

# APOLLO : quand l'IA vient en aide à l'IA

De l'IA pour améliorer l'IA : c'est le sens d'APOLLO. Ce *framework* vise plus précisément à optimiser la conception des puces dédiées à l'apprentissage automatique. Google Research [en est à l'origine](#).

APOLLO doit adapter le plus efficacement possible l'architecture des puces en question (c'est-à-dire la manière dont leurs blocs fonctionnels s'agencent) à des ensembles de tâches. Il utilise pour cela diverses méthodes d'exploration.

Il s'agit de consacrer le moins possible de ressources à cette exploration. Cela implique, pour chaque ensemble de tâches, d'éviter au maximum les combinaisons inadaptées. Mais il faut aussi s'assurer que les combinaisons « valables » soient le plus performantes possible.

Sur ces deux critères associés, la méthode « évolutionniste » – illustrée ci-dessous) se révèle particulièrement efficace. Elle représente les configurations de puces sous forme de séquences correspondant à des « génomes » qu'on peut croiser.



Y coupler une approche d'optimisation basée sur les modèles permet de maintenir un haut niveau d'efficacité à mesure qu'on introduit des contraintes (ici, de surface exploitable).

APOLLO comporte aussi un mécanisme d'apprentissage par transfert. Cela lui permet d'assurer la reproduction de configurations « idéales » entre des architectures cibles soumises à des contraintes différentes. Toutes les méthodes d'optimisation mises à l'épreuve (sur 7 modèles, dont 2 variations de MobileNet) en bénéficient.

Les équipes de Google avaient [déjà](#) appliqué le *machine learning* à la conception de composants, mais à plus bas niveau. En l'occurrence, pour déterminer l'agencement des circuits qui constituent une puce, comme on ferait le plan d'un immeuble.

*Illustration principale © Jakub Jirsk – Fotolia*