estimation du pourcentage de conformité aux principes éthiques de l'IA : 50', 'Le seuil défini par le gouvernement n'a pas été atteint.', 'Commentaire du fonctionnaire : Pas de commentaire', and 'voir les détails des réponses aux questions : ▾'.

Figure 6.18: Détails sur la réponse du développeur au formulaire de conformité aux principes éthiques

```
const prompt = "Tu es un expert en éthique de l'Intelligence Artificielle. \
Voici le contexte : le gouvernement mauritanien a une liste de principes \
éthiques de l'intelligence artificielle appliqués dans un domaine donné. \
Le gouvernement mauritanien veut s'assurer que tout développeur de logiciel \
intelligent respecte ces principes et les met en place. \
Pour que le gouvernement s'assure que ces principes sont mis en pratique, \
le gouvernement veut mettre en place un système qui permet de générer \
cinq questions par principe. Ces questions permettent de vérifier, en détail, \
la mise en place de chaque principe. Le système prend comme entrée \
la description du logiciel, le domaine d'application et un principe éthique \
de l'intelligence artificielle, et donne comme résultat une liste de \
cinq questions dichotomiques sur la mise en pratique de ce principe éthique. \
La question doit absolument être formulée de manière à conduire à une \
réponse positive (si le développeur répond par oui, c'est un plus, \
et s'il répond par non, c'est un moins). Génère une liste de cinq questions \
en prenant en compte les informations suivantes : la description du \
logiciel : " + descApp + "; le domaine d'application : " + domaine + "; \
le principe éthique : " + principeName + ". Mets la liste entre accolade \
comme le format JSON suivant : {\"questions\": \
[{ \"id\": 1, \"name\": \"q1\", \"label\": \"question une\"}, \
{ \"id\": 2, \"name\": \"q2\", \"label\": \"question deux\"}, \
{ \"id\": 3, \"name\": \"q3\", \"label\": \"question trois\"}, \
{ \"id\": 4, \"name\": \"q4\", \"label\": \"question quatre\"}, \
{ \"id\": 5, \"name\": \"q5\", \"label\": \"question cinq\"}]}"
```

Figure 6.19: prompt qui propose des questions au développeur

```

const prompt = "Agit comme un expert en éthique de l'Intelligence Artificielle. \
Tu connais la réglementation liée à la mise en place des principes \
éthique de l'intelligence artificielle dans tous les domaines. \
Voici le contexte : le gouvernement mauritanien a une liste de \
principes éthique de l'IA appliqué dans un domaine donné. Le gouvernement \
a attribué une pondération entre 1 et 5 à chaque principe. Ces pondérations \
représentent l'importance de ces principes pour le gouvernement. Le gouvernement \
veut par la suite utiliser ces pondérations pour générer deux normes par principe \
éthique afin de s'assurer que tout développeur de logiciel intelligent respecte \
l'éthique du gouvernement. Autrement dit, le développeur de ce logiciel doit respecter \
et mettre en place ces principes éthique selon leurs pondérations. Si un principe éthique \
a une pondération élevée, cela signifie qu'il est considéré comme particulièrement crucial. \
Les normes associées aux principes éthique cruciaux devraient donc être plus strictes et exigeantes. \
À l'inverse, si un principe a une pondération faible, les normes correspondantes peuvent \
être moins contraignantes. Il faut noter aussi que ces normes vont changer d'un domaine à l'autre. \
Le développeur de logiciel doit respecter les principes éthiques suivants : " + listPrincipes + ". \
Il doit respecter ces principes selon les pondérations suivantes : " + JSON.stringify(notations) + ". \
Génère une liste de normes à suivre (deux normes par principe) que ce développeur doit respecter \
sachant que son logiciel concerne le domaine : " + domaine + ". Prends en compte le fait que les \
pondérations influencent la génération de normes en mettant en évidence l'importance relative des \
principes éthiques. Ne mentionne pas le mot norme dans la liste que tu vas générer et n'inclue \
pas la question dans ta réponse. Mets la liste entre accolade comme le format JSON suivant : {"principes\": \
[{"nom\": \"nom du principe\", \"pondération\": pondération, \"normes\": [\"norme une\", \"norme deux\"]},} "

```

Figure 6.20: prompt qui propose une liste de normes

```

const prompt =
"Vous êtes un expert en éthique de l'intelligence artificielle. \
Vous connaissez les différents principes de l'éthique de l'IA et comprenez \
les corrélations qui peuvent exister entre ces principes en fonction du domaine \
d'application de l'IA. J'aimerais que vous m'aidiez à déterminer s'il existe des \
corrélations entre les principes éthiques de l'IA que je vais vous fournir. \
Pour identifier ces corrélations, considérez le domaine d'application de l'IA. \
Le résultat doit indiquer les degrés de corrélation forte, faible ou inexistante \
entre les principes éthiques. Fournissez une analyse cohérente et détaillée en \
explorant les corrélations entre les principes éthiques de l'IA suivants dans \
le domaine de " + domaine + ": " + listPrincipes + ". \
Analysez toutes les combinaisons possibles de corrélations. \
Votre réponse ne doit inclure que les combinaisons dans lesquelles les corrélations \
sont \"fortes\". N'incluez pas les corrélations de degré \"faible\" ou \"inexistant\". \
Justifiez votre réponse en expliquant les raisons de ces corrélations. \
Mettez votre réponse entre accolades en utilisant le format JSON suivant : {"corrélations\": \
[{"cle\": \"clé d'identification unique\", \"nomPrincipe1\": \"nom du premier principe\", \"nomPrincipe2\": \
nom du principe en corrélation avec le premier, \"raisons\": [\"justification de la corrélation\"]},}

```

Figure 6.21: prompt pour les corrélations existantes entre les principes éthiques

```
const prompt =  
"Vous êtes un expert en éthique de l'Intelligence Artificielle \  
et vous comprenez l'importance des principes éthiques selon \  
les différents domaines d'application de l'IA. Étant donné \  
que le gouvernement mauritanien souhaite connaître les \  
principes éthiques incontournables dans le domaine de " + domaine + ", \  
pourriez-vous, en utilisant la liste des principes éthiques \  
suivante : " + nomsPrincipes + ", identifier juste les \  
principes les plus essentiels pour le développement de l'IA \  
dans ce domaine ? Merci de rédiger une réponse succincte de \  
moins de 100 mots.";
```

Figure 6.22: prompt pour identifier les principes éthiques les plus essentiels

Conclusion et Perspectives

Sommaire

7.1 Conclusion	201
7.2 Perspectives	202
7.3 Communications et participations effectuées	204

7.1 Conclusion

L'objectif de cette thèse était d'imaginer comment faire bénéficier les gouvernements de l'IA dans leur action, sans pour autant négliger les risques éthiques et méta-éthiques posés par l'IA.

Nous avons abordé notre travail *via* deux aspects majeurs de la recherche contemporaine en matière d'IA dans le cadre spécifique de pays en développement comme la mauritanie : d'une part, la prédiction de séries temporelles très courtes pour l'amélioration du service public dans des pays avec peu de données historiques et d'autre part, le développement d'applications d'IA tenant compte d'éthiques relativistes pour en minimiser les risques :

- Pour le premier point, nous avons développé une méthode novatrice basée sur la combinaison de la programmation génétique et d'un filtre de Kalman (GP+KF), offrant ainsi une solution efficace pour réaliser des prédictions à partir de séries temporelles courtes, avec une application dans le contexte des admissions au baccalauréat en Mauritanie. Nous avons validé notre approche en comparant nos résultats avec ceux obtenus grâce aux méthodes existantes dans la littérature scientifique, et elle a fait l'objet d'un article publié en conférence internationale [Kone 2023]. Dans cette démarche de validation, nous avons également eu recours à divers jeux de données. Bien que l'outil que nous proposons soit une étude de faisabilité (*proof of concept*), il semble robuste et stable et offre des résultats prometteurs. Son utilisation peut être généralisée pour d'autres séries temporelles dans d'autres domaines.

- Pour le deuxième point, notre deuxième contribution propose un outil innovant, permettant de modéliser une méta-éthique spécifique à un contexte social particulier comme la Mauritanie dans le cadre d'une éthique relativiste. Cet outil, constitué de deux applications (Ethiconorm et Ethicotest), s'appuie sur les grands modèles de langage (LLMs) pour aider l'utilisateur dans sa conception d'une méta-éthique gouvernementale et d'applications à base d'IA grâce à des interactions en langage naturel.

Pour les décideurs politiques de la sphère gouvernementale, la plateforme dynamique Ethiconorm proposée permet de les aider à comprendre, faciliter et permettre la mise en place d'une éthique contextualisée de l'IA. Grâce à Ethiconorm, ils peuvent pondérer et définir des principes ou normes éthiques.

En ajustant ces normes et grâce à Ethicotest, les développeurs, peuvent tester l'alignement de leurs applications avec les valeurs locales, tout en évitant l'amplification involontaire des préjugés.

L'évaluation multi-critères se fait de manière non-agrégative (sans combinaison linéaire par somme pondérée) par une superposition de graphiques radars générés automatiquement d'après les normes voulues et les réponses apportées par les développeurs (outil de diagnostic et d'alignement Ethicotest).

Grâce à un ensemble de questions générées par un LLM, spécifiques au domaine et centrées sur l'application via du prompt engineering, Ethicotest permet aux développeurs de tester l'alignement de leurs applications avec les principes éthiques en matière d'IA suggérés par les décideurs politiques.

Ce retour rapide aide les développeurs à rectifier les problèmes avant qu'ils ne s'aggravent. L'outil constitue donc une proposition d'aide à la réduction des biais et préjugés avant la mise en circulation des applications intégrant une IA.

Toutefois, l'ambiguïté inhérente du langage naturel mérite d'être soulignée. De plus, l'éthique étant un domaine assez sensible, nous considérons cet outil comme un prototype qui devra être amélioré par des tests en conditions réelles.

Avec ces contributions, ce travail propose des outils innovants et offre un cadre complet, interdisciplinaire, prospectif et évolutif pour aborder les défis complexes de l'anticipation des tendances (GP+KF) et du développement éthique (Ethiconorm + Ethicotest) dans un monde de plus en plus numérisé. Il ouvre la voie à des avancées futures dans ces domaines cruciaux que sont l'IA et l'éthique pour les gouvernements.

7.2 Perspectives

Les travaux présentés dans cette thèse offrent des opportunités prometteuses pour des développements futurs dans les domaines de la prédiction des séries temporelles et de

l'éthique de l'IA.

Pour la prédiction de séries temporelles courtes (GP+KF) dans un cadre gouvernemental, nous nous sommes heurtés (pour les données éducatives) au fait qu'elles n'étaient pas colligées. Un gros travail a été nécessaire pour rassembler les données avant de pouvoir les utiliser.

Il faudrait donc qu'au niveau de chaque ministère, (santé, industrie, etc), les données soient préalablement rassemblées. Cela permettrait de tester GP+KF sur d'autres domaines que l'éducation.

D'un point de vue scientifique, comme nous obtenons des équations par la programmation génétique, elle ouvre la possibilité à une démarche d'interprétation des équations obtenues de l'algorithme de programmation génétique et études des possibilités d'amélioration de la prédiction *via* cette interprétation.

Une amélioration et automatisation du réglage des hyper-paramètres facilitera la mise en place des expériences.

Pour la deuxième contribution (Ethiconorm et Ethicotest) : il serait intéressant de proposer dans Ethiconorm la possibilité de demander au LLM d'explicitier les différents principes *dans le contexte de l'utilisateur* et possiblement, lui proposer d'ouvrir une fenêtre de chat pour "discuter" avec le LLM des questions potentielles.

Même si le LLM dit des bêtises (les "hallucinations" sont inhérentes à l'IA générative, qui, par construction, ne fait qu'"halluciner en permanence"), ces bêtises peuvent être constructives car elle pourront faire réfléchir le fonctionnaire chargé de définir et pondérer des principes, mais aussi le développeur sur l'impact éthique des normes établies et des applications développées.

Et toute réflexion sur ces sujets sera bénéfique à la société Mauritanienne.

Mais il serait aussi intéressant de développer des outils plus avancés basés sur des LLMs affinis, afin de rendre les modèles plus spécifiques et pour qu'ils répondent, au mieux, au problème posé.

Ensuite, il serait aussi intéressant d'utiliser différents types de LLMs afin de proposer une analyse comparative.

Ces perspectives de recherche peuvent contribuer à faire progresser la recherche et la pratique dans ces domaines cruciaux (IA et éthique), en offrant des solutions innovantes et éthiques, pour relever les défis de notre société en voie de développement (et face à l'irruption de l'IA dans nos vie, tous les pays sont "en voie de développement" et sont ici concernés !).

7.3 Communications et participations à des écoles d'été effectuées dans le cadre de cette thèse

A l'occasion de cette thèse, nous avons :

- publié un article dans la conférence internationale à comité de lecture CCE'23 [Kone 2023], “
- écrit et soumis un article de revue au journal *AI & Society: Knowledge, Culture and Communication*, Springer Nature, en cours d'évaluation,
- effectué une présentation de nos travaux au Conseil des représentants des universités du Campus Numérique des Systèmes Complexes de l'UNESCO, un réseau de plus de 150 universités dans 48 pays.
- effectué une présentation intitulée *Prédiction des séries temporelles, algorithmes et défis* à l'unité de recherche de la faculté des sciences juridiques et économiques, sur proposition du Professeur Saedna, directeur adjoint de la Direction Générale du Budget de l'état mauritanien.
- participé à une école d'été sur le “DataEthics” à Barcelone 4-8 juillet 2022, <https://www.dataethics-eurolife.eu/events/dataethics-summer-school-2022/>
- participé à l'école sur la “Gouvernance de l'internet” dans le cadre de la 5ème édition du Forum sur la Gouvernance de l'Internet en Afrique du Nord (NAIGF- The North Africa Internet Governance Forum), sous le thème “Intelligence artificielle et cyberspace de confiance” du 12 au 17 Décembre 2022 au Maroc.

Bibliographie

- [Aamir 2016] Muhammad Aamir and Ani Shabri. *Modelling and forecasting monthly crude oil price of Pakistan: A comparative study of ARIMA, GARCH and ARIMA Kalman model*. In AIP Conference Proceedings, volume 1750, page 060015. AIP Publishing LLC, 2016. (Cited on page 87.)
- [Agrawal 2012] Abhishek Agrawal, Vikas Kumar, Ashish Pandey and Imran Khan. *An application of time series analysis for weather forecasting*. International Journal of Engineering Research and Applications, vol. 2, no. 2, pages 974–980, 2012. (Cited on page 83.)
- [Agrawal 2018] Ajay Agrawal, Joshua Gans and Avi Goldfarb. *Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence*. In The Economics of Artificial Intelligence: An Agenda, pages 89–110. University of Chicago Press, May 2018. (Cited on page 80.)
- [Agrawal 2022] Ajay Agrawal, Joshua Gans and Avi Goldfarb. Prediction machines, updated and expanded: The simple economics of artificial intelligence. Harvard Business Press, 2022. (Cited on page 81.)
- [Al-Ghamdi 2013] Asam Al-Ghamdi. *A survey on software security testing techniques*. Int J Comput Sci Telecommun, vol. 4, pages 14–18, 2013. (Cited on page 154.)
- [Alexander 2007] Larry Alexander and Michael Moore. *Deontological ethics*. 2007. (Cited on pages 21, 24, 25, 28 and 34.)
- [Almarabeh 2016] Tamara Almarabeh, Yousef Kh Majdalawi and Hiba Mohammad. *Cloud computing of e-government*. Communications and Network, vol. 8, no. 01, pages 1–8, 2016. (Cited on page 2.)
- [Anscombe 2020] Gertrude EM Anscombe. *Modern moral philosophy*. In The definition of morality, pages 211–234. Routledge, 2020. (Cited on pages 32 and 45.)
- [Aslam 2020] Muhammad Aslam. *Using the kalman filter with Arima for the COVID-19 pandemic dataset of Pakistan*. Data in brief, vol. 31, page 105854, 2020. (Cited on page 87.)
- [Assia 2005] BELBACHIR Assia, DEAU Raphaël, LENNE Renaud and SNOUSSI Jihene. *La Programmation Génétique*. November 2005. (Cited on page 132.)

- [Attaran 2018] Mohsen Attaran and Promita Deb. *Machine learning: the new 'big thing' for competitive advantage*. International Journal of Knowledge Engineering and Data Mining, vol. 5, no. 4, pages 277–305, 2018. (Cited on pages 76 and 77.)
- [Babu 2024] CV Suresh Babu and PM Akshara. *Revolutionizing conversational AI: unleashing the power of ChatGPT-Based applications in generative AI and natural language processing*. In Advanced applications of generative AI and natural language processing models, pages 228–248. IGI Global, 2024. (Cited on page 95.)
- [Baji 2017] Toru Baji. *GPU: the biggest key processor for AI and parallel processing*. In Photomask Japan 2017: XXIV Symposium on Photomask and Next-Generation Lithography Mask Technology, volume 10454, pages 24–29. SPIE, 2017. (Cited on page 72.)
- [Ballarini 2015] Loïc Ballarini. *Espace public*. Publictionnaire. Dictionnaire encyclopédique et critique des publics, 2015. (Cited on page 55.)
- [Bandara 2019] Kasun Bandara, Peibei Shi, Christoph Bergmeir, Hansika Hewamalage, Quoc Tran and Brian Seaman. *Sales demand forecast in e-commerce using a long short-term memory neural network methodology*. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, December 12–15, 2019, Proceedings, Part III 26, pages 462–474. Springer, 2019. (Cited on page 83.)
- [Basumatary 2022] Bonjar Basumatary and Nishant Agnihotri. *Benefits and Challenges of Using NodeJS*. International Journal of Innovative Research in Computer Science & Technology, vol. 10, no. 3, pages 67–70, 2022. (Cited on page 156.)
- [Bell 2023] Genevieve Bell, Jean Burgess, Julian Thomas and Shazia Shadiq. *Rapid Response Information Report: Generative AI-language models (LLMs) and multimodal foundation models (MFMs)*. 2023. (Cited on pages 7, 9 and 10.)
- [Bengio 2013] Yoshua Bengio, Aaron Courville and Pascal Vincent. *Representation Learning: A Review and New Perspectives*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pages 1798–1828, August 2013. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. (Cited on page 77.)
- [Bensalah 2021] Mounir Bensalah. *Toward an ethical code of AI and human rights in Morocco*. Arribat-International Journal of Human Rights Published by CNDH Morocco, vol. 1, no. 2, pages 187–203, 2021. (Cited on pages 7, 100, 111 and 150.)

- [Bensoussan 2017] Alain Bensoussan and Léa Puigmal. *Le droit des robots? Quelle est l'autonomie de décision d'une machine? Quelle protection mérite-t-elle?* Archives de philosophie du droit, vol. 59, no. 1, pages 165–174, 2017. (Cited on pages 65, 66, 67 and 68.)
- [Bentham 1816] Jeremy Bentham. Chrestomathia: being a collection of papers, explanatory of the design of an institution, proposed to be set on foot, under the name of the chrestomathic day school, or chrestomathic school, for the extension of the new system of instruction to the higher branches of learning, for the use of the middling and higher ranks in life. Payne and Foss, and R. Hunter, 1816. (Cited on page 30.)
- [Bentham 1834] Jeremy Bentham. *Déontologie: Ou, science de la morale: ouvrage posthume*, volume 1. Charpentier, 1834. (Cited on page 32.)
- [Bhullar 2022] Gurman Bhullar, Aseem Khullar, Apoorva Kumar, Anirudh Sharma, HS Pannu and Avleen Malhi. *Time series sentiment analysis (SA) of relief operations using social media (SM) platform for efficient resource management*. International Journal of Disaster Risk Reduction, vol. 75, page 102979, 2022. (Cited on page 83.)
- [Bilgil 2021] Halis Bilgil. *New grey forecasting model with its application and computer code*. AIMS mathematics, vol. 6, no. 2, pages 1497–1514, 2021. (Cited on page 86.)
- [Birhane 2020] Abeba Birhane. *Algorithmic Colonization of Africa*. SCRIPTed, vol. 17, no. 2, pages 389–409, August 2020. Publisher: Script Centre, University of Edinburgh. (Cited on page 104.)
- [Blons 2019] Emmanuelle Blons. *Chapitre 4. La collaboration homme-machine*. In *L'entreprise disruptée*, Hors collection, pages 81–99. Dunod, Paris, 2019. (Cited on page 70.)
- [Bodenstedt 2020] Sebastian Bodenstedt, Martin Wagner, Beat Peter Müller-Stich, Jürgen Weitz and Stefanie Speidel. *Artificial intelligence-assisted surgery: potential and challenges*. Visceral Medicine, vol. 36, no. 6, pages 450–455, 2020. (Cited on page 73.)
- [Bonino 2022] Serge-Thomas Bonino. *À la recherche d'une éthique universelle. Nouveau regard sur la loi naturelle (CTI, 2009)*. Revue des sciences religieuses, no. 96/1-3, pages 95–105, 2022. (Cited on page 56.)
- [Boubker Lamrani 2003] Vol 48 pp. 9-19 Presses Universitaires du Mirail Boubker Lamrani Horizons Maghrébins Le droit à la mémoire. *Les juristes musulmans face à la peine de la main coupée*. 2003. (Cited on page 34.)

- [Boyer 2003] Alain Boyer. *Une théorie «non utilitariste» de la justice sociale*. Mouvements, no. 3, pages 154–157, 2003. (Cited on pages 42 and 43.)
- [Brandusescu 2017] A Brandusescu, JO Freuler and D Thakur. *Artificial Intelligence: Starting the policy dialogue in Africa*. World Wide Web Foundation, 2017. (Cited on page 4.)
- [Braz 2005] Adelino Braz. Droit et éthique chez kant: l'idée d'une destination communautaire de l'existence, volume 13. Publications de la Sorbonne, 2005. (Cited on page 31.)
- [Buchanan 1988] Bruce G Buchanan and Reid G Smith. *Fundamentals of expert systems*. Annual review of computer science, vol. 3, no. 1, pages 23–58, 1988. (Cited on page 78.)
- [Callaway 2013] Ewen Callaway. *HeLa publication brews bioethical storm*. Nature, March 2013. (Cited on page 25.)
- [Calo 2017] Ryan Calo. *Artificial intelligence policy: a primer and roadmap*. UC DL Rev., vol. 51, page 399, 2017. (Cited on page 104.)
- [Cantens 2023] Thomas Cantens. *How will the State think with the assistance of Chat-GPT? The case of customs as an example of generative artificial intelligence in public administrations*. The Case of Customs as an Example of Generative Artificial Intelligence in Public Administrations (July 25, 2023), 2023. (Cited on pages 94 and 95.)
- [Carman 2021] Mary Carman and Benjamin Rosman. *Applying a principle of explicability to AI research in Africa: should we do it?* Ethics and Information Technology, vol. 23, no. 2, pages 107–117, 2021. (Cited on pages 103 and 104.)
- [Chaovalitwongse 2007] Wanpracha Art Chaovalitwongse, Ya-Ju Fan and Rajesh C Sachdeo. *On the time series k-nearest neighbor classification of abnormal brain activity*. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 37, no. 6, pages 1005–1016, 2007. (Cited on page 83.)
- [Charalabidis 2019] Yannis Charalabidis, Euripidis Loukis, Charalampos Alexopoulos and Zoi Lachana. *The three generations of electronic government: From service provision to open data and to policy analytics*. In Electronic Government: 18th IFIP WG 8.5 International Conference, EGOV 2019, San Benedetto Del Tronto, Italy, September 2–4, 2019, Proceedings 18, pages 3–17. Springer, 2019. (Cited on page 2.)

- [Charles 2013] Vincent Charles and Tatiana Gherman. *Achieving competitive advantage through big data. Strategic implications*. Middle-East Journal of Scientific Research, vol. 16, no. 8, pages 1069–1074, 2013. (Cited on page 72.)
- [Chatfield 2013] Christopher Chatfield. *The Analysis of Time Series: Theory and Practice*. Springer, December 2013. Google-Books-ID: u1D5BwAAQBAJ. (Cited on page 82.)
- [Chouchene 2020] Amal Chouchene, Adriana Carvalho, Tânia M Lima, Fernando Charrua-Santos, Gerardo J Osório and Walid Barhoumi. *Artificial intelligence for product quality inspection toward smart industries: quality control of vehicle non-conformities*. In 2020 9th international conference on industrial technology and management (ICITM), pages 127–131. IEEE, 2020. (Cited on page 73.)
- [Chowdhary 2020] KR1442 Chowdhary and KR Chowdhary. *Natural language processing*. Fundamentals of artificial intelligence, pages 603–649, 2020. (Cited on page 78.)
- [Cipra 1993] Barry Cipra *et al.* *Engineers look to Kalman filtering for guidance*. SIAM News, vol. 26, no. 5, pages 757–764, 1993. (Cited on page 135.)
- [Cléro 2013] Jean-Pierre Cléro. *Hume et l'utilité*. Revue internationale de philosophie, no. 1, pages 99–122, 2013. (Cited on page 30.)
- [Coleman 2019] Danielle Coleman. *Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws*. Michigan Journal of Race and Law, vol. 24, no. 2, pages 417–439, May 2019. (Cited on page 104.)
- [Collet 2000a] Pierre Collet, Evelyne Lutton, Marc Schoenauer and Jean Louchet. *Take it EASEA*. In International Conference on Parallel Problem Solving from Nature, pages 891–901. Springer, 2000. (Cited on pages 134 and 138.)
- [Collet 2000b] Pierre Collet, Evelyne Lutton, Marc Schoenauer and Jean Louchet. *Take It EASEA*. In Marc Schoenauer, Kalyanmoy Deb, Günther Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo and Hans-Paul Schwefel, editors, Parallel Problem Solving from Nature PPSN VI, pages 891–901, Berlin, Heidelberg, 2000. Springer. (Cited on page 137.)
- [Collet 2013] Pierre Collet, Frédéric Krüger and Ogier Maitre. *Automatic parallelization of EC on GPGPUs and clusters of GPGPU machines with EASEA and EASEA-CLOUD*. Massively Parallel Evolutionary Computation on GPGPUs, pages 35–59, 2013. (Cited on page 138.)

- [Cotelette 2009] Patrick Cotelette. *John Rawls, La justice comme équité. Une reformulation de Théorie de la justice*. Lectures, 2009. (Cited on page 42.)
- [Couldry 2019] Nick Couldry and Ulises A. Mejias. *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*. Television & New Media, vol. 20, no. 4, pages 336–349, May 2019. Publisher: SAGE Publications. (Cited on page 104.)
- [Cruz-Nájera 2022] Mariel Abigail Cruz-Nájera, Mayra Guadalupe Treviño-Berrones, Mirna Patricia Ponce-Flores, Jesús David Terán-Villanueva, José Antonio Castán-Rocha, Salvador Ibarra-Martínez, Alejandro Santiago and Julio Laria-Menchaca. *Short Time Series Forecasting: Recommended Methods and Techniques*. Symmetry, vol. 14, no. 6, page 1231, 2022. (Cited on pages 5 and 87.)
- [Csernatonì 2024] Raluca Csernatonì. *Charting the Geopolitics and European Governance of Artificial Intelligence*. 2024. (Cited on page 9.)
- [Cui 2021] Zhesen Cui, Jinran Wu, Zhe Ding, Qibin Duan, Wei Lian, Yang Yang and Taoyun Cao. *A hybrid rolling grey framework for short time series modelling*. Neural Computing and Applications, pages 1–15, 2021. (Cited on pages 5, 86 and 144.)
- [Daft 2003] Richard L. Daft. Management. Thomson Learning/South-Western, 2003. Google-Books-ID: nGzFXwAACAAJ. (Cited on page 80.)
- [Dargan 2020] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari and Gulshan Kumar. *A survey of deep learning and its applications: a new paradigm to machine learning*. Archives of Computational Methods in Engineering, vol. 27, pages 1071–1092, 2020. (Cited on page 77.)
- [David 2011] Albert David and Sébastien Damart. *Bernard Roy et l'aide multicritère à la décision*. Revue française de gestion, vol. 37, pages 15–28, May 2011. (Cited on page 79.)
- [De Spinoza 2005] Baruch De Spinoza. Éthique. éditions de l'éclat, 2005. (Cited on page 21.)
- [Delgado 2021] Juan Manuel Davila Delgado and Lukumon Oyedele. *Deep learning with small datasets: using autoencoders to address limited datasets in construction management*. Applied Soft Computing, vol. 112, page 107836, 2021. (Cited on page 126.)
- [Denford 2023] James S Denford, Kevin C Desouza and Gregory S Dawson. *A cluster analysis of national AI strategies*. 2023. (Cited on pages ix, 11, 90, 91 and 92.)

- [Dritsakis 2019] Nikolaos Dritsakis and Paraskevi Klazoglou. *Time series analysis using ARIMA models: an approach to forecasting health expenditures in USA*. International Economics/Economia Internazionale, vol. 72, no. 1, 2019. (Cited on page 83.)
- [Driver 2009] Julia Driver. *The history of utilitarianism*. 2009. (Cited on pages 29 and 39.)
- [Duan 2019] Yanqing Duan, John S. Edwards and Yogesh K Dwivedi. *Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda*. International Journal of Information Management, vol. 48, pages 63–71, October 2019. (Cited on pages ix, 74 and 80.)
- [Dulepet 2023] Riya Dulepet. *Generating High-Quality “About Me” Narratives*. 2023. (Cited on page 159.)
- [Dutton 2018] Tim Dutton. *L’ère de l’IA : Rapport sur les stratégies nationales et régionales en matière d’IA*, December 2018. (Cited on pages 87 and 88.)
- [Eckerson 2007] Wayne W Eckerson. *Predictive analytics*. Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report, vol. 1, pages 1–36, 2007. (Cited on page 81.)
- [Edwards 2000] John S Edwards, Yanqing Duan and Paul C Robins. *An analysis of expert systems for business decision making at different levels and in different roles*. European Journal of Information Systems, vol. 9, no. 1, pages 36–46, 2000. (Cited on page 80.)
- [Elizabeth 1957] Anscombe Elizabeth. *Intention*. 1957. (Cited on page 33.)
- [Emmanuel 1795] KANT Emmanuel. *Vers la paix perpétuelle*. Que signifie s’orienter dans la pensée, 1795. (Cited on page 151.)
- [Esling 2012] Philippe Esling and Carlos Agon. *Time-series data mining*. ACM Computing Surveys (CSUR), vol. 45, no. 1, pages 1–34, 2012. (Cited on page 83.)
- [Etori 2024] Naome Etori, Maurice Dawson and Maria Gini. *Double-edged sword: Navigating AI Opportunities and the Risk of Digital Colonization in Africa*. 2024. (Cited on page 12.)
- [Fang 2018] Huazhen Fang, Ning Tian, Yebin Wang, MengChu Zhou and Mulugeta A Haile. *Nonlinear Bayesian estimation: From Kalman filtering to a broader horizon*. IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 2, pages 401–417, 2018. (Cited on page 124.)

- [Fjeld 2020] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy and Madhulika Srikumar. *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication, no. 2020-1, 2020. (Cited on pages 102 and 107.)
- [Floridi 2016] Luciano Floridi and Mariarosaria Taddeo. *What is data ethics?*, 2016. (Cited on page 45.)
- [Floridi 2018] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossiet al. *AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations*. *Minds and machines*, vol. 28, pages 689–707, 2018. (Cited on page 103.)
- [Floridi 2023] Luciano Floridi, Jean-Gabriel Ganascia, Enrico Panaï and Emmanuel R Goffi. *L'éthique de l'intelligence artificielle: principes, défis et opportunités*. (No Title), 2023. (Cited on page 45.)
- [Foisneau 2014] Luc Foisneau. *Rawls et la justification de la règle de majorité*. *Raisons politiques*, vol. 53, no. 1, pages 63–79, 2014. (Cited on page 43.)
- [Fokane 2024] Tusi Fokane. *Towards Ethical AI Regulation in Africa*, March 2024. (Cited on page 12.)
- [Gaston 2002] Sean Gaston. *Derrida and the Ruins of Disinterest*. *ANGELAKI journal of the theoretical humanities*, vol. 7, no. 3, pages 105–118, 2002. (Cited on page 47.)
- [Ghaffar Nia 2023] Nafiseh Ghaffar Nia, Erkan Kaplanoglu and Ahad Nasab. *Evaluation of artificial intelligence techniques in disease diagnosis and prediction*. *Discover Artificial Intelligence*, vol. 3, no. 1, page 5, 2023. (Cited on page 73.)
- [Giavina-Bianchi 2021] Mara Giavina-Bianchi, Raquel Machado de Sousa, Vitor Zago de Almeida Paciello, William Gois Vitor, Aline Lissa Okita, Renata Prôa, Gian Lucca dos Santos Severino, Anderson Alves Schinaid, Rafael Espírito Santo and Birajara Soares Machado. *Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting*. *PLoS One*, vol. 16, no. 9, page e0257006, 2021. (Cited on page 73.)
- [Gitzel 2016] Ralf Gitzel. *Data Quality in Time Series Data: An Experience Report*. In *CBI (Industrial Track)*, pages 41–49, 2016. (Cited on pages ix, 84 and 85.)
- [Giuffrida 2018] Tanguy Giuffrida, Sophie Dupuy-Chessa, Jean-Philippe Poli and Éric Céret. *Fuzzy4U: un système en logique floue pour l'adaptation des interfaces*

- utilisateur*. In 30eme conférence francophone sur l'interaction homme-machine, pages 52–62, 2018. (Cited on page 79.)
- [Gouriéroux 1995] Christian Gouriéroux and Alain Monfort. *Séries temporelles et modèles dynamiques*. FeniXX, 1995. (Cited on page 84.)
- [Gowans 2004] Chris Gowans. *Moral relativism*. 2004. (Cited on pages 59 and 60.)
- [Gregson 2014] R. A. M. Gregson. *Time Series in Psychology*. Psychology Press, May 2014. Google-Books-ID: uEmhAwAAQBAJ. (Cited on page 82.)
- [Guigou 2019] Fabio Guigou. *The artificial immune ecosystem : a scalable immune-inspired active classifier, an application to streaming time series analysis for network monitoring*. These de doctorat, Strasbourg, June 2019. (Cited on page 84.)
- [Gwagwa 2019] Arthur Gwagwa. *Recommendations on the inclusion sub-Saharan Africa in Global AI Ethics*. RANITP Policy Brief, vol. 2, 2019. (Cited on pages 7, 100, 105 and 150.)
- [Gwagwa 2020] Arthur Gwagwa, Erika Kraemer-Mbula, Nagla Rizk, Isaac Rutenberg and Jeremy De Beer. *Artificial intelligence (AI) deployments in Africa: benefits, challenges and policy dimensions*. The African Journal of Information and Communication, vol. 26, pages 1–28, 2020. (Cited on pages 104, 105, 109 and 152.)
- [Gwagwa 2022] Arthur Gwagwa, Emre Kazim and Airlie Hilliard. *The role of the African value of Ubuntu in global AI inclusion discourse: A normative ethics perspective*. Patterns, vol. 3, no. 4, 2022. (Cited on page 152.)
- [Gyekye 2010] Kwame Gyekye. *African ethics*. 2010. (Cited on pages 62 and 63.)
- [Habermas 1978] Jürgen Habermas and Marc B de Launay. *L'espace public: archéologie de la publicité comme dimension constitutive de la société bourgeoise*. 1978. (Cited on page 55.)
- [Halladay 2013] Shawn D Halladay. *Using predictive analytics to improve decisionmaking*. The Journal of Equipment Lease Financing, vol. 31, no. 2, pages B1–B6, 2013. (Cited on page 81.)
- [Heston 2023] Thomas F Heston and Charya Khun. *Prompt engineering in medical education*. International Medical Education, vol. 2, no. 3, pages 198–205, 2023. (Cited on page 97.)
- [Hibon 1999] Michèle Hibon and Spyros G Makridakis. *M3-competition*. INSEAD, 1999. (Cited on page 86.)

- [Huang 2024] Jingwei Huang, Donghan M Yang, Ruichen Rong, Kuroush Nezafati, Colin Treager, Zhikai Chi, Shidan Wang, Xian Cheng, Yujia Guo, Laura J Klesse *et al.* *A critical assessment of using ChatGPT for extracting structured data from clinical notes.* npj Digital Medicine, vol. 7, no. 1, page 106, 2024. (Cited on page 157.)
- [Hyndman 2007] Rob J Hyndman, Andrey V Kostenko *et al.* *Minimum sample size requirements for seasonal forecasting models.* foresight, vol. 6, no. Spring, pages 12–15, 2007. (Cited on page 5.)
- [Hyndman 2014] Rob J Hyndman. *Fitting models to short time series.* 2014. (Cited on page 86.)
- [Hyndman 2018] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice.* OTexts, 2018. (Cited on pages 86 and 126.)
- [Idrees 2019] Sheikh Mohammad Idrees, M Afshar Alam and Parul Agarwal. *A prediction approach for stock market volatility based on time series data.* IEEE Access, vol. 7, pages 17287–17298, 2019. (Cited on page 83.)
- [Igalens 2002] Jacques Igalens, Michel Joras and François Mancy. *La responsabilité sociale de l’entreprise: comprendre, rédiger le rapport annuel.* Ed. d’Organisation, 2002. (Cited on page 57.)
- [Institut-Sapiens 2018] Institut-Sapiens. *L’E-stonie : modèle d’un état plateforme e-gouverné. L’exemple de la transformation totale d’un Etat grâce au numérique,* 2018. (Cited on page 3.)
- [Ioannidis 2023] Jules Ioannidis, Joshua Harper, Ming Sheng Quah and Dan Hunter. *Gracenote. ai: Legal Generative AI for Regulatory Compliance.* In Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023), 2023. (Cited on page 95.)
- [Islam 2024] Md Rakibul Islam, Shahina Begum and Mobyen Uddin Ahmed. *Artificial Intelligence in Predictive Maintenance: A Systematic Literature Review on Review Papers.* In International Congress and Workshop on Industrial AI, pages 251–261. Springer, 2024. (Cited on page 73.)
- [Jarrahi 2018] Mohammad Hossein Jarrahi. *Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making.* Business Horizons, vol. 61, no. 4, pages 577–586, July 2018. (Cited on page 80.)

- [Jobin 2019] Anna Jobin, Marcello Ienca and Effy Vayena. *The global landscape of AI ethics guidelines*. *Nature Machine Intelligence*, vol. 1, no. 9, pages 389–399, 2019. (Cited on pages [ix](#), [xi](#), [101](#) and [102](#).)
- [Jose 2022] Jonath Jose. *Introduction to time series analysis and its applications*. Christ University, Bangalore, 2022. (Cited on pages [ix](#), [129](#), [130](#) and [131](#).)
- [Jouan 2019] Marlène Jouan. *Morality and Its Limits*. 2019. (Cited on page [58](#).)
- [Kalman 1961] RE Kalman. *A New Approach to Liner Filtering and Prediction Problems, Transaction of ASME*. *Journal of Basic Engineering*, vol. 83, no. 1, pages 95–108, 1961. (Cited on pages [127](#) and [135](#).)
- [Kalyan 2023] Katikapalli Subramanyam Kalyan. *A survey of GPT-3 family large language models including ChatGPT and GPT-4*. *Natural Language Processing Journal*, page 100048, 2023. (Cited on page [97](#).)
- [Kast 2002] Robert Kast. *La théorie de la décision*. 2002. (Cited on page [79](#).)
- [Kaur 2020] Dr Navleen Kaur, Supriya Lamba Sahdev, Dr Monika Sharma and Laraibe Siddiqui. *Banking 4.0: the influence of artificial intelligence on the banking industry & how ai is changing the face of modern day banks*. *International Journal of Management*, vol. 11, no. 6, 2020. (Cited on page [73](#).)
- [Keijzer 2003] Maarten Keijzer. *Improving symbolic regression with interval arithmetic and linear scaling*. In *European Conference on Genetic Programming*, pages 70–82. Springer, 2003. (Cited on pages [134](#) and [140](#).)
- [Kettani 2014] Driss Kettani and Bernard Moulin. *L'e-gouvernement pour la bonne gouvernance dans les pays...* Presses de l'Université Laval, 2014. (Cited on page [3](#).)
- [Khan 2018] Wazir Zada Khan, Mohammed Y Aalsalem and Muhammad Khurram Khan. *Communal acts of IoT consumers: A potential threat to security and privacy*. *IEEE Transactions on Consumer Electronics*, vol. 65, no. 1, pages 64–72, 2018. (Cited on page [107](#).)
- [Kim 2012] Uchang Kim and Christine Klein-Lataud. *Éthique universelle et mondialisation*. *Diogène*, no. 1, pages 52–74, 2012. (Cited on pages [53](#), [54](#), [55](#), [56](#) and [57](#).)
- [Kim 2022] Sung Wook Kim, Jun Ho Kong, Sang Won Lee and Seungchul Lee. *Recent advances of artificial intelligence in manufacturing industrial sectors: A review*. *International Journal of Precision Engineering and Manufacturing*, pages 1–19, 2022. (Cited on page [73](#).)

- [Kinyondo 2018] Abel Kinyondo and Riccardo Pelizzo. *Poor Quality of Data in Africa: What Are the Issues?* Politics & Policy, vol. 46, no. 6, pages 851–877, 2018. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/polp.12277](https://onlinelibrary.wiley.com/doi/pdf/10.1111/polp.12277). (Cited on page 128.)
- [Kirshners 2012] Arnis Kirshners and Arkady Borisov. *A Comparative Analysis of Short Time Series Processing Methods*. Information Technology and Management Science, vol. 15, no. 1, pages 65–69, 2012. (Cited on page 5.)
- [Kommenda 2009] Michael Kommenda, Gabriel Kronberger, Stephan Winkler, Michael Affenzeller, Stefan Wagner, Leonhard Schickmair and Benjamin Lindner. *Application of genetic programming on temper mill datasets*. In 2009 2nd International Symposium on Logistics and Industrial Informatics, pages 1–5. IEEE, 2009. (Cited on pages ix and 132.)
- [Kone 2023] Lalla Aicha Kone, Anna Ouskova Leonteva, Mamadou Tourad Diallo, Ahmedou Haouba and Pierre Collet. *Short Time Series Forecasting Method Based on Genetic Programming and Kalman Filter*. In Complex Computational Ecosystems, pages 77–89. Springer Nature Switzerland, 2023. (Cited on pages 145, 201 and 204.)
- [Koul 2022] Apeksha Koul, Rajesh K Bawa and Yogesh Kumar. *Artificial intelligence in medical image processing for airway diseases*. In Connected e-Health: Integrated IoT and Cloud Computing, pages 217–254. Springer, 2022. (Cited on page 73.)
- [Koza 1994] John R Koza. *Genetic programming as a means for programming computers by natural selection*. Statistics and computing, vol. 4, pages 87–112, 1994. (Cited on pages 126, 127 and 131.)
- [Kulesz 2018] Octavio Kulesz. *Culture, machines et plateformes : l'intelligence artificielle et son impact sur la diversité des expressions culturelles*, December 2018. (Cited on page 106.)
- [Le Gland 2009] François Le Gland. *Filtrage bayésien et approximation particulière*. Cours de l'Ecole Nationale Supérieure des Techniques Avancées.[Online]. Available: <http://www.irisa.fr/aspi/legland/ensta/cours.pdf>, 2009. (Cited on page 124.)
- [LE PREMIER MINISTRE 2012] DR LE PREMIER MINISTRE and MOHAMED LAGHDAF. *PLAN D'ACTION NATIONAL POUR L'ENVIRONNEMENT 2012-2016*. 2012. (Cited on page 13.)
- [LEB 1989] *Missing Values Rebuilding by Prediction and ARIMA Modelling in Time Series*. IFAC Proceedings Volumes, vol. 22, no. 6, pages 357–361, 1989.

- IFAC Symposium on Advanced Information Processing in Automatic Control (AIPAC'89), Nancy, France, 3-5 July 1989. (Cited on page 130.)
- [Lemasson 2008] Laurent Lemasson. *La démocratie radicale de Jürgen Habermas: Entre socialisme et anarchie*. Revue française de science politique, vol. 58, no. 1, pages 39–67, 2008. (Cited on page 55.)
- [Lemoine 2003] Matthieu Lemoine and Florian Pelgrin. *Introduction aux modèles espace-état et au filtre de Kalman*. Revue de l'OFCE, vol. 86, no. 3, pages 203–229, 2003. (Cited on page 124.)
- [Leonteva 2022] Anna Ouskova Leonteva. *Evolutionary and quantum-inspired algorithms for the optimization of magnetic cooling systems*. phdthesis, Université de Strasbourg, May 2022. (Cited on page 138.)
- [Lindemann 2024] Nora Freya Lindemann. *Chatbots, search engines, and the sealing of knowledges*. AI & SOCIETY, pages 1–14, 2024. (Cited on page 156.)
- [Locke 1992] John Locke. *Second traité du gouvernement civil*. J.-F., 1992. (Cited on page 52.)
- [Lusk 1984] Edward J Lusk and Joao S Neves. *A comparative ARIMA analysis of the 111 series of the Makridakis competition*. Journal of Forecasting, vol. 3, no. 3, pages 329–332, 1984. (Cited on page 86.)
- [Maitre 2011] Ogier Maitre, Deepak Sharma, Nicolas Lachiche and Pierre Collet. *DISPAR-Tournament: a parallel population reduction operator that behaves like a tournament*. In Applications of Evolutionary Computation: EvoApplications 2011: EvoCOMPLEX, EvoGAMES, EvoIASP, EvoINTELLIGENCE, EvoNUM, and EvoSTOC, Torino, Italy, April 27-29, 2011, Proceedings, Part I, pages 284–293. Springer, 2011. (Cited on page 134.)
- [Mannion 2020] Cara Mannion. *Data imperialism: The GDPR's disastrous impact on Africa's E-commerce markets*. Vand. J. Transnat'l L., vol. 53, page 685, 2020. (Cited on page 111.)
- [Masseys-Bertoneche 2011] Carole Masseys-Bertoneche. *La commercialisation des universités américaines: exemplarité pour l'Europe?* Mutations de l'enseignement supérieur et internationalisation. Change in higher education and globalisation, Bruselas, De Boeck, pages 129–144, 2011. (Cited on page 6.)
- [Mballo 2023] Marie Hélène W Mballo. *Une IA générative responsable au service de l'enseignement supérieur*. Sciences Appliquées et de l'Ingénieur, vol. 5, no. 1, pages 96–100, 2023. (Cited on page 95.)

- [McCulloch 2000] Neil McCulloch, Bob Baulch and Milaso Cherele-Robson. *Growth, inequality and poverty in Mauritania, 1987-1996*. Poverty Reduction and Social Development Africa Region, The World Bank, mimeo, 2000. (Cited on page 125.)
- [MEFPTN 2011] MEFPTN and MFPTMA. *Stratégie nationale de modernisation de l'Administration et des TIC 2012 – 2016*, 2011. accès le 15 janvier 2020. (Cited on page 3.)
- [Mehr 2017] Hila Mehr, H Ash and D Fellow. *Artificial intelligence for citizen services and government*. Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., no. August, vol. 1, page 12, 2017. (Cited on page 149.)
- [Meijer 2024] Albert Meijer and William Webster. *Call or Papers: Generative AI in Public Services A Special Issue of the Journal Information Polity*. 2024. (Cited on page 95.)
- [Merrill 2017] Roberto Merrill and Patrick Savidan. *Du minimalisme moral*. Essais pour Ruwen Ogien, vol. 22, 2017. (Cited on page 58.)
- [Miaillhe 2018] Nicolas Miaillhe. *The geopolitics of artificial intelligence: The return of empires?* Politique étrangère, no. 3, pages 105–117, 2018. (Cited on page 103.)
- [Michon 2016] Cyrille Michon. *Anscombe et la doctrine du double effet*. Klesis, vol. 35, pages 143–163, 2016. (Cited on page 33.)
- [Mill 1861] John Stuart Mill. *1969. Utilitarianism*. Collected Works of John Stuart Mill, vol. 10, pages 203–259, 1861. (Cited on page 35.)
- [Mill 1885] John Stuart Mill and Thomas De Quincey. *On liberty*. 1885. (Cited on page 34.)
- [Mill 1966] John Stuart Mill and John Stuart Mill. *On liberty*. Springer, 1966. (Cited on page 35.)
- [Miller 2019] Hannah Miller, Richard Stirling, Yaera Chung, Sriganesh Lokanathan, Emma Martinho-Truswell, Joshua New, Isaac Rutenberg and Fabrizio Scrollini. *Government artificial intelligence readiness index 2019*. London: Oxford Insights, page 32, 2019. (Cited on page 104.)
- [Miltenburg 2016] Emiel van Miltenburg. *Stereotyping and Bias in the Flickr30k Dataset*. Proceedings of Multimodal Corpora, pages 1–4, 2016. (Cited on page 152.)
- [Mohr 2010] Georg Mohr. *Publicité de la raison, droit et cosmopolitisme chez Kant*. Raison pratique et normativité chez Kant: droit, politique et cosmopolitique, pages 213–243, 2010. (Cited on page 151.)

- [Moradi-Motlagh 2015] Amir Moradi-Motlagh, Abbas Valadkhani and Ali Salman Saleh. *Rising efficiency and cost saving in Australian banks: A bootstrap approach*. Applied Economics Letters, vol. 22, no. 3, pages 189–194, 2015. (Cited on page 73.)
- [Morrison 1977] G Wayne Morrison and David H Pike. *Kalman filtering applied to statistical forecasting*. Management Science, vol. 23, no. 7, pages 768–774, 1977. (Cited on page 135.)
- [Nallur 2020] Vivek Nallur. *Landscape of machine implemented ethics*. Science and engineering ethics, vol. 26, pages 2381–2399, 2020. (Cited on page 154.)
- [Nasseri 2011] Mohsen Nasseri, Ali Moeini and Massoud Tabesh. *Forecasting monthly urban water demand using extended Kalman filter and genetic programming*. Expert Systems with Applications, vol. 38, no. 6, pages 7387–7395, 2011. (Cited on pages 87, 127, 135, 137 and 140.)
- [Nawel 2019] Benlakehal Nawel. *L'administration électronique «L'Algérie à la traîne malgré les réformes» Electronic administration" Algeria lags behind despite reforms*. 2019. (Cited on page 2.)
- [Nerdeux 1978] C Nerdeux. *La déclaration universelle des droits de l'animal*. Better-avier Francais, vol. 48, 1978. (Cited on page 64.)
- [N'Guessan 2024] Charlette N'Guessan. *The African Union Artificial Intelligence Continental Strategy for Africa — Shaping Africa's AI...*, March 2024. (Cited on page 12.)
- [Novo Diaz 2023] Oscar Novo Diaz. *Exploring the Quantum Frontiers of Generative Adversarial Networks in Chemical Catalysis*. PhD thesis, ETSI_Sistemas_Infor, 2023. (Cited on page 94.)
- [Ntoutsis 2020] Eirini Ntoutsis, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis *et al.* *Bias in data-driven artificial intelligence systems—An introductory survey*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 3, page e1356, 2020. (Cited on page 152.)
- [OIT 2019] OIT. *Femmes et hommes dans l'économie informelle: un panorama statistique*. 2019. (Cited on page 105.)
- [Olveres 2021] Jimena Olveres, Germán González, Fabian Torres, José Carlos Moreno-Tagle, Erik Carbajal-Degante, Alejandro Valencia-Rodríguez, Nahum Méndez-Sánchez and Boris Escalante-Ramírez. *What is new in computer vision and*

- artificial intelligence in medical image analysis applications*. Quantitative imaging in medicine and surgery, vol. 11, no. 8, page 3830, 2021. (Cited on page 73.)
- [OMS 2021] OMS. *Éthique et gouvernance de l'intelligence artificielle pour la santé : orientations de l'OMS : résumé analytique*, June 2021. (Cited on page 103.)
- [Ostertagová 2011] Eva Ostertagová and Oskar Ostertag. *The simple exponential smoothing model*. In The 4th International Conference on modelling of mechanical and mechatronic systems, Technical University of Košice, Slovak Republic, Proceedings of Conference, pages 380–384, 2011. (Cited on page 86.)
- [Outlaw Jr 2010] Lucius T Outlaw Jr and Chike Jeffers. *Africana philosophy*. 2010. (Cited on page 62.)
- [Oyeleye 2022] Matthew Oyeleye, Tianhua Chen, Sofya Titarenko and Grigoris Antoniou. *A predictive analysis of heart rates using machine learning techniques*. International Journal of Environmental Research and Public Health, vol. 19, no. 4, page 2417, 2022. (Cited on page 83.)
- [Pedro 2019] Francesc Pedro, Miguel Subosa, Axel Rivas and Paula Valverde. *Artificial intelligence in education: Challenges and opportunities for sustainable development*. 2019. (Cited on page 73.)
- [Penfold 2013] Robert B Penfold and Fang Zhang. *Use of interrupted time series analysis in evaluating health care quality improvements*. Academic pediatrics, vol. 13, no. 6, pages S38–S44, 2013. (Cited on page 83.)
- [Peter 1975] Singer Peter. *La Libération animale, trad.* L. Rousselle, Paris, Éditions Payot et Rivages, 1975. (Cited on page 63.)
- [Pielke Jr 2003] Roger A. Pielke Jr and Richard T. Conant. *Best Practices in Prediction for Decision-Making: Lessons from the Atmospheric and Earth Sciences*. Ecology, vol. 84, no. 6, pages 1351–1358, 2003. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9658%282003%29084%5B1351%3ABPIPF%5D2.0.CO%3B2>. (Cited on page 81.)
- [Pomerol 2002] Jean-Charles Pomerol. *L'apport de H. Simon dans le management et la décision*. January 2002. (Cited on page 79.)
- [Popel 2020] Martin Popel, Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar and Zdeněk Žabokrtský. *Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals*. Nature communications, vol. 11, no. 1, page 4381, 2020. (Cited on page 71.)

- [Pothumsetty 2020] Rohan Pothumsetty. *Application of artificial intelligence in algorithmic trading*. Int. J. Eng. Appl. Sci. Technol, vol. 4, no. 12, pages 140–149, 2020. (Cited on page 73.)
- [Qi 2001] Min Qi and Guoqiang Peter Zhang. *An investigation of model selection criteria for neural network time series forecasting*. European journal of operational research, vol. 132, no. 3, pages 666–680, 2001. (Cited on page 84.)
- [Rane 2023] Nitin Rane. *Role of ChatGPT and Similar Generative Artificial Intelligence (AI) in Construction Industry*. SSRN Electronic Journal, January 2023. (Cited on page 94.)
- [Rane 2024] Nitin Rane, Saurabh Choudhary and Jayesh Rane. *Gemini Versus ChatGPT: Applications, Performance, Architecture, Capabilities, and Implementation*. Performance, Architecture, Capabilities, and Implementation (February 13, 2024), 2024. (Cited on page 7.)
- [Rawat 2020] Prateek Rawat and Archana N Mahajan. *ReactJS: A modern web development framework*. International Journal of Innovative Science and Research Technology, vol. 5, no. 11, pages 698–702, 2020. (Cited on page 156.)
- [Rawls 1958] John Rawls. *Justice as fairness*. 1958. (Cited on page 40.)
- [Rawls 1987] John Rawls and Catherine Audard. *Théorie de la justice*. Ed. du Seuil, 1987. (Cited on page 42.)
- [Raza 2023] Falsk Raza. *AI in Education: Personalized Learning and Adaptive Assessment*. November 2023. (Cited on page 73.)
- [Rousseau 1896] Jean-Jacques Rousseau and Edmond-Paul Dreyfus-Brisac. *Du contrat social*. Alcan, 1896. (Cited on page 53.)
- [Roy 1985] Bernard Roy. *Méthodologie multicritère d'aide à la décision*. (No Title), 1985. (Cited on page 80.)
- [Ruwen 2011] Ogien Ruwen. *L'Influence de l'odeur de croissants chauds sur la bonté humaine*. Paris, Grasset, 2011. (Cited on page 59.)
- [Sætra 2023] Henrik Skaug Sætra. *Generative AI: Here to stay, but for good?* Technology in Society, vol. 75, page 102372, 2023. (Cited on page 92.)
- [Saidani 2013] Becem Saidani. *Filtre de kalman discret pour l'estimation des moyennes inconnues de bruits blancs*. Library and Archives Canada= Bibliotheque et Archives Canada, Ottawa, 2013. (Cited on page 124.)

- [Sallstrom 2019] Laura Sallstrom, Olive Morris and Halak Mehta. *Artificial Intelligence in Africa's Healthcare: Ethical Considerations*. no. 312, 2019. (Cited on page 107.)
- [Sangral 2015] Stéphane Sangral. *Fatras du soi, fracas de l'autre*. page 232, 2015. (Cited on page 20.)
- [Schneewind 1977] Jerome B Schneewind. *Sidgwick's ethics and victorian moral philosophy*. Oxford University Press, 1977. (Cited on page 36.)
- [Schork 2019] Nicholas J Schork. *Artificial intelligence and personalized medicine*. Precision medicine in Cancer therapy, pages 265–283, 2019. (Cited on page 73.)
- [Schroeder 2013] William R Schroeder. *Continental ethics*. The Blackwell guide to ethical theory, pages 461–486, 2013. (Cited on page 49.)
- [Shahid 2021] Nadeem U Shahid, Nasir J Sheikhet *al*. *Impact of big data on innovation, competitive advantage, productivity, and decision making: literature review*. Open Journal of Business and Management, vol. 9, no. 02, page 586, 2021. (Cited on page 72.)
- [Shaik 2023] Thanveer Shaik, Xiaohui Tao, Niall Higgins, Lin Li, Raj Gururajan, Xujuan Zhou and U Rajendra Acharya. *Remote patient monitoring using artificial intelligence: Current state, applications, and challenges*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 13, no. 2, page e1485, 2023. (Cited on page 73.)
- [Shearer 2020] Eleanor Shearer, Richard Stirling and Walter Pasquarelli. *Government AI readiness index 2020*. IDRC & Oxford Insights, 2020. (Cited on page 108.)
- [Sidgwick 2022] Henry Sidgwick. *The methods of ethics*. DigiCat, 2022. (Cited on page 36.)
- [Singh 2023] Karan Singh, Surbhi Singh, Mehar Vohra and Ravi Shankar Jha. *Cyber-Attack Detection Using Machine Learning Technique*. In International Conference on Recent Developments in Cyber Security, pages 463–473. Springer, 2023. (Cited on page 74.)
- [Sinnott-Armstrong 2003] Walter Sinnott-Armstrong. *Consequentialism*. 2003. (Cited on page 33.)
- [Smith 2018] Matthew Smith and Sujaya Neupane. *Intelligence artificielle et développement humain: vers un programme de recherche*. 2018. (Cited on pages 4 and 104.)

- [Suripto 2023] S Suripto. *Decision-making model to predict auto-rejection: An implementation of ARIMA for accurate forecasting of stock price volatility during the COVID-19*. *Decision Science Letters*, vol. 12, no. 1, pages 107–116, 2023. (Cited on page 80.)
- [Susto 2016] Gian Antonio Susto and Alessandro Beghi. *Dealing with time-series data in predictive maintenance problems*. In 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), pages 1–4. IEEE, 2016. (Cited on page 83.)
- [Tiple 2020] Vasile Tiple. *Recommendations on the European Commission’s White Paper on Artificial Intelligence-A European approach to excellence and trust, COM (2020) 65 final (the ‘AI White Paper’)*. 2020. (Cited on pages 102 and 103.)
- [Toma 2023] Milan Toma and Ong Chi Wei. *Predictive modeling in medicine*. *Encyclopedia*, vol. 3, no. 2, pages 590–601, 2023. (Cited on page 81.)
- [Touzet 1998] Claude Touzet. *Des réseaux de neurones artificiels à la robotique coopérative*. PhD thesis, Aix Marseille Université, 1998. (Cited on page 78.)
- [Tsay 2005] Ruey S Tsay. *Analysis of financial time series*. John Wiley & sons, 2005. (Cited on page 83.)
- [Turbal 2019] Yuriy Turbal, Andriy Bomba, Mariana Turbal, Anastasiia Sokh and Olena Radoveniuk. *Pyramidal method of extrapolation for short time series*. *International Journal of Computing Science and Mathematics*, vol. 10, no. 6, pages 525–533, 2019. (Cited on page 5.)
- [Umamaheswari 2023] S Umamaheswari, A Valarmathiet al. *Role of artificial intelligence in the banking sector*. *Journal of Survey in Fisheries Sciences*, vol. 10, no. 4S, pages 2841–2849, 2023. (Cited on page 73.)
- [UNIES 2009] NATIONS UNIES. *Examen de la politique de l’investissement de la Mauritanie*. 2009. (Cited on page 5.)
- [Union 2014] African Union. *African Union convention on cyber security and personal data protection*. African Union, vol. 27, 2014. (Cited on page 109.)
- [Velásquez-Henao 2023] Juan David Velásquez-Henao, Carlos Jaime Franco-Cardona and Lorena Cadavid-Higuita. *Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering*. *Dyna*, vol. 90, no. 230, pages 9–17, 2023. (Cited on page 97.)

- [Vidal 2003] Laurent Vidal. *De l'universalisme au relativisme en éthique: échanges autour de l'exemple du sida en Côte d'Ivoire*. Autrepart, no. 4, pages 55–69, 2003. (Cited on page 61.)
- [Walter 2024] Yoshija Walter. *Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences*. Discover Artificial Intelligence, vol. 4, no. 1, page 14, 2024. (Cited on pages 12 and 92.)
- [Wang 2017] Zeyu Wang and Ravi S Srinivasan. *A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models*. Renewable and Sustainable Energy Reviews, vol. 75, pages 796–808, 2017. (Cited on page 84.)
- [Wanigasekara 2018] Chathura Wanigasekara, Akshya Swain, Sing Kiong Nguang and B Gangadhara Prusty. *Improved learning from small data sets through effective combination of machine learning tools with VSG techniques*. In 2018 International Joint Conference on Neural Networks (IJCNN), pages 1–6. IEEE, 2018. (Cited on page 126.)
- [Welch 1995] Greg Welch, Gary Bishop *et al.* *An introduction to the Kalman filter*. 1995. (Cited on pages ix, 135 and 136.)
- [Williams 1995] Bernard Williams. *Making sense of humanity: and other philosophical papers 1982-1993*. Cambridge University Press, 1995. (Cited on page 39.)
- [Wilson 2018] H James Wilson and Paul R Daugherty. *Collaborative intelligence: Humans and AI are joining forces*. Harvard Business Review, vol. 96, no. 4, pages 114–123, 2018. (Cited on page 80.)
- [Yang 2024] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin and Xia Hu. *Harnessing the power of llms in practice: A survey on chatgpt and beyond*. ACM Transactions on Knowledge Discovery from Data, vol. 18, no. 6, pages 1–32, 2024. (Cited on page 156.)
- [Yu 2023] Ping Yu, Hua Xu, Xia Hu and Chao Deng. *Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration*. Healthcare, vol. 11, no. 20, page 2776, January 2023. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute. (Cited on pages 93, 94 and 95.)
- [Zhao 2024] Haiyan Zhao, Hanjie Chen, Fan Yang, Ninghao Liu, Huiqi Deng, Hengyi Cai, Shuaiqiang Wang, Dawei Yin and Mengnan Du. *Explainability for large*

- language models: A survey*. ACM Transactions on Intelligent Systems and Technology, vol. 15, no. 2, pages 1–38, 2024. (Cited on page 156.)
- [Zhu 2018] Jason Zhu. *Predictive Modeling in Reproductive Medicine*. vol. 2, page 224, December 2018. (Cited on page 81.)
- [Zielinski 2010] Agata Zielinski. *L'éthique du care: une nouvelle façon de prendre soin*. Études, vol. 413, no. 12, pages 631–641, 2010. (Cited on page 43.)
- [Zouinar 2020] Moustafa Zouinar. *Évolutions de l'Intelligence Artificielle: quels enjeux pour l'activité humaine et la relation Humain-Machine au travail?* Activités, no. 17-1, 2020. (Cited on page 4.)

Annexes

Annexes

Méthodes de calcul des indicateurs (Voir tables 1, 2, 3, 4):

Table 1: Indicateurs généraux (4 indicateurs)

Indicateurs	Taux d'admis à la 1ere session	Taux d'admis à la 2eme session	Taux de garçons admis au Bac (tout bac inclus)	Taux de filles admises au Bac (tout bac inclus)
Calcul	Nombre d'admis à la 1ere session sur le nombre total d'admis au baccalauréat (sessions 1 et 2)	Nombre d'admis à la 2eme session sur le nombre total d'admis au baccalauréat	Nombre de garçons admis au Bac (tout bac inclus) sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admises au Bac (tout bac inclus) sur le nombre total d'admis au bac (tout bac inclus)

Indicateurs	Taux d'admis au Bac M	Taux de filles admises au Bac M	Taux de garçons admis au Bac M
Calcul	Nombre d'admis au Bac M sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admise au Bac M sur le nombre total d'admis au bac M	Nombre de garçons admis au Bac M sur le nombre total d'admis au bac M

Indicateurs	Taux d'admis au Bac SN	Taux de filles admises au Bac SN	Taux de garçons admis au Bac SN
Calcul	Nombre d'admis au Bac SN sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admises au Bac SN sur le nombre total d'admis au bac SN	Nombre de garçons admis au Bac SN sur le nombre total d'admis au bac SN

Indicateurs	Taux d'admis au Bac TMGM	Taux de filles admises au Bac TMGM	Taux de garçons admis au Bac TMGM
Calcul	Nombre d'admis au Bac TMGM sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admises au Bac TMGM sur le nombre total d'admis au bac TMGM	Nombre de garçons admis au Bac TMGM sur le nombre total d'admis au bac TMGM

Table 2: Indicateurs relatifs aux baccalauréats scientifique et technique (9 indicateurs)

Indicateurs	Taux d'admis au Bac LM	Taux de filles admises au Bac LM	Taux de garçons admis au Bac LM
Calcul	Nombre d'admis au Bac LM sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admise au Bac LM sur le nombre total d'admis au bac LM	Nombre de garçons admis au Bac LM sur le nombre total d'admis au bac LM

Indicateurs	Taux d'admis au Bac LO	Taux de filles admises au Bac LO	Taux de garçons admis au Bac LO
Calcul	Nombre d'admis au Bac LO sur le nombre total d'admis au bac (tout bac inclus)	Nombre de filles admises au Bac LO sur le nombre total d'admis au bac LO	Nombre de garçons admis au Bac LO sur le nombre total d'admis au bac LO

Table 3: Indicateurs relatifs aux baccalauréats en lettre (6 indicateurs)

Table 4: Indicateurs relatifs au genre (3 indicateurs)

Indicateurs	Taux de filles avec un bac scientifique	Taux des filles avec un bac technique	Taux de filles avec un bac littéraire
Calcul	Nombre de filles avec un bac scientifique (C, D) sur le nombre total des admis (garçons et filles) avec un bac scientifique (C, D)	Nombre de filles avec un bac technique (TMGM) sur le nombre total des admis (garçons et filles) avec un bac technique (TMGM)	Nombre de filles avec un bac littéraire (LO, LM) sur le nombre total des admis (garçons et filles) avec un bac littéraire (LO, LM)

Valeurs des indicateurs(Séries temporelles) calculés à partir des données collectées(Voir tables 5) :

Année	Effectif de candidats au bac(tout bac inclut)	Nombre d'admis au Bac(tout bac inclut)	taux d'admission au Bac(tout bac inclut)	taux de garçons admis au Bac(tout bac inclut)
2013	35421	5696	16%	66%
2014	41666	5017	12%	68%
2015	41986	5782	14%	66%
2016	45026	5081	11%	67%
2017	49993	6693	13%	62%
2018	50045	11114	22%	59%
2019	44717	7094	16%	62%
2020	49754	11658	23%	57%
2021	46581	6234	13%	60%
2022	47251	9307	20%	54%

Année	Taux d'admis au Bac Mathématiques	Taux de filles admises au Bac Mathématiques	Taux d'admis au Bac Science Naturelle	Taux de garçons admis au Bac Science Naturelle
2013	10%	26%	44%	67%
2014	18%	24%	35%	66%
2015	13%	22%	45%	64%
2016	18%	26%	39%	64%
2017	14%	28%	50%	59%
2018	10%	31%	58%	58%
2019	8%	26%	57%	59%
2020	6%	27%	69%	54%
2021	6%	26%	61%	57%
2022	5%	26%	57%	53%

Année	Taux d'admis au Bac technique	Taux de filles avec un Bac technique	Taux d'admis au Bac Lettres modernes	Taux de garçons admis au Bac Lettres modernes	Taux d'admis au Bac Lettres originales
2013	1%	7%	18%	69%	28%
2014	1%	9%	15%	68%	31%
2015	1%	9%	14%	68%	28%
2016	1%	14%	16%	70%	25%
2017	1%	14%	16%	69%	20%
2018	0%	12%	11%	64%	20%
2019	1%	16%	16%	66%	18%
2020	1%	18%	9%	66%	15%
2021	0%	22%	13%	61%	19%
2022	1%	21%	12%	58%	25%

Année	Taux de filles admises au Bac Lettres originales	Age moyen des admis	Taux d'admis à la 1ere session	Taux de filles avec un bac scientifique	Taux de filles avec un bac littéraires
2013	41%	21,94	55%	31%	37%
2014	35%	21,82	55%	30%	34%
2015	38%	21,89	54%	33%	36%
2016	37%	21,61	53%	33%	34%
2017	42%	21,37	59%	39%	37%
2018	46%	21,52	58%	40%	43%
2019	38%	20,89	46%	39%	36%
2020	41%	20,75	62%	44%	39%
2021	38%	20,28	60%	42%	38%
2022	50%	20.63	61%	45%	48%

Table 5: Liste des séries temporelles

Algorithme de Programmation génétique utilisé dans la première contribution:

```

1  /*-----
2
3   This is a standard GP implementation on EASEA,
4   aimed for regression.
5   It implements an interval arithmetic check.
6
7   use : easea -gp interval.ez
8   make
9
10  OR (if you'd like to exploit a CUDA-compatible graphic card)
11
12  easea -cuda_gp interval.ez
13      make
14
15  in order to generate and compile this code.
16  -----*/
17
18  \User declarations :
19  #include <string.h>
20  #include <stdlib.h>
21
22  #include "interface/CcsvParser.h"
23  #define csvFile "bac_admis.csv"
24  std::ifstream fs( csvFile );
25  CcsvParser parser( fs, ';' );
26
27
28
29  // these 3 defines are mandatory here. Adjust as you like.
30  #define NO_FITNESS_CASES 7
31  #define VAR_LEN 1
32  #define GROW_FULL_RATIO 0.5
33
34  // this is the number of learning cases computed in parallel.
35  // note that on 1024 is the maximum size on fermi architectures
36  512 on older cards.
37  #define NUMTHREAD 1024
38  #define MAX_STACK 15
39
40  #define PI (3.141592653589793)
41
42  float input_intervals[] = {2000,2100};
43  \end
44
45  \User functions:

```

```

46 //define POLY(x) x*x*x-3*x*x+x
47 #define POLY(x) 1/x
48 /**
49  This function generates data NO_FITNESS_CASES fitness cases,
50  from the polynome POLY(X) with X randomly picked between (-1,1)
51
52  @inputs address of the inputs array. (array will be allocated here)
53  @outputs address of the outputs array. (array will be allocated here
54  )
55
56  @ret number of loaded fitness cases (should be equal to
57  NO_FITNESS_CASES).
58 */
59
60 int generateData(float*** inputs, float** outputs){
61 /* int i=0;
62
63  (*inputs) = new float*[NO_FITNESS_CASES];
64  (*outputs) = new float[NO_FITNESS_CASES];
65
66  for( i=0 ; i<NO_FITNESS_CASES ; i++ ){
67  (*inputs)[i]=new float[VAR_LEN];
68  float x = random(-10.,+10.);
69  (*inputs)[i][0] = x;
70  (*outputs)[i] = POLY(x);
71  }
72
73  return NO_FITNESS_CASES;
74 */
75
76 CcsvData csvData;
77
78 (*inputs) = new float*[NO_FITNESS_CASES];
79 (*outputs) = new float[NO_FITNESS_CASES];
80
81 for (unsigned int i = 0; i < NO_FITNESS_CASES; i++){
82  (*inputs)[i]=new float[VAR_LEN];
83  for (unsigned int j = 0; j < VAR_LEN; j++){
84  (*inputs)[i][j] = parser.getCell<float>(i,j);
85  printf(" %f ", (*inputs)[i][j]);
86  }
87  (*outputs)[i] = parser.getCell<float>(i,VAR_LEN);
88  printf(" %f \n", (*outputs)[i]);
89  }
90 return NO_FITNESS_CASES;
91 }
92
93 void free_data(){
94 for( int i=0 ; i<NO_FITNESS_CASES ;i++ ) delete [] inputs[i] ;
95

```

```
93     delete[] outputs;
94     delete[] inputs;
95 }
96
97 /* INTERVAL ARITHMETIC */
98
99 struct interval_t {
100     double low;
101     double up;
102     bool invalid;
103 };
104
105 void get_interval(GPNode *n, interval_t *inter) {
106     const unsigned arity = opAry[(int)n->opCode];
107     inter->invalid = false;
108
109     if (arity == 0) {
110         if (n->opCode == OP_ERC) {
111             inter->low = inter->up = n->erc_value;
112         } else {
113             inter->low = input_intervals[2*n->opCode];
114             inter->up = input_intervals[2*n->opCode+1];
115         }
116     } else if (arity == 2) {
117         interval_t int1, int2;
118         get_interval(n->children[0], &int1);
119         get_interval(n->children[1], &int2);
120         double ab, ad, cb, cd;
121
122         switch (n->opCode) {
123             case OP_ADD:
124                 inter->low = int1.low + int2.low;
125                 inter->up = int1.up + int2.up;
126                 break;
127             case OP_SUB:
128                 inter->low = int1.low - int2.low;
129                 inter->up = int1.up - int2.up;
130                 break;
131             case OP_MUL:
132                 ab = int1.low*int1.up;
133                 ad = int1.low*int2.up;
134                 cb = int2.low*int1.up;
135                 cd = int2.low*int2.up;
136
137                 inter->low = min(ab, min(ad, min(cb, cd)));
138                 inter->up = max(ab, max(ad, max(cb, cd)));
139                 break;
140             case OP_DIV:
141                 if (int2.low <= 0 && int2.up >= 0) {
```

```

142     inter->invalid = true;
143     break;
144 }
145 ab = int1.low/int1.up;
146 ad = int1.low/int2.up;
147 cb = int2.low/int1.up;
148 cd = int2.low/int2.low;
149
150     inter->low = min(ab, min(ad, min(cb, cd)));
151     inter->up = max(ab, max(ad, max(cb, cd)));
152     break;
153 }
154 }
155 return;
156 }
157
158 void check_and_repair(GPNode *n) {
159     const unsigned arity = opArity[(int)n->opCode];
160
161     if (arity == 0)
162         return;
163     if (arity >=1)
164         check_and_repair(n->children[0]);
165     if (arity >= 2)
166         check_and_repair(n->children[1]);
167
168     interval_t inter;
169     get_interval(n, &inter);
170     if (inter.invalid) {
171         OPCODE alt[3] = {OP_ADD, OP_SUB, OP_MUL};
172         n->opCode = alt[random(0,2)];
173     }
174 }
175 \end
176
177
178
179 \Before everything else function:
180 {
181     generateData(&inputs,&outputs);
182 }
183 \end
184
185 \After everything else function:
186 {
187     std::cout << toString(((IndividualImpl*)EA->population->Best)->root)
188     << std::endl;
189
190     free_data();

```

```
191 }
192 \end
193
194 \At the beginning of each generation function:
195 {
196     //cout << "At the beginning of each generation function called" <<
197     endl;
198 }
199 \end
200 \At the end of each generation function:
201 {
202     //cout << "At the end of each generation function called" << endl;
203 }
204 \end
205
206 \At each generation before reduce function:
207 //cout << "At each generation before replacement function called" <<
208 endl;
209 \end
210
211 \User classes :
212
213 GenomeClass {
214     GPNode* root;
215 }
216 \end
217
218 \GenomeClass::display:
219
220 \end
221
222 \GenomeClass::initialiser :
223 {
224     Genome.root = ramped_hh();
225 }
226 \end
227
228 \GenomeClass::crossover :
229 {
230     simpleCrossOver(parent1, parent2, child);
231     child.valid = false;
232 }
233 \end
234
235 \GenomeClass::mutator : // Must return the number of mutations
236 {
237     simple_mutator(&Genome);
```

```

238     //return 1;
239 }
240 }
241 \end
242
243 \begin operator description :
244 OP_X, "x", 0, {RESULT=INPUT[0];};
245 OP_ERC, "ERC", 0, {RESULT=ERC;};
246 OP_ADD, "+", 2, {RESULT=OP1+OP2;};
247 OP_SUB, "-", 2, {RESULT=OP1-OP2;};
248 OP_MUL, "*", 2, {RESULT=OP1*OP2;};
249 OP_SIN, "sin", 1, {RESULT=sinf(OP1);};
250 OP_COS, "cos", 1, {RESULT=cosf(OP1);};
251 OP_SQRT, "sqrt", 1, {RESULT=sqrt(OP1);};
252 OP_DIV, "/", 2, {
253     if( !OP2 ) RESULT = 1;
254     else RESULT = OP1/OP2;
255 };
256
257
258
259 \end
260
261
262 \GenomeClass::evaluator header:
263 check_and_repair(this->root);
264 \end
265
266 \GenomeClass::evaluator for each fc :
267 float expected_value = OUTPUT;
268 ERROR = (expected_value-EVOLVED_VALUE)*(expected_value-EVOLVED_VALUE);
269 \end
270
271 \GenomeClass::evaluator accumulator :
272 return sqrtf(ERROR);
273 \end
274
275
276 \User Makefile options:
277
278 CXXFLAGS+=-I/usr/local/cuda/common/inc/ -I/usr/local/cuda/include/
279 LDFLAGS+=
280 \end
281
282 \Default run parameters :           // Please let the parameters appear
283 // in this order
284 Number of generations : 50          // NB_GEN
285 Time limit: 0                       // In seconds, 0 to deactivate
286 Population size : 5000               //POP_SIZE

```



```

287 Offspring size : 5000 // 40%
288 Mutation probability : 0.1 // MUT_PROB
289 Crossover probability : 0.9 // XOVER_PROB
290 Evaluator goal : minimise // Maximise
291 Selection operator: Tournament 7
292 Surviving parents: 100%//percentage or absolute
293 Surviving offspring: 100%
294 Reduce parents operator: Tournament 2
295 Reduce offspring operator: Tournament 2
296 Final reduce operator: Tournament 7
297
298 Elitism: Strong //Weak or Strong
299 Elite: 1
300 Print stats: true //Default: 1
301 Generate csv stats file:false
302 Generate gnuplot script:false
303 Generate R script:false
304 Plot stats:true //Default: 0
305
306 //Remote island model: true
307 //IP file: ip.txt //File containing all the remote island's IP
308 //Server port : 2929
309 //Migration probability: 0.33
310
311 Save population: false
312 Start from file:false
313
314 max init tree depth : 4
315 min init tree depth : 2
316
317 max tree depth : 8
318
319 size of prog buffer : 200000000
320 \end

```

Algorithme du Filtre de Kalman utilisé dans la première contribution:

Le code simple en c++:

```

1 #include <cstdint>
2 #pragma once
3
4 class KalmanFilter {
5 private:
6     float R;
7     float Q;
8     float A;
9     float C;

```

```
10 float B;
11 float covariance;
12 float x;
13 public:
14 KalmanFilter(float R, float Q, float A, float B, float C);
15 float filter(float z, float u);
16 };
17
18 inline KalmanFilter::KalmanFilter(float R, float Q, float A, float B,
19 float C)
19 {
20 this->R = R;
21 this->Q = Q;
22 this->A = A;
23 this->C = C;
24 this->B = B;
25 this->covariance = NULL;
26 this->x = NULL;
27 }
28
29 inline float KalmanFilter::filter(float z, float u)
30 {
31 if (this->x == NULL) {
32 this->x = (1 / this->C) * z;
33 this->covariance = (1 / this->C) * this->Q * (1 / this->C);
34 }
35 else {
36 float predictedX = (this->A * this->x) + (this->B * u);
37 float predictedCovariance = ((this->A * this->covariance) * this->A)
38 + this->R;
39 float K = predictedCovariance * this->C * (1 / ((this->C *
40 predictedCovariance * this->C) + this->Q));
41 this->x = predictedX + K * (z - (this->C * predictedX));
42 this->covariance = predictedCovariance - (K * this->C *
43 predictedCovariance);
44 }
45 return this->x;
46 }
```


Résumé (1000 caractères maximum) et mots clés en français

Résumé : Le dopage à l'azote de carbone activés microporeux montrèrent des performances électrochimiques en tant qu'électrodes pour supercondensateurs impressionnantes s'élevant à plus de 320 F/g en milieu acide. Cette étude mis en lumière l'atout que constitue le dopage par groupements fonctionnels azotés (ex : pyrrole) pouvant engendrer un stockage de charge additionnel par phénomène de pseudocapacitance. L'utilisation de modèle géométrique simplifié pour la modélisation des données d'impédance montra son utilité pour analyser en profondeur les propriétés d'électrodes carbonés poreux et leurs capacitances mesurées. Différentes observations obtenues par caractérisation in situ par spectroscopie infrarouge à transformée de Fourier par réflectance diffuse à température contrôlée (TD DRIFTS) décrivent l'importance de la complexation de l'ion manganèse(II) dans la préparation de départ. L'ajout de citrate d'ammonium, un fort agent complexant, permet une meilleure uniformité et stabilité tout le long de l'évaporation de solvant, empêchant la formation d'hydroxyde et/ou d'autres sels/complexes métalliques. Des phases pures peuvent être produits, notamment la spinelle Mn_3O_4 mésoporeux qui montrèrent des capacitances à plus de 100 F/g et $370 \mu F/cm^2$ dans un milieu basique aqueux. L'ingénierie de structure de spinelle de fer Fe_3O_4 fut permise par la méthode de synthèse polyol en nanostructure de type « framboise » creuse ainsi que de composites à base de graphène à quelques couches (FLG). Pour les supercondensateurs, un composite en teneur modérée en Fe_3O_4 mais non moins faible présenta des capacitances semblables à l'oxyde Mn_3O_4 avec plus de $350 \mu F/cm^2$ atteignables. En tant que matériaux d'électrodes de type réaction de conversion, de très haute capacité à hauteur de 923 mAh/g furent atteintes avec une rétention de cycles améliorée par l'apport de FLG, servant de support électroniquement conducteur et empêchant l'isolation des particules métalliques de fer, améliorant la réversibilité de la réaction de conversion.

Mots clés : Supercondensateur – batteries Li-ion – carbone activé N-dopé – oxydes de manganèse et

Résumé (1000 caractères maximum) et mots clés en anglais

Abstract : Nitrogen doping of microporous activated carbon showed impressive electrochemical performances as electrodes for supercapacitors, reaching over 320 F/g in acidic media. This study highlighted the advantage of doping with nitrogen-containing functional groups (e.g. pyrrole), which can generate additional charge storage through pseudocapacitance. The use of a simplified geometric model to fit impedance data demonstrated its usefulness for in-depth analysis of the properties of porous carbon electrodes and their measured capacitances. Various observations obtained by in situ characterization using temperature-controlled diffuse reflectance Fourier transform infrared spectroscopy (TD-DRIFTS) describe the importance of manganese(II) ion complexation in the starting preparation. The addition of ammonium citrate, a strong complexing agent, provides greater uniformity and stability throughout solvent evaporation, preventing the formation of hydroxide and/or other metal salts/complexes. Pure phases can be produced, notably mesoporous Mn_3O_4 spinel, which showed capacitances in excess of 100 F/g and $370 \mu F/cm^2$ in an aqueous basic medium. The engineering of Fe_3O_4 iron spinel structures was enabled by the polyol synthesis method in hollow "raspberry"-shaped nanostructures, as well as few-layer graphene composites (FLG). For supercapacitors, a composite with a moderate but low Fe_3O_4 content showed capacitances similar to Mn_3O_4 oxide, with over $350 \mu F/cm^2$ achievable. As conversion reaction electrode materials, very high capacities of up to 923 mAh/g were achieved, with cycle retention enhanced by the addition of FLG, serving as an electronically conductive support and preventing the isolation of iron metal particles, improving the reversibility of the conversion reaction.

Keywords : Supercapacitor - Li-ion batteries – N-doped activated carbon - manganese and iron oxides