

IA : Alibaba dévoile sa puce Hanguang 800

[Après Intel](#), le groupe chinois [Alibaba](#) a dévoilé une nouvelle puce d'inférence basée sur l'intelligence artificielle (IA), mercredi lors de sa [conférence Apsara](#) à Hangzhou.

La puce, nommée Hanguang 800, intègre une unité de traitement neuronal (neural processing unit ou NPU) pour accélérer les travaux d'inférence propres à l'apprentissage automatique (machine learning).

Le projet a été développé par T-Head, l'unité de recherche en charge des processeurs pour le cloud et le calcul à la périphérie du réseau (edge computing) au sein d'Alibaba DAMO Academy.

Les chercheurs ont développé l'architecture matérielle de la puce et ses algorithmes. Des algorithmes optimisés pour des applications métiers, de la logistique au commerce.

Alibaba Group dit utiliser actuellement le processeur Hanguang 800 pour ses propres opérations. La traduction automatique, la recherche de produits et les recommandations sur ses sites de e-commerce sont concernées. La publicité et le support client le sont aussi.

Informatique avancée

La puce peut permettre à une machine d'accomplir « en cinq minutes » plutôt qu'en une heure auparavant certaines tâches informatiques, selon les promoteurs de l'offre.

Hanguang 800 « améliore les capacités de calcul qui stimuleront nos activités actuelles et émergentes, tout en augmentant l'efficacité énergétique », a ainsi déclaré le CTO d'Alibaba Group et président d'[Alibaba Cloud Intelligence](#), Jeff Zhang (en photo).

« Dans un avenir proche, a précisé le dirigeant, nous prévoyons de donner à nos clients les moyens d'accéder à de l'informatique avancée, grâce à cette puce, à tout moment et en tout lieu, par l'intermédiaire de notre activité cloud. »

AWS, Microsoft Azure et Google Cloud sont prévenus.