

Introduction à l'Intelligence Artificielle

Eric Gaussier

AMA/LIG - MIAI
Université Grenoble Alpes
Eric.Gaussier@imag.fr

Table des matières

- 1 Intelligence artificielle : survol historique
- 2 Apprentissage automatique
- 3 Quelles données ?
- 4 Limites et questions
- 5 Conclusion

Des définitions qui ont varié au cours du temps (1)

1. Les débuts de l'informatique : Test de Turing – *Computing Machinery and Intelligence* (1950)

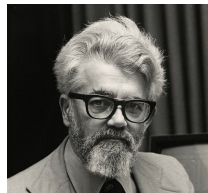
ELIZA (J. Weizenbaum - 1966) - PARRY (K. Cosby - 1972, paranoïd AI)

2. Conférence Dathmouth (1956) *J. McCarthy, M. Minsky (org.)*

Adoption du terme IA (Machine Intelligence, D. Michie

Faire en sorte que les machines réalisent des tâches que les humains font aisément (voir, parler, conduire, manipuler des objets, planifier les activités quotidiennes, ...)

From Artificial Intelligence, Y. Wilks, 2019



Des définitions qui ont varié au cours du temps (2)

3. *Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique, M. Minsky*



Faire faire à des machines des tâches qui seraient qualifiées d'intelligentes si elles étaient faites par des humaines

Une volonté sous-jacente : analyser et comprendre le cerveau humain

Des définitions qui ont varié au cours du temps (2)

3. *Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique, M. Minsky*



Faire faire à des machines des tâches qui seraient qualifiées d'intelligentes si elles étaient faites par des humaines

Une volonté sous-jacente : analyser et comprendre le cerveau humain

Des définitions qui ont varié au cours du temps (2)

3. *Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique, M. Minsky*



Faire faire à des machines des tâches qui seraient qualifiées d'intelligentes si elles étaient faites par des humaines

Une volonté sous-jacente : analyser et comprendre le cerveau humain

Des définitions qui ont varié au cours du temps (3)

4. Dernière en date (High-Level Expert Group in AI)

L'intelligence artificielle désigne des systèmes construits par des humains qui, étant donné un objectif complexe, agissent dans le monde physique ou numérique en percevant les signaux issus de leur environnement, en interprétant les données collectées (que celles-ci soient structurées ou non), en raisonnant sur les informations et connaissances extraites de ces données, et en choisissant les actions à mener de façon à accomplir le but fixé.

Analyse/interprétation/raisonnement interviennent à différents niveaux.

Domaines concernés

Les grands domaines "cœur"

- Raisonnement (programmation logique, programmation par contraintes, démonstration, ...), aide à la décision, représentation des connaissances, web sémantique
- Apprentissage automatique (optimisation, apprentissage par renforcement, clustering, apprentissage profond, ...)
- Robotique et perception (vision, traitement automatique des langues et de la parole, planification, interactions, ...)
- Intelligence collective (systèmes multi-agents)

Les domaines associés

- Contributeurs : informaticiens, mathématiciens, sciences cognitives, psychologie, traitement du signal, ...
- Domaines d'applications : santé, transports, industrie 4.0, environnement, énergie, défense, ...

Une dichotomie "logique/numérique"

- Quel type de raisonnement ?
Logique ou numérique (statistique)
- Quel type de représentations ?
Faits logiques, comptes/vecteurs numériques

Illustration

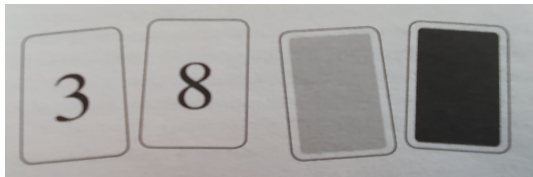
Corpus parallèle Hansard (parlement canadien)

No ifs, no buts, we want the truth

Pas de si, pas de mais, nous voulons la vérité

Est-ce que la logique est à la base de notre raisonnement ?

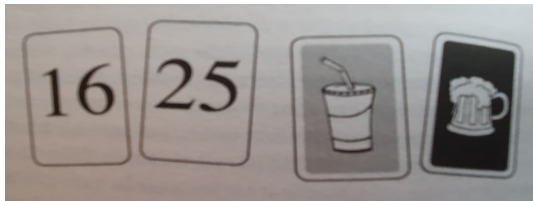
Quelles cartes doit-on retourner pour tester la phrase ...
"si une carte est paire sur une face, elle est grise sur l'autre" ?



P. Wason, quoted in Artificial Intelligence, Y. Wilks

Est-ce que la logique est à la base de notre raisonnement ?

Quelles cartes doit-on retourner pour tester la phrase ...
"si vous buvez de l'alcool, vous devez avoir au moins 18 ans" ?



P. Wason, quoted in Artificial Intelligence, Y. Wilks

Est-ce que la logique est à la base de notre raisonnement ?

D. Kahneman, *Thinking, fast and slow*, 2011



Et les statistiques ?



Persée triomphant, sculpture d'A. Canova (1801)

Dichotomie logique/numérique

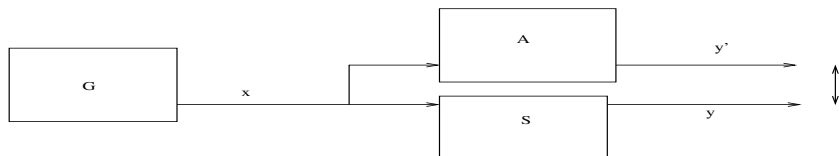
- Les deux types de raisonnement sont bien sûr intéressants (couplage ?)
- De nombreuses tâches semblent finalement ne pas requérir des raisonnements poussés
- Une logique naturelle ? Des comparaisons numériques plus simples ?
- Les approches numériques ont pris le pas sur les approches logiques dans beaucoup de domaines mais sont loin d'être entièrement satisfaisantes (explicabilité, nombre de données, ...)

Domaine de recherche encore ouvert

Table des matières

- 1 Intelligence artificielle : survol historique
- 2 Apprentissage automatique**
- 3 Quelles données ?
- 4 Limites et questions
- 5 Conclusion

Qu'est-ce que l'apprentissage ?

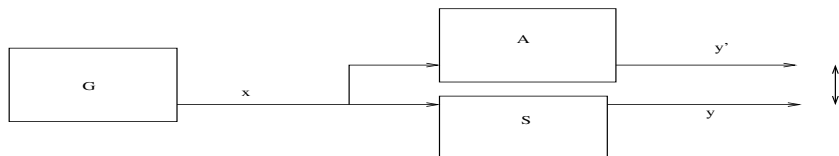


On cherche une fonction f dans une famille de fonctions \mathcal{F} telle $y' = f(x)$ soit le plus proche possible de y , la sortie fournie par S à partir de l'entrée x

- Apprentissage non supervisé
- Apprentissage supervisé (faiblement, semi-supervisé)
- Apprentissage par renforcement

Focus sur l'apprentissage supervisé

Qu'est-ce que l'apprentissage ?

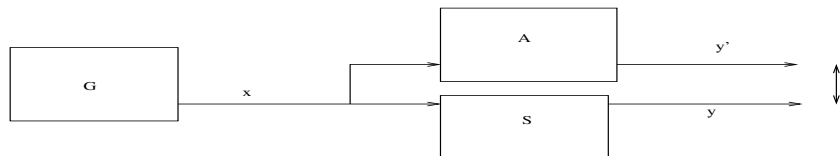


On cherche une fonction f dans une famille de fonctions \mathcal{F} telle $y' = f(x)$ soit le plus proche possible de y , la sortie fournie par S à partir de l'entrée x

- Apprentissage non supervisé
- Apprentissage supervisé (faiblement, semi-supervisé)
- Apprentissage par renforcement

Focus sur l'apprentissage supervisé

L'apprentissage supervisé



- ❶ Ensemble d'apprentissage : $\mathcal{D} = ((x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)}))$
- ❷ $x \in \mathbb{R}^p$, $y \in \mathcal{Y}$ (catégorisation binaire : $\mathcal{Y} = \{0, 1\}$)
- ❸ **Apprenant A** sélectionne la fonction $f \in \mathcal{F}$ la plus appropriée

Comment sélectionner f ?

Mesure de la qualité d'une fonction apprise

Fonction de coût (*loss function*)

$L : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}^+$, telle que $L(y, y') > 0$ pour $y \neq y'$

Illustration

- Coût 0 – 1 :

$$L(y^{(i)}, f(x^{(i)})) = \begin{cases} 0 & \text{si } y^{(i)} = f(x^{(i)}), \\ 1 & \text{sinon.} \end{cases}$$

- \mathcal{F} = polymômes de degré ≤ 5 : $f(x) = a_0 + a_1x + \dots + a_5x^5$
Paramètres : (a_0, \dots, a_5)

Sélection de $f \in \mathcal{F}$

- ① Minimisation du risque fonctionnel (erreur de généralisation) :

$$\arg \min_{f \in \mathcal{F}} R(f), \quad R(f) = \mathbb{E}_{P(x,y)} L(y, f(x)) = \int_x \int_y P(x, y) L(y, f(x)) dx dy$$

- ② Minimisation du risque empirique :

$$\arg \min_{f \in \mathcal{F}} \text{Remp}(f; \mathcal{D}), \quad \text{Remp}(f; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^n L(y^{(i)}, f(x^{(i)}))$$

Sélection de $f \in \mathcal{F}$

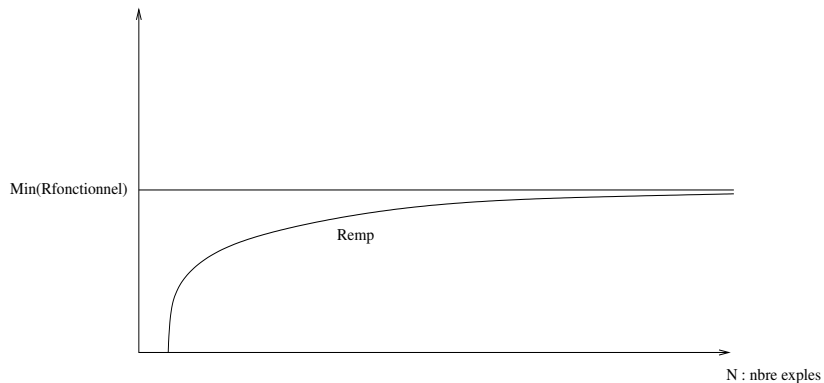
- ① Minimisation du risque fonctionnel (erreur de généralisation) :

$$\arg \min_{f \in \mathcal{F}} R(f), \quad R(f) = \mathbb{E}_{P(x,y)} L(y, f(x)) = \int_x \int_y P(x, y) L(y, f(x)) dx dy$$

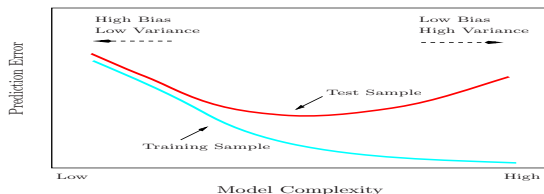
- ② Minimisation du risque empirique :

$$\arg \min_{f \in \mathcal{F}} \text{Remp}(f; \mathcal{D}), \quad \text{Remp}(f; \mathcal{D}) = \frac{1}{n} \sum_{i=1}^n L(y^{(i)}, f(x^{(i)}))$$

Justification de la minimisation du risque empirique



Mais en pratique !

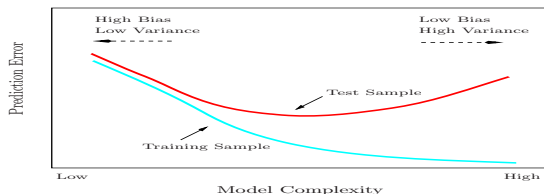


Solution : $\arg \min_{f \in \mathcal{F}} \text{Remp}(f) + \lambda \Omega(f)$

$\Omega(f)$ représente une mesure de la complexité de f

Image tirée de Elements of statistical learning. Hastie, Tibshirani, Friedman. Springer

Mais en pratique !



Solution : $\arg \min_{f \in \mathcal{F}} \text{Remp}(f) + \lambda \Omega(f)$

$\Omega(f)$ représente une mesure de la complexité de f

Image tirée de *Elements of statistical learning*. Hastie, Tibshirani, Friedman. Springer

Régularisation : complexité, connaissances et contraintes

$$\min_{f \in \mathcal{F}} \text{Remp}(f) + \underbrace{\lambda \Omega(f)}_{\text{paramètre de régularisation}}$$

régularisation

La régularisation permet :

- De rendre compte de la complexité de la fonction choisie
- D'intégrer des connaissances et des contraintes

Data Scientist

Un *data scientist* :

- Détermine la famille \mathcal{F} (réseau avec 1 couche cachée de 10 neurones)
- Détermine les contraintes/connaissances à prendre en compte
- Détermine les données à utiliser pour *apprendre* la fonction $f \in \mathcal{F}$:

$$\mathcal{D} = \{(x^{(1)}, y^{(1)}), \dots, (x^{(i)}, y^{(i)}), \dots, (x^{(n)}, y^{(n)})\}$$

- Établit la fonction objectif à optimiser :

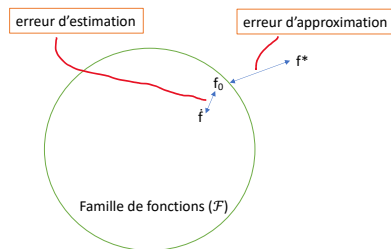
$$G(f, \mathcal{D}) = \underbrace{\sum_{i=1}^n (f(x^{(i)}) - y^{(i)})^2}_{\text{fonction de coût}} + \underbrace{\lambda \|\theta\|_2^2}_{\text{régularisation}}$$

- Choisit/développe l'algorithme d'apprentissage :

$$\arg \min_{f \in \mathcal{F}} G(f, \mathcal{D})$$

Compromis estimation-approximation (1)

f^* désigne la fonction réelle, que l'on ne connaît pas mais que l'on souhaite approcher ; f_0 est la fonction de \mathcal{F} la plus proche de f^* et \hat{f} est la fonction obtenue par résolution du problème de minimisation ($\hat{f} \in \mathcal{F}$). L'erreur de la fonction apprise peut se décomposer en erreur d'estimation et erreur d'approximation :



Compromis estimation-approximation (2)

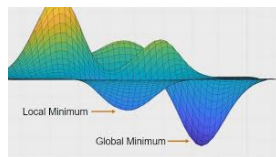
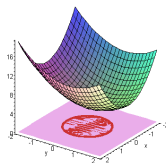
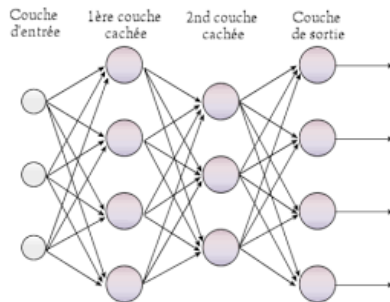


Figure – Minimisation erreur estimation (gche), erreur approximation (dte)

Apprentissage profond : perceptron multicouche (1)

Terminologie : *Multilayer Perceptron (MLP), Feedforward Neural Network (FFNN), fully connected network*

- $\mathbf{y} \in \mathbb{R}^4, \mathbf{x} \in \mathbb{R}^3; \mathbf{y} = f(\mathbf{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\mathbf{x})))$
- Profondeur du réseau, dimensionnalité des couches



Apprentissage profond : perceptron multicouche (2)

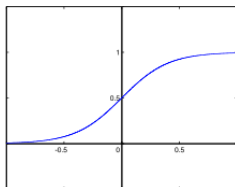
Quelles fonctions f^i intermédiaires ?

Soit \mathbf{h}^{i-1} l'entrée de f^i ($\mathbf{h}^0 = \mathbf{x}$) :

$$f^i(\mathbf{h}^{i-1}) = \sigma(\mathbf{W}^i \mathbf{h}^{i-1} + \mathbf{b}^i)$$

avec $\mathbf{h}^{i-1} \in \mathbb{R}^{p_i}$, $\mathbf{W}^i \in \mathbb{R}^{p_{i+1} \times p_i}$, $\mathbf{b}^i \in \mathbb{R}^{p_{i+1}}$

La fonction σ est une fonction *en général non linéaire* d'activation (sigmoïde, tanh, RELU, identité)

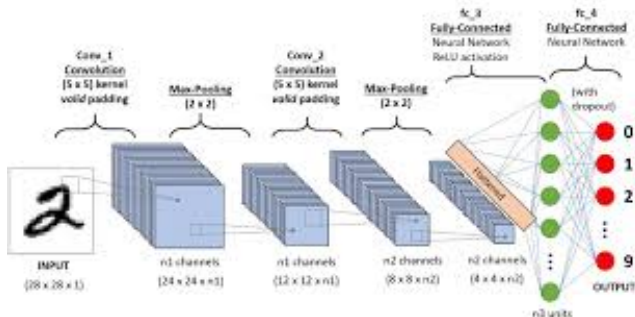


Apprentissage profond : perceptron multicouche (3)

- Un MLP est un approximateur universel
- Famille de fonctions riche : bonne approximation mais estimation plus délicate a priori
- Nombre de paramètres important (exemple ci-dessus)
- Nécessité d'un grand nombre d'exemples d'apprentissage
- Compréhension (très) incomplète (qualité des minima locaux, ...)

Apprentissage profond : réseaux convolutionnels

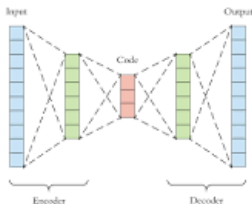
Généralisation des perceptrons



Apprentissage profond : autoencodeurs

Apprentissage de représentation non supervisé

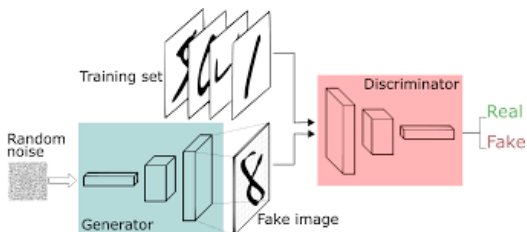
- Encodeur et décodeur miroirs l'un de l'autre
- Fonction de coût mesure l'erreur de reconstruction



Apprentissage profond : GAN

Apprentissage de modèles génératifs : Generative Adversarial Networks (GAN)

- Générateur produit des exemples proches des exemples réels
- Discriminateur différencie exemples réels et générés



Apprentissage profond

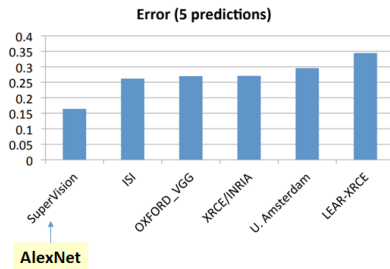
Remarques générales

- ❶ Différents types de réseaux
 - MLP, réseaux convolutionnels, autoencodeurs, GAN
 - Réseaux récurrents
 - Transformers (mécanisme d'attention)
- ❷ Cadre standard de l'apprentissage :
 - Minimisation d'une fonction objectif (descente du gradient : algorithme backpropagation)
 - Hyperparamètres : architecture du réseau (nombre et dimension des couches cachées), type de gradient (taille des batchs, du pas de gradient, ...)
- ❸ Nombre de paramètres : plusieurs centaines de millions en vision et traitement automatique des langues (capacité de calcul, données)
- ❹ Approches *bout-en-bout* avec des entrées et sorties variées

Apprentissage profond : performances

- Des résultats inégalés sur certaines tâches (vision 2012, traitement automatique des langues 2016, ...)

Ranking of the best results from each team



Apprentissage profond : un monde de *légos*

- Des approches de bout-en-bout permettant de résoudre des problèmes complexes
 - Bibliothèques (Tensorflow, Pytorch) qui permettent de construire des applications complexes
 - Composition de réseaux *en légo* (descente de gradient avec backpropagation - dérivabilité des fonctions)
 - Question-réponse, compréhension de textes : de l'intégration de services à des réseaux intégrés (code simplifié de manière très significative)
- Accessible à n'importe quelle personne ayant des compétences en programmation !

Table des matières

- 1 Intelligence artificielle : survol historique
- 2 Apprentissage automatique
- 3 Quelles données ?**
- 4 Limites et questions
- 5 Conclusion

L'annotation : un processus plus ou moins complexe

La disponibilité de données annotées (ou facilement annotables) dépend de la tâche considérée

- Traduction automatique
- Systèmes de question-réponse
- Pertinence des pages web pour une requête
- Objets dans des images, actions dans des vidéos

Table des matières

- 1 Intelligence artificielle : survol historique
- 2 Apprentissage automatique
- 3 Quelles données ?
- 4 Limites et questions**
- 5 Conclusion

Limites (1)

- Nombre d'exemples : quelques exemples sont nécessaires à un enfant de 3 ans pour reconnaître un chien par exemple (des dizaines de millions pour un ordinateur, à la hauteur du nombre de paramètres)
- Certifiabilité, explicabilité, équité (CEE - *FAT*)
 - Biais dans les données, mais aussi dans les usages
 - Certifiabilité difficile (voire impossible ?)
 - Sécurité informatique a ses limites
 - Détournement des utilisations prévues



Limites (2)

- Impacts socio-économiques
 - Une utilisation systématique et généralisée (ce ne sont que des outils !)
 - Impact sur les emplois
- Impacts environnementaux

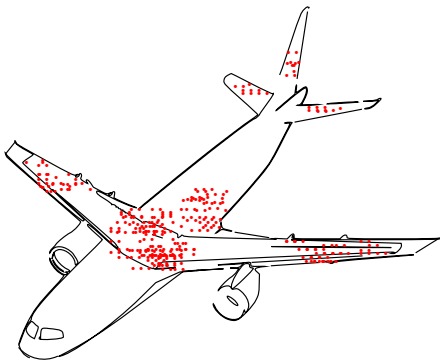
Les émissions pour l'entraînement d'un modèle vont de 18kg eqCO₂ à 284T eq CO₂. 652kg eqCO₂ pour le modèle le plus courant, soit environ un aller Paris-Hong Kong en avion, ou 2500km en voiture... (Strubell et al., 2019)

 - Quelles architectures (cloud, edge, ...) ?, quels modèles (parcimonie) ?

Weapons of Math Destruction, C. O'Neil, Crown, 2016
ecoinfo.cnrs.fr/2019/10/01/impact-environnemental-de-lia/

Limites (3)

Au-delà des corrélations : abstraction, causalité



Les algorithmes sont-ils *intelligents* ?

- Non ... mais encore

- Des algorithmes *bêtes* qui paraissent intelligents (Cambridge Analytica)
- Des comportements/stratégies qui *ressemblent* à ceux des humains

Les algorithmes peuvent-ils créer ?



Art and Artificial Intelligence Laboratory, Rutgers University

Table des matières

- 1 Intelligence artificielle : survol historique
- 2 Apprentissage automatique
- 3 Quelles données ?
- 4 Limites et questions
- 5 Conclusion**

Conclusion

- L'IA : un domaine en pleine effervescence, notamment grâce à l'apprentissage profond et à l'accès *facile* aux technologies
- Un mouvement qui s'inscrit dans le développement numérique (*nil novi sub sole*)
- Régulation juridique et éthique *délicate*
- Quelle en sera la suite ? Il est difficile de prédire ...

... surtout le futur !

Conclusion

- L'IA : un domaine en pleine effervescence, notamment grâce à l'apprentissage profond et à l'accès *facile* aux technologies
- Un mouvement qui s'inscrit dans le développement numérique (*nil novi sub sole*)
- Régulation juridique et éthique *délicate*
- Quelle en sera la suite ? Il est difficile de prédire ...

... surtout le futur !

No ifs, no buts, we want the truth

Cessons de tourner autour du pot ... de chambre

No ifs, no buts, we want the truth

Cessons de tourner autour du pot ... de chambre

Quelques Références (1)

- *Artificial Intelligence : Modern Magic or Dangerous Future ?*, Y. Wilks, Icon Books, 2019
- *Le temps des algorithmes*, S. Abiteboul and G. Dowek, Éditions Le Pommier, 2017
- *Le mythe de la singularité*, J.-G. Ganascia, Éditions du Seuil, 2017
- *Weapons of Math Destruction*, C. O'Neil, Crown, 2016
- *Thinking, Fast and Slow*, D. Kahneman, Penguin Books, 2011

Quelques Références (2)

- *Deep Learning*, I. Goodfellow, Y. Bengio and A. Courville, MIT Press, 2016
- *Elements of statistical learning*, T. Hastie, R. Tibshirani and J. Friedman, Springer, 2008
- *Causality : Models, Reasoning and Inference*, J. Pearl, Cambridge University Press, 2000
- *Artificial Intelligence : A Modern Approach*, S. Russel and P. Norvig, Prentice Hall, 1995 (1st edition)

Les auteurs à éviter

- Discours simplificateur voire erroné, scientifiquement non fondé, jouant sur les peurs (promotion d'un livre)