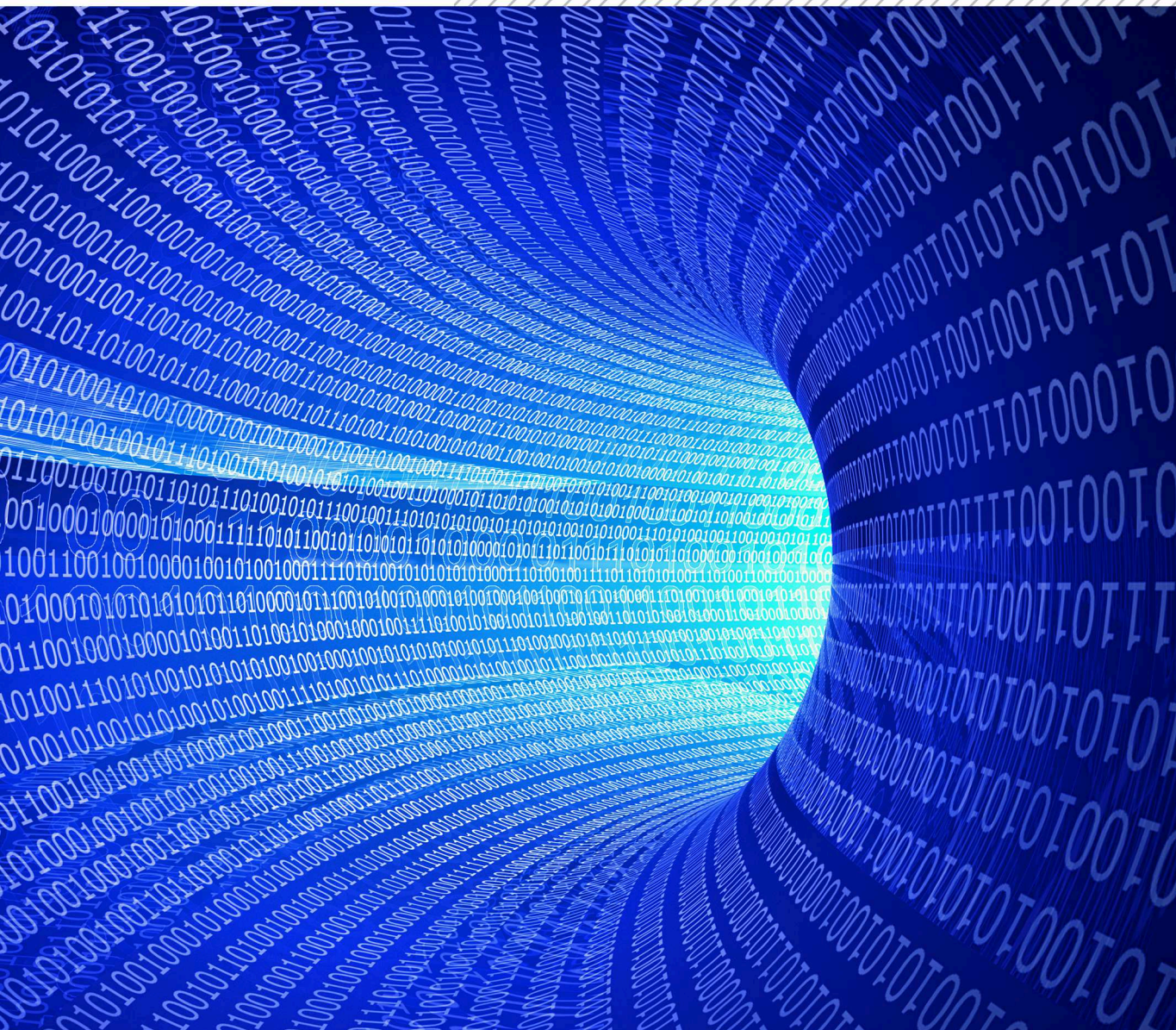




TECHNIQUES
DE L'INGÉNIEUR

LES FOCUS
TECHNIQUES DE L'INGÉNIEUR



TECHNOLOGIES LOGICIELLES
ARCHITECTURES DES SYSTÈMES
INTRODUCTION À L'INTELLIGENCE
ARTIFICIELLE



**TECHNIQUES
DE L'INGÉNIEUR**

Réf. : **H3720 V1**

Introduction à l'intelligence artificielle

Date de publication :
10 août 2021

Cet article est issu de : **Technologies de l'information | Technologies logicielles
Architectures des systèmes**

par **Jean-Paul HATON**

Mots-clés

apprentissage | intelligence artificielle | reconnaissance de forme | réseaux neuronaux

Résumé L'intelligence artificielle (IA) s'attache à résoudre des problèmes qui relèvent d'activités humaines de nature variée (perception, prise de décision, planification, diagnostic, interprétation de données, compréhension du langage, conception). Ces problèmes nécessitent de mettre en jeu une grande quantité de données et de connaissances, soit exploitées directement, soit codées sous différentes formes (distributions de probabilités, poids synaptiques, etc.). Nous présentons les différents modèles développés depuis le début de l'IA, ainsi que les grands domaines d'application. Les aspects éthiques liés à l'utilisation de systèmes d'IA sont également étudiés.

Keywords

learning | artificial intelligence | pattern recognition | neural networks

Abstract Artificial Intelligence (AI) aims at solving problems involved in various types of human activities (perception, decision making, planning, diagnosis, data interpretation, design, language understanding). Solving such problems need to use large amount of data and knowledge, either directly exploited or coded into various forms (probability distributions, synaptic weights, etc.). The different models developed since the beginning of AI are presented. Aspects of ethics related to the use of AI systems are also tackled.

Pour toute question :

Service Relation clientèle
Techniques de l'Ingénieur
Immeuble Pleyad 1
39, boulevard Ornano
93288 Saint-Denis Cedex

Par mail :

infos.clients@teching.com

Par téléphone :

00 33 (0)1 53 35 20 20

Document téléchargé le : **28/12/2022**

Pour le compte : **7200106152 - éditions ti // céline BLONBOU // 2.59.188.28**

© Techniques de l'Ingénieur | tous droits réservés

Introduction à l'intelligence artificielle

par **Jean-Paul HATON**

Professeur émérite

LORIA – Institut Universitaire de France – Université de Lorraine – Nancy, France

1. Intelligence naturelle ... et artificielle	H 3 720 - 2
1.1 Définition de l'intelligence.....	– 2
1.2 Les grands modèles de l'IA.....	– 2
2. Bref historique	– 3
3. Apprentissage	– 4
3.1 Introduction	– 4
3.2 Apprentissage symbolique	– 4
3.3 Apprentissage numérique.....	– 4
3.4 Supervisé vs non-supervisé.....	– 5
3.5 Apprentissage par renforcement.....	– 5
4. Réseaux neuronaux profonds	– 5
5. Domaines d'application	– 7
6. Aspects éthiques	– 8
6.1 Introduction	– 8
6.2 Protection des données personnelles	– 8
6.3 Transparence des algorithmes	– 9
6.4 Quelques domaines.....	– 9
6.5 Premier bilan	– 10
7. Conclusion	– 10
Pour en savoir plus	Doc. H 3 720

Le but de l'intelligence artificielle (IA) est double. D'une part, l'IA s'attache à résoudre des problèmes qui relèvent d'activités humaines ou animales de nature variée : perception, planification, interprétation de données, diagnostic, prise de décision, compréhension du langage, conception. D'autre part, l'IA cherche à mieux comprendre et modéliser l'intelligence. Elle se rapproche ainsi des sciences cognitives dont elle s'inspire par ailleurs pour la conception de modèles (mémoire, raisonnement, apprentissage). La nécessité de restreindre l'activité à un champ d'application limité et de s'appuyer sur des connaissances de nature diverse est apparue rapidement en IA. Cette approche symbolique de l'IA a donné lieu aux systèmes à base de connaissances.

Une autre approche, dite connexionniste, tente de s'inspirer du fonctionnement du cortex cérébral. Un réseau neuronal est formé par l'interconnexion d'un grand nombre de neurones artificiels. Il présente des propriétés intéressantes, notamment la capacité d'apprendre à partir de grandes quantités d'exemples.

Des succès récents ont médiatisé l'IA : jeu d'échecs, jeu de go, poker, robots martiens, le jeu américain de questions-réponses Jeopardy, reconnaissance d'images, reconnaissance de la parole, jeux vidéo et autres.

Cet article présente les différents modèles de l'IA, ainsi que les méthodes d'apprentissage associées. Il décrit aussi un vaste ensemble d'applications (médecine, industrie, militaire, banque, droit, etc.) et aborde aussi les aspects éthiques liés à ces applications.

1. Intelligence naturelle ... et artificielle

1.1 Définition de l'intelligence

L'intelligence est une notion multiforme et difficile à préciser. Philosophes et scientifiques se sont attachés à la définir depuis des millénaires. Pour les besoins de l'IA, nous nous contenterons de qualifier l'intelligence par un ensemble de capacités, notamment de mémorisation, de structuration de la connaissance et de conceptualisation, de perception, de raisonnement, de prise de décision, d'apprentissage, de communication et de dialogue.

Ces capacités se retrouvent, à des degrés divers, dans les systèmes d'IA actuels. Ces systèmes se nourrissent des avancées dans ces différents domaines. Inversement, comme il a été dit, l'IA contribue à ce que nous comprenions mieux l'intelligence.

1.2 Les grands modèles de l'IA

Dès le début de l'IA dans les années 1950, trois grands types de modèles ont été proposés par les chercheurs pour concevoir des machines intelligentes :

- les **modèles symboliques**, correspondant à une approche que l'on peut qualifier d'IA symbolique, permettent de doter les systèmes d'IA de mécanismes de raisonnement capables de manipuler les données symboliques qui constituent les connaissances d'un domaine. Cette approche fait appel aux modèles et méthodes de la logique. Elle a donné lieu aux systèmes à bases de connaissances. A. Newell considérait le niveau de connaissance (*knowledge level*) comme le trait d'union entre l'homme et la machine intelligente, permettant de rationaliser le comportement d'un agent qui, à l'instar de l'être humain, peut prendre une décision en menant un raisonnement fondé sur ses connaissances. Un aspect important est la capitalisation et la diffusion du savoir et de l'expérience, éléments constituant la mémoire d'une organisation. Une base de connaissances se construit à partir d'ontologies qui la structurent et la contraignent. De telles bases interviennent dans de nombreux champs d'application comme l'intelligence économique, le droit, la production industrielle, la médecine, etc.

Pour être utiles, ces bases doivent inclure des modèles opérationnels du monde et du bon sens. Le projet Cyc lancé en 1984 par D. Lenat a cet objectif ambitieux. Il s'agit du plus grand projet existant visant à formaliser les connaissances relatives à notre vie quotidienne (la version actuelle comprend des millions de faits et de règles formalisés dans un langage fondé sur le calcul logique des prédicats) et à les utiliser pour mener des raisonnements. Cyc comporte également des relations de causalité, reconnues très importantes dans le comportement humain. En 2018, Paul Allen (co-fondateur de Microsoft) a lancé dans son *Institute for Artificial Intelligence* le projet Alexandria, avec la même ambition d'apprendre le bon sens à une machine. Ce thème très important pour l'avenir de l'IA a également été repris par l'agence américaine DARPA sous la forme d'une proposition de projet *Machine Common Sense* ;

- les **modèles neuromimétiques** correspondent à une approche que l'on peut qualifier d'IA connexionniste par analogie métaphorique. Ils reviennent à s'inspirer du fonctionnement du cortex cérébral. L'entité de base est un modèle formel très simplifié du

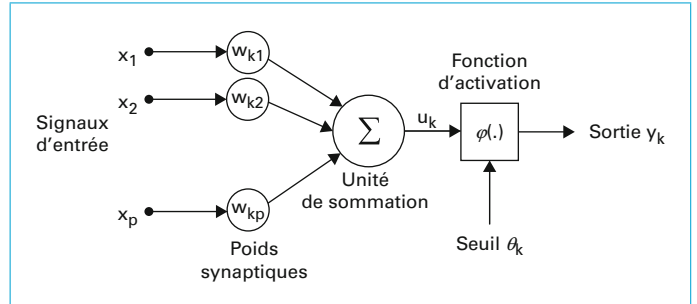


Figure 1 – Exemple de perceptron multicouche

neurone proposé par McCulloch et Pitts en 1943. Chaque neurone possède un certain nombre d'entrées synaptiques, chacune assortie d'un poids ; le neurone effectue la somme pondérée de ses entrées et active sa sortie, reliée à d'autres neurones, si la somme atteint un seuil prédéterminé, comme le montre la figure 1.

Un système est formé par l'interconnexion d'un grand nombre de tels neurones en couches successives. Cette approche a donné lieu aux réseaux neuromimétiques actuels, avec une grande variété des modèles. La plupart de ces modèles, comme le perceptron multicouche de la figure 2, sont de type *feedforward*, ce qui signifie que les informations transitent dans un sens unique, de la couche d'entrée vers la couche de sortie. Comme d'autres modèles, les réseaux neuronaux sont capables de calculer toute fonction, d'où leur utilité en IA.

Ce perceptron très simple, conçu pour la reconnaissance de 3 mots parlés, possède 3 couches : une couche d'entrée recevant les paramètres de l'entité à reconnaître, une couche cachée et une

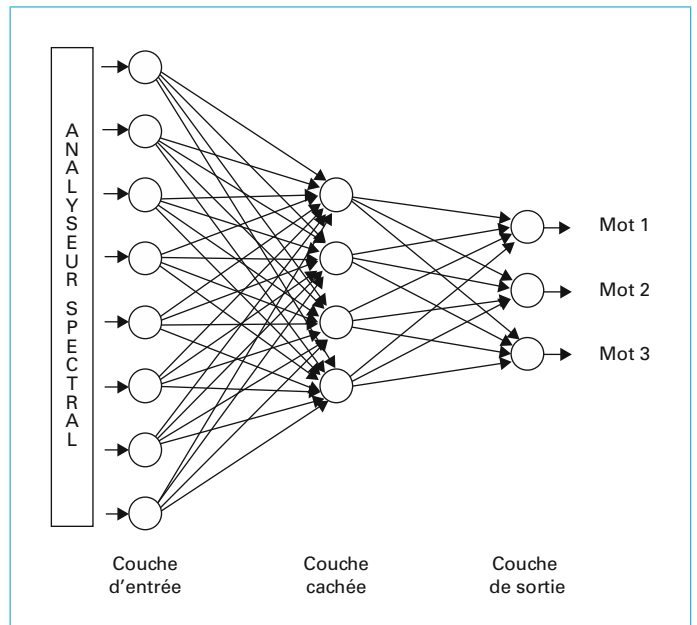


Figure 2 – Exemple de perceptron multicouche

couche de sortie donnant la réponse du perceptron. Chaque rond représente un neurone. Les flèches reliant les neurones possèdent toutes un poids, appris lors d'une phase préalable d'apprentissage, représentant la force de la connexion entre deux neurones ;

- enfin, les **modèles probabilistes et statistiques** présentent un cadre formel intéressant pour capturer la variabilité inhérente au monde réel et en rendre compte. Ces modèles, tout comme les modèles neuromimétiques, sont capables d'apprendre à partir d'exemples. L'apprentissage revient ici à mémoriser des distributions de probabilités à l'aide d'algorithmes souvent complexes mais dont les propriétés sont parfaitement connues.

Un modèle probabiliste largement utilisé est celui des réseaux bayésiens. Ces réseaux sont des graphes constitués de nœuds représentant les concepts d'un domaine et d'arcs représentant des relations de causalité probabilisées entre deux concepts (par exemple, tel état physiopathologique d'un patient peut être la cause de tel symptôme, avec telle plausibilité). Un réseau bayésien permet de mener un raisonnement probabiliste sur des faits multiples grâce à des mécanismes de propagation de coefficients de probabilité à travers le réseau. Il est ainsi très intéressant dans des problèmes à choix multiples comme le diagnostic, notamment médical et industriel. Pour des applications en vraie grandeur, ces réseaux de dépendance conditionnelle peuvent atteindre des dimensions considérables. La mise au point et l'exploitation de tels réseaux sont des questions bien maîtrisées.

Le temps est une dimension essentielle dans de nombreuses activités en IA. De ce fait, les modèles statistiques intégrant la variable temporelle, ou modèles stochastiques, comptent parmi les plus utilisés en intelligence artificielle. Le modèle stochastique le plus répandu est le modèle de Markov caché, ou MMC (*Hidden Markov Model*, *HMM*). C'est le cas en reconnaissance de la parole [H 3 728] où chaque entité à reconnaître (mot, unité phonétique) est représentée par une source de Markov capable d'émettre le signal vocal correspondant à cette entité. La reconnaissance revient alors à calculer la vraisemblance de la suite d'observations acoustiques constituant l'entité à reconnaître par rapport à chacun des modèles appris. Le modèle présentant la plus grande vraisemblance d'avoir émis cette suite d'observations fournit la réponse.

Les MMC ont également été utilisés avec succès dans d'autres domaines que la parole, en particulier l'interprétation d'images, la reconnaissance de l'écriture, l'interprétation de signaux (radar, sonar, biologiques, etc.) ou la robotique.

Les trois modèles présentés ci-dessus se rencontrent, parfois simultanément, dans les systèmes d'IA actuels. Les systèmes de traitement d'images et de reconnaissance de la parole, par exemple, sont le plus souvent fondés sur la complémentarité entre des modèles stochastiques MMC et des modèles neuronaux. Les modèles statistiques jouent également un rôle fondamental dans le traitement des grandes masses de données et leur exploitation, en particulier pour la découverte de régularités ou de connaissances.

Une caractéristique commune à tous ces types de modèles est leur capacité d'apprentissage à partir d'exemples. L'apprentissage, capacité fondamentale de l'intelligence, joue ainsi un rôle majeur dans le bon fonctionnement des systèmes d'IA.

2. Bref historique

L'IA est née dans les tout premiers temps de l'informatique. En 1950, Alan Turing publie un article [2] dans lequel il pose la question « Les machines peuvent-elles penser ? » L'auteur propose un test, connu désormais sous le nom de *Test de Turing*, destiné à répondre à la question posée. Dans ce test, inspiré du jeu de l'imitation, une personne interroge une machine et un être humain qu'il ne voit pas. Lorsque, sur la base des réponses fournies, l'interrogateur ne peut plus distinguer l'être humain de la machine, la machine est déclarée intelligente. Les questions peuvent être de

tout type. Pour l'instant, aucun programme n'a pu tromper un interrogateur pendant un temps suffisamment long. Les agents conversationnels (*chatbots*, et *callbots* via le téléphone) qui se sont multipliés ces dernières années augmentent progressivement leur compétence à mener un dialogue avec un être humain, comme le montre le prix Alexa lancé par Amazon.

Le terme *Artificial Intelligence* apparaît en 1956 lors d'une école d'été (*Darmouth Conference*) réunissant un ensemble de jeunes chercheurs dont certains allaient devenir de grands noms du domaine : J. McCarthy, M. Minsky, etc. L'ambitieuse conjecture (toujours ouverte) énoncée par ces chercheurs est que : « *every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it* ». Les travaux pionniers ont ainsi abordé les domaines des jeux (échecs, dames), de la reconnaissance de formes (parole, caractères écrits), de la résolution de problèmes (systèmes *General Problem Solver*, *GPS* de A. Newell et A. Simon, et Alice de J.L. Laurière).

La cybernétique, mouvement scientifique de l'après-Second Guerre mondiale autour de N. Wiener, W.R. Ashby, L. Couffignal et beaucoup d'autres, a joué également un rôle dans la genèse de l'IA, notamment avec la notion de régulation et le concept clé de rétroaction (*feedback*), aussi bien chez l'animal que dans la machine, toujours aussi important pour les réseaux neuronaux ou la robotique, spécialement pour des tâches de pilotage.

À la même époque, F. Rosenblatt inventait le perceptron monocouche, réseau neuronal classifieur linéaire, ancêtre des réseaux neuronaux profonds actuels qui sont toujours fondés sur la modélisation du neurone proposée en 1943. L'intérêt pour ces modèles a décliné vers 1969 avec la publication d'un livre sur leurs limites par M. Minsky et S. Papert [7]. Il faudra attendre le perceptron multicouche et la redécouverte de l'algorithme d'apprentissage associé de rétropropagation du gradient d'erreur pour constater un regain d'activité dans ce domaine jusqu'aux réseaux neuromimétiques profonds actuels qui occupent presque exclusivement la scène médiatique.

La conception de systèmes à bases de connaissances, et notamment, de systèmes experts, représente néanmoins toujours un domaine important de l'IA. Les systèmes experts sont conçus pour atteindre les performances d'experts humains dans des domaines limités en raisonnant sur un ensemble de connaissances acquises pour l'essentiel auprès de ces experts. Apparus vers 1975 (cf. le système de diagnostic des maladies du sang MYCIN), ils ont eu un impact certain sur l'IA, et aussi un retentissement médiatique parfois exagéré car ils n'ont pas tenu toutes les promesses qui leur furent associées. Le terme de système expert disparaît peu à peu, au profit du concept plus général de système à bases de connaissances. Ce concept est fondé sur une séparation entre d'une part les connaissances nécessaires pour résoudre un problème et d'autre part les mécanismes de raisonnement qui exploitent ces connaissances.

Parmi ces travaux, les systèmes multi-agents occupent une part importante. L'idée est de parvenir à une décision commune à un ensemble d'entités par fusion d'informations et de points de vue, dans un cadre de coopération ou de compétition. Les modèles développés (tableau noir, acteurs, société d'experts) ont donné lieu à de nombreuses applications.

Dès les premières années de l'IA, la langue naturelle écrite a fait l'objet d'un ensemble de projets en traduction automatique (avec des résultats fort limités pendant longtemps, la difficulté de la tâche ayant été initialement sous-estimée) et en dialogue humain-machine. Le système SHRDLU, développé par T. Winograd, permettait un dialogue avec un robot manipulateur dans un monde de blocs de formes (cubes, cônes, sphères) et de couleurs différents. Le niveau de « compréhension » de la langue atteint par SHRDLU était essentiellement lié à la simplicité de son monde de blocs : nombre et de types de blocs, nombre et niveau de complexité des actions possibles. Il a été impossible d'étendre le

système à un univers plus complexe, sans parler du monde réel. Toutefois, son influence a été notable dans les domaines de la langue naturelle et de la planification d'actions.

En ce qui concerne la langue parlée, les travaux en reconnaissance automatique de la parole ont débuté dès la fin des années 1950. Comme on l'a vu, l'utilisation de modèles stochastiques (modèles de Markov cachés), puis leur association à des modèles neuronaux ont permis d'atteindre des niveaux de performance remarquables en reconnaissance de mots ou de phrases, même pour des vocabulaires de dizaines de milliers de mots. En revanche, la compréhension de la parole demeure un sujet de recherche actif [H 3 728].

3. Apprentissage

3.1 Introduction

L'apprentissage, caractéristique fondamentale de l'intelligence, est partie prenante de la plupart des systèmes d'IA, permettant d'optimiser leurs performances.

L'apprentissage automatique (*machine learning*) est une discipline très active et multiforme, selon les modèles sous-jacents. Toutes les méthodes impliquent l'existence de grandes bases de données d'exemples et parfois de contre-exemples, le plus souvent dûment étiquetés, sur lesquels se fonde l'apprentissage. Un système apprend à partir de chaque exemple, avec l'idée d'être capable de généraliser son comportement à de nouveaux cas non encore rencontrés, grâce aux bonnes propriétés des modèles appris.

On distingue deux grands types d'apprentissage en IA, l'apprentissage symbolique et l'apprentissage numérique. Nous résumons ci-dessous les grandes caractéristiques de ces méthodes.

3.2 Apprentissage symbolique

Ce type d'apprentissage, lié aux systèmes à base de connaissances, concerne l'acquisition de concepts et de connaissances structurées. Un exemple fondateur bien connu est Dendral, système expert en chimie organique, capable d'apprendre des règles d'explication à partir de données de spectrographie de masse. Ajouter de nouvelles règles dans une base de connaissances pose des problèmes de gestion des bases, de maintien de cohérence et, éventuellement, de généralisation des connaissances apprises par recherche d'explications.

Les nombreuses méthodes proposées relèvent de deux grands types d'apprentissage :

- l'**apprentissage par détection de similarités** (*Similarity-based learning*) dans lequel on apprend en détectant des similarités et des dissemblances dans une base d'apprentissage d'exemples et de contre-exemples. Cette recherche d'associations et de liens de causalité significatifs permet d'extraire des pépites de connaissances à l'aide de mécanismes de fouille de données associés à une évaluation par un expert du domaine. Cette coopération IA-expertise humaine se retrouve souvent dans les applications pratiques ;

- l'**apprentissage par recherche d'explications** (*Explanation-based learning*) dans lequel on apprend à partir des explications extraites d'exemples et de contre-exemples.

Il faut également adjoindre à ces deux types d'autres méthodes telles que l'apprentissage par analogie, la construction automatique de taxonomies et d'ontologies, l'inférence inductive de connaissances ou encore le regroupement automatique de concepts, ainsi que les arbres de décision. Un arbre de décision est une structure arborescente qui permet de classer une certaine entité (objet, cas, etc.) par un ensemble de questions. À chaque feuille d'un arbre est

associée une question portant sur les caractéristiques de l'entité étudiée. Les feuilles terminales correspondent aux différentes classes définies. L'apprentissage revient à apprendre l'ensemble des questions relatives aux caractéristiques des objets présentés. Plusieurs logiciels de construction d'arbres ont été développés (par exemple CART, C4.5, ID3, etc.).

Les travaux en apprentissage symbolique ont donné lieu à de nombreuses réalisations. En revanche, les méthodes proposées demeurent spécifiques et liées à un domaine particulier. Depuis le début des années 2000, ces méthodes sont moins étudiées que les méthodes relevant de l'apprentissage numérique.

3.3 Apprentissage numérique

La majorité des progrès réalisés en IA depuis une dizaine d'années sont dus à l'utilisation de méthodes numériques d'apprentissage.

Les **modèles statistiques** (modèles de Markov cachés et réseaux bayésiens) ont déjà été brièvement présentés. Un des intérêts de ces modèles réside dans l'automatisation de l'apprentissage des différents paramètres et distributions de probabilité du modèle à partir de données représentatives de l'application considérée (appelées observations). Le principe est de trouver un modèle qui maximise la probabilité d'un ensemble (ou dans le cas des MMC d'une séquence) d'observations, c'est-à-dire de déterminer le modèle qui explique le mieux cette séquence. Il n'est pas possible de trouver un tel modèle de façon analytique. L'apprentissage est alors assuré par des algorithmes itératifs d'estimation des paramètres, notamment l'algorithme de Baum-Welch, cas particulier de l'algorithme Espérance-Maximisation, dit EM.

Dans les années 1990, un autre modèle numérique statistique, les **machines à vecteurs supports**, appelées aussi Séparateurs à Vaste Marge (*Support Vector Machines, SVM*) ont contribué de façon notable au succès des méthodes numériques d'apprentissage. Une SVM [TE 5 255] est essentiellement un classifieur discriminant à deux classes (qui peut être étendu à une SVM multiclassée) dont le critère d'optimisation est la largeur de la marge entre les deux classes, c'est-à-dire la zone vide autour de la surface de décision définie par les formes les plus proches. Ce modèle est intéressant, mais les réseaux neuronaux profonds se sont révélés plus performants que les SVM pour de nombreuses applications.

Les **modèles neuronaux** utilisent également un apprentissage numérique. Le principe est d'optimiser les poids des connexions entre les neurones des différentes couches du modèle. L'algorithme de rétropropagation du gradient d'erreur a permis le développement dans les années 1990 de modèles neuronaux comportant un nombre limité (quelques unités) de couches cachées. L'idée de ce type d'apprentissage est la suivante : si la réponse du système à une entrée donnée est incorrecte, un gradient d'erreur est calculé en fonction des réponses de la couche de neurones de sortie. Ce gradient est rétropropagé de couche en couche depuis la couche de sortie jusqu'à la couche d'entrée en modifiant les poids des connexions inter-neurones de façon adéquate. On démontre que, moyennant un nombre suffisant d'exemples d'apprentissage, l'algorithme converge vers une configuration stable de poids pour l'ensemble des neurones. Récemment, le nombre de couches cachées a été considérablement augmenté, tout en conservant la capacité d'apprentissage à partir d'exemples, ce qui a donné naissance aux réseaux neuronaux profonds (*Deep Neural Nets, DNN*) présentés au § 4.

Après avoir appris sur un grand nombre d'exemples étiquetés, un réseau neuronal est non seulement capable de classer correctement ces exemples, mais peut aussi traiter de nouveaux objets de même catégorie qu'il n'a pas vus durant la phase d'apprentissage ; cette capacité de généralisation est une des propriétés très intéressantes de ces modèles.

3.4 Supervisé vs non-supervisé

Les méthodes, symboliques et numériques, décrites ci-dessus sont toutes de type **apprentissage supervisé**, ou avec professeur [H 5 010]. Elles nécessitent, comme il a été dit, de grandes quantités d'exemples étiquetés avec la bonne réponse associée (portion de signal vocal et phonème ou syllabe correspondant, description d'un cas de panne et diagnostic associé, image et sa description, etc.). La réalisation de telles bases d'exemples annotés est très coûteuse (les bases d'images utilisées pour entraîner un système d'identification d'objets comportent des millions d'images indexées ; les bases liées à la reconnaissance de la parole contiennent des centaines d'heures de parole étiquetées phonétiquement). Une tendance apparue avec Internet est de sous-traiter aux utilisateurs eux-mêmes l'étiquetage des données comme une tâche, parfois invisible, annexe à l'utilisation d'un système (*crowdsourcing*).

Une autre classe de méthodes concerne l'**apprentissage non supervisé**, dans lequel l'opération se fait de façon totalement autonome [H 5 012]. Des données non étiquetées sont présentées en entrée sans indiquer les réponses attendues en sortie. C'est le mécanisme d'apprentissage lui-même qui propose des catégories pour regrouper les réponses possibles. Cette solution semble idéale dans la mesure où elle ne nécessite pas de grandes bases de données étiquetées. En fait, les deux types d'apprentissage ne sont pas destinés aux mêmes tâches. L'apprentissage non supervisé cherche à partitionner de lui-même, sans intervention extérieure d'un « professeur », les données qui lui sont présentées en catégories homogènes au sens d'un certain critère ou d'un ensemble de caractéristiques communes. Les méthodes de regroupement, parfois appelées **coalescence** (*clustering*), ont également donné lieu à de nombreuses études et au développement de méthodes et de logiciels variés (partitionnement, hiérarchisation, etc.), notamment dans le domaine de la reconnaissance des formes [AF 1 510]. Le succès de l'apprentissage supervisé pour les réseaux neuronaux a quelque peu occulté l'importance de l'apprentissage non supervisé. Ce dernier est certainement appelé à se développer, par référence à l'être humain et à l'animal. Une autre tâche de cet apprentissage est la recherche de liens de causalité qui sont d'une grande importance dans de nombreuses activités telles que le diagnostic, la prise de décision, comme il a été dit plus haut. Ce type d'apprentissage peut être utilisé simultanément avec un apprentissage supervisé dans lequel les ensembles de données étiquetées sont remplacés par des entités qui changent au cours du temps telles que des vidéos : la trame du temps t de la vidéo est utilisée comme un prédicteur pour la trame $t+1$, sans aucun étiquetage préalable.

Une solution intermédiaire a également été proposée, qualifiée de semi-supervisée. Comme l'étiquetage complet d'une base d'apprentissage est une tâche lourde et coûteuse, l'idée est de restreindre le volume de données étiquetées nécessaires à un bon apprentissage. Un exemple de tel apprentissage est le co-apprentissage, dans lequel deux classificateurs apprennent un ensemble de données, mais en utilisant chacun un ensemble de caractéristiques différentes, si possible indépendantes. Le classifieur le mieux capable de bien traiter un exemple d'apprentissage va jouer le rôle de « professeur » pour l'autre. Des approches de ce type ont été utilisées pour la classification de documents HTML et en traitement de la langue naturelle. Une idée similaire se trouve dans les réseaux antagonistes génératifs (*Generative Adversarial Networks, GAN*). Dans un GAN, deux réseaux sont placés en compétition. Le premier réseau, le générateur, génère un échantillon d'apprentissage, tandis que le second, le discriminateur, essaie de détecter si un échantillon est réel ou bien s'il est le résultat du générateur.

3.5 Apprentissage par renforcement

Parmi les méthodes d'apprentissage, l'**apprentissage par renforcement** occupe une place à part. Le principe est d'apprendre par essais et erreurs comment se comporter de manière optimale dans des environnements incomplètement connus, situation très commune dans la réalité. Cet apprentissage, très développé en robo-

tique, comme l'algorithme de *Q-Learning*, s'est inspiré au départ de théories de psychologie animale et humaine ; il modélise le comportement d'apprentissage optimal par essais et erreurs permettant de s'adapter à un environnement. Imaginons un système, ou un agent, situé dans un certain environnement, changeant et mal connu de l'agent, comme un robot se déplaçant dans un univers qu'il cartographie au fur et à mesure. L'agent peut décider de mener un ensemble d'actions qui lui apporteront éventuellement une récompense. Une action de l'agent conduit à un nouvel état de l'environnement dans lequel il peut mener une nouvelle action qui conduit à un nouvel état, etc. L'environnement est le plus souvent stochastique, ce qui signifie que le nouvel état est aléatoire. Ce processus d'essais et erreurs est conduit jusqu'à obtenir la meilleure réponse possible. La façon efficace la plus utilisée pour raisonner dans de telles conditions est le formalisme des processus de décision markovien (*Markov Decision Process, MDP*) et celui des processus de décision markovien partiellement observables (*Partially Observable Markov Decision Process, POMDP*), ces derniers étant plus proches de la réalité des applications. Il existe de très nombreuses variantes de l'apprentissage par renforcement, en général formalisées dans un cadre probabiliste. Cet apprentissage est largement utilisé, y compris dans les réseaux neuronaux profonds, notamment dans le vaste domaine des jeux : backgammon (1994), Atari (2013) et, plus récemment, go et poker.

La combinaison de l'apprentissage profond et de l'apprentissage par renforcement semble également prometteuse. De très bons résultats ont déjà été obtenus en classification d'images, et aussi en traitement de la langue naturelle.

4. Réseaux neuronaux profonds

Une avancée majeure de l'IA depuis 2010 s'est produite dans le domaine des réseaux neuronaux profonds (*Deep Neural Networks, DNN*) et des algorithmes d'apprentissage associés [H 1 098]. Déjà, dès 2006, les modèles acoustiques de reconnaissance de la parole avaient été améliorés de façon importante grâce aux modèles neuronaux profonds. Depuis, les DNN ont montré leur efficacité dans des domaines très variés : jeux, traitement de textes, vision par ordinateur, diagnostic, robotique, banque, etc. Comme on l'a vu, un réseau neuronal est un classifieur capable d'apprendre des fonctions de décision. Globalement, un réseau neuronal est capable d'apprendre une mise en correspondance (*mapping*) entre ses entrées et ses sorties, ce qui permet d'utiliser de tels systèmes pour des tâches de classification permettant d'identifier la classe d'appartenance de l'entité placée en entrée. Ces entités peuvent être de natures extrêmement diverses : mots prononcés par un locuteur, image, diagnostic d'un patient, place d'un pion sur un jeu de go, etc. Le perceptron monocouche de Rosenblatt avait des performances très limitées puisqu'il ne pouvait apprendre que des fonctions linéaires. L'introduction de couches cachées en nombre restreint (comme le perceptron de la figure 2 à une couche cachée) a permis d'augmenter les performances. Les DNN sont caractérisés par le fait que leur profondeur (c'est-à-dire le nombre de couches cachées de neurones) est augmentée de façon très importante pour atteindre jusqu'à un millier de couches, ce qui leur donne la capacité d'apprendre des fonctions de mise en correspondance beaucoup plus complexes, d'où leur succès actuel.

La rapide émergence des réseaux profonds est due à la conjonction de trois conditions :

- l'existence de très grandes bases de données étiquetées nécessaires à l'apprentissage de ces modèles. Il s'agit d'un exemple du phénomène récent de *Big Data*. En reconnaissance de la parole, les grands opérateurs du domaine disposent ainsi de millions d'heures de parole, ce qui permet de disposer de

systèmes de reconnaissance dans de nombreuses langues (plus d'une centaine pour Google). Ces bases s'enrichissent quotidiennement. 80 % des données qu'elles renferment sont non structurées, ce qui nécessite une analyse sémantique préliminaire. Le RGPD (Règlement général sur la protection des données) mis en place en 2018 par l'Union européenne, notamment dans la suite de la loi française de 1978 « Informatique et libertés » et des travaux de la CNIL, aura une influence sur l'exploitation de ces grands entrepôts de données. Il importe en effet d'assurer la protection des données personnelles, tout spécialement les données sensibles comme les informations médicales ;

- la disponibilité de capacités de calcul en constante augmentation (notamment à l'aide de cartes additionnelles GPU et du calcul parallèle haute performance [H 1 013]). Des exemples sont le processeur TPU (*Google's Tensor Processing Unit*) ou les puces NVIDIA qui accélèrent notablement le temps d'apprentissage d'un DNN. L'apprentissage d'un DNN peut également être implémenté sur un FGPA (*Field Programmable Gate Array*). Sur le plan des logiciels, citons Theano, bibliothèque logicielle écrite en langage Python pour le développement de systèmes d'apprentissage profond. Ces moyens sont nécessaires en pratique, sachant que le nombre de paramètres à apprendre dans un DNN peut être de l'ordre du milliard.

La loi de Moore ayant à peu près atteint ses limites, il ne faut plus compter seulement sur l'augmentation intrinsèque de puissance des circuits intégrés pour améliorer les performances d'apprentissage, mais aussi sur des solutions innovantes, plus créatives, tant logicielles que matérielles. Ceci constitue un changement important par rapport à ce que nous avons vécu lors des dernières décennies. Les algorithmes actuels d'apprentissage fonctionnent aussi bien sur le « nuage » (cf. les solutions *Open Source* proposées par Google avec son *Cloud Machine Learning*, ou d'autres) que localement là où s'effectue la capture des données. La technologie des circuits intégrés proposera certainement à l'avenir des architectures d'accélérateurs d'apprentissage, adaptées aux différents lieux de traitement des données ;

- l'amélioration des algorithmes d'apprentissage profond (*Deep Learning*). Initialement, les algorithmes d'apprentissage utilisaient des sigmoïdes comme fonctions non linéaires pour transmettre le résultat du traitement de l'erreur d'une couche à la suivante. L'utilisation de fonctions ReLU (*Rectified Linear Unit*) a permis un gain en temps très important. Par ailleurs, la technique de *dropout* qui revient à éliminer des neurones inutiles durant l'apprentissage permet d'éviter l'écueil du sur-apprentissage, susceptible d'apparaître dans les réseaux neuronaux.

La complexité de l'apprentissage d'un DNN a conduit les chercheurs à essayer de réduire cette complexité. Une solution initialement proposée par K. He est celle des réseaux résiduels, *ResNet*. L'idée est d'introduire des raccourcis dans les connexions entre les neurones de couches cachées, ce qui facilite grandement l'apprentissage.

L'apprentissage fondé sur la rétropropagation du gradient d'erreur peut aussi parfois être couplé à l'apprentissage par renforcement. Ainsi le logiciel *AlphaGo Zero* de *DeepMind* qui s'est révélé le meilleur au jeu de go combine DNN et recherche arborescente de type Monte Carlo. Son apprentissage comporte une phase d'apprentissage profond supervisé fondé sur une base de parties jouées par des experts (plus de 100 000) et une phase d'apprentissage par renforcement partant simplement des règles du jeu de go.

L'apprentissage de tels systèmes nécessite, comme on l'a dit, des quantités de données et des moyens de calcul considérables. Les systèmes résultants sont compétents dans la tâche unique sur laquelle a porté leur apprentissage. Une tendance actuelle est de concevoir des environnements d'apprentissage multitâches. Le système IMPALA (*Importance Weighted Actor-Learner Architecture*) de *DeepMind* en est un exemple.

Outre les perceptrons, il existe d'autres types de réseaux neuronaux profonds, avec leurs propres forces et faiblesses :

- **les réseaux récurrents**. L'état d'un réseau est ici fonction de l'entrée du réseau à un instant donné et de son état à l'instant précédent. Sa structure permet une mémoire des entrées précédentes qui persiste dans ses états internes et peut en conséquence avoir un effet sur ses sorties futures. Les réseaux récurrents sont ainsi les plus profonds des réseaux dits profonds. Ils permettent en outre d'exploiter de façon naturelle et efficace le parallélisme des programmes. Ces réseaux peuvent être entraînés par une variante de l'algorithme de rétropropagation, mais leur apprentissage est délicat. Ils sont bien adaptés au traitement de séquences temporelles telles que celles rencontrées en reconnaissance de la parole, reconnaissance de l'écriture, traduction automatique, etc. Le type de réseau récurrent le plus utilisé, tout spécialement en reconnaissance de la langue et en traduction, est le réseau de neurones récurrents à mémoire court et long terme (*Long short-term memory, LSTM*). Un réseau LSTM peut mémoriser une information à plus ou moins long terme, un peu comme le fait notre cerveau. Par exemple, lors de l'analyse d'une phrase, le réseau peut se souvenir du début lorsqu'il arrive à la fin ;

- **les réseaux convolutifs** (*Convolutional Neural Networks, CNN ou ConvNet*). Ces réseaux dont un des initiateurs est Y. Le Cun, tirent leur inspiration du travail de Hubel et Wiesel sur le cortex visuel du chat [3]. Un réseau convolutif profond est formé d'un ensemble hiérarchique de couches de cellules formant un tuilage couvrant l'ensemble du champ visuel (figure 3). Chaque couche fournit un ensemble de paramètres convolutifs de niveau d'abstraction de plus en plus haut à partir des données d'entrée du réseau. Chaque cellule agit comme un filtre local permettant d'exploiter les corrélations spatiales présentes dans une image. L'extraction automatique et la hiérarchisation des paramètres représentatifs des données en entrée sont une des grandes forces de ce modèle. En effet, dans un système d'IA « traditionnel » (reconnaissance de formes, diagnostic), les données brutes présentées en entrée d'un système sont transformées en vecteurs de paramètres (*features*) sur lesquels est prise la décision finale. Une succession de niveaux de traitement permet de transformer les données brutes en représentations de plus en plus abstraites.

Par exemple, une image est présentée en entrée sous forme d'un ensemble de pixels. Le premier niveau de traitement va détecter et localiser des bords ou des angles, le second des arrangements particuliers de bords que le troisième niveau va identifier comme des objets, et ainsi de suite jusqu'à obtenir une représentation abstraite des données adéquate pour la tâche poursuivie.

De façon similaire, la parole se décompose en phrases, mots, syllabes, phones, traits acoustico-phonétiques (bruits, formants, etc.) jusqu'au niveau initial du signal acoustique. La définition de ces niveaux, ainsi que l'extraction automatique des paramètres associés, requiert une importante expérience du domaine d'application.

Le fait que, dans les CNN, la hiérarchie des niveaux ainsi que les paramètres associés sont déterminés automatiquement, sans intervention humaine lors de l'apprentissage, est un des gros avantages de ces modèles. Plus précisément, un CNN présente une structure comportant deux types de couches, directement inspirée des connaissances actuelles des neurosciences de la vision :

- des couches de convolution destinées à détecter les arrangements pertinents de paramètres provenant de la couche précédente ;

- des couches de regroupement (sous-échantillonnage) chargées de rassembler des paramètres semblables.

Les successions de couches « Convolutions/Sous-échantillonnage » peuvent être dupliquées plusieurs fois en fonction de la complexité du problème, comme précisé dans le texte.

Les CNN se sont révélés très efficaces dans la représentation de données complexes structurées et dans leur traitement, tout spécialement dans les domaines de la classification d'images dans la compétition annuelle ILSVRC dont le but est de localiser et d'identifier des

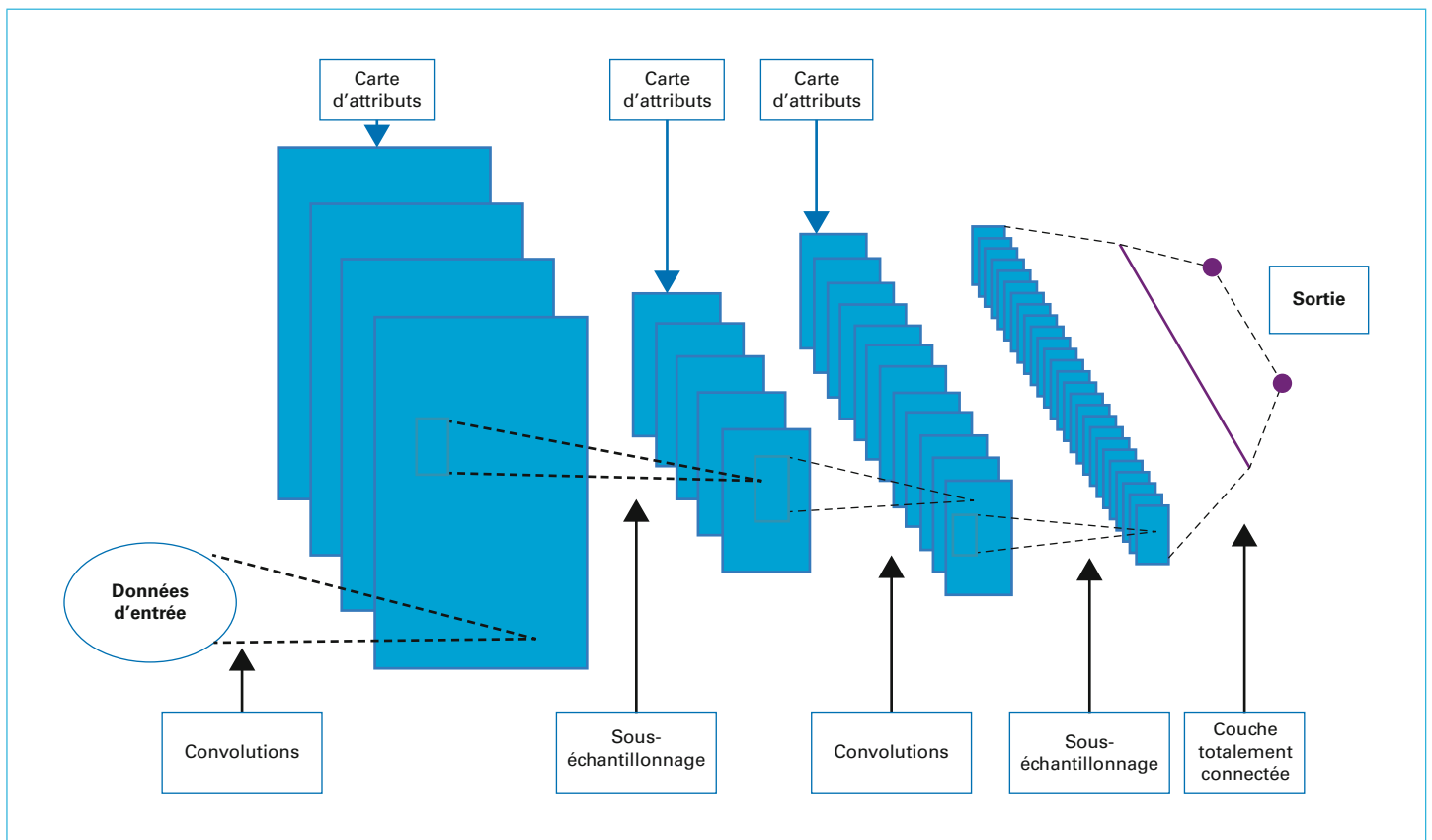


Figure 3 – Exemple de réseau neuronal convolutif

objets dans des scènes naturelles, et récemment dans la reconnaissance de visages. Les CNN obtiennent les meilleurs résultats, de l'ordre de 2 % d'erreur de classification d'images pour ILSVRC, alors que l'erreur humaine sur les mêmes données est de l'ordre de 5 %. Il en est de même pour la reconnaissance de la parole et de l'écriture. D'excellents résultats ont également été obtenus dans une grande variété de domaines tels que la physique (analyse de résultats d'expériences d'accélération de particules), la biologie (traitement de mutations d'ADN), le traitement de la langue naturelle et la traduction automatique. Dans ces deux derniers champs d'activité, les traditionnels modèles de langage à base de n-grammes (courtes séquences de lettres et de symboles de longueur 2 à 5, voire plus) sont remplacés par des modèles de langage neuronaux.

Tous les réseaux neuromimétiques actuels présentent une structure plane bidimensionnelle, les différentes couches de neurones étant organisées horizontalement. Il n'en va pas de même dans le cortex humain qui possède une structure tridimensionnelle où les neurones sont organisés en colonnes. Nous avons réalisé à Nancy une modélisation informatique du modèle de colonne corticale d'Y. Burnod. J. Hinton a récemment proposé un modèle similaire de *capsules*. Ces modèles sont intéressants dans leur principe et doivent être améliorés pour devenir compétitifs avec les DNN actuels.

Malgré leurs propriétés remarquables ayant conduit aux spectaculaires performances évoquées ci-dessus, les réseaux profonds sont intrinsèquement limités et ne résoudront donc pas tous les problèmes qui se posent à l'IA. Les limitations les plus importantes concernent :

- la nécessité d'énormes quantités de données d'apprentissage. Contrairement à l'être humain, ces réseaux ne possèdent pas encore de capacité d'abstraction ;

- le manque de transparence, c'est-à-dire l'incapacité pour ces systèmes d'expliquer leurs résultats (aspect boîte noire). Des efforts sont faits en ce sens, comme le projet *Explainable AI* lancé par l'agence de recherche américaine DARPA ;

- la difficulté d'intégrer des connaissances explicites telles que celles utilisées en IA symbolique : ainsi, un DNN peut apprendre aisément des corrélations entre ses entrées et ses sorties, mais sans pouvoir expliciter les relations de causalité éventuelles entre ces données. La causalité demeure un thème de recherche important en IA symbolique.

5. Domaines d'application

Depuis plusieurs décennies, l'IA a conduit au développement d'applications opérationnelles dans de nombreux domaines tels que le diagnostic [H 7 217] et l'aide à la décision avec les systèmes à bases de connaissances ou les réseaux neuronaux, ainsi que la reconnaissance de formes et de la parole avec les modèles stochastiques. L'introduction des DNN a permis d'atteindre des niveaux de performance inégalés qui ont valu une couverture médiatique sans précédent dans quasiment tous les secteurs d'activité. Comme toujours, le transfert de la recherche en laboratoire vers les applications réelles nécessite des investissements conséquents en temps et en argent. Le cheminement est long entre une première annonce et un véritable produit commercialisé. En médecine, notamment, le déploiement d'applications implique un mécanisme d'évaluation complexe, notamment clinique. Enfin, le succès d'une application implique une véritable confiance dans l'IA ainsi qu'une claire

identification des responsabilités (cf. la voiture autonome en cours d'expérimentation et de mise au point).

Sans prétention d'exhaustivité, car la liste s'allonge continuellement, on peut citer :

- dans le domaine des jeux, avec l'évolution de la technologie et des algorithmes, les systèmes fondés sur l'IA sont devenus les meilleurs : morpion (1952), dames (1994), échecs (1997), go (2016), poker (2016), jeux vidéo (2017) ;
- en aide au diagnostic et à la recommandation d'actions (médical, spatial, industriel...) et aide à la décision (banques, assurances, conduite de procédés, domaine militaire) ;
- en aide à la conception, notamment de puces électroniques (*Electronic Design Automation*, EDA) ;
- en reconnaissance et synthèse de la parole. La conjonction HMM-DNN conduit à des taux de reconnaissance proches de l'humain (annonce récente de Google) et fait le succès des assistants personnels vocaux ;
- en identification de locuteurs et détection d'émotions ;
- en traitement de la langue naturelle : systèmes de questions-réponses, traduction, dialogue, description d'une image ou d'une scène. Le système Jeopardy d'IBM est un bel exemple du niveau atteint dans le domaine ;
- en interprétation de signaux (surveillance, conduite, cybercriminalité) ;
- en robotique : robots autonomes (exploration, intervention en milieu hostile), robots de compagnie (pour enfants malades, personnes âgées ou astronautes), voitures autonomes sans chauffeur ;
- en traitement d'images : diagnostic à partir d'images médicales (rayons X, IRM...), reconnaissance de l'écriture, reconnaissance de visages (avec les aspects éthiques associés et les dérives potentielles d'identification d'individus sans leur consentement...), télédétection, détection, classification et localisation d'objets, identification d'actions, etc ;
- en biologie, notamment pour le traitement des séquences d'ADN ;
- en finance : évaluation du risque, *trading*, *marketing* prédictif ;
- en droit : justice prédictive ;
- en médecine : les applications citées ci-dessus préfigurent une évolution vers une médecine prédictive et personnalisée, ainsi qu'une évolution des métiers (radiologie, dermatologie, anatomopathologie...).

6. Aspects éthiques

6.1 Introduction

Les succès récents ont fait entrer l'IA dans notre vie quotidienne. Mais l'IA suscite aussi des inquiétudes, parfois alimentées par des représentations relevant de scénarios de science-fiction. Elle demande ainsi un encadrement réglementaire pour garantir son caractère éthique.

L'éthique rappelle le fait que ce qui est technologiquement possible n'est pas toujours humainement souhaitable. Il s'agit donc d'identifier les risques et de s'assurer que le développement de l'IA s'effectue réellement au service de l'humanité. Le problème n'est pas nouveau et s'est posé de même pour l'énergie nucléaire, les manipulations génétiques et le clonage, etc.

L'Europe joue un grand rôle sur le plan de l'éthique. L'UE a mis en place le *High Level Group on Artificial Intelligence* qui a publié un document pour une IA de confiance (*Trustworthy AI*), utile et au service de l'être humain. Le livre blanc publié en février 2020 confirme cette position de l'UE en matière d'IA.

L'UNESCO s'est, de son côté, attachée à l'élaboration d'une norme mondiale sur l'éthique professionnelle de l'IA [8].

En février 2020, le Département de la défense (DoD) américain a défini des principes éthiques qui devront être respectés lors du développement de technologies intégrant de l'IA. La France, l'Allemagne, l'Europe, le Japon ont également publié des rapports sur cette question.

6.2 Protection des données personnelles

Le premier enjeu éthique est la protection des données personnelles manipulées : données génétiques, données biométriques, données concernant des infractions, des condamnations pénales et des mesures de sûreté ainsi que toutes données sur l'origine raciale ou ethnique, les opinions politiques, l'appartenance syndicale, etc.

La protection des données personnelles pose le problème de l'anonymisation des données collectées, afin d'empêcher l'identification des personnes concernées [H 5 537]. Un compromis est à trouver pour préserver l'anonymat, sans mettre en cause la pertinence du contenu.

L'Europe dispose du dispositif le plus avancé au monde. S'inspirant directement de la loi française Informatique et Libertés (1978), l'Union Européenne a introduit en 2018 le Règlement général de l'UE sur la protection des données personnelles, RGPD, déjà évoqué, qui est une référence en matière de protection des données à caractère personnel. Il renforce et unifie la protection des données pour les individus au sein de l'Union européenne. Son but est de permettre aux entreprises européennes de développer leurs activités numériques dans un contexte juridique égalitaire et compétitif. Comme c'est un règlement, il ne nécessite aucune transposition juridique dans les États membres. Les entreprises européennes deviennent responsables des données à caractère personnel qu'elles collectent et de leurs usages. Le RGPD vise trois objectifs :

- renforcer les droits des personnes ;
- responsabiliser les acteurs traitant des données ;
- crédibiliser la régulation (notamment par le biais d'amendes).

Le RGPD a déjà conduit à des amendes importantes infligées à des entreprises.

La justice américaine est également active. Le respect de la vie privée et la protection des données personnelles sont ainsi devenus des sujets d'actualité, notamment après les scandales liés à des opérateurs comme Google, Twitter, Alibaba ou Facebook (piratage de données personnelles, croisement de fichiers de clients, etc.).

Le Japon a rejoint l'Europe sur la protection des données et la Californie envisage une loi similaire.

Une tendance existe pour permettre aux forces de l'ordre d'accéder à des données personnelles pour la lutte contre la criminalité et le terrorisme. Ainsi, aux USA, le *Clarifying Lawful Overseas Use of Data Act (CLOUD Act)*, loi fédérale adoptée en 2018, fait l'objet de critiques car elle ouvre aux forces de l'ordre américaines un large accès aux données personnelles d'un individu sans que celui-ci en soit informé.

Le RGPD impose la nomination d'un DPO (*Digital Protection Officer*). Au sein de l'entreprise, le DPO joue le rôle d'un conseiller et d'un chef d'orchestre. C'est à lui de superviser la conformité de l'entreprise avec les engagements imposés par le RGPD. En complément du DPO, le *digital ethics officer* (DEO) disposera d'une double compétence numérique et juridique. Son rôle sera d'accompagner un déploiement éthique et respectueux des droits fondamentaux de l'IA au sein de son entreprise. Le DEO est pour le moment quasi inconnu, surtout en France, mais cela va rapidement évoluer.

6.3 Transparence des algorithmes

La question de la propriété des données et de l'ouverture de leur accès se pose également. Il s'agit d'atteindre un équilibre entre la protection et la valorisation de cette nouvelle richesse que sont les données et l'accès libre (*open data*), favorisant l'innovation. En France, la loi pour une République numérique a décidé de l'ouverture des données publiques, sans intervenir sur d'autres types de données. Cette loi, sans concerner spécifiquement l'IA, impose en particulier une exigence de transparence aux administrations quant aux traitements algorithmiques servant à prendre des décisions concernant les citoyens.

Le RGPD donne également à une personne le droit de demander l'explication et la justification d'une décision la concernant. Si les décisions en question sont des décisions de la justice ou de la police, on comprend que la transparence des systèmes d'IA est indispensable. Or il est actuellement très difficile d'expliquer comment un réseau neuronal profond parvient à une décision. Cela pose donc problème, et est l'une des raisons justifiant l'étude de modèles d'IA capables d'expliquer leurs résultats, un domaine de recherche en plein développement. Cela est utile en cas de défaillance, pour garder le sens de ce qui se passe dans un système, établir la confiance des utilisateurs et faciliter la communication humain – machine (cf. le projet américain *Explainable AI* évoqué au § 4).

6.4 Quelques domaines

■ Santé

La France a créé en 2018 le *Health Data Hub*, afin de centraliser les données provenant de sources multiples (les bases, les dossiers patients, la médecine de ville et les patients eux-mêmes) accessibles par une grande variété d'acteurs : professionnels de santé, chercheurs, acteurs économiques et enfin citoyens. Tout en permettant une meilleure connaissance, sans cesse actualisée, de l'état de santé des patients, cette collecte massive de données à caractère personnel présente des risques éthiques et juridiques. L'enjeu de l'anonymisation des données est bien sûr fondamental, pour que le suivi des comportements affectant la santé ne devienne pas un instrument de contrôle.

De même aux USA, le *Health Insurance Portability and Accountability Act* est une loi qui impose à tous les intervenants du secteur des soins de santé de protéger les informations détenues sur les patients. Cette loi édicte des normes dans 3 domaines : protection de la vie privée, encodage des données médicales et protection de la confidentialité, de l'intégrité et de la disponibilité des données médicales

■ Parole et image

La voix et l'image sont des données personnelles à part entière, comme l'avait précisé la CNIL dès 1996. À ce titre, elles relèvent du RGPD qui assure ainsi une protection juridique aux individus. Cette protection est bienvenue dans la mesure où le traitement de la parole et des images est devenu courant en IA.

On a vu qu'une caractéristique des systèmes d'IA est de nécessiter de très grandes quantités de données pour leur apprentissage. À chaque utilisation, la voix du locuteur est ainsi enregistrée dans le but louable d'améliorer les modèles utilisés, mais à son insu.... Là encore ressort l'importance de garantir l'anonymat des données, ainsi que du degré de confiance que l'on a dans les fournisseurs de ces services.

L'utilisation de l'IA pour l'identification de personnes à partir d'images peut poser problème en ce qui concerne le respect de la vie privée. La reconnaissance faciale est ainsi utilisée dans les aéroports (notamment en France), mais elle peut aussi servir à des fins de surveillance, de recherche de personnes portées disparues et de suivi d'individus en Chine (plusieurs dizaines de millions de caméras installées dans des lieux publics) et dans

différentes villes dans le monde (Los Angeles depuis 2009, Londres, etc.). Cela permet de réduire la criminalité, le vol ou la fraude et d'augmenter la sécurité des personnes. En Chine, le paiement par reconnaissance faciale est également opérationnel sur des plateformes de paiement électronique comme Alipay et Wechat.

Plusieurs États et villes ont décidé d'interdire l'utilisation de la reconnaissance faciale dans des lieux publics. L'État de Californie vient de décider un moratoire de trois ans en la matière. En France, l'application d'« Authentification en ligne certifiée sur mobile » (Alicem) lancée en mai 2019, vise à créer une « identité numérique » pour faciliter l'accès à certains services administratifs ou commerciaux sur Internet pour les détenteurs d'un passeport biométrique. L'objectif est de créer un document virtuel officiel d'identité pour « s'identifier électroniquement et s'authentifier auprès d'organismes publics ou privés ». La vérification de l'identité est effectuée par un dispositif de reconnaissance faciale. Les données personnelles stockées sont codées avec un maximum de sécurité. Alicem a engendré un certain nombre de réticences sur les plans éthiques et juridiques, notamment en ce qui concerne le délai de conservation des données et la notion de *consentement libre et non-imposé* prévue par le RGPD mais non respectée par ce système.

D'autres travaux de recherche tendent à conjuguer parole et image pour détecter l'état émotionnel d'une personne. L'idée est de reconnaître quelques émotions fondamentales (joie, peur, tristesse, colère, dégoût et surprise) à partir de signaux expressifs, après une phase d'apprentissage, fondée comme toujours sur un ensemble important de données d'exemples. Les résultats sont encore modestes mais soulèvent déjà des réticences sur l'utilisation de systèmes qui concerneraient l'intimité affective des personnes. En Chine, l'annonce de l'utilisation d'un casque pour contrôler l'état émotionnel d'une personne à partir de signaux EEG a été faite récemment.

Il ne faut pas condamner en bloc ces technologies, car elles sont d'une grande utilité dans de nombreux domaines : indexation d'archives visuelles (films, vidéos, émissions TV) facilitant la recherche, aide aux enquêtes policières, authentification de personnes, lutte contre le piratage et l'usurpation d'identités. Mais il faut lutter contre les applications contraires à la loi et aux droits des personnes.

■ Robotique

Depuis les premiers robots humanoïdes, la robotique a accompli d'énormes progrès, et des robots dits intelligents, plus ou moins autonomes, sont apparus dans divers domaines d'activité : médecine, transports, industrie, agriculture, militaire, etc. En 1942, I. Asimov avait énoncé trois « lois de la robotique », selon lesquelles un robot ne peut notamment pas attenter à la sécurité d'un être humain et doit obéir à un être humain, sauf en cas de conflit avec le principe précédent. Ces lois semblent supposer que les robots comprennent le monde qui les entoure et sont dotés de conscience, ce qui n'est pas le cas, même si des laboratoires travaillent sur ce thème.

Pour l'instant, il faut établir un cadre éthique des robots et des normes intégrant la protection des valeurs humaines, afin de combler le vide législatif actuel et d'éviter de possibles dérives. Le problème n'est pas simple, car les valeurs éthiques et les sensibilités varient selon les pays et les civilisations alors que les normes et les standards doivent être universels... Les armes autonomes à base d'IA présentent un danger majeur. Certains pays financent des programmes pour développer l'armement à base d'IA : drones de combat, chars, armes létales autonomes, etc. Plusieurs appels ont été lancés pour un contrôle strict de tels programmes, malgré l'opposition de certains pays. Mais l'IA est appelée à jouer un rôle grandissant dans les activités militaires dans de nombreux secteurs : traitement et sélection des données, aide à la décision, prédiction de comportements, combat collaboratif, avec des principes rappelés récemment par la ministre française des

armées : respect du droit international et permanence du contrôle humain.

Une démarche précurseuse se trouve dans la résolution du Parlement européen sur la robotique de 2017, concernant des règles de droit civil sur la robotique. On y insiste « *sur le principe de transparence, à savoir qu'il devrait toujours être possible de fournir la justification rationnelle de toute décision prise avec l'aide de l'IA* ». La résolution préconise qu'un robot soit doté d'une « boîte noire » contenant les données sur chaque opération réalisée. Elle envisage également l'adoption de règles spéciales dans des domaines tels que les véhicules autonomes, les drones, les robots de compagnie et de soins à la personne, et les robots médicaux. Cela rejoint une tendance nouvelle de conception de robots éthiques, capables d'évaluer les conséquences de leurs actions [S 7 900].

■ Véhicule autonome

Les aspects scientifiques et techniques liés à la conception de véhicules autonomes font l'objet de nombreuses études théoriques et d'expérimentations en situation réelle : automobile, camion, train, bateau, tracteur agricole, avion... [S 7 819]

Tout aussi importants sont les aspects éthiques, psychologiques, politiques et juridiques. La convention de Vienne a été adaptée en 2016 de façon à autoriser les systèmes de conduites automatisées « *si ces technologies sont conformes aux réglementations de l'ONU ou peuvent être contrôlées et désactivées par le conducteur* ». La législation a déjà été modifiée dans certains États des USA (Nevada, Californie), en Grande-Bretagne, etc., mais le chemin est encore long vers un système international cohérent en termes de règles de conduite et de partage de responsabilités. La question de la détermination des degrés de responsabilité est en effet centrale. Si une voiture autonome provoque un accident, qui est considéré comme responsable ? Le constructeur du véhicule, l'ingénieur qui a développé le système de conduite autonome, l'entraîneur qui a réalisé l'apprentissage du système d'IA en lui fournissant les données adéquates, le propriétaire de la voiture ou la personne assise à la place du conducteur ? La responsabilité sera partagée entre ces personnes, selon des modalités qui restent à définir. Il faudra modifier le droit pour prendre en compte ce type de partage.

Pour ce qui concerne la **responsabilité civile**, une adaptation du dispositif est nécessaire. Sur l'aspect financier, on escompte que la conduite autonome permettra de diminuer la fréquence et la gravité des accidents. Maintenir le principe d'une couverture obligatoire par les assurances est donc tout à fait envisageable.

Pour ce qui concerne la **responsabilité pénale**, la question demande également une réponse, mais elle n'est au fond pas si éloignée de questions déjà jugées, relatives à la responsabilité pour faute des concepteurs et fabricants de systèmes automatisés quand ces systèmes présentent des défauts. De tels systèmes sont déjà présents à bord des véhicules, comme l'ABS ou le freinage automatique d'urgence, et ont fait l'objet de procès.

6.5 Premier bilan

La réflexion sur l'éthique des machines est essentielle. Le projet *Moral Machine* du MIT est un exemple intéressant. Ce projet tend à étudier les réactions du grand public devant le problème complexe de la distribution du risque par les véhicules autonomes. L'irruption de l'IA dans le quotidien soulève des questions importantes relatives à la nuisance des systèmes d'IA (cf. lois d'Asimov pour les robots), au statut moral des machines, à la détermination de responsabilités et aux propriétés requises d'un système du fait de son rôle social et médical potentiel : prédictibilité, transparence à l'inspection...

On peut imaginer une IA éthique « par conception » qui prendrait en compte les considérations éthiques dans toute la chaîne, depuis la conception d'un produit jusqu'à l'utilisateur. Pour cela il faut intégrer la dimension éthique dans la formation des les

écoles d'ingénieurs et les universités, en complément des volets scientifiques et techniques de l'IA.

La technologie nous a rendus, parfois à notre insu, éminemment observables, voire manipulables. Les enjeux éthiques de l'IA sont donc clairs, et justifient un encadrement des pratiques de recherche et des développements industriels. L'éthique est devenue une nécessité pour le développement de l'IA, mais les règles ne sont pas les mêmes partout. On peut schématiquement distinguer trois approches qui résument les difficultés à adopter des principes éthiques de l'IA :

- les États-Unis considèrent les données comme un élément commercial ;
- la Chine utilise les données afin de surveiller et définir les « bons citoyens » ;
- l'Europe, avec le RGPD et d'autres initiatives, entreprend de protéger ses citoyens et leurs données.

En coordonnant les efforts de chaque État afin d'organiser une filière européenne de l'IA et de créer des pôles de compétences, l'Europe a un rôle à jouer, comme elle l'a déjà montré avec le RGPD.

7. Conclusion

Depuis plusieurs décennies, l'IA a conduit au développement d'applications opérationnelles dans de nombreux domaines tels que le diagnostic et l'aide à la décision avec les systèmes à bases de connaissances ou les réseaux neuronaux, ainsi que la reconnaissance de formes et de la parole avec les modèles stochastiques. L'introduction des DNN a permis d'atteindre des niveaux de performance inégalés qui ont valu une couverture médiatique sans précédent dans quasiment tous les secteurs d'activité. Comme toujours, le transfert de la recherche en laboratoire vers les applications réelles nécessite des investissements conséquents en temps et en argent. Le cheminement est long entre une première annonce et un véritable produit commercialisé. En médecine, notamment, le déploiement d'applications implique un mécanisme d'évaluation complexe, notamment clinique. Enfin, le succès d'une application implique une véritable confiance dans l'IA ainsi qu'une claire identification des responsabilités (cf. la voiture autonome en cours d'expérimentation et de mise au point).

On assiste actuellement de ce fait à un accroissement considérable des investissements privés en intelligence artificielle par les industriels du domaine : Google, Facebook, IBM, Microsoft, Amazon, Adobe, Yandex, Baidu... Les startups ont également proliféré dans le monde entier. Elles sont ainsi en 2021 14 fois plus nombreuses qu'en l'an 2000 en Amérique du Nord, selon l'index de l'IA de l'Université de Stanford. Leur nombre total est de plus de 1 700 dans le monde. Le chiffre d'affaires de l'IA se situe autour de 2 milliards de dollars américains en 2018 (et pourrait atteindre près de 90 milliards en 2024, d'après le cabinet d'analyse Tractica), tandis que plus de 55 000 brevets, relatifs à l'IA au sens large, ont été déposés et que plus de 130 000 publications scientifiques ont été publiées cette même année, selon l'Organisation Mondiale de la Propriété Industrielle, OMPI.

Parallèlement, les acteurs institutionnels de la recherche lancent de grands projets : en Europe, aux États-Unis (cf. le projet *Quest for Intelligence* du MIT tendant à coupler la compréhension de ce que sont l'intelligence et la réalisation de nouveaux systèmes utiles à l'humanité, ou encore la création du *Joint Artificial Intelligence Center* par le Ministère de la Défense dans le cadre des *National Mission Initiatives*, grands projets s'attaquant aux grands défis actuels, intégrant les aspects éthiques et humanitaires des recherches).

La montée en puissance de l'IA pose la question des conséquences en matière d'emploi. L'IA permet d'automatiser de nombreuses tâches répétitives. Nous assistons à une évolution importante du travail car l'IA supprimera ou transformera très certainement des métiers dans de nombreux secteurs d'activité, tout en nécessitant une évolution forte des qualifications (dans la maintenance, la gestion des données, la création assistée, etc.) et des compétences, y compris pour des métiers qui sont encore à imaginer.

Introduction à l'intelligence artificielle

par **Jean-Paul HATON**

Professeur émérite

LORIA – Institut Universitaire de France – Université de Lorraine – Nancy, France

Sources bibliographiques

- [1] HATON (J.-P.) *et al.* – *Le raisonnement en intelligence artificielle - Modèles, techniques et architectures pour les systèmes à bases de connaissances*, InterEditions (1991).
- [2] TURING (A.). – *Computing Machinery and Intelligence*, *Mind*, 49, pp. 433-460 (1950).
- [3] LE CUN (Y.) *et al.* – *Deep learning*, *Nature*, vol. 521, pp. 436-444 (2015).
- [4] HATON (J.-P.). – *La parole numérique*, Académie Royale de Belgique, Collection Poche, n° 79 (2016).
- [5] CHANDRA (V.) et HAREENDRAN (A.). – *Artificial Intelligence and Machine Learning*, PHI Learning, 2014.
- [6] RUSSELL (S.) et NORWIG (P.). – *Artificial Intelligence: A Modern Approach (3d edition)*, Pearson (2015).
- [7] MINSKY (M.) et PAPERT (S.). – *Perceptrons : an introduction to computational geometry*, Cambridge (1969).
- [8] European Commission. – *On Artificial Intelligence – A European approach to excellence and trust*. Livre blanc (2020). https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

À lire également dans nos bases

- FAUGEROUX (O.), GRIEU (S.), TRAORE (A.), BODNAR (J.-L.) et CLAUDET (B.). – *Outils de l'intelligence artificielle appliqués au CND* [R 1 403] Mesures-Analyses, Contrôle non destructif (2013).
- HATON (J.P.). – *Reconnaissance automatique de la parole* [H 3 728] Technologies de l'information. Documents numériques, gestion de contenu (2018).
- CANU (S.). – *Machines à noyaux pour l'apprentissage statistique* [TE 5 255] Technologies de l'information, Technologies logicielles, architecture des systèmes (2007).
- MOUTARDE (F.). – *Apprentissage statistique supervisé* [H 5 010] Innovations technologiques, Innovations en électronique et TIC (2019).
- MOUTARDE (F.). – *Apprentissage statistique non supervisé* [H 5 012] Innovations technologiques, Innovations en électronique et TIC (2020).
- ARTIERES (T.). – *Reconnaissance de formes* [AF 1 510] Technologies de l'information. Technologies logicielles. Architecture des systèmes (2011).
- DEFOUR (D.) et ETIEMBLE (D.). – *Processeurs graphiques totalement programmables (GPU)* [H 1 013] Technologies de l'information. Technologies logicielles. Architecture des systèmes (2011).
- SONDECK (L.P.). – *Anonymisation des données, une nécessité à l'ère du RGPD* [H 5 537] Technologies de l'information. Sécurité des systèmes d'information (2019).
- TESSIER (C.). – *Conception et usage des robots : quelques questions éthiques* [S 7 900] Automatique-robotique. Robotique (2016).
- BONNIFAIT (P.) et ZINOUNE (C.). – *Introduction aux techniques de navigation autonome pour les véhicules intelligents* [S 7 819] Technologies de l'information. Technologies radar et applications (2021).
- ETIEMBLE (D.). – *Supports matériels pour les réseaux neuronaux profonds* [H 1 098] Technologies de l'information. Technologies logicielles. Architecture des systèmes (2021).

Revues

- AI Magazine (USA)
 Artificial Intelligence (NL)
 Bulletin de l'AFIA, Association Française d'Intelligence Artificielle (F)
 IEEE Transactions on Knowledge and Data Engineering (USA)
- Journal of Intelligent Manufacturing (GB)
 IEEE Transactions on Neural Networks and Learning Systems (USA)
 Neural Networks (GB)
 IEEE Transactions on Pattern Analysis and Machine Intelligence (USA)

Sites Internet

- AFIA Association Française d'IA : <https://afia.asso.fr/>
 ECCAI European Coordinating Committee for Artificial Intelligence : <https://eccai.org/>
 AAAI Association for the Advancement of Artificial Intelligence : <https://www.aaai.org/>
- ENNS European Neural Network Association : <https://e-nns.org/>
 International Neural Network Society INNS : <https://www.inns.org/>

Gagnez du temps et sécurisez vos projets en utilisant une source actualisée et fiable



RÉDIGÉE ET VALIDÉE
PAR DES EXPERTS




MISE À JOUR
PERMANENTE



100 % COMPATIBLE
SUR TOUS SUPPORTS
NUMÉRIQUES



SERVICES INCLUS
DANS CHAQUE OFFRE

- > + de 340 000 utilisateurs chaque mois
- > + de 10 000 articles de référence et fiches pratiques
- > Des Quiz interactifs pour valider la compréhension 

SERVICES ET OUTILS PRATIQUES



Questions aux experts*

Les meilleurs experts techniques et scientifiques vous répondent



Articles Découverte

La possibilité de consulter des articles en dehors de votre offre



Dictionnaire technique multilingue

45 000 termes en français, anglais, espagnol et allemand



Archives

Technologies anciennes et versions antérieures des articles



Info parution

Recevez par email toutes les nouveautés de vos ressources documentaires

*Questions aux experts est un service réservé aux entreprises, non proposé dans les offres écoles, universités ou pour tout autre organisme de formation.

Les offres Techniques de l'Ingénieur

INNOVATION

- Éco-conception et innovation responsable
- Nanosciences et nanotechnologies
- Innovations technologiques
- Management et ingénierie de l'innovation
- Smart city – Ville intelligente

MATÉRIAUX

- Bois et papiers
- Verres et céramiques
- Textiles
- Corrosion – Vieillessement
- Études et propriétés des métaux
- Mise en forme des métaux et fonderie
- Matériaux fonctionnels. Matériaux biosourcés
- Traitements des métaux
- Élaboration et recyclage des métaux
- Plastiques et composites

MÉCANIQUE

- Frottement, usure et lubrification
- Fonctions et composants mécaniques
- Travail des matériaux – Assemblage
- Machines hydrauliques, aérodynamiques et thermiques
- Fabrication additive – Impression 3D

ENVIRONNEMENT – SÉCURITÉ

- Sécurité et gestion des risques
- Environnement
- Génie écologique
- Technologies de l'eau
- Bruit et vibrations
- Métier : Responsable risque chimique
- Métier : Responsable environnement

ÉNERGIES

- Hydrogène
- Ressources énergétiques et stockage
- Froid industriel
- Physique énergétique
- Thermique industrielle
- Génie nucléaire
- Conversion de l'énergie électrique
- Réseaux électriques et applications

GÉNIE INDUSTRIEL

- Industrie du futur
- Management industriel
- Conception et production
- Logistique
- Métier : Responsable qualité
- Emballages
- Maintenance
- Traçabilité
- Métier : Responsable bureau d'étude / conception

ÉLECTRONIQUE – PHOTONIQUE

- Électronique
- Technologies radars et applications
- Optique – Photonique

TECHNOLOGIES DE L'INFORMATION

- Sécurité des systèmes d'information
- Réseaux Télécommunications
- Le traitement du signal et ses applications
- Technologies logicielles – Architectures des systèmes
- Sécurité des systèmes d'information

AUTOMATIQUE – ROBOTIQUE

- Automatique et ingénierie système
- Robotique

INGÉNIERIE DES TRANSPORTS

- Véhicule et mobilité du futur
- Systèmes aéronautiques et spatiaux
- Systèmes ferroviaires
- Transport fluvial et maritime

MESURES – ANALYSES

- Instrumentation et méthodes de mesure
- Mesures et tests électroniques
- Mesures mécaniques et dimensionnelles
- Qualité et sécurité au laboratoire
- Mesures physiques
- Techniques d'analyse
- Contrôle non destructif

PROCÉDÉS CHIMIE – BIO – AGRO

- Formulation
- Bioprocédés et bioproductions
- Chimie verte
- Opérations unitaires. Génie de la réaction chimique
- Agroalimentaire

SCIENCES FONDAMENTALES

- Mathématiques
- Physique Chimie
- Constantes physico-chimiques
- Caractérisation et propriétés de la matière

BIOMÉDICAL – PHARMA

- Technologies biomédicales
- Médicaments et produits pharmaceutiques

CONSTRUCTION ET TRAVAUX PUBLICS

- Droit et organisation générale de la construction
- La construction responsable
- Les superstructures du bâtiment
- Le second œuvre et l'équipement du bâtiment
- Vieillessement, pathologies et réhabilitation du bâtiment
- Travaux publics et infrastructures
- Mécanique des sols et géotechnique
- Préparer la construction
- L'enveloppe du bâtiment
- Le second œuvre et les lots techniques

OFFRE



Technologies logicielles Architectures des systèmes

Traiter, acheminer et sécuriser l'information : un enjeu économique et stratégique d'actualité pour l'entreprise

Ref : TIP402WEB

PRÉSENTATION

Une analyse détaillée des **systèmes d'exploitation, des langages de programmation et des techniques de production du logiciel,**

Les **concepts et fondements** en matière de gestion de **bases de données** et les **stratégies de conception des systèmes d'information,**

Un panorama complet des **protocoles de communication, de l'architecture de systèmes** et des **techniques de sécurisation des réseaux,**

Des outils pour **améliorer la performance des processeurs et des systèmes.**

VOTRE COMMANDE :

Référence	Titre de l'ouvrage	Prix unitaire H.T	Qté	Prix total H.T
TIP402WEB	Technologies logicielles Architectures des systèmes	2 250 €	1	2 250 €
Total H.T en €				2 250 €
T.V.A : 5,5%				123,75 €
Total TTC en €				2 373,75 €

VOS COORDONNÉES :

Civilité M. Mme

Prénom _____

Nom _____

Fonction _____

E-mail _____

Raison sociale _____

Adresse _____

Code postal _____

Ville _____

Pays _____

Date :

Signature et cachet obligatoire

CONDITIONS GÉNÉRALES DE VENTE

Conditions générales de vente détaillées sur simple demande ou sur www.technique-ingenieur.fr

Si vous n'êtes pas totalement satisfait, vous disposeriez d'un délai de 15 jours à compter de la réception de l'ouvrage pour le retourner à vos frais par voie postale. Livraison sous 30 jours maximum.