

# Les machines comprennent enfin le français, sans magie ni boîte noire

*Les machines d'aujourd'hui sont en passe de résoudre des problèmes complexes majeurs grâce aux avancées extraordinaires que nous faisons depuis plus de 20 ans en informatique. Les pères fondateurs de l'intelligence artificielle avaient vu juste sur ce point, les machines peuvent aujourd'hui interpréter des images et comprendre du texte.*

Temps de lecture : minute

---

27 décembre 2020

Republication du 7 juillet 2020

Pour faire avancer la compréhension automatique du langage naturel en français, nous avons choisi des voies technologiques et éthiques différentes des grands acteurs américains. Les entreprises ont besoin de nos outils pour se transformer, mais nous devons être attentifs pour éviter la désillusion face à des solutions imaginées comme magiques, peu respectueuses de nos utilisateurs, et de nos valeurs.

## L'intelligence artificielle n'existe pas !

Luc Julia le dit très bien dans son dernier livre : *“L'intelligence artificielle n'existe pas”* . Ce que l'on appelle intelligence artificielle aujourd'hui n'est pas très intelligent. Il s'agit le plus souvent de faire des rapprochements entre des millions d'octets de données à notre disposition, et de nouvelles données sur lesquelles on souhaite faire une prédiction. Ce sont des mathématiques et des statistiques, rien de plus. Donnez des millions de

photos de chat pour entraîner un algorithme à reconnaître des chats, il sera capable de reconnaître à peu près correctement les prochains chats qu'on lui proposera. Présentez lui ensuite une photo de chien, il sera uniquement en mesure de vous dire qu'il ne s'agit pas d'un chat.

Beaucoup d'approches en vogue fonctionnent sur le même principe pour comprendre le langage, en faisant des rapprochements statistiques sur le positionnement des mots entre eux, au sein d'une phrase et au sein d'un paragraphe ou d'un corpus de texte plus large. Ce n'est pas suffisant pour espérer comprendre réellement le langage et tirer le maximum de ce texte brut qui représente une véritable mine d'or, notamment pour les entreprises.

## Des algorithmes pensés, conçus et alimentés par des humains

Nous avons repris les bases de la langue française pour imaginer une nouvelle méthode adaptée à sa compréhension par des algorithmes. Avec une nouvelle approche de la technologie appliquée au traitement de la langue, nous sommes fiers de proposer aujourd'hui une solution complète pour répondre à un besoin primordial des entreprises françaises : tirer profit des millions de lignes de texte brut non exploitées dans leurs systèmes d'information.

La solution n'est pas magique, elle repose sur des méthodes éprouvées à base de dictionnaires d'ontologies catégorisées par des vrais humains. Mais attention, pas des micro travailleurs payés à la tâche. Nous travaillons avec des équipes dédiées via une plateforme d'annotation responsable créée et maîtrisée par nos développeurs afin de nous assurer de la qualité de nos bases de connaissances sémantiques. Une démarche contrôlée tout au long de la chaîne, en France, pour le français.

Cette approche fine de l'enrichissement par des humains qualifiés est complétée par l'apport de solutions d'apprentissage machine (*machine learning*) et d'apprentissage profond (*deep learning*) pour traiter un volume de données toujours plus important. Nous croyons dans la complémentarité de ces approches plutôt que dans les guerres de clochers qui agitent les communautés de l'intelligence artificielle depuis leurs débuts. Une approche hybride qui met en valeur l'expertise et l'inventivité de nos ingénieurs et de nos spécialistes du langage.

## Non à la boîte noire, pensons des algorithmes auditables, aux décisions explicables !

Pour proposer une solution efficace, robuste et digne de confiance aux entreprises françaises et à leurs clients, il n'est pas concevable de se reposer sur des boîtes noires statistiques. Nos algorithmes de compréhension du langage naturel sont auditables et nous pouvons en toute circonstance expliquer le cheminement qui aboutit à la décision.

C'est un pré-requis que j'appelle de mes vœux pour toutes les solutions embarquant des algorithmes de prédiction et de prise de décision. Nous devons, à tout moment, avoir la main sur les résultats fournis par nos systèmes automatisés. La confiance se gagne là où les fantasmes et les peurs s'arrêtent.

Il est temps d'arrêter de vouloir mettre de l'intelligence artificielle partout, pour se concentrer sur des cas concrets qui apportent un véritable service. Le traitement automatique du français est un des sujets les plus importants du champ de l'intelligence artificielle pour les entreprises, et nous sommes prêts à passer à l'étape suivante pour intégrer cette brique de manière industrielle.

La compréhension du texte et de la donnée brute deviennent une commodité, une nouvelle marche vers une utilisation business toujours plus fine et plus intelligente. Exploisons vite tout son potentiel avec pragmatisme et ambition pour transformer les données brutes inexploitées en véritable mine d'or !

*Charles Borderie est co-fondateur de Lettria*

---

Article écrit par Charles Borderie