

L'intelligence artificielle s'empare du dark data

Ce n'est pas sans raison que les GAFAMI s'intéressent autant à nos données personnelles.

Tout le monde convient que les données sont importantes aujourd'hui et le seront encore plus à l'avenir. La donnée, souvent qualifiée de nouvel or noir, a pourtant une autre utilité que celle d'être simplement un actif de l'entreprise : exploitée et valorisée par une IA éthique, elle devient aussi un levier de croissance pour l'entreprise.

Le dark data, cette donnée inexploitée

De 80 à 90% des données générées et collectées par les organisations ne sont pas structurées ⁽¹⁾; inexploitées, elles grossissent jour après jour les rangs du dark data.

Gartner définit le dark data comme les actifs d'information que les organisations collectent, traitent et stockent au cours des activités commerciales normales, mais qui échouent à les utiliser à d'autres fins (par exemple, l'analyse ou l'amélioration des relations commerciales).

Ainsi, les organisations conservent souvent les dark data à des fins de conformité uniquement. Pourtant, le stockage et la sécurisation des données entraînent généralement plus de dépenses (et parfois plus de risques) que de valeur, car elles ne sont pas utilisées..

Ces données non structurées et inexploitées, qui n'ont pas encore été analysées ou traitées, sont toutefois un véritable gisement de productivité pour l'entreprise. Aujourd'hui, seulement 31% des entreprises arrivent à exploiter les données non structurées et à en tirer de la valeur⁽²⁾.

La donnée, un levier de croissance ignoré

On peut s'étonner de l'absence d'intérêt des entreprises pour ces données non structurées. Bien souvent, les entreprises sont hésitantes, effrayées par le temps de traitement de données qu'elles considèrent être dans un format inaccessible. Ces dernières se méfient également de la gestion de l'aspect légal lié à l'utilisation de ces données, en particulier depuis [la mise en place du RGPD](#).

Les données sont parfois jugées trop anciennes ou obsolètes pour fournir de la valeur. Les français sont d'ailleurs parmi les plus sceptiques, selon une étude Splunk ⁽³⁾ : la majorité des répondants français (64%) reconnaissent la valeur des données comme extrêmement ou très précieuses pour le succès global de leur organisation, comparé à 81% pour la moyenne mondiale. Les raisons de ce désamour sont à la fois techniques et organisationnelles.

Les bons outils au bon moment

Pourtant, ces doutes sont vite balayés dès que les entreprises engagent les bonnes approches avec les bons outils. Il existe d'autres approches que la méthode machine learning, souvent jugée longue, fastidieuse (apprentissage), parfois obscure (effet boîte noire) à mettre en œuvre.

L'automatisation de certaines tâches est ainsi possible au travers de l'IA symbolique, parangon de transparence, qui, du fait de son explicabilité by design, va rassurer les utilisateurs et [le délégué à la protection des données](#) (DPO) en particulier.

Les exemples sont nombreux : pour un bid manager, ce sera d'automatiser l'extraction d'informations pour permettre aux gestionnaires de décupler le nombre d'appels d'offres traités; pour un compliance officer dans le secteur bancaire, ce sera d'assurer [la preuve de la conformité](#), au travers de l'analyse de données non structurées comme des textes ou des conversations téléphoniques. Le nombre de ces données disponibles est exponentiel; le caractère d'immédiateté dans leur traitement prend, lui aussi, de plus en plus d'importance.

Un flux nouveau et continu de données

L'essor des solutions de communication et de collaboration, depuis le développement du télétravail et les périodes de confinement, est sans commune mesure. Nous vivons une période de multiplication des types de communications, qui deviennent de plus en plus asynchrones (ou fast asynchrones), mêlant plusieurs médias au même moment.

On constate, ainsi, une accélération des applications mobiles de type messageries instantanées, au détriment des SMS ⁽⁴⁾, qui voient leur recul s'accroître fortement depuis le deuxième trimestre 2020, avec un effet plus important durant les deux confinements : -21% au quatrième trimestre, après -15% au troisième et -23% au deuxième. Les applications de messagerie sont d'ailleurs autant utilisées pour un usage professionnel que privé.

Le 12 février 2020, WhatsApp a annoncé avoir dépassé la barre symbolique des 2 milliards d'utilisateurs dans le monde entier, avec 1,4 milliard d'appels vocaux et vidéo effectués le soir du Nouvel An 2020 dans le monde. Cette donnée nouvelle rarement exploitée, issue de ces médias divers, va enrichir l'expérience client. Sa nécessaire utilisation dans le secteur de la relation client va là aussi bénéficier des prouesses de l'IA symbolique, seule capable de reconnaître plusieurs intentions dans un même message.

L'objectif est bien de se servir de la donnée, toute la donnée disponible, pour décider et mieux décider avec moins d'erreurs. L'intelligence artificielle va permettre un passage à l'échelle de l'analyse de cette donnée inexploitée, qui serait impossible avec les méthodes traditionnelles manuelles. C'est une des missions du Chief Data Officer que de favoriser l'implication des métiers, par l'acculturation de l'ensemble des parties prenantes de l'entreprise au bon usage des données et de l'IA.

—

(1) [Unstructured Data, MongoDB](#)

(2) [Managing unstructured data and extracting value.](#) A-Team, 2020

(3) [The state of dark data.](#) Spunk, 2019.

(4) [Marché des communications électroniques en France \(T4 2020\).](#) Arcep, avril 2021.