

Préface

L'intelligence artificielle (IA) est un domaine très vaste, et ce livre est lui-même volumineux. Nous avons tenté de couvrir la totalité de ce domaine, qui embrasse la logique, les probabilités et les mathématiques du continu, mais aussi la perception, le raisonnement, l'apprentissage et l'action, ainsi que l'équité, la confiance, le bien-être social et la sécurité, et une large gamme d'applications, allant des dispositifs microélectroniques jusqu'aux robots qui explorent des planètes, en passant par les services en ligne utilisés par des milliards d'utilisateurs.

Cet ouvrage est sous-titré « Une approche moderne », car nous considérons le domaine de l'IA d'un point de vue actuel. Nous présentons l'état des connaissances dans un cadre commun, en reformulant les travaux antérieurs à l'aide des idées et de la terminologie qui prévalent aujourd'hui. Nous nous excusons auprès des personnes dont les domaines spécifiques sont moins identifiables qu'ils ne l'auraient été avec une présentation plus conventionnelle.

Nouveautés de cette édition

Cette nouvelle édition reflète l'évolution du domaine de l'IA depuis la parution de la dernière édition en 2010 :

- ◆ Nous mettons davantage l'accent sur l'apprentissage automatique (*machine learning*) que sur l'ingénierie des connaissances créées manuellement, en raison de la disponibilité croissante de données, de ressources de calcul et d'algorithmes nouveaux.
- ◆ L'apprentissage profond (*deep learning*), les langages de programmation probabiliste et les systèmes multiagents sont traités plus longuement et font chacun l'objet d'un chapitre particulier.
- ◆ Les thèmes de la compréhension du langage naturel, de la robotique et de la vision par ordinateur ont été revus pour prendre en compte l'impact de l'apprentissage profond.
- ◆ Le chapitre sur la robotique traite désormais des robots en interaction avec les humains et des applications à la robotique de l'apprentissage par renforcement (*reinforcement learning*).
- ◆ Dans les éditions antérieures, nous définissions l'objectif de l'IA comme la création de systèmes tentant de maximiser l'utilité espérée, où l'information spécifique sur l'utilité (l'objectif à remplir) était fournie par les humains concepteurs du système. Nous ne supposons plus désormais que le but est fixé, ni connu du système d'IA ; au contraire, le système peut avoir à agir dans l'incertitude des véritables objectifs des humains pour le compte desquels il opère. Il doit apprendre ce qu'il faut maximiser et se comporter de manière adéquate, même s'il est incertain quant à l'objectif.
- ◆ L'impact de l'IA sur la société est plus longuement traité, y compris les problèmes fondamentaux d'éthique, d'équité, de confiance et de sécurité.
- ◆ Les exercices ne sont plus à la fin des chapitres, mais disponibles en ligne sur aima.cs.berkeley.edu, en anglais. Nous pourrions ainsi en proposer régulièrement de nouveaux, les mettre à jour et les améliorer, de façon à mieux servir les besoins des enseignants et refléter les progrès du domaine et des outils logiciels associés.
- ◆ Au total, 25 % du contenu de ce livre est complètement nouveau. Les 75 % restants ont été largement revus de manière à proposer une vision unifiée du domaine. 22 % des références de cette édition concernent des travaux postérieurs à 2010.

Présentation de l'ouvrage

La notion d'**agent intelligent** est le thème unificateur principal. Nous définissons l'IA comme l'étude des agents qui reçoivent des percepts de l'environnement et qui entreprennent des actions. Chaque agent de ce type implémente une fonction qui fait correspondre des séquences de percepts à des actions. Nous montrons les différentes manières de représenter ces fonctions sous la forme d'agents réactifs, de systèmes de planification en temps réel, de systèmes de prise de décision (rationnels) et de systèmes avec apprentissage profond. Nous présentons l'apprentissage à la fois comme une méthode de construction de systèmes compétents et comme une manière d'étendre la portée du concepteur à des environnements inconnus. La robotique et la vision sont traitées non pas comme deux problèmes distincts, mais comme des outils contribuant à des objectifs. Nous insistons sur l'importance de l'environnement des tâches dans la détermination de la conception d'agent appropriée.

Notre principal objectif est de transmettre les *idées* qui ont émergé au cours des soixante-dix dernières années de recherche en IA et des deux millénaires de travaux connexes qui les ont précédées. Nous avons essayé d'éviter toute formalisation excessive dans la présentation de ces idées tout en restant précis. Nous avons inclus des formules mathématiques et des algorithmes en pseudocode afin de rendre les idées clés plus concrètes; les concepts mathématiques et les notations sont décrits à l'annexe A, le pseudocode est décrit à l'annexe B.

Cet ouvrage a essentiellement été conçu pour servir de support dans le cadre d'une formation de licence. Il se compose de 28 chapitres, chacun correspondant à environ une semaine de cours; deux semestres sont donc nécessaires pour traiter l'ensemble des sujets abordés, même si un cours d'un semestre peut être envisagé sur la base de chapitres sélectionnés selon l'intérêt des étudiants ou de l'enseignant. Ce livre peut également être utilisé dans le cadre d'un master ou d'un doctorat (en ajoutant peut-être certaines des références majeures suggérées dans les notes bibliographiques), ou pour l'autoapprentissage, ou plus simplement comme ouvrage de référence.

La lecture de ce livre ne demande que des connaissances de base en informatique accessibles en premier cycle d'université (algorithmes, structures de données, complexité). Quelques bases en analyse et en algèbre linéaire sont souhaitables pour comprendre en détail certains sujets.

Ressources en ligne

Des ressources en ligne sont disponibles (en anglais) sur [pearsonhighered.com/cs-resources](https://www.pearsonhighered.com/cs-resources) ou sur le site web du livre, aima.cs.berkeley.edu. Vous y trouverez :

- ♦ des exercices, des projets de programmation et des projets de recherche. Ceux-ci ne se trouvent plus à la fin de chaque chapitre, mais uniquement en ligne. Dans cet ouvrage, nous faisons référence aux exercices en ligne de la manière suivante : « exercice 6.NARY »;
- ♦ les instructions vous permettant de repérer les exercices par nom ou par thème;
- ♦ les implémentations des algorithmes du livre en Python, Java et d'autres langages de programmation (actuellement hébergées sur github.com/aimacode);
- ♦ une liste de plus de 1 400 établissements ayant utilisé ce livre, proposant souvent des liens vers des supports de cours en ligne et des plans de cours;
- ♦ du matériel éducatif et des liens supplémentaires pour les étudiants et les instructeurs;
- ♦ des instructions sur la façon de signaler les erreurs présentes dans l'édition américaine, dans le cas probable où il en reste.

À propos de la couverture

La couverture représente la position finale de la sixième partie, décisive, de la rencontre de 1997 entre Garry Kasparov (jouant les noirs) et le programme DEEP BLUE, qui s'est soldée pour la première fois par la victoire de l'ordinateur sur l'homme au jeu d'échecs. Kasparov est représenté en haut. À sa droite se trouve une position critique de la seconde partie du match historique de go entre l'ancien champion du monde Lee Sedol et le programme ALPHAGo programme de go de la société DeepMind. Le 37^e coup joué par ALPHAGo était une

insulte à des siècles d'orthodoxie du jeu et a été jugé en direct par des experts humains comme une erreur grossière, avant de s'avérer gagnant. En haut à gauche, on peut voir le robot humanoïde Atlas construit par Boston Dynamics. L'image d'une voiture sans conducteur analysant son environnement se trouve entre Ada Lovelace, la première personne de l'histoire à avoir programmé une machine, et Alan Turing, dont le travail fondateur a défini l'intelligence artificielle. En bas de l'échiquier se trouvent le robot Mars Exploration Rover et une statue d'Aristote, le pionnier de la logique ; son algorithme de planification proposé dans *De Motu Animalium* apparaît derrière le nom des auteurs. Nous avons placé derrière l'échiquier un exemple de programme probabiliste utilisé par l'ONU dans le cadre du Traité d'interdiction complète des essais nucléaires, qui identifie les explosions nucléaires parmi les signaux sismiques.

Remerciements

Ce livre a mobilisé un village planétaire. Plus de 600 personnes y ont apporté leur expertise et ont suggéré des améliorations. La liste complète se trouve sur le site aima.cs.berkeley.edu/ack.html ; nous remercions chacune d'entre elles. Nous ne pouvons mentionner ici qu'une poignée de contributeurs, d'une importance particulière. D'abord, les coauteurs :

- ◆ Judea Pearl (section 13.5. Réseaux causaux) ;
- ◆ Vikash Mansinghka (section 15.4. Modèles probabilistes sous forme de programmes) ;
- ◆ Michael Wooldridge (chapitre 18. Prise de décision multiagent) ;
- ◆ Ian Goodfellow (chapitre 21. Apprentissage profond) ;
- ◆ Jacob Devlin et Mei-Wing Chang (chapitre 24. Apprentissage profond en traitement du langage naturel) ;
- ◆ Jitendra Malik et David Forsyth (chapitre 25. Vision par ordinateur) ;
- ◆ Anca Dragan (chapitre 26. Robotique).

Ensuite, les collaborateurs cruciaux :

- ◆ Cynthia Yeung et Malika Cantor (gestion du projet) ;
- ◆ Julie Sussman et Tom Galloway (conseils sur l'écriture et révision) ;
- ◆ Omari Stephens (illustrations) ;
- ◆ Tracy Johnson (éditrice) ;
- ◆ Erin Ault et Rose Kernan (couverture et passage à la couleur) ;
- ◆ Nalin Chhibber, Sam Goto, Raymond de Lacaze, Ravi Mohan, Ciaran O'Reilly, Amit Patel, Dragomir Radiv, et Samagra Sharma (développement de code et tutorat en ligne) ;
- ◆ Google Summer of Code students (développement de code).

Stuart aimerait remercier sa femme, Loy Sheflott, pour sa patience infinie et sa sagesse illimitée. Il espère que Gordon, Lucy, George et Isaac liront bientôt cet ouvrage après lui avoir pardonné d'y avoir consacré autant de temps. Le cercle des étudiants de Russell a apporté une aide inestimable, comme toujours.

Peter aimerait remercier ses parents, Torsten et Gerda, pour lui avoir montré le chemin, ainsi que sa femme, Kris, ses enfants, Bella et Juliet, et ses amis, pour leurs encouragements et la compréhension dont ils ont fait preuve malgré les longues heures consacrées à l'écriture et les heures plus nombreuses encore dévolues à la réécriture.

Table des matières

I Intelligence artificielle

1	Introduction	1
1.1	Qu'est-ce que l'IA ?	1
1.2	Fondements de l'intelligence artificielle	5
1.3	Histoire de l'intelligence artificielle	15
1.4	État de l'art	24
1.5	Risques et bénéfices de l'IA	28
	Résumé, notes bibliographiques et historiques	31
2	Agents intelligents	33
2.1	Agents et environnements	33
2.2	Bons comportements : le concept de rationalité	35
2.3	Nature des environnements	38
2.4	Structure des agents	43
	Résumé, notes bibliographiques et historiques	53

II Résolution de problèmes

3	Résolution de problèmes par exploration	57
3.1	Agents de résolution de problèmes	57
3.2	Exemples de problèmes	60
3.3	Algorithmes d'exploration	64
3.4	Stratégies d'exploration non informées	69
3.5	Stratégies d'exploration informées (heuristiques)	77
3.6	Fonctions heuristiques	88
	Résumé, notes bibliographiques et historiques	95
4	Exploration en environnements complexes	101
4.1	Exploration locale et problèmes d'optimisation	101
4.2	Exploration locale d'espaces continus	109
4.3	Exploration avec des actions non déterministes	111
4.4	Exploration en environnement partiellement observable	115
4.5	Agents d'exploration en ligne et environnements inconnus	123
	Résumé, notes bibliographiques et historiques	128

5	Exploration antagoniste et jeux	133
5.1	Théorie des jeux	133
5.2	Décisions optimales dans les jeux	134
5.3	Exploration alpha-bêta heuristique	141
5.4	Exploration d'arbre Monte-Carlo	146
5.5	Jeux stochastiques	149
5.6	Jeux partiellement observables	152
5.7	Limitations des algorithmes d'exploration pour le jeu	156
	Résumé, notes bibliographiques et historiques	158
6	Problèmes de satisfaction de contraintes	163
6.1	Définition des problèmes de satisfaction de contraintes	163
6.2	Propagation de contraintes : inférence dans les CSP	168
6.3	Exploration par <i>backtracking</i> pour les CSP	172
6.4	Exploration locale pour les CSP	178
6.5	Structure des problèmes	179
	Résumé, notes bibliographiques et historiques	183
 III Connaissances, raisonnement et planification		
7	Agents logiques	187
7.1	Agents fondés sur les connaissances	188
7.2	Le monde du wumpus	189
7.3	Logique	192
7.4	La logique propositionnelle : une logique très simple	195
7.5	Démonstration de théorèmes en logique propositionnelle	199
7.6	Vérification efficace de modèles en logique propositionnelle	208
7.7	Agents fondés sur la logique propositionnelle	213
	Résumé, notes bibliographiques et historiques	220
8	Logique du premier ordre	225
8.1	Retour sur la représentation	225
8.2	Syntaxe et sémantique de la logique du premier ordre	229
8.3	Utiliser la logique du premier ordre	238
8.4	Ingénierie des connaissances en logique du premier ordre	243
	Résumé, notes bibliographiques et historiques	248
9	Inférence en logique du premier ordre	251
9.1	Inférence propositionnelle <i>versus</i> inférence du premier ordre	251
9.2	Unification et inférence en premier ordre	253
9.3	Chaînage avant	257
9.4	Chaînage arrière	263
9.5	Résolution	268
	Résumé, notes bibliographiques et historiques	278
10	Représentation des connaissances	283
10.1	Ingénierie ontologique	283
10.2	Catégories et objets	285
10.3	Événements	290
10.4	Objets mentaux et logique modale	294
10.5	Systèmes de raisonnement pour les catégories	296
10.6	Raisonnements avec informations par défaut	300
	Résumé, notes bibliographiques et historiques	303

11 Planification classique	309
11.1 Définition de la planification classique	309
11.2 Algorithmes pour la planification classique	313
11.3 Heuristiques pour la planification	317
11.4 Planification hiérarchique	320
11.5 Planification et action dans des domaines non déterministes	327
11.6 Temps, ordonnancement et ressources	336
11.7 Analyse des méthodes de planification	339
Résumé, notes bibliographiques et historiques	339
IV Connaître et penser l'incertain	
12 Quantification de l'incertitude	345
12.1 Agir dans l'incertitude	345
12.2 Probabilités : notations de base	348
12.3 Inférence utilisant des distributions conjointes complètes	354
12.4 Indépendance	356
12.5 La règle de Bayes et son utilisation	357
12.6 Modèles bayésiens naïfs	360
12.7 Le monde du wumpus revisité	362
Résumé, notes bibliographiques et historiques	364
13 Raisonnement probabiliste	369
13.1 Représentation des connaissances dans un domaine incertain	369
13.2 Sémantique des réseaux bayésiens	371
13.3 Inférence exacte dans les réseaux bayésiens	382
13.4 Inférence approchée dans les réseaux bayésiens	389
13.5 Réseaux causaux	402
Résumé, notes bibliographiques et historiques	406
14 Raisonnement probabiliste temporel	413
14.1 Temps et incertitude	413
14.2 Inférence dans les modèles temporels	417
14.3 Modèles de Markov cachés	423
14.4 Filtres de Kalman	428
14.5 Réseaux bayésiens dynamiques	434
Résumé, notes bibliographiques et historiques	444
15 Programmation probabiliste	447
15.1 Modèles probabilistes relationnels	448
15.2 Modèles probabilistes en univers ouvert	453
15.3 Suivre l'évolution d'un monde complexe	459
15.4 Modèles probabilistes sous forme de programmes	463
Résumé, notes bibliographiques et historiques	467
16 Prise de décision simple	473
16.1 Désirs, croyances et incertitude	473
16.2 Concepts de base de la théorie de l'utilité	474
16.3 Fonctions d'utilité	477
16.4 Fonctions d'utilité multiattribut	483
16.5 Réseaux de décision	487
16.6 La valeur de l'information	489
16.7 Préférences inconnues	495
Résumé, notes bibliographiques et historiques	498

17	Prise de décision complexe	503
17.1	Problèmes de décision séquentiels	503
17.2	Algorithmes pour les PDM	512
17.3	Les problèmes de bandit	519
17.4	PDM partiellement observables	526
17.5	Algorithmes de résolution des PDMPO	528
	Résumé, notes bibliographiques et historiques	532
18	Prise de décision multiagent	537
18.1	Propriétés des environnements multiagents	537
18.2	Théorie des jeux non coopératifs	542
18.3	Théorie des jeux coopératifs	560
18.4	Prise de décision collective	565
	Résumé, notes bibliographiques et historiques	576
 V Apprentissage		
19	Apprendre à partir d'exemples	581
19.1	Les différentes formes d'apprentissage	581
19.2	Apprentissage supervisé	583
19.3	Apprentissage d'arbres de décision	586
19.4	Évaluation et choix de la meilleure hypothèse	594
19.5	Théorie de l'apprentissage	600
19.6	Régression et classification avec des modèles linéaires	603
19.7	Modèles non paramétriques	612
19.8	Méthodes d'apprentissage par ensemble	620
19.9	L'apprentissage automatique en pratique	628
	Résumé, notes bibliographiques et historiques	636
20	Apprentissage de modèles probabilistes	643
20.1	Apprentissage statistique	643
20.2	Apprentissage avec données complètes	646
20.3	Apprentissage avec variables cachées : l'algorithme EM	657
	Résumé, notes bibliographiques et historiques	665
21	Apprentissage profond	669
21.1	Réseaux simples à propagation avant	670
21.2	Graphes de calcul en apprentissage profond	674
21.3	Réseaux convolutifs	677
21.4	Algorithmes d'apprentissage	682
21.5	Généralisation	685
21.6	Réseaux de neurones récurrents	689
21.7	Apprentissage non supervisé et apprentissage par transfert	692
21.8	Applications	697
	Résumé, notes bibliographiques et historiques	699
22	Apprentissage par renforcement	705
22.1	Apprendre de ses récompenses	705
22.2	Apprentissage par renforcement passif	707
22.3	Apprentissage par renforcement actif	712
22.4	Généralisation et apprentissage par renforcement	717
22.5	Recherche de politique	723
22.6	Apprentissage par démonstration et apprentissage par renforcement inverse	725
22.7	Applications de l'apprentissage par renforcement	728
	Résumé, notes bibliographiques et historiques	730

VI Communiquer, percevoir et agir

23 Traitement du langage naturel	735
23.1 Modèles de langue	735
23.2 Grammaire	745
23.3 Analyse syntaxique	747
23.4 Grammaires augmentées	752
23.5 Complexité des langues naturelles réelles	756
23.6 Tâches de TALN	759
Résumé, notes bibliographiques et historiques	760
24 Apprentissage profond en traitement du langage naturel	765
24.1 Plongement lexical	765
24.2 Réseaux de neurones récurrents pour le TALN	769
24.3 Modèles séquence à séquence	772
24.4 Architecture de transformateur	776
24.5 Préentraînement et apprentissage par transfert	778
24.6 État de l'art	781
Résumé, notes bibliographiques et historiques	785
25 Vision par ordinateur	789
25.1 Introduction	789
25.2 Formation des images	790
25.3 Attributs d'image élémentaires	795
25.4 Classification des images	802
25.5 Détection des objets	805
25.6 Reconstruction du monde en 3D	807
25.7 Utilisation de la vision	811
Résumé, notes bibliographiques et historiques	823
26 Robotique	829
26.1 Les robots	829
26.2 Aspects matériels	830
26.3 Les tâches de la robotique	833
26.4 Perception robotique	834
26.5 Planification et commande	840
26.6 Planification de mouvements incertains	856
26.7 Apprentissage par renforcement en robotique	858
26.8 Humains et robots	860
26.9 Autres cadres pour la robotique	866
26.10 Domaines d'application	869
Résumé, notes bibliographiques et historiques	872

VII Conclusions

27 Philosophie, éthique et sécurité de l'IA	879
27.1 Les limites de l'IA	879
27.2 Les machines peuvent-elles vraiment penser?	882
27.3 L'éthique de l'IA	884
Résumé, notes bibliographiques et historiques	900
28 Avenir de l'IA	907
28.1 Composants des agents	907
28.2 Architectures d'IA	913

Annexe A Rappels mathématiques	917
A.1 Analyse de la complexité et notation $O()$	917
A.2 Vecteurs, matrices et algèbre linéaire	919
A.3 Distributions de probabilités	920
Annexe B Notes sur les langages et les algorithmes	923
B.1 Définition de langages sous forme de Backus-Naur (BNF)	923
B.2 Description d'algorithmes en pseudocode	924
B.3 Ressources supplémentaires en ligne	925
Bibliographie	927
Index	965