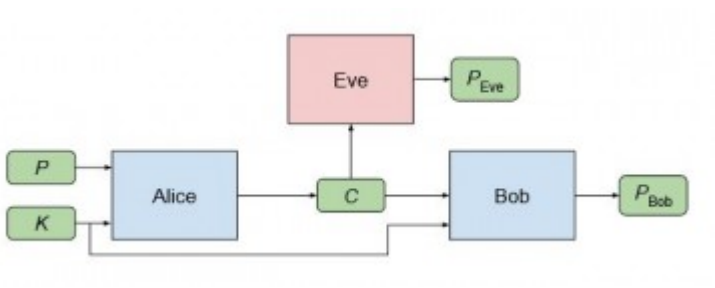
Une IA est capable de concevoir son propre chiffrement

A la fois une technique très prometteuse pour sécuriser les communications et une démonstration très pratique des questions éthiques que pose l'IA. Dans un article de recherche, une équipe de Google montre comment elle a poussé deux intelligences artificielles à mettre en œuvre des techniques de chiffrement pour échapper aux tentatives d'espionnage d'une troisième IA, chargée elle de percer au jour les communications.

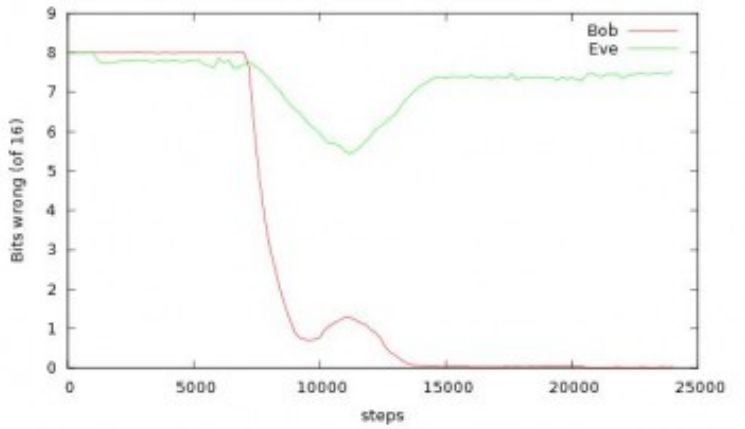


Pour ce faire, l'équipe de Google Brain, qui travaille sur les sujets d'apprentissage profond (Deep Learning), a programmé trois réseaux de neurones – baptisés Alice, Bob et Eve. Alice devait envoyer un message à Bob, que ce dernier devait déchiffrer tandis qu'Eve était programmé pour intercepter cette communication et tenter d'en décoder le contenu. Si Alice et Bob partageaient d'entrée une clef de chiffrement (K dans le graphique ci-dessus), ils n'avaient pas été programmés pour savoir chiffrer des communications. Simplement, si le message déchiffré par Bob était trop éloigné de l'original, l'IA avait été programmée pour le considérer comme un échec. Alice, elle, associait la notion d'échec au fait qu'Eve parvienne, dans ses tentatives d'interception du message, à un résultat supérieur à celui que fournirait le hasard. Enfin, pour Eve, un écart trop important entre le message envoyé par Alice et le résultat qu'elle parvenait à décoder était vu comme un revers.

Les IA découvrent seules le chiffrement

Sur la base de cette programmation assez simple, les chercheurs Martin Abadi et David G. Andersen ont lancé plusieurs essais. Si certains se sont soldés par des flops, Bob se montrant incapable de reconstruire le message d'Alice, dans la plupart des cas, les deux IA ont su mettre au point un système leur permettant d'échanger avec très peu d'erreurs. Et quand, lors de certains tests, Eve s'approchait d'un déchiffrement du message, les deux autres IA ont su réagir en améliorant leurs techniques cryptographiques jusqu'à repousser les tentatives de l'espionne.

Comme le montre le graphique ci-contre, si Bob commence à comprendre Alice après quelque 7 500 itérations (le taux d'erreurs chutant rapidement), Eve y parvient également peu à peu mettant en danger le secret échangé par les deux autres IA. Mais celles-ci prennent alors des contre-mesures, en changeant leurs techniques de chiffrement pour laisser Eve dans le brouillard le plus total à partir de quelque 15 000 itérations. Ces résultats montrent que les IA peuvent donc « *découvrir des formes de chiffrement et déchiffrement, sans qu'on leur ait enseigné des algorithmes spécifiques pour ce faire* », écrivent les deux chercheurs.



A lire aussi :

L'IA prédit les verdicts de la Cour européenne des droits de l'homme

IA : comment IBM Watson aide les conseillers clientèle du Crédit Mutuel

L'IA va devenir une priorité des DSI, dit le Gartner

Crédit photo : agsandrew / shutterstock