Quand les IA miment l’activité cérébrale

Grâce à la technique de l’apprentissage auto-supervisé, les réseaux de neurones artificiels sont désormais capables d’identifier par eux-mêmes ce qui est important. Une performance qui pourrait être à la base du succès de notre propre cerveau.

Depuis une décennie déjà, les plus impressionnants systèmes d’intelligence artificielle sont majoritairement formés à l’aide d’énormes inventaires de données « étiquetées ». En effet, on étiquette une image, par exemple « chat tigré » ou « chat-tigre » (comme l’oncille en Amérique du Sud ou en le serval Afrique) pour « entraîner » un réseau de neurones artificiel à distinguer correctement un chat tigré d’un chat-tigre. Cette stratégie s’est révélée à la fois spectaculairement gagnante et terriblement déficiente.

Un tel entraînement, dit « supervisé », nécessite des données laborieusement étiquetées par des humains, et les réseaux neuronaux prennent souvent de mauvais raccourcis, apprenant à associer les étiquettes à des informations minimales voire, parfois, superficielles. Par exemple, un réseau de neurones pourrait utiliser la présence d’herbe pour reconnaître la photo d’une vache, du fait que les vaches sont généralement photographiées dans des champs.

LIRE L’ARTICLE