

Formation : Salesforce affine le machine learning de Trailhead

Salesforce a annoncé l'intégration de nouveaux modules dans [sa plate-forme de formation Trailhead](#).

La mise à jour en question permet aux développeurs, aux rédacteurs et aux cadres qui animent la plate-forme d'utiliser l'IA de façon responsable et les sensibiliser à l'impact qu'elle peut avoir sur les utilisateurs et l'entreprise.

Il faut dire que le fonctionnement d'une IA reste bien souvent totalement opaque pour l'utilisateur.

Mieux comprendre le fonctionnement des algorithmes

C'est d'autant plus le cas en raison de la présence d'algorithmes d'apprentissage automatique qui évoluent au fur et à mesure du traitement des données. Et justement comment évoluent-ils ? Restent-ils dans l'éthique convenue, sont-ils réellement partiels ?

Pour limiter les déviations, la mise à jour ajoute donc ces nouveaux modules Trailhead, conçus pour aider les clients à mieux comprendre le contenu d'un modèle d'IA.

Avec eux, ils pourront exclure certaines données qui parasitent l'IA et identifier ses déviations facilement.

Autrement dit, les utilisateurs vont pouvoir surveiller l'évolution des algorithmes afin qu'ils évitent de biaiser les résultats par « préjugés ».

□ New module alert! We're so excited to announce the Responsible Creation of Artificial Intelligence module to empower anyone to learn how to use and build AI in a responsible and trusted way. Read more about the importance of this new module today: <https://t.co/ArSkduoXZ3> pic.twitter.com/xOWPXuTxUM

— Trailhead (@trailhead) [28 mai 2019](#)

Les inévitables biais du machine learning

Dernièrement, [une étude](#) de quatre chercheurs français [a révélé](#) que les algorithmes ont tendance à reproduire les valeurs implicites des humains. Et donc à guider parfois vers des décisions biaisées par rapport à un comportement rationnel ou à la réalité.

Une partie des biais repérés sont dits « cognitifs ». Ils sont liés à la manière dont les algorithmes sont écrits. Le programmeur peut avoir suivi des modélisations populaires sans s'assurer de leur exactitude (cas du « mouton de Panurge »). Il peut aussi favoriser sa vision du monde (biais d'anticipation et de confirmation) ou encore déceler des corrélations entre deux événements indépendants (biais de « corrélations illusoires »).

D'autres biais sont dits « statistiques » concernent en premier lieu les données d'entrée – en d'autres termes, celles sur lesquelles les algorithmes s'entraînent.

Les biais des algorithmes peuvent également avoir une motivation financière. Ce sont les biais économiques. Google [favorisant ses propres services dans son moteur de recherche](#) en est une parfaite illustration.