

# APPRENTISSAGE CONTINU ET ESTIMATION DU GRADIENT INSPIRÉS DE LA BIOLOGIE POUR LE CALCUL NEUROMORPHIQUE

—  
PAR **AXEL LABORIEUX** (2014),  
CHERCHEUR EN NEUROSCIENCES COMPUTATIONNELLES  
EN POSTDOC À L'INSTITUT FRIEDRICH MIESCHER DE BÂLE

Les algorithmes d'IA accomplissent maintenant des tâches qui semblaient auparavant réservées aux êtres dotés de systèmes nerveux. Ces algorithmes sont tous basés sur le concept de réseau de neurones artificiels. Dans le but de baisser la consommation énergétique de ces algorithmes d'IA, Axel Laborieux s'est appuyé sur le calcul neuromorphique qui prend pour modèle le fonctionnement du cerveau humain.

Depuis une dizaine d'années, le domaine de l'intelligence artificielle (IA) a connu une révolution. Les ordinateurs sont désormais capables de réaliser automatiquement des tâches cognitives difficiles comme la détection d'objets dans des images ou vidéos en haute définition, la compréhension du langage naturel à partir de son ou de texte écrit, et même de battre le champion du monde de go, un jeu de société millénaire qui avait jusqu'alors résisté aux algorithmes à cause de son trop grand nombre de possibilités à chaque tour.

## Les récents développements de l'intelligence artificielle et ses limites

Dans un réseau de neurones artificiels, des unités qui s'apparentent à des neurones biologiques très simplifiés effectuent des opérations. Ces unités sont reliées entre elles par des connexions qui s'apparentent aux synapses connectant les neurones biologiques entre eux. Le réseau

peut alors réaliser un calcul à partir de données, par exemple une image, comme s'il s'agissait d'une entrée sensorielle, et produire un résultat.

Pour que ce résultat soit utile, il faut d'abord entraîner le réseau sur une grande base de données d'exemples dans un processus itératif où la force des connexions entre les neurones est modifiée petit à petit pour que le réseau s'améliore. Il est fascinant que les algorithmes d'IA puissent alors généraliser à des données jamais vues lors de l'entraînement et ainsi accomplir des tâches similaires à ce que ferait un cerveau, tout cela à partir de calculs inspirés initialement des réseaux de neurones biologiques !

Toutefois, les algorithmes d'IA possèdent plusieurs limitations qui les rendent très différents du cerveau sous beaucoup d'aspects. Tout d'abord, la consommation énergétique d'un algorithme d'IA, aussi bien pendant l'entraînement que pendant son utilisation ultérieure, est supérieure à celle du cerveau de plusieurs ordres de grandeur. La raison de cette surconsommation est d'abord architecturale : l'architecture de J. von Neumann des ordinateurs sépare physiquement le processeur où les calculs ont lieu, et la mémoire où les données sont stockées. Dans le contexte de l'IA, cette séparation donne lieu à un goulot d'étranglement où la plus grande partie de l'énergie est dépensée pour transporter la donnée d'un point à un autre. Dans le cerveau, une telle séparation n'existe pas. Le calcul neuromorphique est le domaine de recherche qui consiste à développer des architectures où la mémoire et les unités de calcul sont au plus proche, comme dans le cerveau.

## Le calcul neuromorphique pour contourner les limites des réseaux de neurones actuels

Au lieu de simuler un réseau de neurones sur ordinateur, pourquoi ne pas directement construire le réseau de neurones artificiels avec des composants physiques innovants ?

L'idée n'est pas nouvelle et remonte aux années 80 avec les travaux de Carver Mead. Des neurones et synapses artificiels reproduisant très fidèlement les propriétés observées en biologie ont été construits avec des composants physiques. Mais il n'est pas suffisant d'en assembler un grand nombre pour qu'un calcul utile se produise : le problème est aussi algorithmique car les connexions doivent changer de manière cohérente à l'échelle du réseau.

En effet, un composant neuromorphique ne devrait avoir besoin que de l'information qui lui est localement accessible dans l'espace et dans le temps pour limiter les mouvements de données, comme dans le cerveau. Cependant, les réseaux de neurones actuels n'ont pas été pensés pour être compatibles avec les contraintes physiques des composants neuromorphiques. Ainsi, la création de puces électroniques qui implémenteraient directement les algorithmes actuels d'IA à faible coût énergétique est difficile.

L'objectif de ma thèse a donc été d'adapter des principes présents dans le cerveau pour concevoir des algorithmes d'IA compatibles avec une implémentation sur puce électronique dédiée, et de participer à la conception de ces puces. Plus précisément, ma thèse se décompose en trois projets indépendants autour de cette thématique.

## Les réseaux de neurones artificiels ont la mémoire courte

Dans mon premier projet, je me suis intéressé au problème de l'oubli catastrophique dans les réseaux de neurones artificiels. L'oubli catastrophique est une limitation qui empêche les réseaux de neurones artificiels d'apprendre plusieurs tâches à la suite.

Si l'on entraîne un réseau de neurones à reconnaître des chiffres manuscrits, et qu'ensuite on souhaite utiliser ce même réseau pour apprendre à reconnaître des lettres, le réseau apprendra cette seconde tâche en oubliant très rapidement comment reconnaître des chiffres. Ce comportement est très différent du cerveau humain et il serait bien sûr préférable que le réseau continue à apprendre quand de nouvelles données sont disponibles ou qu'une nouvelle tâche doit être apprise, sans avoir besoin de réapprendre les anciennes données en même temps que les nouvelles.

*“Le calcul neuromorphique est le domaine de recherche qui consiste à développer des architectures où la mémoire et les unités de calcul sont au plus proche, comme dans le cerveau.”*

Le problème de l'oubli catastrophique s'explique par le fait qu'un réseau de neurones doit à la fois modifier ses connexions pour apprendre une tâche comme expliqué plus haut, mais aussi les empêcher de changer afin de conserver ce qui a été appris dans le passé, ce qui semble contradictoire à première vue.

Dans le cerveau, les synapses sont dites plastiques, car leurs forces peuvent être modifiées pour l'apprentissage, mais elles sont aussi métaplastiques : il existe des processus biologiques qui les rendent plus ou moins plastiques, de manière à consolider certains souvenirs et en oublier d'autres. Ce concept a été beaucoup étudié en neurosciences computationnelles, un domaine de recherche visant à modéliser les mécanismes des réseaux de neurones biologiques.

Dans un article publié dans le journal *Nature Communications*<sup>1</sup> dans le cadre de l'élaboration de ma thèse, nous avons montré comment ce concept de métaplasticité pouvait être naturellement appliqué aux réseaux de neurones binaires (c'est-à-dire pouvant prendre deux valeurs), une catégorie de réseaux de neurones particulièrement étudiée pour les applications de calcul neuromorphique.

Alors que les approches précédentes pour réduire l'oubli catastrophique dans les réseaux de neurones artificiels reposaient sur des calculs supplémentaires à effectuer entre les différentes tâches pour identifier les connexions à consolider, notre approche permet de les identifier directement pendant l'apprentissage, ce qui la rend compatible avec des composants neuromorphiques. Il est en effet envisageable de réaliser un tel comportement métaplastique en utilisant les propriétés physiques de nanomatériaux magnétiques. →

→ Cependant, bien que cette nouvelle règle de consolidation soit locale, le processus par lequel la connexion sait dans quelle direction évoluer ne l'est pas, ce qui constitue une autre limitation majeure des réseaux de neurones artificiels actuels pour l'implémentation avec des composants physiques.

### Un apprentissage très différent du cerveau

Mettons de côté l'apprentissage de plusieurs tâches successives et concentrons-nous sur l'apprentissage d'une seule tâche. La chaîne d'événements qui séparent une entrée sensorielle (par exemple la vue d'une taupe sortant d'un trou dans le jeu de la taupe) et la réaction qui en résulte (le mouvement du bras pour frapper la taupe avec le marteau) sont constituées de nombreuses étapes intermédiaires. L'information arrive par les yeux et traverse les différentes aires du cortex visuel avant de déclencher le geste. Cette idée de hiérarchie est aussi présente dans les réseaux de neurones artificiels utilisés actuellement : on dit qu'ils sont profonds. Dans les deux cas, l'entrée et la sortie du système sont séparées par plusieurs couches de neurones et de synapses.

Le défi qui se présente aussi bien aux réseaux de neurones biologiques qu'aux réseaux artificiels est le suivant : comment une synapse située loin en amont de la sortie peut-elle être modifiée pour améliorer les performances futures ? Il s'agit d'un problème difficile, car modifier une connexion loin de la sortie aura un effet en cascade sur le reste de la hiérarchie.

Bien que l'on ne sache pas comment fait le cerveau pour résoudre ce problème, des expériences en neurosciences montrent qu'une synapse est modifiée sur la base de l'activité locale des neurones qu'elle relie. Dans les réseaux artificiels, ce problème est résolu avec une méthode en apparence très différente qui consiste à calculer explicitement l'influence d'une modification synaptique sur le résultat en sortie (le gradient). Cette opération n'est plus locale mais effectuée par un système parallèle ou auxiliaire.

Jusqu'à présent, cette méthode fonctionne bien pour les réseaux de neurones codés. D'ailleurs, cela fonctionne beaucoup mieux que toutes les alternatives qui sont pourtant plus similaires à ce que semble faire le cerveau. Et encore une fois, cette différence avec la biologie rend difficile l'implémentation directe avec des composants physiques, car le système devrait être capable de réaliser deux types de calculs distincts avec les mêmes composants. Au lieu de calculer explicitement le gradient, les méthodes alternatives cherchent à l'estimer en fonction de quantités locales à chaque connexion. Cependant, même si ces méthodes sont plus plausibles du point de vue de la

*“L’oubli catastrophique est une limitation qui empêche les réseaux de neurones artificiels d’apprendre plusieurs tâches à la suite.”*

biologie, elles ne fonctionnent en général que sur des tâches simples et ne passent pas à l'échelle sur des tâches d'IA plus réalistes.

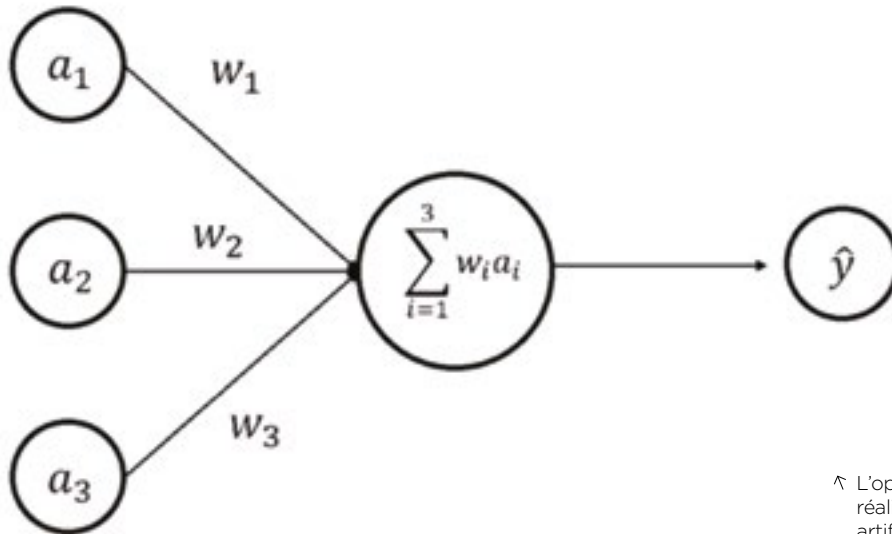
Dans une collaboration avec le Mila de Montréal et Thales à Saclay effectuée dans le cadre de ma thèse, nous avons montré qu'une de ces méthodes alternatives peut passer à l'échelle, ce qui ouvre la possibilité de concevoir des systèmes physiques capables d'apprendre en utilisant directement la physique de leurs composants, sans avoir besoin de connaître leurs caractéristiques individuelles. Une telle façon de réaliser l'apprentissage sera beaucoup plus efficace en termes de besoins énergétiques que l'utilisation de centres de calcul, pour les raisons expliquées précédemment. Cela permettra de déployer des systèmes intelligents dans des contextes embarqués comme les voitures autonomes, la robotique, etc.

À l'heure actuelle, des puces électroniques implémentant directement des réseaux de neurones artificiels commencent à voir le jour, mais elles sont encore à des stades préliminaires et se limitent à l'utilisation d'un réseau préentraîné sur ordinateur et transféré par la suite sur un circuit électronique dédié.

### Des matériaux innovants permettent la création de réseaux de neurones physiques

Les neurones artificiels simulés par les IA sont très simples comparés aux neurones biologiques. L'opération essentielle qu'ils réalisent se réduit à l'addition des signaux provenant d'autres neurones, préalablement multipliés par la force des connexions synaptiques, et à la comparaison de cette somme avec un seuil. Le neurone émet alors un résultat qui sera utilisé par les neurones suivants. Cependant, cette opération doit être réalisée un très grand nombre de fois à cause du nombre élevé de neurones et de connexions.

En utilisant des composants mémoire pour encoder les connexions et en les disposant sous forme de tableau où chaque ligne est connectée à un neurone de l'entrée, et chaque colonne à un neurone de sortie, il est possible d'implémenter ces opérations de manière très efficace du point de vue énergétique en utilisant par exemple les



^ L'opération essentielle réalisée par les neurones artificiels.

lois de conservation du courant électrique. Le défi est alors d'avoir suffisamment de précision et peu de variabilité entre les composants pour obtenir un résultat précis.

Pour ce type d'implémentation, les réseaux de neurones binaires évoqués plus haut sont particulièrement intéressants car les connexions et les neurones ne peuvent prendre que deux valeurs (par exemple 1 et -1). Les opérations à faire sur puce deviennent donc plus simples et plus robustes aux erreurs de lecture. Les performances d'un réseau de neurones binaires sur les tâches d'IA sont bien sûr moins élevées qu'un réseau utilisant des valeurs continues, mais restent convenables par rapport aux gains obtenus pour l'implémentation physique.

Des membres de mon équipe au C2N avaient conçu une telle puce pour encoder un réseau de neurones binaires. Dans une collaboration avec le CEA-Leti de Grenoble et l'université d'Aix-Marseille effectuée dans le cadre de ma thèse, nous avons montré qu'en utilisant la puce dans un régime particulier la même architecture pouvait être utilisée pour encoder une troisième valeur de connexion synaptique.

Cela permet au réseau d'être plus performant tout en conservant une implémentation robuste aux variations entre les composants et aux erreurs de lecture. Des estimations ont montré que ce type d'implémentation physique réduit l'énergie nécessaire pour utiliser le réseau de plusieurs ordres de grandeur comparée à l'utilisation d'ordinateurs actuels. En effet, dans ce type d'architecture, les composants mémoire codant les connexions synaptiques se trouvent au cœur du circuit effectuant le calcul et n'ont donc pas besoin d'être déplacés comme dans l'architecture de J. von Neumann des ordinateurs classiques.

L'interdisciplinarité est au cœur du développement de systèmes intelligents.

Ces trois projets menés pendant ma thèse m'ont permis de prendre la mesure des défis à surmonter pour la création de systèmes reposant sur l'utilisation d'IA nécessitant peu d'énergie. L'interdisciplinarité est probablement le plus important d'entre eux. Construire de tels systèmes met en jeu des connaissances de physique, d'électronique, de neurosciences et d'informatique.

Les approches développées indépendamment dans chaque discipline ont tendance à ne pas prendre en compte les contraintes et les avancées des autres domaines. Par exemple, créer un système qui ne ferait qu'imiter les neurones biologiques dans toute leur complexité sans prendre en compte les avancées algorithmiques dans le domaine de l'apprentissage machine ne produira pas de calcul utile. De même, une approche consistant à implémenter tels quels les algorithmes de l'apprentissage machine sur une puce est susceptible de ne pas fonctionner.

L'interaction entre disciplines est bénéfique : la révolution récente de l'IA a eu pour origine l'étude du cerveau, et les méthodes récentes d'IA permettent à leur tour de définir de nouvelles directions de recherche en neurosciences en posant de nouvelles questions. C'est pour moi la preuve que seul un effort commun dans toutes ces disciplines nous permettra de comprendre et répliquer les principes d'une intelligence nécessitant peu d'énergie. ×

1. Lien : <https://www.nature.com/articles/s41467-021-22768-y>



**AHEL  
LABORIEUX  
(2014)**

Après ma thèse sur le calcul neuromorphique, j'ai décidé d'enchaîner sur un postdoctorat en neurosciences computationnelles dans le groupe de Friedemann Zenke à l'Institut Friedrich Miescher pour la recherche biomédicale à Bâle, en Suisse.