

Préface

Branche de la philosophie, l'éthique détermine les règles auxquelles nous devrions tous soumettre nos comportements. Que faire ? Comment se conduire en société ? Quelles maximes adopter pour s'orienter dans le monde ? Qu'entend-on par bien agir ? L'éthique aide – ou tout au moins le prétend-elle – à répondre à ces questions. Originellement, elle reposait sur les leçons tirées d'habitudes de vie ou de préceptes anciens ou encore sur la volonté de perfection des individus, leur aspiration à la vertu. À partir du siècle des Lumières, on voulut trouver de grands principes généraux, à l'égal de ceux en vigueur dans les sciences physiques, auxquels toute recommandation se ramènerait. Ainsi en va-t-il du « déontisme », qui commande de respecter un ensemble de règles admissibles par tous, ou de l'« utilitarisme », qui propose de maximiser une mesure d'utilité. Dans le courant du XX^e siècle, ces principes entrèrent en crise : la complexité des enjeux faisait que l'individu ne savait plus évaluer ni les conséquences, ni le sens même de ses actes et, par suite, ne parvenait plus à s'orienter. Les philosophes engagèrent alors de multiples réflexions pour sauver ce qu'ils pouvaient de l'éthique. Ainsi, en alla-t-il des travaux de Jurgen Habermas qui portèrent sur les conditions d'une délibération sincère et désintéressée, en vue de s'accorder sur des normes de comportement acceptables par tous, ou des réflexions de Hans Jonas sur l'« heuristique de la peur » et la responsabilité vis-à-vis des générations futures.

À l'heure du numérique, des réseaux sociaux et de la publicité ciblée, les habitudes de vie et les échanges interhumains se trouvent bousculés. Les liens qui tissent la trame du tissu social, l'amitié, la confiance, la réputation, se réécrivent. Et, la globalisation exige de se référer à des valeurs partagées à l'échelle planétaire. Dans ce contexte, il apparaît plus délicat encore qu'autrefois de s'accorder sur une morale commune. Néanmoins, même s'il est toujours difficile, aujourd'hui comme hier, d'établir le fondement des idées du bon et du juste à partir desquelles nous décidons de régler

nos comportements, nous restons libres et donc responsables de nos agissements. À moins d'être dépossédés de nos moyens d'agir ou de perdre l'usage de la raison, nous assumons toujours les conséquences de nos actes.

Or, avec les dispositifs programmés au moyen de techniques d'intelligence artificielle, en particulier au moyen d'algorithmes d'apprentissage machine entraînés sur de grandes masses de données, nous sommes désormais confrontés à des situations inédites : des automates semblent décider à notre place, ou tout au moins, nous soufflent-ils les décisions à prendre, sans que nous soyons toujours en mesure de comprendre ce qui les motive. On qualifie parfois ces dispositifs d'« agents », voire d'« agents autonomes », pour signifier qu'ils agissent par eux-mêmes. Toutefois, cette dénomination est abusive, et ce pour deux raisons. D'un côté, ce ne sont pas des agents en ce qu'ils ne sont pas à l'initiative de l'action, même s'ils contribuent à des prises de décision automatiques, c'est-à-dire sans intervention humaine entre l'acquisition d'information et l'action. D'un autre côté, ils ne sont pas autonomes, au sens propre, car ils n'ont pas de volonté. Il arrive que l'on s'efforce de préserver une place à l'humain dans la prise de décision, mais dans cette éventualité, cette présence de l'humain est trop souvent de pure forme. C'est tout particulièrement le cas lorsque le verdict de la machine demeure aussi opaque qu'un oracle. On a alors d'autres choix que de s'y soumettre, ou de le rejeter en bloc. Il n'en demeure pas moins que nous agissons de plus en plus par le truchement de tels dispositifs, soit par délégation, soit sous leur empire. Dans ce contexte, nous ne sommes plus toujours en mesure de répondre de nos actes.

Dès lors, pour assumer nos responsabilités, il convient de prendre des mesures en amont, lors de la conception, de la réalisation ou du suivi des machines, par exemple de vérifier qu'elles préservent la liberté de choix des individus et protègent la vie privée en ne laissant pas fuiter des données personnelles, qu'elles s'affranchissent du poids des préjugés, ou de ce que l'on appelle aujourd'hui des biais, qu'elles indiquent les éléments d'informations qui conduisent à la décision dans chaque cas particulier, afin de permettre aux opérateurs humains d'exercer leur jugement, sans se soumettre au dictat de la machine, etc. Lesdits « agents autonomes » soumis à toutes ces exigences ne sauraient, à proprement parler, être qualifiés de moraux ou d'éthiques. Pour reprendre la vieille distinction kantienne, ils n'agissent aucunement « par devoir », mais « selon les règles du devoir » que nous leur imposons.

En revanche, s'ils sont bien conçus, ils nous aideront à nous comporter nous-mêmes en tant qu'agents moraux, en ce que les différentes décisions que nous prendrons par leur entremise se conformeront aux prescriptions morales que nous nous imposerons. L'enjeu apparaît donc essentiel. Et, cet ouvrage porte justement sur ce point : comment

faire, en pratique, pour concevoir de tels dispositifs ? Et, quels critères éthiques devons-nous retenir dans leur conception ? On songe à la conformité aux lois et aux règlements, aux biais que nous avons évoqués, à la résolution des conflits de normes et aux transformations que cela induit tant dans les métiers que dans la vie personnelle.

Jean-Gabriel GANASCIA
Professeur Sorbonne Université

Enjeux et enquêtes sur l'éthique et la transition numérique

L'éthique, définie dans l'Antiquité comme « principes moraux » (Aristote 2019 ; Signer 1986) dictant les vertus du comportement (Hursthouse 1999), s'identifie de plus en plus à des principes de déontologie et de règles sociales liées aux conséquences des actions (Bonhoeffer 2012 ; Siau *et al.* 2020). La dimension de l'évaluation de l'activité et des systèmes est devenue alors importante afin de répondre à ces principes. Par ailleurs, la transition numérique de la société tend à exploiter des systèmes émanant de l'intelligence artificielle (IA) et de traitement de données. Or, les techniques d'intelligence artificielle ont tendance à simuler le comportement et surtout à reproduire la pensée et les actions. Les conséquences de ces actions doivent ensuite être évaluées en tenant compte des règles sociales et de la déontologie. En fait, les techniques d'IA sont principalement basées sur l'échantillonnage et l'analyse de données d'un côté et les règles et procédures cognitives d'un autre côté. Les résultats de ces approches modifient notre comportement au quotidien en introduisant de nouveaux éléments générés par les connexions du traitement massif des connaissances et des algorithmes. *Deep et machine learning*, ChatGPT¹ sont parmi les principaux exemples de cette invasion dans notre activité.

Introduction rédigée par Nada MATTÀ, Marie-Hélène ABEL, Hedi KARRAY* et Inès SAAD. (* *The views expressed in this publication (book) are the responsibility of the authors and do not necessarily reflect the views of the European Commission nor of the European Innovation Council and SMEs Executive Agency. The European Commission or the European Innovation Council and SMEs Executive Agency are not liable for any consequence stemming from the reuse of this publication (book).*)

1. Voir : https://openai.com/blog/chatgpt?utm_source=bdmtools&utm_medium=siteweb&utm_campaign=chatgpt.

La principale question débattue dans cet ouvrage peut être : « Quels sont les différents aspects à considérer pour évaluer les principes de l'éthique dans les approches de la transition numérique et notamment les systèmes intelligents ? » Afin de répondre à cette question, nous procédons d'abord à expliquer les principales technologies des techniques et d'usage des systèmes intelligents. Nous explorons ensuite les travaux abordant l'éthique dans la transition numérique afin de mettre en avant les dimensions d'éthique à prendre en compte dans le développement de ces systèmes, extraites à partir d'une enquête réalisée auprès des chercheurs en numérique. Enfin, nous introduisons un résumé des huit chapitres de cet ouvrage répondant à ces questionnements.

Ces investigations sont réalisées dans le cadre de l'activité du chapitre français IEEE SMC² où plusieurs actions sont menées afin d'étudier la relation entre le numérique et l'activité humaine.

1.1. La transition numérique et ses enjeux

Les technologies de l'information offrent de plus en plus des approches de traitement permettant une compréhension de l'écosystème socioéconomique interne et externe. Ces techniques permettent d'une part de capturer et d'exploiter les données et les informations produites par une activité et/ou existantes dans l'environnement et d'autre part, de fournir des outils supports d'aide à la décision. Les acteurs socio-économiques sont alors amenés à appréhender ces technologies et à les intégrer dans leurs organisations. La transition numérique est définie comme l'intégration des technologies de traitement de l'information, tout en véhiculant un changement profond des habitudes, afin de permettre une compréhension de l'écosystème, amenant par là, une meilleure performance de l'organisation (Hess *et al.* 2016 ; Zacklad 2020). Nous pouvons citer, comme exemple, l'exploitation massive des outils supports au télétravail (Zoom[®], Microsoft Teams[®], Webex[®]), notamment lors de la crise sanitaire de la Covid-2019. De même, la compréhension de l'environnement permet actuellement une prise de conscience sociétale, poussant vers des actions soutenables. L'avancée des technologies de traitement de l'information apporte des changements radicaux dans les activités et les comportements, notamment au niveau de la communication et de la prise de décision (figure I.1) (Zacklad 2020 ; Vial 2021). Ces technologies se basent en grande partie sur des approches de l'IA qui ont fait leurs preuves comme support à la compréhension de l'écosystème et à la prise de décision.

2. Voir : <https://r8.ieee.org/france-smc/>.

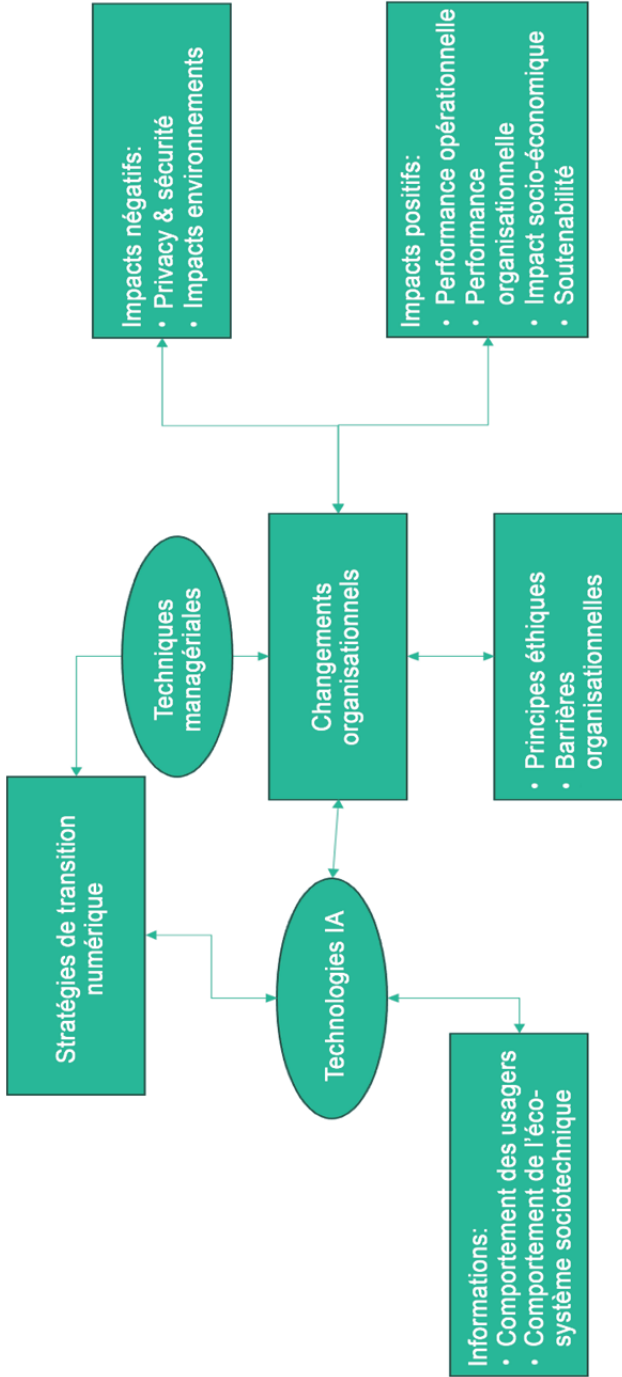


Figure I.1. La transition numérique des organisations (Vial, 2021)

1.1.1. Les techniques de l'intelligence artificielle

Les principes de base des approches d'intelligence artificielle (IA) sont de représenter le raisonnement et le comportement du vivant dans les techniques de calcul (Fetzer *et al.* 1990 ; Dick 2019). Par exemple, les systèmes fondés sur des règles tendent à illustrer principalement la déduction, le raisonnement à base de cas représente des inférences par analogie, les algorithmes d'apprentissage automatique tendent à simuler l'induction. Nous citons également les systèmes multi-agents qui représentent la coopération dans les communautés des abeilles et des fourmis. Cette similitude de comportements du vivant conduit au-delà des aides à la prise de décision et des assistances à de réelles collaborations entre l'humain et les algorithmes d'IA.

Certaines approches de l'IA utilisent les dimensions cognitives et proposent un raisonnement logique fondé sur les connaissances de retour d'expérience (les ontologies, les systèmes à base de règles et des cas). D'autres techniques utilisent le traitement statistique des données, à partir de lacs de données *Data Lake* pour agréger des caractéristiques et générer des règles et des modèles de raisonnement (les réseaux neuronaux, l'apprentissage profond) (Hunt 2014). Ces deux approches s'interconnectent selon des apprentissages supervisés, nommées actuellement IA hybride (figure I.2).

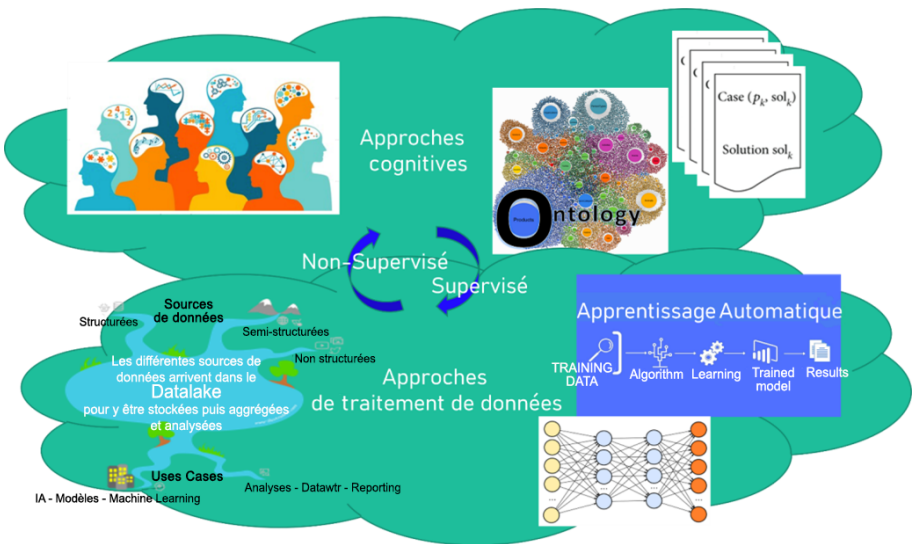


Figure I.2. Principales techniques de l'intelligence artificielle

Des questionnements importants peuvent se poser sur l'influence et la pertinence de ces techniques pour la compréhension de l'écosystème :

- l'expertise et les données exploitées sont-elles aussi exhaustives pour suggérer des modèles de raisonnement et de prise de décision efficace ? Est-ce que ces données sont suffisamment globales pour représenter des situations réelles ?
- ces modèles peuvent-ils être aussi complets afin de prétendre représenter différents aspects du raisonnement du vivant ?
- ces approches peuvent-elles reconnaître et éviter les données erronées ainsi que des expériences incomplètes ?

1.1.2. Les principales applications de l'IA

Depuis 1955 (la naissance de la notion d'IA et le test de Turing) jusqu'à nos jours, l'application de l'IA augmente exponentiellement surtout avec l'augmentation des capacités de calcul des machines. En termes d'applications, au début, l'IA a été utilisée principalement dans la santé et l'industrie en tant que systèmes à base de connaissances et de cas (Dick 2019). D'autres techniques, comme la logique floue et les réseaux de neurones, sont utilisées dans la reconnaissance de forme comme le traitement d'image et de la parole (Hunt 2014) et la robotique. De même, les systèmes multi-agents sont utilisés dans les réseaux et l'informatique des nuages Cloud Computing.

Nous pouvons citer des applications de ces systèmes dans certains domaines (figure I.3) :

- traitement du langage naturel : la traduction, la recherche d'information, la génération de texte exploitées dans des activités médicales, industrielles, marketing, et juridiques ;
- traitement des images : la supervision, la reconnaissance culturelle et archéologique, la reconnaissance faciale, le diagnostic médical, les changements climatiques, la réalité augmentée, les jumeaux numériques, etc. ;
- traitement numérique des données : la prédiction de problèmes, la maintenance, la prédiction de comportements, la supervision, les relations clients-marché, la recommandation, etc. ;
- Web sémantique : le Web de données, la recherche d'informations, la génération de texte, chatbot, les réseaux sociaux et les entraides, la formation à distance, etc.

Ces outils imposent une transformation du comportement des utilisateurs et un changement organisationnel important, notamment en considérant l'influence de ces

approches dans le raisonnement des acteurs sociotechniques, en considérant spécialement la génération des caractérisations de données et la prédiction du comportement de l'écosystème. Ces outils ne représentent pas seulement les comportements du vivant, comme déjà mentionné dans leur définition, mais soulignent les influences mutuelles avec l'environnement. Nous pouvons même mentionner la possibilité d'une auto-évolution des systèmes d'IA, à l'instar des raisonnements qu'ils représentent. L'appréhension de ces techniques pose, de plus en plus, des questionnements éthiques indispensables à l'intégration de ces approches dans l'environnement socio-économique.

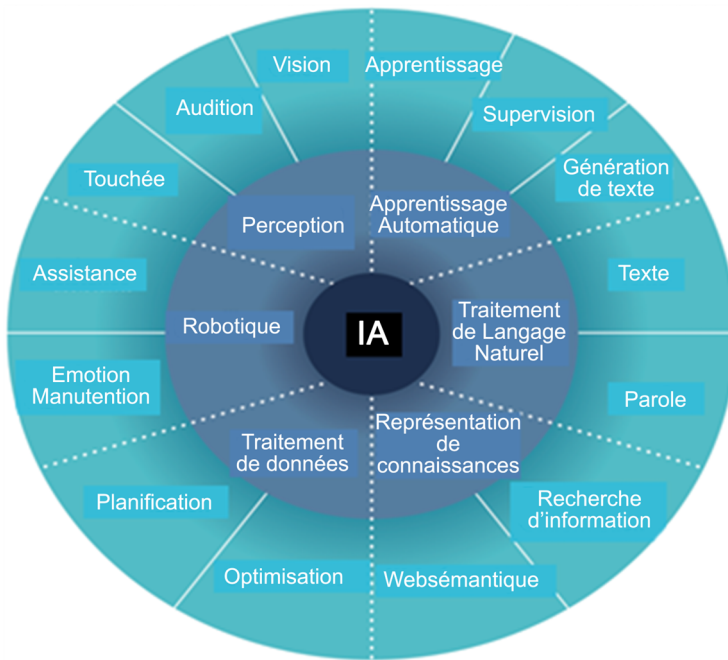


Figure I.3. Exemples de domaines d'application de l'IA

I.2. Principes de l'éthique

L'éthique, par sa définition, est reliée à la moralité, aux vertus du comportement et aux règles sociales (Hursthouse 1999). La question d'application des théories de la moralité et des vertus comme principes de l'éthique (Siau *et al.* 2020) est abordée dans plusieurs sciences, notamment le juridique, la médecine, le commerce et l'ingénierie (Frey *et al.* 2018). Par conséquent, des principes ont ensuite été définis, où les conséquences de ces sciences dans la société sont principalement étudiées. Nous pouvons noter, que, par exemple, dans les sciences médicales, certains principes ont été

prescrits comme les objectifs communs, les devoirs fiduciaires, les normes et la responsabilité juridique et professionnelle ainsi que des méthodes de transformation de ces principes dans la pratique. De même, la notion d'éthique environnementale et le comportement durable de l'homme dans son écosystème sont étudiés dans les sciences pour l'ingénieur (Palmer *et al.* 2014). Dans cette science, les concepts d'éthique sont introduits en tant que responsabilité, autonomie, vertu, droit et statut moral (Powers *et al.* 2020).

1.3. L'éthique et l'intelligence artificielle

L'application des définitions éthiques théoriques dans la transition numérique mentionne principalement l'analyse de la nature et de l'impact social de cette transition et la justification de cet impact (Newell et Marabelli 2015 ; Majchrzak *et al.* 2016 ; Mittelstadt *et al.* 2016). Les entreprises sont invitées à gérer un compromis entre leurs performances et les principes de l'éthique (Vial 2021). Ces compromis devraient être aux niveaux opérationnel et stratégique de l'entreprise (Zacklad 2020 ; Vial 2021). L'application des principes de l'éthique dans la transition numérique est fortement liée à l'intégration des applications de traitement de données et des technologies de l'IA.

Actuellement, plusieurs sociétés savantes comme l'ACM³ et l'IEEE⁴ ont défini certains principes éthiques pour la transition numérique et l'IA. L'OCDE (OCDE 2019) et le Groupe d'experts d'intelligence artificielle (HLEG 2019) de la Commission européenne ont proposé quatre principes concernant l'autonomie humaine, la prévention des préjudices, l'équité et l'explicabilité. Ces principes sont issus de ceux définis dans certaines sciences et surtout dans les sciences médicales (Mittelstadt *et al.* 2016). Bendt Mittelstadt défend que ces principes ne puissent pas être appliqués directement dans la transition numérique. En effet, une confusion existe dans la définition de la responsabilité qui doit être partagée entre les multiples développeurs des technologies de l'IA et les comportements des usagers de ces technologies. Il reste donc des grands enjeux pour définir les normes et les règles éthiques spécifiques à l'application de ces technologies notamment afin d'assurer des développements et des usages durables. Il sera essentiel de développer des systèmes d'IA qui interagissent éthiquement, notamment avec l'humain et la société. Thilo Hagendorff présente dans son article (Hagendorff 2020) une étude de différentes directives sur l'éthique de l'IA. Les principes essentiels sont mis en évidence dans cette étude tels que la responsabilité pour l'IA explicable, l'équité, la vie privée et la discrimination liée au Data Mining, les préjugés et la robustesse dans l'apprentissage automatique.

3. Voir : <https://www.acm.org/>.

4. Voir : <https://www.ieee.org/>.

De même, les techniques d'IA peuvent influencer les recherches et les découvertes en sciences, surtout lorsqu'on utilise l'exploration des données et l'apprentissage automatique. Powers et Ganascia (Powers et Ganascia 2020) mentionnent « l'expansion des connaissances due à l'IA semble être une ressource d'étude épistémique, et en même temps, nous ne pouvons pas comprendre pleinement ce que nous obtenons réellement ». Les principes éthiques en tant que justice et moralité sont donc nécessaires pour faire face à cette exploration afin d'éviter la discrimination et les préjugés. L'autonomie et l'évolution dynamique des agents et des algorithmes d'IA ont tendance à simuler le comportement. L'équité, la robustesse et l'explicabilité doivent jouer un rôle important dans l'évolution de ces techniques afin de tenir compte des règles sociales, de la durabilité et des impacts environnementaux (Palmer *et al.* 2014).

Nous pouvons noter que les études sur l'application de l'éthique dans la transition numérique, et notamment les approches de l'IA, soulèvent encore un certain nombre de défis. Dans notre étude, nous avons réalisé une enquête sur ces enjeux auprès de la communauté des chercheurs français en IA. Quelques résultats de cette enquête sont explorés dans la section suivante.

I.4. Enquêtes dans le cadre de la recherche en France

Des enquêtes sur l'éthique dans l'IA ont été menées auprès des chercheurs français en IA et systèmes numériques, par le chapitre français de la communauté scientifique SMC de l'IEEE. Plusieurs questions ont été posées sur ces sujets à travers la liste de diffusion du groupe de recherche CNRS I3 « Modélisation et interopérabilité des entreprises et des systèmes d'information »⁵ :

- 1) selon vous, quels sont les aspects de l'éthique que les systèmes intelligents devraient-ils prendre en compte ?
- 2) comment modéliser ces aspects afin de les prendre en compte par ces systèmes intelligents ?
- 3) l'apprentissage automatique peut-il être considéré comme éthique et sous quelles contraintes ?
- 4) l'éthique peut être envisagée à différents phases : audit, acquisition de connaissances et identification des besoins, collecte de données, traitement de ces connaissances et/ou données, développement de règles et de systèmes d'aide à la décision, exploitation et usage de ces systèmes. Comment assurer son respect ?
- 5) l'identification d'un organisme tiers de confiance peut-elle être l'une des solutions à l'identification des responsabilités ? Quels organismes pourraient jouer ce rôle ?

5. Voir : <http://crinfo.univ-paris1.fr/ModESI/index.htm>.

6) selon vous, quelles disciplines devraient faire partie de la recherche en éthique dans la transition numérique ?

Un groupe de 26 chercheurs ont répondu à ces questions. Nous avons résumé leurs réponses dans ce qui suit.

1.4.1. Les aspects de l'éthique à considérer dans les systèmes numériques

Différents aspects de l'éthique ont été identifiés dans les systèmes intelligents. Certains de ces aspects sont liés à la définition philosophique de l'éthique comme la morale et la valeur, mais d'autres correspondent à des règles et des comportements sociaux. Nous notons par exemple la prudence, la loyauté et l'éthique des groupes. Les responsabilités sont également mentionnées en lien direct avec les lois et règles sociales (figure I.4).



Figure I.4. Les dimensions de l'éthique à considérer en IA

1.4.2. Comment considérer l'éthique dans les technologies numériques ?

En tant que sciences de l'ingénieur, les techniques de développement de systèmes numériques proposent principalement des approches de modélisation pour le développement de systèmes supports d'aide à la décision. Ainsi, plusieurs dimensions liées à l'explicabilité et la transparence de ces techniques sont mentionnées comme des guides dans la définition et le développement de ces techniques. Nous pouvons en citer principalement :

- pour les approches de modélisation :
 - règles de conception et de modélisation ;
 - expression des contraintes et des limites des modèles ;
 - recherche d'auteurs et d'institutions métadonnées ;
 - obtention de compromis normatifs et techniques ;
 - erreurs de réduction dues au biais et aux abstractions ;
- pour le développement de systèmes d'aide à la décision :
 - le partage des responsabilités entre les acteurs humains et les systèmes numériques ;
 - la décision finale sera prise par un humain ;
 - des spécification et explication des règles ;
- les acteurs qui doivent être impliqués :
 - des législateurs ;
 - les utilisateurs finaux ;
 - des chercheurs en sciences humaines et sociales ;
 - les concepteurs et développeurs de systèmes.

1.4.3. L'apprentissage automatique éthique

L'apprentissage automatique propose des algorithmes de prise de décision ou de prédiction et une représentation du monde en évolution basée sur la reconnaissance et l'agrégation automatiques des données. Le manque d'exhaustivité des sources de données et la représentativité des modèles d'apprentissage, la découverte et la recherche peuvent largement influencer le résultat, que ce soit la prédiction ou la prise

de décision. Une sensibilisation et des alertes ont été identifiées afin de réduire ces problèmes (figure I.5). En particulier, sur les biais, les données discriminées, le manque de ressources, et sur les utilisations non contrôlées. Des recommandations ont été formulées pour réduire ces dimensions, en particulier sur l'expressivité et l'explication de l'apprentissage automatique en utilisant des représentations sémantiques, afin de mettre en avant une transparence de ces traitements et une validation éthique.

Positif	Statuts sémiotiques de données
	Supports d'aide à la décision
	Outils au service des utilisateurs
	Usages éthiques
	Aspect équitable et ouvert (Open Source)
	Propriété d'équité, de loyauté, de transparence et d'explicabilité
	Validation par une expertise indépendante
	Validation par la population
Supervision par des principes	
Négatif	Biais
	Biais de discrimination
	Ressources imparfaites
	Subjectivité
	Usages non maîtrisés
	Non-prise en compte du contexte social
	Décision automatique

Figure I.5. Des recommandations pour un apprentissage automatique éthique

I.4.4. Comment assurer le respect des principes de l'éthique dans les systèmes intelligents ?

En réponse à cette question, deux dimensions principales ont été notées :

- les personnes impliquées dans le développement de ces systèmes devraient être intègres et montrer une complète transparence de leurs traitements. De même, il faudrait

y avoir une séparation de préoccupations et de finalités de manière à réduire les intérêts financiers sous-jacents à ces développements ;

– un collectif social, basé sur des débats démocratiques et des réflexions citoyennes, devrait être mis en place afin de valider les usages potentiels de ces systèmes notamment le respect de la vie privée, de la propriété intellectuelle, des certifications et des réglementations juridiques.

I.4.5. Organisations de confiance

Comme garantie du respect de ces connaissances et recommandations, les chercheurs proposent d'établir des organisations de confiance, notamment des experts juridiques, gouvernementaux et scientifiques (figure I.6).

CNIL, CSA Comités scientifiques et juridiques (par branche) Tiers de confiance Comités d'experts et de citoyens
États, Union européenne Tout le monde
Plateformes d'audibilité

Figure I.6. Organismes de confiance

I.4.6. Les disciplines à mobiliser

Tel que noté en figure I.6, les développeurs de systèmes numériques ne sont pas en mesure d'appliquer les principes éthiques. Ils ont besoin de collaborations avec des chercheurs d'autres disciplines pour introduire ces aspects dans leurs produits, surtout lorsque l'on considère l'impact important de ces produits dans la société et le comportement scientifique. La figure I.7 montre que le droit, les sciences sociales et la philosophie sont nécessaires pour identifier les règles et les principes d'évaluation en collaboration avec les sciences de l'ingénierie qui développent et utilisent principalement ces systèmes.

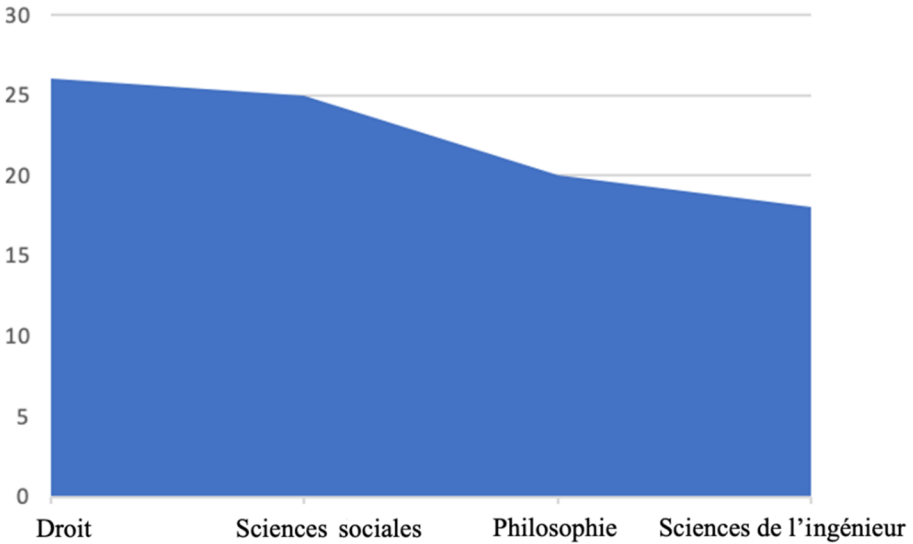


Figure I.7. Disciplines impliquées dans la validation éthique des systèmes numériques

I.5. Organisation de l'ouvrage

Cet ouvrage est organisé en 8 chapitres, explorant plusieurs questionnements de l'éthique dans la transition numérique. D'abord en chapitre 1, la dimension philosophique de l'éthique dans la technique est discutée par Bruno Bachimont, en mettant en avant la notion de non-neutralité éthique de la technique et la conception de l'éthique (*ethical design*). Florence Sèdes met en avant, dans le chapitre 2, les biais volontaires et involontaires dans les calculs et les approches numériques, support à la prise de décision, notamment les algorithmes d'apprentissage numérique répandus actuellement et influençant des stratégies de prise de décision. Les dimensions de transparence et de soutenabilité sont alors discutées à ce propos. La question d'usage des techniques numériques, notamment des approches de l'intelligence artificielle, est analysée dans le chapitre 3 par Alain Mille. Des propriétés sont alors énoncées comme guides aux concepteurs des outils numériques pour répondre aux notions éthiques. Au chapitre 4, Grégory Bonnet, Nadjet Bourdache, Abdel-Allah Mouaddib et Mihail Stojanovski discutent l'intégration de l'éthique dans la conception des systèmes autonomes où la difficulté de fournir à ces agents la capacité de jugement sur leurs propres actions est argumentée. L'utilisabilité des approches numériques dans l'industrie du futur est ensuite présentée dans le chapitre 5 par Emmanuel Caillaud et Lou Grimal, soulignant les enjeux de ces approches face aux défis de la soutenabilité

et de l'écologie des systèmes de production. Des principes sont alors définis comme guides aux acteurs de l'industrie pour relever ces défis. Damien Trentesaux, Lamia Berrah et Karine Samuel complètent, au chapitre 6, ces réflexions par une discussion sur les risques éthiques liés aux différents usages des techniques numériques en industrie. Le lien entre ces risques et la performance industrielle est mis en avant afin de proposer un cadre opérationnel d'intégration des notions d'éthique dans l'industrie du futur. La réglementation juridique liée à l'éthique est mise en exergue, dans le chapitre 7, par Nathalie Nevejans, expliquant d'une part, la volonté des organisations juridiques et gouvernementales de fournir un cadre pour la transition numérique et d'autre part, les préoccupations sociales sur les usages des approches d'intelligence artificielle. Anne Dourgnon, Eunika Mercier Laurent et Alain Antoine présentent, dans le chapitre 8, une application de techniques numériques au sein de l'industrie illustrant la transformation induite par la transition numérique dans les actions quotidiennes des acteurs dans plusieurs organisations comme l'EDF, FRAMATOME, SPIE Nucléaire et TECHNICATOME, Boost Conseil et ASSYSTEM. Les défis futurs de l'éthique dans la transition numérique sont discutés par Hedi Karray, en conclusion de cet ouvrage.

I.6. Remerciements

Cet ouvrage est le fruit des actions menées au sein du chapitre français IEEE SMC. Nous tenons particulièrement à remercier Jean-Paul Barthès, Jean-Paul Jamont, Mickaël Coustaty et François Rauscher pour leur soutien.

I.7. Bibliographie

- Aristotle, A. (2019). *The ethics of Aristotle*. BoD–Books on Demand.
- Boden, M.A. (1996). *Artificial intelligence*. Elsevier, Amsterdam.
- Bonhoeffer, D. (2012). *Ethics*. Simon and Schuster, New York.
- Dick, S. (2019). Artificial intelligence. *Harvard Data Science Review*, 1(1), 1–8.
- European Commission (2019). High Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Rapport, European Commission.
- Fetzer, J.H. (1990). *What is Artificial Intelligence?*. Springer Netherlands, Amsterdam.
- Frey, R.G., Wellman, C.H. (2008). *A companion to applied ethics*. John Wiley & Sons, Londres.

- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1), 99–120.
- Hesse, A. (2018). Digitalization and leadership – how experienced leaders interpret daily realities in a digital world. Dans *Hawaii International Conference on System Sciences*, Waikoloa Beach, 1854–1863.
- Hunt, E.B. (2014). *Artificial intelligence*. Academic Press, Cambridge.
- Hursthouse, R. (1999). *On virtue ethics*. OUP Oxford, Oxford.
- Mittelstadt, B.D. *et al.* (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2) 2053951716679679.
- Newell, S., Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of ‘datification’. *J. Strateg. Inf. Syst.*, 24(1), 3–14.
- OECD (2019). Forty-two countries adopt new OECD Principles on Artificial Intelligence - OECD. Forty-two countries adopt new OECD Principles on Artificial Intelligence [En ligne]. Disponible à l’adresse : <http://www.oecd.org/science/forty-two-countries-adopt-new-oecdprinciples-on-artificial-intelligence.htm>.
- Palmer, C., McShane, K., Sandler, R. (2014). Environmental ethics. *Annual Review of Environment and Resources*, 39, 419–442.
- Powers, T.M., Ganascia, J.G. (2020). *The Ethics of the Ethics of AI*. Oxford Handbook of Ethics of AI, Oxford.
- Siau, K., Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. *Journal of Database Management (JDM)*, 31(2), 74–87.
- Singer, P. (1986). *Applied ethics*. Oxford readings in philosophy, Oxford.
- Vial, G. (2021). Understanding digital transformation: A review and a research agenda. *Managing Digital Transformation*, 13–66.
- Zacklad, M. (2020). Les enjeux de la transition numérique et de l’innovation collaborative dans les mutations du travail et du management dans le secteur public. Dans *Les transformations du travail dans les services publics*, Gillet, A. (dir.). Presses de l’EHESP, Paris.