

intelligence artificielle et sémantique : un couple sur le long terme ?

L'intelligence artificielle a le vent en poupe et s'infiltre progressivement dans la plupart des outils que nous utilisons quotidiennement. Sur ce terrain, la sémantique a-t-elle un rôle à jouer ou appartient-elle au passé ? Dans quel cas peut-elle s'avérer utile ?



« notre particularité, c'est que nous sommes capables d'extraire le sens d'une requête, de la comprendre et d'y apporter une réponse »

Christophe Clot, Davi

Certains en sont persuadés : l'intelligence artificielle, ou plus largement l'information cognitive, sera la future révolution « *post internet* ». D'autres, comme Microsoft, l'affirment sans détour : « *2017 sera l'année de l'IA* ». Applications, services, industrie, entreprises... La déferlante de l'IA, qui a fait des progrès majeurs l'an passé, est en passe de s'accélérer, et pourquoi pas d'amorcer cette fameuse « *quatrième révolution industrielle* » dont on entend tant parler. Pour l'y aider, les technologies cognitives, en particulier la sémantique, semblent avoir plus d'un atout dans leur manche. Notamment dans l'objectif de créer des machines intelligentes capables de comprendre, d'interpréter et de restituer le langage naturel humain. Amélioration des échanges clients, robots-rédacteurs ou encore chatbots... Elles savent se montrer indispensables.

C'est l'avis de Christophe Clot, le fondateur de Davi, une start-up française qui travaille depuis sa création en 2000 sur les agents virtuels. À ses débuts, les robots conversationnels relevaient encore du fantasme. Davi commence donc par s'intéresser à l'animation 3D de la partie « *incarnée* » de ces agents. Après avoir lancé un moteur conversationnel avec différents prestataires

français, qui a rapidement montré ses limites, la société monte en 2010 une équipe de R&D dédiée à la compréhension et au traitement automatique du langage naturel. Cela donne Rétorik, une plateforme de web services disponible en Saas.

Rétorik embarque un moteur d'animation 3D (gestuelle, émotion, expression) et un moteur conversationnel basé sur de l'intelligence artificielle (traitement du langage et des émotions) dans des logiciels métier et sur tous types de médias (hologramme, site web, objet connecté, téléphone ou borne interactive, etc.). « *Notre particularité*, explique Christophe Clot ; *c'est que nous sommes capables d'extraire le sens d'une requête, de la comprendre et d'y apporter une réponse* ».

Pour ce faire, Davi a développé un dictionnaire auto-apprenant de plus d'un million d'entrées, qui se met à jour toutes les 48 heures sur internet. Breveté en 2016, c'est grâce à lui que la société peut donner du sens aux mots et aux contextes. À celui-ci est associé une ontologie permettant à un agent conversationnel d'être personnalisé pour un environnement métier précis et de faire de lui un « *chatbot expert* ».

En pratique, Davi a déployé ses agents dans le secteur du médical (service Vidal), dans les assurances (pour informer et orienter les

clients sur différents contrats), ou encore auprès d'offices de tourisme (sur des bornes ou des applications). « *La technologie est commune à l'ensemble de nos clients*, explique Christophe Plot ; *mais la base de connaissance est personnalisée selon les environnements et les ambiguïtés à lever. L'objectif est que l'agent conversationnel ne confonde pas deux demandes qui se ressemblent : "je voudrais le prêt d'une voiture" et "je voudrais un prêt pour une voiture"* ».

Cinq séquences sont donc nécessaires pour traiter la requête d'un utilisateur : débruitage (le chatbot est capable de comprendre le langage SMS ou les mots inutiles d'une requête), lématisation (le chatbot étudie le registre du langage et en extrait des concepts forts), compréhension du sens des concepts (grâce à l'ontologie), construction de la réponse et définition de la bonne stratégie de dialogue, et enfin la réponse. Demain, la technologie mise au point par Davi pourrait même être capable de comprendre les émotions de l'interlocuteur à travers sa voix et son image.

Même son de cloche chez Syllabs, qui aime se définir comme une « *vieille start-up de dix ans* », et dans laquelle ses deux fondateurs, tous deux experts en sémantique pour la Commission européenne, ont réuni leurs compétences respectives en machine learning et en

linguistique. Leur créneau ? La production automatique de textes personnalisés et actualisés pour leurs clients. Chez Syllabs, les ingénieurs linguistes côtoient les informaticiens, les data scientists et les scénaristes pour qu'un algorithme puisse collecter, trier et organiser des millions de sources afin de les transformer, en temps réel, en informations publiables.

Le réseau Guy Hoquet l'Immobilier fait partie de ses clients. Conscient de l'importance de développer une stratégie de contenus web produits en masse (inbound marketing), son ambition était de créer, pour chacune de ses agences, un média local dédié à l'actualité de son quartier. Un projet démesuré compte tenu des 450 médias locaux qu'il aurait fallu créer. « Syllabs a donc relevé le défi, explique l'un de ses fondateurs, Claude de Loupy ; plusieurs centaines d'articles sont en effet écrits chaque jour à partir de données structurées provenant de l'open data, de données web et d'informations sourcées par les agents Guy Hoquet : données de l'Insee, secteur immobilier, vie locale, sport, commerce, etc. » Toutes ces données hétérogènes sont ensuite fusionnées dans une



« un outil qui enrichit la connaissance et l'actualité d'un secteur ou d'un territoire, en offrant un accès à des milliers de données existantes »

Claude de Loupy, Syllabs

outil d'information pertinent qui enrichit la connaissance et l'actualité d'un secteur ou d'un territoire, en offrant un accès éditorialisé à des milliers de données existantes ». Pourtant, force est de constater que la technologie de Syllabs a toute sa place dans l'univers des médias. La liste de ces clients, dont *Le Monde*, *Le Parisien*, *L'Express* ou encore Radio France, en atteste.

« Vous ne le savez pas encore, mais l'intelligence artificielle peut révolutionner votre business ». Comme le prouve son site internet, l'éditeur Antidot, acteur français de réfé-

nous faisons hier, explique-t-il ; nous avons fait le tri dans les technologiques et les marchés, nous avons massivement investi en R&D pour nous attaquer à de nouveaux challenges, dont le deep learning ».

L'éditeur, qui considère « qu'on ne peut pas résoudre les problèmes du futur avec les technologies du passé », s'est donc recentré sur deux angles : son moteur de recherche (actuellement en test bêta chez ses clients), qu'il affirme avoir totalement réinventé, et sur deux technologies de text mining jugées indispensables pour classifier et extraire des méta-données destinées à alimenter la base d'apprentissage dont il a besoin pour s'attaquer au marché du deep learning. C'est tout juste s'il accepte de reconnaître utiliser une forme de sémantique dans sa nouvelle technologie de moteur de recherche afin de répondre à des traitements de langage naturel, mais en retirant les technologies sémantiques. « Plus les volumes de données sont importants, plus on est obligé d'enlever les modèles linguistiques, poursuit Fabrice Lacroix ; la sémantique va donc être retirée petit à petit, car elle constitue un carcan pour les technologies. J'estime par exemple que plus les technologies des agents conversationnels vont avancer, plus les agents seront entraînés, moins ils auront besoin de modèles sous-jacents ».

Difficile encore de comparer la pertinence et la performance de technologies dont seul le temps, qui fonctionne à une vitesse exponentielle dans ce domaine, pourra trancher. D'ailleurs, preuve que celui-ci s'accélère toujours plus, le site internet d'Antidot présente encore son produit Antidot Finder Suite comme « le moteur de recherche sémantique le plus avancé et le plus complet du marché ». ■

Clémence Jost



« plus les volumes de données sont importants, plus on est obligé d'enlever les modèles linguistiques »

Fabrice Lacroix, Antidot

immense base de connaissances. Elles sont alors liées entre elles et c'est à ce moment-là que des scénarios, bâtis par des linguistes, pourront être utilisés par l'intelligence artificielle pour créer un texte. « Les scénaristes ne rédigent pas des textes à proprement parler, poursuit Claude de Loupy ; il s'agit plutôt de micro segments de textes, constitués de vocabulaires spécifiques à notre client, qu'il valide toujours en amont ».

Bien sûr, ces contenus n'ont pas vocation à concurrencer la presse. Selon Claude de Loupy, « ils constituent un

rence en moteurs de recherche d'entreprise, a su prendre le virage de l'IA. Son objectif : améliorer les performances de l'accès à l'information à l'aide du machine learning, qui rend possible de nombreux projets jusqu'à présent irréalisables grâce notamment à la classification automatique d'immenses corpus de documents. Mais surtout, ne lui parlez plus de sémantique ! Son président Fabrice Lacroix tient à prendre le contrepied de ce qu'il considère être une technologie d'un autre âge : « Ce que nous proposons aujourd'hui n'a plus rien à voir avec ce que