

LA REVANCHE DES NEURONES

L'invention des machines inductives et la controverse de l'intelligence artificielle

[Dominique Cardon](#), [Jean-Philippe Cointet](#), [Antoine Mazières](#)

La Découverte | « Réseaux »

2018/5 n° 211 | pages 173 à 220

ISSN 0751-7971

ISBN 9782348040689

DOI 10.3917/res.211.0173

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-reseaux-2018-5-page-173.htm>

Distribution électronique Cairn.info pour La Découverte.

© La Découverte. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

LA REVANCHE DES NEURONES

L'invention des machines inductives
et la controverse de l'intelligence artificielle

Dominique CARDON
Jean-Philippe COINTET
Antoine MAZIÈRES

L'épisode est en passe de devenir légendaire dans l'histoire de l'informatique. Nous sommes en octobre 2012 à la conférence ECCV qui réunit les chercheurs spécialisés en *computer vision*¹.

« Alors à la compétition de 2012, qui débarque ? C'est Hinton [*le "père" du renouveau des réseaux de neurones*] et c'est le séisme. Il ne connaît rien au domaine de la vision par ordinateur et il prend deux petits gars pour tout faire sauter ! Un [*Alex Krizhevsky*] qu'il a enfermé dans une boîte et il lui a dit : "Tu ne sors pas tant que ça ne marche pas !" Il a fait tourner des machines énormes, qui avaient des GPU à l'époque qui n'étaient pas ultra, mais qu'il faisait communiquer entre eux pour les booster. C'était un truc de machinerie complètement dingue. Sinon, ça n'aurait jamais marché, un savoir-faire de geek, de programmation qui est hallucinant. À l'époque, les mecs de *computer vision* s'excitaient sur ImageNet depuis deux trois ans [*une base de données de 1,2 million d'images étiquetées avec 1 000 catégories servant de benchmark pour comparer les résultats en classification des différents compétiteurs*]. Le number one, il était à 27,03 % d'erreur, le number 2 à 27,18 %, le number 3 à 27,68 %. Et Hinton, il envoie son mec sorti de nulle part : "on a fait tourner un gros deep, on est à 17 !" Il met 10 points à tout le monde ! Comme ça, le jeune geek, il arrive, il annonce le résultat, la salle bondée à craquer. Enfin, il comprend rien à rien, genre il a 17 ans ! Il ne sait pas pourquoi les trucs sont là. Lui, il était enfermé dans sa boîte, il ne connaissait rien au domaine. Et là, il est face à Fei Fei ! Et tu as LeCun qui est assis au fond de la salle qui se lève pour répondre aux questions [*Li Fei Fei, professeur d'informatique qui dirige SAIL le laboratoire historique d'intelligence artificielle de Stanford ; Yann LeCun, aujourd'hui directeur de FAIR, le laboratoire d'intelligence artificielle de Facebook et un des acteurs centraux du renouveau des réseaux de neurones*]. Et tu as tous les grands manitous du *computer vision* qui essayent de réagir : "Mais en fait c'est pas possible. Ça va pas marcher pour la reconnaissance d'objet quand il faut..." Enfin, les mecs étaient tous par terre parce que *grosso modo* cela foutait en l'air 10 ans d'intelligence, de *tuning*, de sophistication.

1. Cette enquête a été conduite dans le cadre du projet ALGODIV (ANR-15-CE38-0001). Les auteurs souhaitent remercier Telmo Menezes pour ses conseils. Dans cette enquête, nous exploitons trois entretiens réalisés avec des chercheurs français en informatique qui ont participé à la renaissance des réseaux de neurones. Afin de conserver le caractère brut de leurs propos, ils ont été anonymisés.

Figure 1. Machine hypothético-déductive (1) et machine inductive (2)

Source : auteurs.

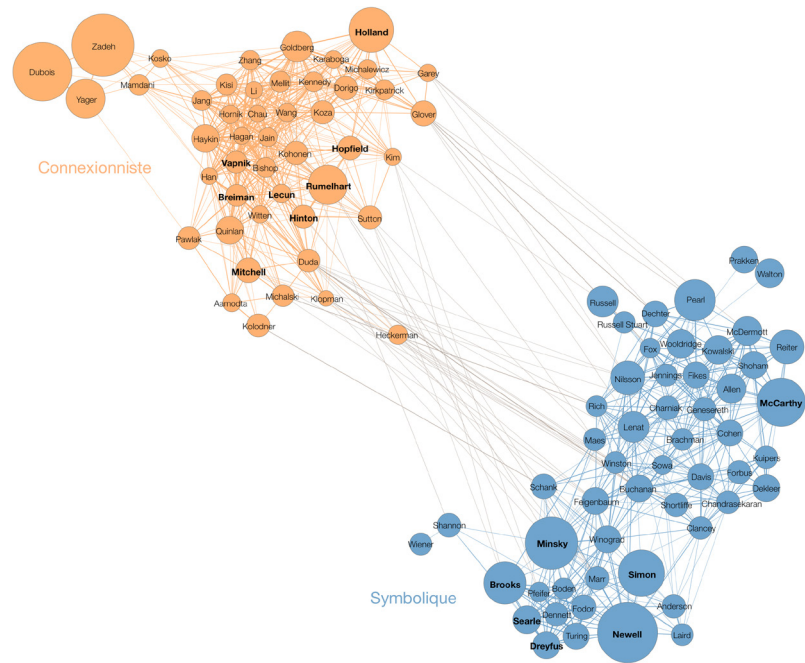
données « brutes » et les naïvetés d'un calcul sans théorie (Gitelman, 2013). Si une telle mise en garde est certainement nécessaire pour relativiser certains discours imprudents assurant que les « données parlent d'elles-mêmes », elle ne rend cependant pas justice au travail résolu et intensément artificiel entrepris par les promoteurs des techniques de *deep learning* pour imposer la seconde architecture de calcul, celle que nous appellerons dans cet article *machine inductive* et, plus précisément encore, *machine connexionniste* afin de mettre en évidence le type particulier d'induction dont elle se réclame. La fabrication d'artefacts susceptibles de produire un calcul inductif sur de grandes masses de données est le résultat d'une histoire conflictuelle et d'une série de constructions d'une très grande ingéniosité. L'induction est quelque chose qu'il faut constamment amener à la machine, défendre contre des opposants, produire au moyen de calculs spécifiques, déployer dans des architectures propres, calibrer avec des données adaptées. Les concepteurs de ces machines ne sont pas les naturalistes candides auxquels les sciences sociales constructivistes aiment parfois les réduire. L'idée de confier aux machines le soin de produire des prédictions pertinentes en apprenant des données – *i.e.* le calcul inductif – est un projet, une théorie et, surtout, un dispositif qui a une histoire agitée. Pour être mis en place et produire ses effets, il a exigé un patient travail de reconfiguration de l'architecture des machines « intelligentes » dont cet article voudrait rendre compte.

Symbolique vs Connexionnisme

La méthode des réseaux de neurones que l'on vient de voir triompher à ECCV'12 n'est en rien nouvelle. Profitant de l'augmentation de la puissance de calcul des ordinateurs et de l'accessibilité de gigantesques bases de données, elle parvient aujourd'hui à honorer une promesse qui avait été faite au début de la cybernétique. De façon surprenante, le terme récemment retenu pour qualifier ces stupéfiantes prouesses calculatoires est celui d'*intelligence artificielle* (IA). Le retour sur le devant de la scène de ce vocable forgé en 1956 par John McCarthy constitue une intéressante énigme pour l'histoire des sciences et

ce mouvement ainsi que les principaux contributeurs aux multiples méthodes du *machine learning*, comme Breiman, Mitchell ou Vapnik. Du côté « symbolique », on retrouve le noyau fondateur de l'IA (McCarthy, Minsky, Simon et Newell) dans une disposition qui reflète leurs proximités et leurs divergences, entouré par les principaux contributeurs à la production des modélisations cognitives, des systèmes experts et même à la critique de l'IA symbolique (Dreyfus, Searle, Brooks).

Figure 2. Réseau de co-citations des 100 auteurs les plus cités par les publications scientifiques mentionnant « Artificial Intelligence »⁴



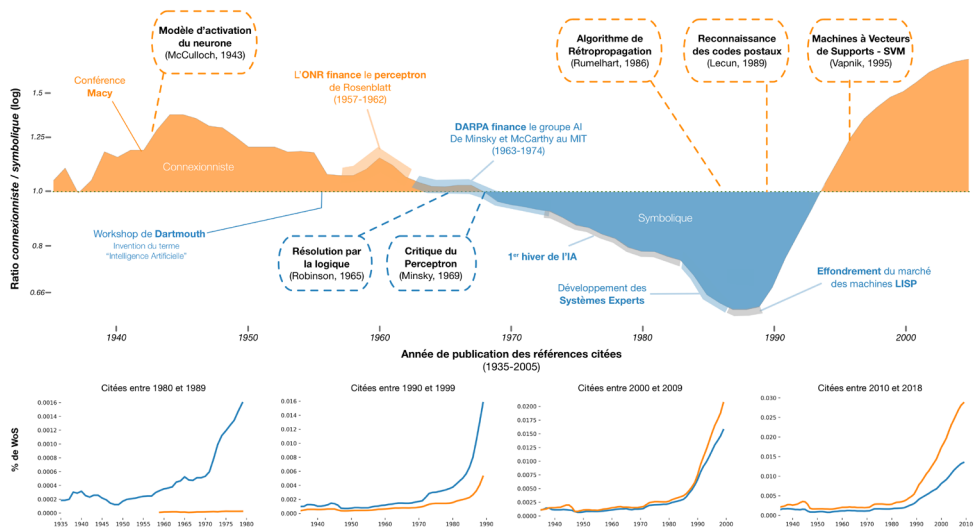
Source : auteurs.

Mais la controverse entre les deux communautés de l'IA apparaît mieux si l'on observe la chronologie de l'impact académique des publications

4. Le corpus « Intelligence Artificielle » contient 27 656 publications rassemblées en février 2018 sur *Web of Science* par la requête : TS=(“artificial intelligence”). La taille des nœuds dépend de la fréquence avec laquelle l’auteur apparaît. Les auteurs régulièrement co-cités dans les mêmes publications sont liés dans le réseau. Un algorithme de détection de communauté révèle la bipartition du réseau en deux communautés cohésives.

machine connexionniste n'est donc pas la conséquence d'une mutation de l'histoire des idées, ou de la validité d'un modèle scientifique sur un autre, mais le résultat d'une controverse qui a conduit les acteurs à déplacer, transformer et redéfinir profondément la forme donnée à leurs artefacts. Le processus auquel cette grille de lecture nous permet d'être attentifs est un long travail historique de recomposition des alliances et des paradigmes entre communautés scientifiques en compétition. Celui-ci affecte les techniques de calcul, mais aussi et surtout la forme donnée à ces machines, leurs objectifs, les données dont elles traitent et les questions qu'elles adressent (Latour, 1987). Pour le dire dans une formule que l'on précisera tout au long de l'article : alors que les concepteurs des machines symboliques cherchaient à insérer dans le *calculateur* et le *monde* et l'*horizon*, la réussite actuelle des machines connexionnistes tient au fait que de façon presque opposée, ceux qui les fabriquent vident le *calculateur* pour que le *monde* se donne à lui-même son propre *horizon*.

Figure 3. Évolution de l'influence académique des approches connexionniste et symbolique



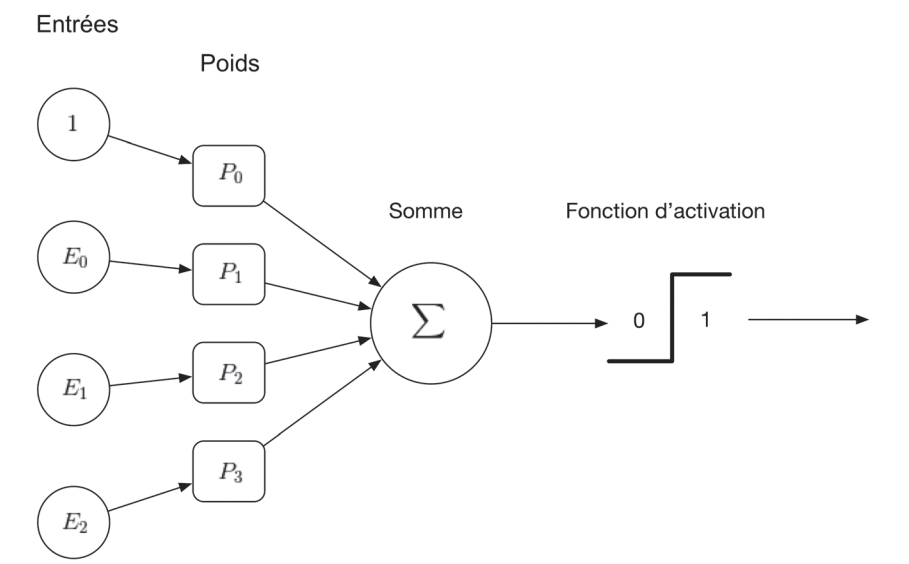
La courbe principale (en haut) représente l'évolution du ratio entre le nombre de publications citées dans le corpus connexionniste (en clair) et le nombre correspondant dans le corpus symbolique (en foncé), tous deux normalisés par le nombre total de publications dans WoS. Les courbes annexes (en bas) représentent pour chacun des corpus, le nombre de publications citées au cours d'une période donnée.

Source : auteurs.

LA CYBERNÉTIQUE ET LE PREMIER CONNEXIONNISME

Les réseaux de neurones trouvent leur origine dans l’histoire pionnière de l’informatique et de la première cybernétique. Bien que l’étiquette soit postérieure, celle-ci peut en effet être dite « connexionniste »⁷ et ne cessera de se référer à la proposition de modéliser mathématiquement un réseau de neurones faite par le neurophysiologiste Warren McCulloch et le logicien Walter Pitts en 1943. Cet article fondateur continue, jusque dans les citations actuelles des articles de *deep learning*, à être donné comme le point de départ de l’aventure connexionniste. La frise chronologique de l’activité scientifique en IA (figure 3) fait clairement apparaître la domination de l’approche connexionniste pendant la période de la première cybernétique. L’article *principles* de McCulloch et Pitts propose un modèle formel (figure 4) dans lequel le neurone prend des variables en entrées, y applique un poids pour produire une somme qui, si elle dépasse un certain seuil, déclenche l’activation du neurone.

Figure 4. Modèle formel d’un neurone artificiel à seuil binaire

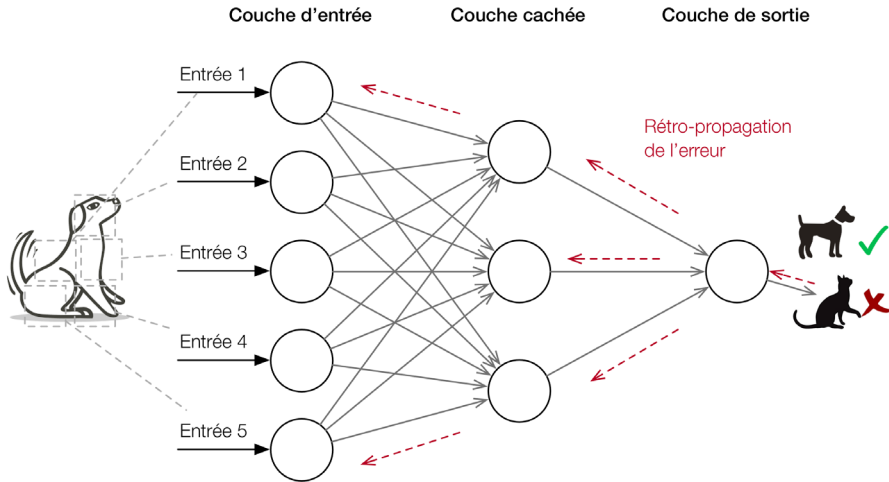


Source : auteurs.

7. Le premier emploi du terme de « connexionnisme » apparaît chez D. Hebb en 1949 et sera ensuite repris par F. Rosenblatt en 1958 (Andler, 1992).

esprit proche des machines cybernétiques, l'erreur en sortie est « propagée » vers les entrées (figure 5).

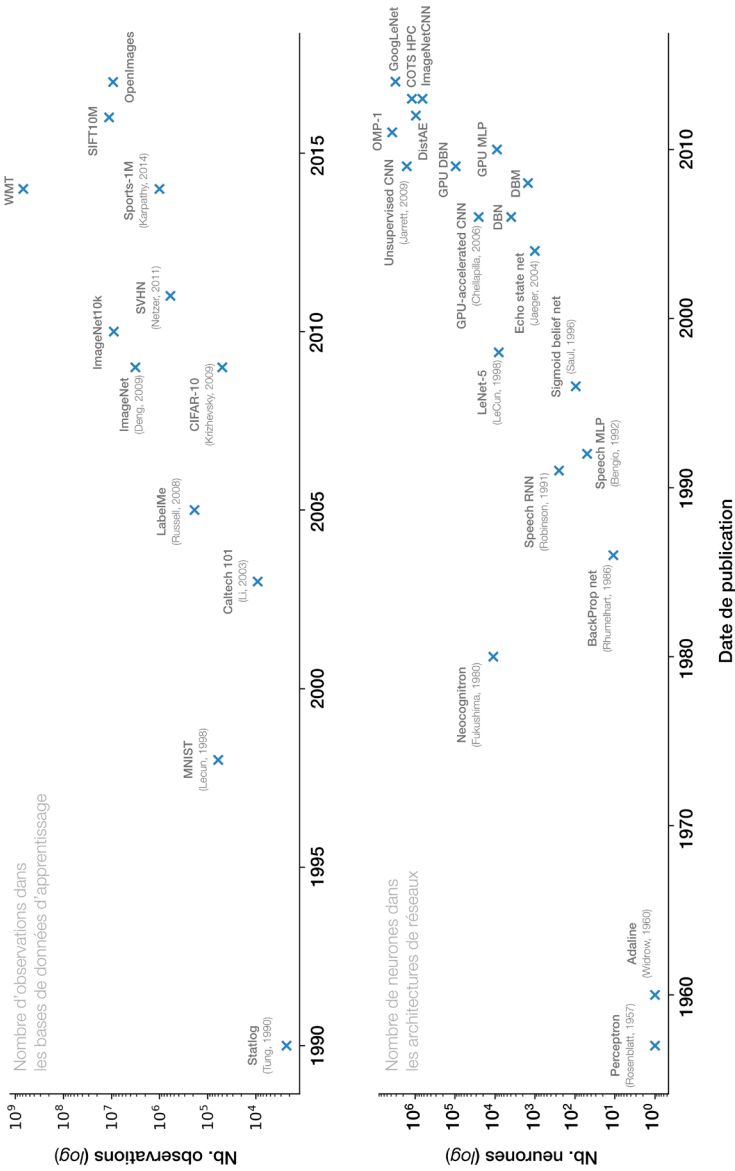
Figure 5. Fonctionnement d'un réseau de neurones simple



Source : auteurs.

Fortes d'un algorithme à vocation générale permettant d'optimiser tout type de réseaux de neurones, les années 1980 et 1990 constituent une remarquable période d'inventivité et marquent le renouveau du connexionnisme. Un des premiers succès est l'application faite par Yann Lecun sur la reconnaissance des codes postaux effectuée à AT&T Bell Labs (Lecun *et al.*, 1989) qui « invente » la technique de la convolution. En utilisant une base de données de l'*US Postal Service*, il parvient à entraîner un réseau multicouche pour reconnaître les chiffres du code postal inscrit sur les colis. Le succès de son approche devient une des premières utilisations industrielles généralisées des réseaux de neurones dans les secteurs bancaire (vérification des montants des chèques) et postal. Suivront ensuite toute une série de propositions pour accueillir un plus grand nombre de couches cachées, complexifier la carte des connexions (encodeurs), diversifier les fonctions d'optimisation (RELU), intégrer de la mémoire dans les couches du réseau (*réseaux récurrents et LSTM*), mixer selon les parties du réseau apprentissage non supervisé et supervisé (*beliefs network*), etc. (Kurenkov, 2015). De façon très créative, de nombreuses architectures câblant différemment les relations entre les neurones sont alors testées pour en explorer les propriétés.

Figure 6. Croissance du nombre de données dans les *datasets* de recherche de 1990 à 2015 (en haut) et du nombre de neurones dans les architectures de calcul mises en place de 1960 à 2015.



Source : Ces données ont été partiellement extraites de Goodfellow *et al.* (2016, pp. 21 et 24) et complétées par l'article de Wikipédia « List of dataset for Machine Learning »

la conférence NIPS, un papier remarqué est un article qui propose une nouvelle architecture, à laquelle, comme pour les planètes, les chercheurs donnent systématiquement des noms composant ainsi un curieux bestiaire (figure 7). En déplaçant du modèle à l'architecture, le lieu où s'exprime l'inventivité des chercheurs, ce sont aussi les compétences et les qualités que requiert leur conception qui se transforment en permettant, notamment en raison de la disponibilité d'outils ouverts et faciles à manipuler, à une nouvelle population de *datascientists*, de bidouilleurs et de programmeurs d'entrer dans le champ précédemment très fermé des producteurs d'IA. En transformant l'architecture des machines prédictives, les connexionnistes ont ainsi contribué à déplacer les mondes sociaux de l'IA : d'abord, parce que les données « réelles », notamment celles venues des industries du numérique, se sont (partiellement) substituées aux *dataset* « jouets » des laboratoires académiques, ensuite parce que les savoir-faire requis pour fabriquer les machines connexionnistes appellent des compétences en développement informatique qui n'étaient pas celles des précédentes générations de l'IA.

LE TRAVAIL DE L'INDUCTION

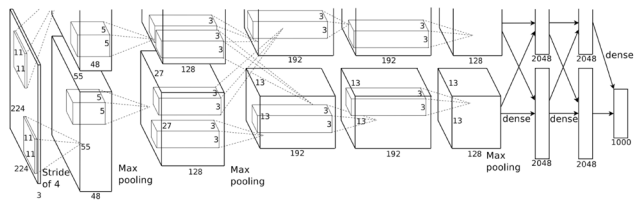
La trajectoire des machines intelligentes dont on vient de résumer l'histoire en quatre configurations successives témoigne de la profonde transformation de leur architecture (tableau 1 ci-dessous). Le *monde*, le *calculateur* et l'*horizon* de ces dispositifs ont été très profondément remaniés et les articulations entre ces composantes façonnent des dispositifs qui proposent des définitions sensiblement différentes de l'intelligence, du raisonnement et de la prédiction.

Tableau 1. Les quatre âges des machines prédictives

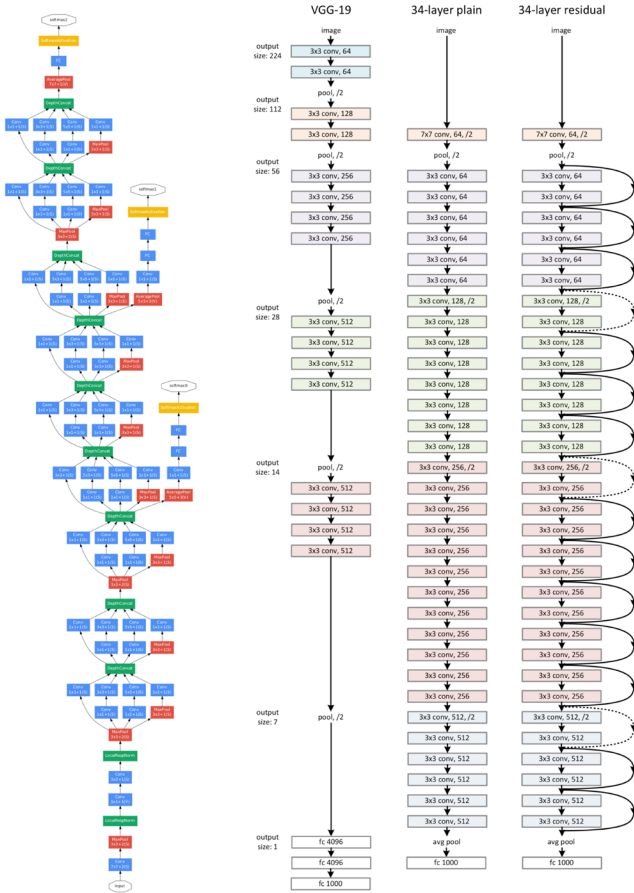
<i>Machine</i>	<i>Monde</i>	<i>Calculateur</i>	<i>Horizon</i>
Cybernétique (connexionniste)	<i>Environnement</i>	« Boîte noire »	<i>Negative feedback</i>
IA Symbolique (symbolique)	<i>Monde « jouet »</i>	<i>Raisonnement logique</i>	<i>Résolution de problème</i>
Système expert (symbolique)	<i>Monde de connaissances expertes</i>	<i>Sélection des hypothèses</i>	<i>Exemples/ contre-exemples</i>
Deep learning (connexionniste)	<i>Le monde comme vecteur de données massives</i>	<i>Réseau de neurones profond</i>	<i>Optimisation de l'erreur sur objectif</i>

Source : auteurs.

Figure 7. Exemples de trois architectures de réseaux de neurones victorieuses du challenge ILSVRC de 2012 à 2015



(a) 9 couches du réseau AlexNet (ImageNet 2012)



(b) 40 couches de GoogLeNet (ImageNet 2014)

(c) 152 couches par Microsoft (ImageNet 2015)

Source : auteurs.