



**HAL**  
open science

## IA et neurosciences

Frédéric Alexandre

► **To cite this version:**

Frédéric Alexandre. IA et neurosciences. Sébastien Konieczny; Henri Prade. L'intelligence artificielle: De quoi s'agit-il vraiment ?, 117, Cepadues-Editions, pp.15 - 43, 2020, 9782364938502. hal-02891480

**HAL Id: hal-02891480**

**<https://inria.hal.science/hal-02891480>**

Submitted on 7 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## IA et neurosciences

Frederic Alexandre  
Inria Bordeaux Sud-Ouest; Labri UMR 5800; IMN UMR 5293  
146 rue Leo Saignat; 33076 Bordeaux, France  
Frederic.Alexandre@inria.fr

Concernées par l'étude du système nerveux, les neurosciences entretiennent des relations profondes avec l'intelligence artificielle, par des contributions réciproques. Comme pour d'autres disciplines, l'intelligence artificielle a fourni aux neurosciences des outils pour donner du sens aux connaissances qu'elles accumulent et traiter les données\* qu'elles génèrent. Ceci est particulièrement critique pour un domaine étudiant le cerveau, généralement considéré comme le plus complexe des systèmes existants. Ses caractéristiques anatomiques et fonctionnelles abordées de façon multi-échelles et transdisciplinaire (chimie, biologie, physique, psychologie), produisent des connaissances dont la validité est régulièrement remise en cause par de nouveaux faits et de nouvelles méthodes d'investigation. Les progrès récents des neurosciences doivent beaucoup au développement de nouvelles technologies (neurochimie, électrophysiologie, imagerie) qui génèrent des masses de données\* dans lesquelles il faut rechercher des indices très diffus. Sur tous ces aspects, l'intelligence artificielle est aujourd'hui un partenaire essentiel des neurosciences.

L'étude du système nerveux est aussi singulièrement l'étude du système biologique responsable de la faculté d'intelligence chez les animaux. Il était donc naturel que l'intelligence artificielle s'intéresse aux neurosciences, y fasse référence pour qualifier les fonctions qu'elle se propose de modéliser et s'en inspire pour les réaliser. Certains précurseurs de l'intelligence artificielle (Alan Turing, Norbert Wiener, John von Neumann) faisaient des références fortes aux neurosciences et se montraient au moins autant intéressés par comprendre le cerveau que par créer une « machine pensante ». La confrontation de ce positionnement avec celui d'autres pères fondateurs (John McCarthy, Alan Newell, Herbert Simon), plutôt orientés vers la logique\* et l'hypothèse du système formel (physical symbol system hypothesis), a donné naissance à une dualité fondamentale de l'intelligence artificielle, reprise par l'expression des frères Dreyfus<sup>1</sup> « Making a mind versus modelling the brain ». Ceci résume deux pistes majeures pour émuler\* une intelligence artificielle, en créant un esprit ou en modélisant le cerveau, auxquelles s'est ajoutée la piste de l'intégration neuro-symbolique, visant à hybrider le meilleur des deux approches. Un débat important associé à cette dualité est le problème de l'ancrage du symbole, défini par Stevan Harnad (comment les symboles acquièrent leur signification), et de son émergence possible à partir de traitements sub-symboliques réalisés par des neurones (comment des traits extraits par des traitements perceptifs ou moteurs

---

<sup>1</sup> Making a Mind versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint. Hubert Dreyfus et Stuart Dreyfus. Daedalus, Vol. 117, No. 1, pp. 15-43. 1988.

élémentaires peuvent être combinés pour établir cette sémantique).

L'apport des neurosciences à l'intelligence artificielle a été notable pour répondre à certaines limitations d'une approche purement logique\* et abstraite de l'intelligence et considérer, en contraste, que l'intelligence du vivant reposait fortement sur le rôle du corps et des émotions, singulièrement absents de cette approche. Mieux comprendre grâce aux neurosciences les liens entre corps et cerveau et le rôle des émotions dans la cognition a joué un rôle clé dans le développement de l'intelligence artificielle incarnée\* , également avec le concours de la robotique qui permettait de fournir un substitut de corps. L'apport des neurosciences a été également décisif pour modéliser l'apprentissage, même si d'autres approches numériques ou logiques\* ont également développé cette faculté. Les premiers modèles neuronaux ont mis l'accent sur l'apprentissage et les performances se sont encore fortement améliorées avec les modèles ultérieurs, dits connexionnistes. Ceci explique que la plausibilité biologique a pu décroître voire disparaître de certains modèles connexionnistes dont le but n'était plus de rendre compte de l'apprentissage des neurones biologiques mais de fournir une capacité d'adaptation à divers traitements d'information. Les modèles neuronaux centrés sur la description biophysique des mécanismes (potentiels d'actions, transmission synaptique) sont peu utilisés en intelligence artificielle aujourd'hui. On peut enfin mentionner un rôle important des neurosciences lors de tentatives de réalisation matérielle d'une intelligence artificielle puisque par leurs études anatomiques, les neurosciences fournissent des pistes pour réaliser des circuits émulant certaines propriétés de traitement d'information.

Enfin, la modélisation du cerveau et les travaux en intelligence artificielle ont déjà donné lieu à de belles fertilisations croisées. On peut mentionner notamment :

- les progrès impressionnants de l'apprentissage par renforcement\* , et en miroir l'étude du codage de l'erreur de prédiction de la récompense par le système dopaminergique du cerveau ;
- la description des codages procédural et déclaratif pour représenter l'information dans le cerveau et en intelligence artificielle ;
- ou encore la compréhension de mécanismes de résolution de problèmes par des aller-retour constructifs entre neurosciences et intelligence artificielle.

Sans préjuger où cette dynamique pourrait nous conduire, elle devrait en tout cas participer au progrès de ces deux disciplines.