

Livre blanc coordonné par Bertrand Braunschweig, directeur du centre de Saclay (Île-de-France), Institut national de recherche en informatique et en automatique (Inria), septembre 2016

La cellule « Veille et prospective » de l'Inria décrit, d'une manière très pédagogique et en 80 pages seulement, les différentes technologies de l'[intelligence artificielle](#) (IA) et aborde les problématiques majeures liées à ces progrès. « *Plus nous développons l'intelligence artificielle, plus le risque est grand d'étendre uniquement certaines capacités intelligentes (par exemple optimisation et fouille par l'apprentissage) au détriment d'autres, peu susceptibles de générer un retour sur investissement immédiat ou de présenter un quelconque intérêt pour le créateur de l'agent (par exemple : morale, respect, éthique, etc.)* » avertit l'Inria.

À ce propos, il cite les chercheurs Thomas G. Dietterich, et Eric J. Horvitz qui recommandent « *d'accorder plus d'attention aux cinq risques à court terme auxquels sont exposés les systèmes basés sur l'IA, à savoir : les bugs dans les logiciels ; les cyberattaques ; la tentation de jouer à "l'Apprenti Sorcier", c'est-à-dire donner la capacité aux systèmes d'IA de comprendre ce que veulent les utilisateurs au lieu d'interpréter littéralement leurs ordres ; "l'autonomie partagée", à savoir la coopération fluide des systèmes d'IA avec les utilisateurs de façon que les utilisateurs puissent toujours reprendre le contrôle en cas de besoin et enfin, les impacts socio-économiques de l'IA : en d'autres termes l'IA doit être bénéfique pour l'ensemble de la société et pas seulement pour quelques privilégiés* ».

À lire sur inria.fr ou [télécharger directement le livre blanc](#).

N°41 Hiver 2016-2017