

## Voyage au cœur de l'IA – Hors-série Libération (10/2017)

### **p. 004 :**

- déclaration des droits de l'Homme et de la Machine ;
- société de surveillance généralisée ;
- algorithmes faussés par des biais sociaux ou ethniques.

### **p. 005 :**

- IA : termes introduits par les chercheurs américains John McCarthy et Marvin Minsky en 1955 pour une demande de subvention à la Fondation Rockefeller ;
- promesse d'un traducteur universel en 1965.

### **p. 006 :**

- IA à remplacer par « compétences artificielles » ;
- car échecs, jeu de go, imagerie, traduction : compétences spécialisées ;
- toute compétence nouvelle implique un nouvel apprentissage long et onéreux ;
- intelligence humaine infiniment adaptable acceptant l'imprécision ;
- est liée au corps et à ses interactions avec un environnement collectif ;
- réseaux neuronaux impossibles à contrôler ligne à ligne : risque d'erreurs liés à des biais dus à la nature des données fournies ;
- de sorte que l'on ne peut pas accorder une plus grande confiance à l'IA qu'à l'être humain.
- monopoles mondiaux (GAFA : Google, Apple, Facebook, Amazon) en contradiction avec une société libérale politiquement et économiquement.

### **p. 007 :**

- ces monopoles transcendent les états en peine de les assujettir à la fiscalité ;
- éventuelle destruction massive d'emplois ou emplois « nouveaux » ;
- la crainte n'est pas la meilleure façon de faire face...

### **p. 011 :**

- l'IA chamboulerait tout : accorder des prêts immobiliers, traductions diverses, conduire un véhicule ;
- elle serait donc magique et on en accepterait les prouesses sans en comprendre les mécanismes ;
- les chercheurs honnêtes savent que ceci n'est encore qu'un mirage et parlent « d'apprentissage ».

### **p. 012 :**

- l'intelligence qui permet de comprendre le monde qui nous entoure, de s'adapter ou anticiper les changements est l'objectif à atteindre ;
- cet objectif n'est pas irréaliste mais encore lointain et cette chimère prospère dans un inconscient nourri de science-fiction ;
- entraînés par une base de données pertinente, les logiciels typés IA basés sur l'apprentissage profond commencent à faire leur preuve à grande échelle (erreur de quelques % au concours « ImageNet challenge »).

### **p. 013 :**

- recherche ouverte ; conférences annuelles : International Conference on Machine Learning (ICML) et Neural Information Processing Systems (NIPS) ;
- travaux utilisés par les industriels dans plusieurs années.

### **p. 014 :**

- apprentissage par renforcement pour « résoudre l'intelligence » ;
- apprentissage faiblement ou non supervisé ;
- réseaux à capsules car somme astronomique de données pour un apprentissage ;
- différence à faire avec une recherche appliquée qui tente de rendre concrètes les possibilités d'apprentissage profond dans tous les domaines ;
- DeepMind veut produire une vraie intelligence artificielle (délai de 10 à 15 ans) ;
- « les tâches pour lesquelles on construit des IA sont actuellement confiées à des humains. Quand ces tâches sont automatisées, le travail humain devient superflu conduisant à un chômage de masse nécessitant une redistribution des bénéfices » (Blaise Agüera y Arcas de Google Brain).

**p. 015 :**

- les masses de données pour entraîner les réseaux neuronaux ne sont pas neutres ;
- la validité de l'outil informatique ne vient pas de la programmation logicielle mais de la justesse des données d'apprentissage.

**p. 017 :**

- avec l'IA, on évoque presque toujours l'apprentissage machine grâce à des réseaux de neurones ;
- rétropropagation du gradient ;
- réseau convolutif pour diminuer le volume d'informations (filtres et découpage d'une image, par ex.).

**p. 022 :**

- comment dépasser les limites d'apprentissage actuelles ?
- un progrès dans les produits vient d'un progrès technologique quelques années auparavant et d'une avancée scientifique 10 à 20 ans plus tôt ;
- les règles de déduction logiques sont incompatibles avec l'apprentissage ;
- on remplace la logique par des opérations sur des vecteurs.

**p. 023 :**

- la plupart du savoir issu des données réside dans les ordinateurs, pas dans la tête des gens ;
- l'apprentissage profond permet de déléguer le traitement des données au système d'apprentissage lui-même ;
- oui, beaucoup de données sont biaisées car issues d'actions humaines ;
- aux USA, des études ont montré que les décisions d'un juge pouvaient être influencées par sa situation affective, des résultats sportifs, l'heure de la journée : plus d'IA réduirait ces biais humains ;
- il y a deux types de systèmes entraînés : ceux qui prennent des décisions rapides et très nombreuses (conduite autonome) où la machine accomplit des tâches qui relèvent du subconscient chez l'homme et ceux nécessitant des explications car pouvant affecter la vie des gens. Pour l'instant, les systèmes ayant nécessité un apprentissage prennent des décisions rapides.

**p. 024 :**

- la presse fait état d'employés japonais d'une agence d'assurance-vie sont remplacés par un robot.

**p. 025 :**

- la première révolution industrielle était fondée sur l'automatisation des muscles humains et animaux mais avec l'IA, il s'agit des cerveaux et les machines deviennent elles-mêmes des travailleurs : robots dans les usines et algorithmes dans les bureaux ;
- économistes et historiens considèrent que l'homme s'adaptera comme par le passé : si la croissance des besoins n'a pas de fin, alors la nécessité de travail non plus ;
- « useless people » : nouvelle classe d'inutiles : pas seulement sans emploi mais inemployables ;
- Bill Gates est favorable à une répartition des richesses produites par les machines via une taxe sur les robots tandis qu'Elon Musk (Tesla) appuie le principe d'un revenu universel.

**p. 026 :**

- il semble probable (à plus de 90%) que les télémarketeurs soient remplacés par des « chatbots » et que la quasi-totalité des postes de caissières et caissiers disparaisse ;
- le phénomène de relocalisation dû à l'utilisation de robots pourrait priver d'emplois des pays tels que Inde, Chine, Ethiopie, petites mains de l'industrie textile en Asie ;
- dans l'histoire humaine, la création l'a toujours emporté sur les destructions (d'emplois) ;
- les cols blancs seraient aussi menacés : 600 traders remplacés par le trading électronique à haute fréquence (99% des achats et de vente chez Goldman Sachs) et 200 ingénieurs pilotant des algorithmes pour des salaires cinq fois inférieurs.

**p. 027 :**

- l'enjeu serait d'inverser le mouvement actuel d'appauvrissement du contenu des emplois créés : activités de service peu qualifiées ;
- la plupart des métiers nécessite des aptitudes à l'empathie ; seuls ne déclineraient donc que les emplois ne nécessitant ni matière grise (d'où éducation) ni empathie (travaux de manutention, par ex.).

**p. 030 :**

- l'IA a des dimensions politique, sociétale, éthique, culturelle ;
- seuls les Chinois défient aujourd'hui la suprématie américaine ;
- l'Europe est attentive à une approche équilibrée de l'IA en se souciant dorénavant et déjà des effets ravageurs que peuvent avoir certaines technologies sur l'emploi et le tissu démocratique ainsi que sur le respect de la vie privée et pas uniquement sur l'efficacité des marchés.

**p. 031 :**

- avoir un grand marché améliore le produit car les données sont centrales dans l'IA ;
- quant à la place de l'humain dans l'IA, une analogie peut être faite avec les MOOC pour la formation en ligne qui prétendait résoudre les problèmes d'éducation en la rendant plus accessible (en Afrique...) ;
- l'IA va impacter tous les citoyens et tous les domaines : politique, santé, éducation, industrie, police, culture.

**p. 032 :**

- intelligence : faculté humaine. Il faut donc trouver un autre mot.

**p. 033 :**

- outil : il n'y a pas plus de compétition entre l'homme et la machine (utilisant l'IA) qu'entre le trompettiste et la trompette ou entre le maçon et la truelle ;
- nécessité d'ajouter des prothèses à notre intelligence parfois défaillante : bibliothèques, machines à calculer, ordinateurs, algorithmes ;
- la frontière entre ce que font les humains et les machines ne cesse de se déplacer ; d'où la transformation de concepts structurants : travail, priorité de vie, vie privée, croyance en l'existence d'un « propre de l'homme » qui le distinguerait de l'animal et de la machine.

**p. 040 :**

- science-fiction : le premier androïde est porté à l'écran dans un court-métrage de Méliès en 1900 (Coppelia, la poupée animée) ;
- on trouve de nouveau une femme robot en 1927 dans Métropolis (Fritz Lang).

**p. 042 :**

- en 1921, l'écrivain tchèque Karel Kapek invente le mot robot : de « robota », travail forcé en tchèque ;
- une importance est accordée aux androïdes en répliquant l'image de l'homme considérés comme un menace (Blade Runner de Ridley Scott en 1982) ;
- le terme « robotique » est utilisé en 1942 par Asimov dans sa nouvelle Cercle Vicieux ;
- dans la série Les Robots, Asimov imagine des principes des rapports entre l'Homme et la machine ;
- premier robot industriel (cerveau = ordinateur ; muscles = servomécanismes) mis en place en 1961 chez General Motors pour la manipulation de pièces chaudes et pour les entasser. Il est à noter qu'au tout début, il n'y avait même pas d'ordinateur mais un simple disque magnétique.

**p. 044 :**

- avec 2001, l'Odyssée de l'Espace, se poursuivent les scénarios manichéens où l'IA échappe au contrôle de l'humanité et devient destructrice ;
- en 2015, avec Ex Machina, Alex Garland offre une autre vision : Ava, une androïde, a-t-elle ou non une conscience ?

**p. 046 :**

- la loi de Moore postule un doublement de la puissance de calcul tous les dix-huit mois ;
- la singularité technologique fait l'hypothèse que l'évolution de l'informatique atteindra un point au-delà duquel il ne sera plus possible de l'appréhender ;
- c'est ce qui est décrit en 2007 par l'écrivain Vernor Vinge dans Rainbows end ;
- Ghost in the shell fait la part belle au transhumanisme où l'homme est amélioré par la machine allant jusqu'à des humains greffés de prothèses artificielles et autres androïdes sexués.

**p. 047 :**

- la spéculation la plus outrancière du transhumanisme va jusqu'au Uploading où le contenu du cerveau humain pourrait être transféré sur un support greffé sur un corps robotique inaltérable ;

- les prémices d'un tel transfert pourraient se réaliser en 2040 et l'homme pourrait ainsi s'affranchir des limites assignées au corps.

**p. 049 :**

- les Etats-Unis sont pilotés par des recherches de ruptures et de nouvelles frontières : conquête de l'Ouest, de l'Espace, Internet et transformation de l'homme ;
- ils pensent que la technologie – si on n'influe pas sur son développement - peut résoudre tous les problèmes et sont moins sensibles aux effets négatifs ;
- c'est une approche individualiste qui ne se préoccupe guère du reste de la société.

**p. 054 :**

- déjà, les conséquences d'une disparition brusque du numérique : avions qui tombent ; ascenseurs bloqués ; que devient l'argent ?
- dans la littérature, le futur est une zone méprisée car ce qui compte c'est le « moi, je » et le passé.

**p. 055 :**

- on vit la fin d'un monde qui a débuté en 1989 avec la chute du mur de Berlin suivie de dix ans de globalisation (ou mondialisation) et le XXIème siècle est arrivé avec le 11 septembre : l'Amérique s'était créé son nouvel ennemi.

**p. 056 :**

- responsabilité des erreurs commises par les robots ou pour une utilisation imprudente ou criminelle de ces robots ?

**p. 057 :**

- première loi : « Un robot ne peut porter atteinte à un être humain ni, restant passif, laisser cet être humain exposé au danger » ;
- reconnaître qu'un être humain est exposé à un danger exige une capacité d'analyse et une compréhension générale des situations que n'ont pas les robots qui agissent de manière spécialisée ;
- nécessité de programmer des échelles (ou systèmes) de valeur : choix entre une collision avec un arbre ou un piéton, par exemple, pour une conduite d'un véhicule autonome (règle pour traiter le dilemme) ;
- seconde loi formulée en 1940 par l'écrivain Isaac Asimov : « Un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi » ;
- troisième loi : « Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi ».

**p. 060 :**

- soumis à dix taches d'encre constituant le test de Rorschach, les réponses du système d'interprétation d'images DenseCap sont analysées ensuite comme typiques : « d'un adolescent aux processus cognitifs rudimentaires ».

**p. 061 :**

- en 2015, l'équipe de recherche Google Brain (qui compte 300 personnes) a rendu libre le moteur d'apprentissage profond TensorFlow.

**p. 062 :**

- Greg Corrado, l'un des fondateurs de cette entreprise, présente l'IA non pas comme une fonctionnalité mais comme une nouvelle matière brute telle le verre ou l'aluminium ;
- en collaboration avec des hôpitaux en Inde, son équipe travaille sur le dépistage précoce de la rétinopathie diabétique grâce à l'analyse de clichés d'yeux ;
- aux USA, la santé repose sur l'assurance privée et est confrontée à un manque de médecins : pour éviter des dépenses éventuellement inutiles, une application (Doc.AI) peut aider à faire le tri pour savoir s'il faut consulter (renseignements sur la personne et résultats d'analyses sanguines et génétiques) ;
- si Google et Facebook font la course en tête, c'est qu'ils disposent de quantités colossales de données pour leurs systèmes d'apprentissage automatique.

**p. 063 :**

- Greg Corrado reconnaît que « les avancées actuelles posent des questions sociales importantes. Il faut des discussions sur les usages de ces technologies entre législateurs, philosophes et techniciens.

**p. 064 :**

- demain, avec les progrès de l'IA, il y aura sans doute de nouveaux emplois mais aussi plus de chômage. C'est ce que pense Peter Eckersley à l'Electronic Frontier Foundation qui ajoute que tout est piratable.
- à San Francisco, la poussée de Techies a fait exploser le prix de l'immobilier et, de l'autre côté de l'avenue Van Ness, des exclus du système transportent le peu qu'ils possèdent dans des caddies de supermarché ;
- dans la Silicon Valley, l'idée d'un revenu universel est populaire.

**p. 066 :**

- un des projets de Google Brain, Magenta, est le développement d'algorithmes qui génèrent de l'art et de la musique ;
- la limite est celle que nous nous fixerons.

**p. 069 :**

- apprendre à détecter des événements dans des articles de presse ou ceux de l'agence AFP : le texte d'origine est passé dans un logiciel qui sépare les mots et identifie leur nature grammaticale. Puis à chaque mot est affecté un ensemble de vecteurs (word embeddings) transmis à un réseau de neurones. Il reste alors à l'entraîner à extraire des événements de manière supervisée : suivant les réponses, on atténue ou on renforce le « raisonnement » ;
- extraction de données médicales : par exemple, parmi toutes les publications, un tel système peut aider un oncologue à extraire les informations utiles pour lui ;
- dans le domaine juridique, cela est utile pour tenir compte de la jurisprudence ;
- actuellement, dans le domaine du Machine Learning (c'est principalement la presse généraliste qui parle d'IA), il ne s'agit que de gagner du temps dans l'exploitation de données.

**p. 070 :**

- aide à la décision dans un environnement de catastrophe naturelle ou d'évènement inattendu (dans un stade).

**p. 071 :**

- détection des comportements non verbaux (gestes, expression faciale) d'un être humain afin qu'un agent virtuel conversationnel autonome puisse adapter son apparence (sourire) et sa conversation ;
- un agent doit avoir l'air de savoir de quoi il parle (compétence) et inspirer confiance ;
- le projet est de développer un agent conversationnel qui jouerait le guide dans les musées.

**p. 071 :**

- dans l'apprentissage supervisé, on fournit des bases de données (images, par ex.) aux réseaux de neurones ;
- ces données labellisées réclament beaucoup de temps de préparation car elles sont étiquetées ('chien', 'chat'...) : les classes sont donc prédéfinies ;
- dans le Machine Learning non supervisé, des classes ne sont pas prédéfinies et les données sont utilisées sans label ;

**p. 072 :**

- une fois les vecteurs des mots obtenus, on peut faire la somme des vecteurs d'une phrase et détecter l'émotion : discours de haine dans les messages postés en ligne, commentaires positifs ou négatifs (restaurant), etc.

**p. 073 :**

- étude de l'acquisition du langage chez l'humain car, encore actuellement, des transcriptions (faites par des experts humains) sont synchronisées avec les pistes audio ;
- le but est de réduire au maximum la quantité de transcription lors de l'apprentissage par une machine (réseaux de neurones) d'une langue, y compris en tenant compte des changements dans la voix.

**p. 074 :**

- cela fait longtemps que les machines nous dépassent : automobiles qui se déplacent plus vite que nous courrons ; avions alors que nous ne pouvons nous élever à grand peine que de deux mètres ; téléphone et radio pour la transmission des sons à distance ; puissance de calcul des ordinateurs ; diagnostics à partir de photos (mélanomes), etc.
- les ingénieurs et mathématiciens n'ont jamais acquis un grand pouvoir politique même s'ils sont à l'origine d'empires industriels ;
- les compétences simulées par l'IA demeurent spécialisées dans des tâches spécifiques ;

**p. 075 :**

- cependant, par facilité ou « démission », il pourrait y avoir délégation aux machines : primes d'assurance-vie, sélection d'étudiants à l'entrée à l'université, évaluation des sentences pénales ;
- ceux qui prédisent que l'IA aura remplacé l'être humain pour les traductions l'écriture de romans, la chirurgie n'ont aucune expérience dans ces domaines ;
- il se peut que des machines traduisent des courriers professionnels, des articles scientifiques et des brèves de presse mais *quid* des romans ou de la poésie ;
- la crainte d'une prise de pouvoir des machines provient de la tradition judéo-chrétienne qu'en fabriquant un être animé à son image, l'homme égalerait Dieu ;
- c'est pourquoi la culture japonaise accueille la présence de robots avec plus de sérénité car aucun péché n'est attaché à la réalisation d'un double de l'homme ;
- difficulté quand la présence humaine est impliquée (coiffure) ou lors d'activités de conception et d'imagination (musique).

**p. 079 :**

- la médecine de demain doit être préventive, prédictive, personnalisée et participative (P4) ;
- mise en place de bio-banques pour diagnostics très tôt, traitements personnalisés, facteurs prédictifs, surveillance au plus près ;
- en France, l'Assurance maladie possède un gigantesque fichier, le Système national d'information interrégimes (Sniiram) qui regroupe l'ensemble des remboursements des frais de santé.

**p. 080 :**

- l'IA intervient à toutes les étapes du fonctionnement d'une voiture autonome : à partir d'informations sur l'environnement fournies par des capteurs (chaussée, signalisation, présence de piétons - et pas de personnages sur un panneau publicitaire...- et autres véhicules, il faut distinguer ce qui est important ;

**p. 081 :**

- si un tel véhicule parvient à se mouvoir sur les voies rapides et les autoroutes, il est à la peine en milieu urbain ou sur une départementale ; en présence de troupeaux d'animaux, par exemple ;
- dans les situations inhabituelles ou de crise, le véhicule s'arrête, éventuellement après s'être garé ;
- sinon, il doit être en permanence connecté à des bases de données mémorisant ces situations ;
- une voiture autonome doit pouvoir être reprise en mains.

**p. 081 :**

- les compagnies aériennes économiseraient, chaque année, 30 milliards d'Euros de rémunération si elles pouvaient se passer des deux navigants de bord pour les 20 000 avions de ligne actuels ;
- les aides nouvelles, plus ou moins liées à l'IA, seraient une aide à la décision pour les pilotes en cas de changement de certains paramètres (plan de vol), d'avaries (détournement vers un autre aéroport intégrant les infrastructures hôtelières et hospitalières) ;
- un allègement de la charge de travail permettrait de revoir les effectifs à la baisse en augmentant légalement la durée de travail aux commandes.

**p. 082 :**

- les données relatives aux passagers seraient utilisées pour leur proposer – via l'écran dans le siège devant eux et moyennant commission – des programmes et services correspondant à leur intérêt ;
- une maintenance prédictive pour équipements et moteurs : « on vend des heures de vol... ».

**p. 084 :**

- dans certaines rédactions, des « moteurs de rédaction » sont capables de produire de courts articles à partir de données basiques : résultats sportifs, prévisions météo, chiffres financiers d'entreprises ;
- c'est de l'IA car il y a automatisation de tâches « intellectuelles » mais pas de l'intelligence ;
- les méthodes de l'IA ne se limitent pas à l'apprentissage de type Machine Learning.

**p. 085 :**

- rédaction de newsletters personnalisées en fonction des données acquises sur le destinataire ;
- l'IA, par la reconnaissance de textes et d'images pourrait servir pour débusquer les fake news ;
- Google et Facebook sont incontournables mais, même chez eux, rien ne laisse penser qu'il y aura une intelligence forte – la machine ayant une conscience de soi – dans les 50 à 100 ans qui viennent.

**p. 085 :**

- les enceintes vocales connectées, dopées à l'IA, sont capables de répondre à des requêtes vocales ou/et d'agir en conséquence.

**p. 086 :**

- cela a commencé avec les algorithmes de recommandation orientant la consommation des utilisateurs et l'arrivée des assistants vocaux va amplifier le phénomène ; on pourra dire : « donne-moi les dernières nouvelles » ;
- toutes les commandes vocales enregistrées par Google Home, Echo et Home Pod sont transférées dans les clouds de Google et Amazon sans que l'on sache comment elles seront utilisées.

**p. 087 :**

- pour chaque note de musique enregistrée, la machine va observer dans quel environnement elle se situe et quelle note vient après ;
- elle suit un algorithme de prédiction qui analyse le passé et recherche les régularités ;
- après avoir « ingéré » ces données, elle va pouvoir essayer de créer des sons mélodieux.

**p. 088 :**

- le Japon est en avance pour la création et le déploiement – dans la société – de robots « sociaux » capables de simuler des émotions : aides aux enfants autistes, assistance aux personnes âgées, accueil dans les hôtels.

**p. 089 :**

- l'humain projette des relations affectives avec des robots qui « aident », même non humanoïdes ;
- les hommes n'aiment pas voir « souffrir » les artefacts et les machines ; ce qui explique le succès du jeu Tamagotchi, un animal de compagnie virtuel qui « disparaît » (sorte de mort) si on ne s'en occupe pas ;
- la capacité des robots à s'adapter, à partager une histoire, fait basculer les comportements des humains.
- les machines basées sur le partage de données peuvent manipuler les personnes qui « conversent » avec elles et communiquer ce qu'elles ont appris à d'autres.

**p. 092 :**

- dans les rues, ce qui importe, ce n'est pas le nombre de caméras mais l'IA qu'il y a derrière ;
- il est indéniable qu'ont eu lieu ces dernières années des progrès spectaculaires de l'apprentissage automatique en matière de reconnaissance d'images, de textes et de sons ;
- l'efficacité du logiciel PredPol (Predictive Policing) – utilisant une technologie dite prédictive provenant de la sismologie et implanté dans 60 villes américaines – est très discutée et ne se distingue guère de la statistique quelque peu améliorée ;

**p. 093 :**

- une telle technologie est aussi utilisée pour évaluer le risque de récidive de détenus et accorder ou non des remises de peines ;
- en France, est en développement un algorithme qui suggère des montants d'indemnités de licenciement ou de pension alimentaire à partir de 1000 à 20000 arrêts selon les contentieux.

**p. 094 :**

- cela peut être une façon de soulager les tribunaux mais il y a un risque d'identification bien que noms, prénoms et adresses des parties et des témoins ne soient pas communiqués ;
- afin d'en maîtriser tous les aspects, la gendarmerie a développé en interne un outil censé prévoir les « points chauds » mais reste consciente du risque d'alors moins regarder ce qui se passe ailleurs ;
- le logiciel AnaCrim permet de déterminer des incohérences, des manques d'informations et de générer des hypothèses de travail à partir de données officielles : auditions, photos, écoutes ; excluant, pour le moment, les données non structurées : extractions de blogs, réseaux sociaux, etc. ;
- le privé, prévoit, lui, de surveiller les interactions sur les réseaux sociaux « susceptibles de révéler une menace terroriste » ; la société Sinequa qui compte pour clients des agences de renseignements ainsi que les ministères français de la Défense, des Finances et de l'Intérieur ;

**p. 095 :**

- la loi renseignement laisse quasi libre le passage à la surveillance de masse moyennant la promesse d'un contrôle renforcé ;
- au détriment d'un Etat de droit, l'IA va établir des corrélations qui vont devenir normatives.

**p. 096 :**

- la CNIL (Commission nationale de l'informatique et des libertés) a amorcé en janvier 2017 une discussion autour de risques de discriminations et de questions éthiques soulevées par les algorithmes : moteurs de recherche, recommandation, médecine prédictive, analyse des actes de récidive, etc.

**p. 097 :**

- la question de transparence n'est pas simple : si on vous dit ce qu'il y a dans un moteur thermique d'une automobile, ce n'est pas pour cela que vous comprenez son fonctionnement ;
- un sondage de l'Ifop montre qu'un tiers des sondés ne sait pas ce qu'est un algorithme et une crainte de que l'IA restreigne les choix en les dépolitisant et soit utilisée pour tracer et surveiller ;
- la gestion des malades et la politique pénale, par exemple, nécessitent un pacte social collectif et il n'est pas souhaitable de se retrancher derrière une gestion par des machines censées être plus efficaces.

**p. 098 :**

- des experts de l'IA ont publié en 2015 une lettre ouverte appelant l'ONU à encadrer le développement de robots armés (Stephane Hawking, Elon Musk, Noam Chomsky, Steve Wozniak) ;
- ces « armes létales autonomes » pourraient se déplacer et prendre seules la décision d'éliminer une cible en fonction d'un algorithme interne : tuer pourrait devenir un acte banal ;

**p. 101 :**

- après la poudre et le nucléaire, ce serait la troisième révolution de guerre ;
- la partie logicielle est aisément copiable et la partie mécanique n'est pas compliquée à produire ;
- par contre, il semble qu'une IA ne puisse élaborer, sans intervention humaine, un plan tel que : entrer dans un bâtiment, explorer quelques pièces et ressortir après avoir éventuellement éliminé une cible ;
- début 2017, des dizaines de chercheurs en IA se sont réunis au sud de San Francisco et ont élaboré une charte en 23 points : le 18<sup>ème</sup> établit qu'une course à l'armement des armes létales autonomes doit être évitée.

**p. 102 :**

- depuis qu'un drone Predator – piloté à distance – a tué un lieutenant de Ben Laden en novembre 2001, les Américains sont persuadés que la guerre du futur se gagnera avec des armes autonomes : Russes et Chinois tentent de suivre ;
- sont déjà déployés des robots télécommandés pour le déminage (PackBot) ou dotés de mitrailleuses et de lance-grenades (Swords et Maars) ;
- les développements futurs portent sur des Laws (Lethal autonomous weapons systems) pilotés par l'IA opérant sur terre, dans les airs et en mer, sans conscience et sans famille (à indemniser) ;
- l'avion furtif sans pilote X-47B serait la préfiguration de ces nouveaux drones tueurs ;
- 18 milliards de dollars (sur 600 milliards de budget annuel de défense américaine à comparer aux 32 milliards d'Euros en France) ont été consacrés aux Laws.



**p. 103 :**

- actuellement, les robots coopèrent avec les soldats pour servir d'éclaireurs, transporter les munitions mais les Etats-Unis seraient à moins de 10 ans de disposer de robots autonomes.

**p. 104 :**

- une étude a été menée sur des IA existantes – IA spécialisées avec un mécanisme de l'OTAN- quant aux possibilités de dérives malveillantes ou d'instabilités conduisant à des situations de crise militaire ;  
- l'imaginaire associé à l'IA s'appuie sur une imagerie issue des films de science-fiction américains ;

**p. 105 :**

- le concept de Singularité technologique est fondateur de la pensée transhumaniste ;  
- les craintes sont plus prononcées chez les croyants d'une religion monothéiste et se réfèrent à une IA forte qui aurait conscience de ses capacités ;  
- les craintes « économique » et « sociale » (perte massive d'emplois) s'inscrivent dans la séquence des résistances aux progrès technologiques (telles que le luddisme en Angleterre et la révolte des Canuts au XIX<sup>e</sup> en France) ;  
- le premier facteur de risque est la cyber-insécurité des IA en les détournant de leurs fonctionnalités ;  
- l'autonomie des systèmes civils et militaires qui s'appuient sur l'IA induit donc de nouveaux risques ;  
- le chef d'état-major russe a déclaré en 2016 que 30% des systèmes d'armes seront robotisés en 2025 ;  
- la société Kalachnikov a présenté sa gamme de robots armés autonomes ;  
- l'agence américaine de recherche en armement (Darpa) teste depuis 2016 un navire transocéanique autonome (le SeaHunter, long de 45 mètres !) chasseur de sous-marins ;  
- quelques heures d'apprentissage de thèses racistes ont suffi à l'IA conversationnelle TAY développée par Microsoft pour tenir des propos xénophobes...

