

LUCIANO FLORIDI

L'ÉTHIQUE DE L'INTELLIGENCE ARTIFICIELLE

Principes, défis et opportunités

Préface de Jean-Gabriel Ganascia

Édition française éditée et traduite par
Enrico Panaï et Emmanuel R. Goffi

SOMMAIRE

NOTE DES TRADUCTEURS	13
PRÉFACE	
<i>Jean-Gabriel Ganascia</i>	17
PRÉFACE DE L'AUTEUR	23
REMERCIEMENTS	35

PREMIÈRE PARTIE COMPRENDRE L'IA

1. LE PASSÉ : L'ÉMERGENCE DE L'IA	43
1.1 Introduction : la révolution numérique et l'IA	44
1.2 Le pouvoir clivant du numérique : couper-coller la modernité	48
1.3 Nouvelles formes d'agir	54
1.4 L'IA : un domaine de recherche en quête d'une définition	56
1.5 Conclusion : éthique, gouvernance et conception	58
2. LE PRÉSENT : L'IA, UNE NOUVELLE FORME D'AGENCE ET NON D'INTELLIGENCE	61
2.1 Introduction : qu'est-ce que l'IA ? « Je le sais quand je le vois »	62
2.2 L'IA comme contrefactuel	66
2.3 Les deux âmes de l'IA : l'ingénierie et le cognitif	70
2.4 IA : un divorce réussi dans l'infosphère	75
2.5 L'utilisation des humains et des interfaces	80
2.6 Conclusion : qui s'adaptera à qui ?	83
3. L'AVENIR : LE DÉVELOPPEMENT PRÉVISIBLE DE L'IA	87
3.1 Introduction : regarder dans les germes du temps	88
3.2 Données historiques, hybrides et synthétiques	89
3.3 Règles contraignantes et constitutives	96

3.4 Les problèmes difficiles, les problèmes complexes et la nécessité de l'enveloppe	99
3.5 L'avenir du design	106
3.6 Conclusion : l'IA et ses saisons	108

DEUXIÈME PARTIE ÉVALUER L'IA

4. UN CADRE UNIFIÉ DE PRINCIPES ÉTHIQUES POUR L'IA	115
4.1 Introduction : trop de principes ?	116
4.2 Un cadre unifié de cinq principes pour l'IA éthique	117
4.3 La bienfaisance : promouvoir le bien-être, préserver la dignité et assurer la pérennité de la planète	121
4.4 Non-malfaisance : vie privée, sécurité et « prudence en matière de capacité »	122
4.5 Autonomie : le pouvoir de « décider de décider »	123
4.6 Justice : promouvoir la prospérité, préserver la solidarité, éviter l'injustice.	124
4.7 Explicabilité : permettre les autres principes par l'intelligibilité et la responsabilité	125
4.8 Une vue synoptique	126
4.9 L'éthique de l'IA : d'où vient-elle et pour qui ?	127
4.10 Conclusion : des principes aux pratiques	128
5. DES PRINCIPES AUX PRATIQUES :	
LES RISQUES DU MANQUE D'ÉTHIQUE	131
5.1 Introduction : les traductions à risque	131
5.2 Le shopping éthique	133
5.3 Le « Bluwashing » éthique	135
5.4 Lobbying éthique	137
5.5 Le dumping éthique	139
5.6 L'esquive éthique	142
5.7 Conclusion : l'importance de mieux connaître	144
6. L'ÉTHIQUE DOUCE ET LA GOUVERNANCE DE L'IA	145
6.1 Introduction : de l'innovation numérique à la gouvernance du numérique	145
6.2 Éthique, réglementation et gouvernance	148
6.3 Conformité : nécessaire mais insuffisante	152

6.4	Éthique dure et douce	152
6.5	L'éthique douce comme cadre éthique	156
6.6	Analyse de l'impact éthique	161
6.7	La préférence numérique et la cascade normative	162
6.8	Le double avantage de l'éthique numérique	164
6.9	Conclusion : l'éthique comme stratégie	166
7.	CARTOGRAPHIER L'ÉTHIQUE DES ALGORITHME	169
7.1	Introduction : une définition pratique de l'algorithme	169
7.2	Carte de l'éthique des algorithmes	172
7.3	Preuves non concluantes conduisant à des actions injustifiées	175
7.4	Des preuves insondables conduisant à l'opacité	177
7.5	Des preuves erronées entraînant un biais indésirable	183
7.6	Résultats injustes conduisant à la discrimination	187
7.7	Effets transformateurs menant à des défis pour l'autonomie et la vie privée informationnelle	190
7.8	Traçabilité menant à la responsabilité morale	194
7.9	Conclusion : le bon et le mauvais usage des algorithmes	198
8.	MAUVAISES PRATIQUES :	
	L'UTILISATION ABUSIVE DE L'IA POUR LE MAL SOCIAL	201
8.1	Introduction : l'utilisation criminelle de l'IA	201
8.2	Préoccupations	205
8.2.1	Émergence	207
8.2.2	Responsabilité légale	208
8.2.3	Surveiller	211
8.2.4	Psychologie	212
8.3	Menaces	213
8.3.1	Commerce, marchés financiers et insolvabilité	213
8.3.2	Drogues nocives ou dangereuses	217
8.3.3	Infractions contre les personnes	218
8.3.4	Infractions sexuelles	223
8.3.5	Vol et fraude, falsification et personnalisation	225
8.4	Solutions possibles	228
8.4.1	Lutter contre l'émergence	229
8.4.2	Le traitement de la responsabilité	231
8.4.3	Vérification de la surveillance	234
8.4.4	Traiter de la psychologie	237
8.5	Développements futurs	238

8.5.1 Domaines du CpIA	238
8.5.2 Double usage	239
8.5.3 Sécurité	240
8.5.4 Personnes	241
8.5.5 Organisations	241
8.6 Conclusion : des mauvais usages de l'IA à l'IA socialement bonne	242
9. BONNES PRATIQUES :	
L'UTILISATION APPROPRIÉE DE L'IA POUR LE BIEN SOCIAL	245
9.1 Introduction : l'idée de l'IA pour le bien social	246
9.2 Une définition de l'IApBS	251
9.3 Sept facteurs essentiels pour une IApBS réussie	255
9.3.1 Falsifiabilité et déploiement incrémental	255
9.3.2 Garanties contre la manipulation des prédicteurs	259
9.3.3 Intervention adaptée au contexte du récepteur	261
9.3.4 Explication adaptée au contexte du récepteur et objectifs transparents	264
9.3.5 Protection de la vie privée et consentement de la personne concernée	270
9.3.6 L'équité situationnelle	273
9.3.7 Sémantisation adaptée aux humains	276
9.4 Conclusion : facteurs d'équilibre pour l'IApBS	278
10. COMMENT CRÉER UNE SOCIÉTÉ POUR UNE BONNE IA :	
QUELQUES RECOMMANDATIONS	283
10.1 Introduction : quatre façons de réaliser une société pour une bonne IA	283
10.2 Qui nous pouvons devenir : permettre l'épanouissement de l'être humain sans dévaloriser ses capacités	286
10.3 Ce que nous pouvons faire : renforcer l'action humaine sans supprimer la responsabilité humaine	287
10.4 Ce que nous pouvons réaliser : augmenter les capacités de la société sans réduire le contrôle humain.	288
10.5 Comment interagir : cultiver la cohésion sociétale sans éroder l'autodétermination humaine	289
10.6 Vingt recommandations pour une société pour une bonne IA	290
10.7 Conclusion : la nécessité de politiques concrètes et constructives	299

11. LE GAMBIT :	
L'IMPACT DE L'IA SUR LE CHANGEMENT CLIMATIQUE	301
11.1 Introduction : le pouvoir à double tranchant de l'IA	301
11.2 L'IA et les « transitions jumelles » de l'UE	306
11.3 IA et changement climatique : défis éthiques	308
11.4 IA et changement climatique :	
l'empreinte carbone numérique	310
11.5 Treize recommandations en faveur de l'IA contre le changement climatique	315
11.5.1 Promouvoir l'IA éthique dans la lutte contre le changement climatique	316
11.5.2 Mesurer et auditer l'empreinte carbone de l'IA : chercheurs et développeurs	317
11.5.3 Mesurer et contrôler l'empreinte carbone de l'IA : les décideurs politiques	318
11.6 Conclusion : une société plus durable et une biosphère plus saine	319
12. L'IA ET LES OBJECTIFS	
DE DÉVELOPPEMENT DURABLE DES NATIONS UNIES	321
12.1 Introduction : l'IApBS et les ODD de l'ONU	321
12.2 Évaluer les preuves de l'IAxODD	323
12.3 L'IA au service de mesures relatives à la lutte contre les changements climatique	328
12.4 Conclusion : un programme de recherche pour l'IAxODD	331
13. CONCLUSION : LE VERT ET LE BLEU	333
13.1 Introduction : du divorce entre l'agir et l'intelligence au mariage entre le vert et le bleu	333
13.2 Le rôle de la philosophie comme design conceptuelle	336
13.3 Retour aux « germes du temps »	338
13.4 La nécessité des cols verts	341
13.5 Conclusion : l'humanité comme une belle anomalie	343
RÉFÉRENCES	345