



INTELLIGENCE ARTIFICIELLE & VALORISATION DES DONNÉES TEXTUELLES



Un livre blanc QWAM

© QWAM CONTENT INTELLIGENCE 2020

PRÉFACE



Françoise Soulié-Fogelman
Conseiller scientifique Hub France IA

L'Intelligence artificielle est la troisième vague dans la mise en place de l'économie numérique, après l'accès aux données (arrivée des sources de données abondantes des mobiles et du web), puis le développement des outils analytiques (big data et machine learning). Cette troisième vague a de fait souvent pris la forme d'un tsunami en balayant beaucoup de travaux antérieurs. Prenons le cas d'ImageNet, un concours de reconnaissance d'images, organisé chaque année depuis 2010 où chaque édition voyait les chercheurs du domaine apporter des améliorations de performance de quelques pour cent. En 2012, Geoff Hinton avec ses co-auteurs propose une méthode à base d'apprentissage par deep learning et écrase la concurrence en gagnant plus de 9% de performance. Les années suivantes, tous les candidats utilisent du deep learning et ont d'ailleurs atteint des performances surpassant celles des humains.

Pour le traitement des données textuelles, il s'est passé quelque chose de similaire : en 2009, un papier de Google parle de « The Unreasonable Effectiveness of Data ». Plutôt que d'utiliser des grammaires complexes, il suffirait, pour beaucoup de tâches, d'apprendre à partir de (très) grands

corpus les combinaisons de mots (statistiques des n-grammes). L'arrivée des modèles deep learning, après 2012 et leur victoire à ImageNet, n'a fait que renforcer ces positions : ils permettent de construire des espaces de représentations extrêmement efficaces en exploitant de grands corpus de textes non annotés largement disponibles sur le web. Ainsi Google a déployé en 2018 le modèle BERT pour construire ses modèles de langage (en anglais).

Alors, les jeux sont-ils joués ? Doit-on penser que les modèles d'apprentissage IA purs (deep learning) ont définitivement enterré les méthodes des linguistes pour le traitement des langues ? En fait, pas encore. Les ontologies sont des composants clés en traitement de la langue naturelle (TLN) pour représenter la connaissance d'un domaine. Pour l'instant, les modèles d'IA ne savent pas apprendre automatiquement des relations, ces ontologies. Il faut donc, pour l'instant en tous cas, être capable de mixer les approches classiques de TLN et les modèles d'apprentissage IA, en exploitant, avec pragmatisme, les avantages de chacune.

C'est cette approche que QWAM a mise en œuvre et les cas d'usage présentés ici illustrent sa très grande efficacité. Bonne lecture !

SOMMAIRE

INTRODUCTION , Christian Langevin	5
IA & VALORISATION DES DONNÉES TEXTUELLES	6
Panoramas des cas d'usage	
ÉTAT DE L'ART , Vincent Guigue	8
IA, NLP et apprentissage statistique	
QWAM ET L'IA , Ivan Monnier	13
Comment l'IA s'impose chez QWAM	
CAS APPLICATIFS	15
CRM-Ventes	16
Études Marketing	19
Ressources humaines	22
Analyse de risques et d'incidents	24
Compliance	29
IA, Défense et Sécurité	32
Moteur de recherche, Information & Media	34
Analyse de l'Information web	37
L'IA au service de la démocratie	39
CONCLUSION , Fanny Maffre	41
CONTACTS / CONTRIBUTIONS	44

INTRODUCTION



Christian Langevin
Directeur Général QWAM

Depuis quelques années, l'Intelligence Artificielle (IA) s'est imposée comme un des thèmes majeurs de la transformation numérique des entreprises et des organisations. Ce terme d'IA recouvre de très nombreuses technologies : reconnaissance d'images, reconnaissance de la parole, traduction automatique, modélisation de systèmes complexes... et bien sur le traitement automatique du langage naturel (TALN, TLN ou TAL ou en anglais NLP - Natural Language Processing), sujet de ce livre blanc.

Ce domaine devient de plus en plus important pour les entreprises et les organisations sachant que la majorité des données produites ou circulant dans les entreprises ou sur le web sont textuelles. Le NLP associe de nombreuses technologies, par ailleurs en plein essor, et plusieurs disciplines : linguistique, sémantique, big data, analyse de données, apprentissage, machine learning, deep learning ... Les avancées récentes, notamment dans le deep learning et les réseaux de neurones (Word2Vec, Bert et ses variantes, GPT-3, etc.), font que le domaine évolue très rapidement. Les performances s'améliorent en permanence, de même que le niveau de per-

tinence des résultats obtenus, entraînant une spectaculaire multiplication des cas d'usage dans de multiples secteurs d'activités dont certains jusqu'ici pas ou peu concernés par le traitement automatique du langage naturel.

L'IA génère des fantasmes et soulève des avis extrêmes, des espoirs exagérés ou à l'inverse des craintes infondées. Le traitement automatique de la langue n'y échappe pas. Nous avons donc choisi de présenter un panorama de cas d'usage du NLP parmi les plus représentatifs à ce jour (accès à l'information, moteur de recherche, ventes et marketing, ressources humaines, analyse de risques, compliance). Certains d'entre eux sont illustrés par des exemples de réalisations en entreprise menées par QWAM.

Au regard de ces cas concrets, nous avons recueilli l'avis d'experts pour éclairer certains usages et thématiques dans lesquels l'IA nous paraît particulièrement bien adaptée. Nous prévoyons d'aborder dans un second tome d'autres cas d'usage (sécurité intérieure, fraudes, analyse de données scientifiques et santé, parcours clients, etc.).

IA & VALORISATION DES DONNÉES TEXTUELLES

Panoramas des cas d'usage

MÉDIAS ET CONTENUS NUMÉRIQUES

Génération automatique de méta-données, Moteur de recherche, Recommandation et SEO



GESTION ET VALORISATION DES CONTENUS

Gestion et enrichissement des données textuelles, Ingénierie et reprises d'archives, Exploitation de silos documentaire, Dématérialisation - BPM



ANALYSE DES AVIS CLIENTS

Sites internet
Forums
Mails
Service clients
Réseaux sociaux



ANALYSE ET CLASSIFICATION DES DONNÉES

Finance (rapports internes, contreparties, ...)
Santé (comptes-rendus d'interventions, pharmacovigilance, ...)
Assurance (sinistres, contrats, ...)



HYBRIDATION DES DONNÉES

Analyse de textes couplés aux données structurées pour décupler les capacités d'exploration



CRIBLAGE ET ROUTAGE DE L'INFORMATION

Distribution des mails, messages et demandes aux interlocuteurs concernés, Intégration d'informations ciblées dans les CRM, ...



ENQUÊTES ET DÉBATS

Réalisation d'études d'opinions, Analyse d'enquêtes, Navigation dans les verbatims, Visualisation des suggestions clients, Enquêtes RH,...



DÉTECTION D'INCIDENTS, MAINTENANCE PRÉDICTIVE, ...



ANALYSE DE RISQUES, KYC, E-RÉPUTATION, ENQUÊTES FINANCIÈRES, DÉTECTION DE FRAUDES



BUSINESS, MARKET & SALES INTELLIGENCE

Observatoire innovation, start-ups, ... Compréhension des marchés et des concurrents, Analyse de l'information web en masse



MOTEUR DE RECHERCHE, ANALYSE DE DOCUMENTATION TECHNIQUE, KM, SCIENTIFIC INTELLIGENCE



COMPLIANCE ET RÉGLEMENTATION



QWAM

CONTENT INTELLIGENCE

Retrouvez nos cas d'usage sur blog.qwamci.com

ÉTAT DE L'ART



Vincent Guigue
Professeur associé
Maître de conférences
UPMC LIP 6

IA, NLP ET APPRENTISSAGE STATISTIQUE : 70 ANS DE R&D

Le Traitement automatique du langage ou NLP en anglais (Natural Language Processing) constitue l'un des domaines applicatifs les plus vastes de l'IA couvrant un large spectre, traduction, analyse thématique, extraction de connaissances, décomposition grammaticale, génération de résumé... L'apprentissage automatique ou statistique, (en anglais, machine learning), se fonde sur une double approche mathématique et statistique. Cette approche, qui s'applique à de nombreux domaines et secteurs, permet à l'ordinateur d'apprendre à partir de données et de résoudre des tâches complexes.

FONDATION (1955-1985)

Cinq grandes périodes permettent de structurer les évolutions en apprentissage statistique depuis l'apparition de cette discipline au cours de la seconde moitié du XXe siècle. L'intelligence artificielle, comme l'informatique, est généralement considérée comme l'héritage des travaux du mathématicien et cryptologue britannique Alan Turing. L'expression « Intelligence artificielle »

apparaît pour la première fois lors de la conférence de Dartmouth en 1956. Du côté de l'apprentissage statistique, qui ne constitue alors qu'une petite sous-communauté de l'IA, les contributions sont principalement théoriques avant 1985. Parmi les propositions célèbres, on peut mentionner l'algorithme du perceptron de Rosenblatt, le travail de Tikhonov sur les formulations régularisées et les premiers modèles non paramétriques d'estimation de densité. Ces travaux ont posé les bases des algorithmes qui sont encore utilisés en traitement de données.

POC / FAISABILITÉ (1985-1995)

La décennie suivante correspond à l'émergence de multiples PoC, visant à vérifier la faisabilité des concepts développés en matière d'apprentissage statistique d'un point de vue marketing, économique, technique... C'est également l'occasion pour la communauté scientifique de développer et de tester les premières applications à visée industrielle. Ainsi, l'algorithme de la rétro-propagation du gradient et les architectures de réseaux de neurones à convolution ont permis le développement et l'implantation, notamment

dans les secteurs d'activités de la Poste et de la banque, de systèmes de reconnaissance de chiffres manuscrits. Du côté de la transcription et de la traduction automatique, les chaînes de Markov cachées permettent de franchir un palier de performances et ouvrent de nouvelles perspectives. Au début des années 1990, de nouvelles formulations convexes simplifient grandement l'accès à l'apprentissage. Cette simplification permet par ailleurs la multiplication des expériences et l'augmentation des performances.

INDUSTRIALISATION (1995-2005)

L'intérêt de l'apprentissage automatique ayant été démontré, la décennie 1995-2005 correspond à un essor et une structuration tant de la communauté que des outils. De nombreuses toolboxes voient le jour, permettant un changement d'échelle des expériences. Les propositions de modèles continuent

à façonner la communauté scientifique (boosting, forêts aléatoires...). Les premières propositions correspondant à des prédictions structurées datent également de cette période. Les applications se multiplient rapidement dans de multiples domaines : maintenance prédictive, filtres spam, traitement du signal, classification de documents sur le web, médecine, bioinformatique...

Parmi les nouvelles problématiques applicatives, il faut mentionner la recommandation et l'introduction de la NMF (non Negative Matrix Factorization) qui permet de simplifier l'apprentissage de représentations des utilisateurs et la personnalisation dans de nombreux algorithmes.

C'est au cours de cette même décennie que débute la bataille pour la recherche d'information sur le web entre Google et Microsoft. Tout l'enjeu de ce conflit industriel à l'échelle de deux leaders mondiaux du numérique porte sur la



qualité des résultats obtenus avec leurs outils de recherche respectifs en matière de cohérence, fiabilité, pertinence de remontée de l'information... Et donc, par voie de conséquence, la capacité à comprendre les messages ou, a minima, la capacité à faire le lien entre une requête et un document pertinent.

RÉSEAU SOCIAUX & BIG-DATA : EXPLOSION DE L'APPRENTISSAGE (2005-2015)

Les années qui suivent voient l'explosion des réseaux sociaux et l'entrée dans l'ère du big data. Du point de vue algorithmique, la spécificité des données sociales se traduit notamment par une remise en cause de l'hypothèse classique i.i.d. (données indépendantes et identiquement distribuées) et de nouvelles problématiques autour des graphes comme la modélisation de la diffusion d'information ou la découverte des influenceurs dans un réseau. Ces travaux posent les bases de l'essor futur des algorithmes d'apprentissage par renforcement.

D'un point de vue applicatif, la généralisation du web participatif (ou « web-2.0 ») a permis de récolter des avis en temps réel à travers des blogs, des revues, des tweets, des pages Facebook. Il a également ouvert des possibilités innovantes pour effectuer des sondages, gérer la e-réputation des entreprises, prévenir les buzz négatifs... L'analyse des données textuelles progresse alors radicalement, via successivement l'introduction de nouveaux espaces vectoriels sémantiques ou encore la mo-

délisation des phrases dans les textes. C'est au cours de cette même décennie 2005-2015 que les techniques de profiling se généralisent ainsi que la recherche d'informations contextuelles voire personnelles. Le secteur du e-commerce et les médias généralisent également l'utilisation des systèmes de recommandation. Les nouvelles architectures profondes de réseaux de neurones révolutionnent le domaine de la vision et vont jusqu'à remettre en cause le modèle de l'industrie automobile en basculant vers des véhicules (de plus en plus) autonomes.

Sur le plan industriel, cette décennie correspond à une prise de conscience de la valeur des données manipulées au quotidien (big data). Le stockage et la valorisation sont rendus possibles grâce à de nouvelles plateformes de base de données et de traitements distribués comme Hadoop, puis Spark. L'explosion du domaine est ressentie en termes de ressources humaines, les GAFAM (Google, Apple, Facebook, Amazon, Microsoft) investissent et recrutent massivement.

La conjonction entre big data, natural language processing et Machine Learning, nécessite alors le développement d'outils spécifiques capables de traitement à très large échelle. Jusqu'alors, les outils industriels de NLP ne faisaient que rarement appel à des techniques de machine learning et se basaient sur des règles donnant des résultats certes précis mais difficiles à évaluer en termes de couverture.



Y a-t-il un avion sur la photo ? Ce signal présente-t-il une anomalie ? Cette revue est-elle positive ?

RE-FONDATION ET CONVERGENCES (2015-2020)

Le domaine de l'apprentissage automatique est actuellement en cours de refondation autour de l'apprentissage de représentation avec une attention particulière autour des nouvelles architectures de réseaux de neurones. Globalement, l'enjeu porte sur la représentation de l'ensemble des éléments du problème (signaux, item, individu, ...) sous forme de vecteurs continus (souvent abstraits) et l'exploitation de ces profils dans différentes tâches. De manière générale, multiplier les problématiques autour de concepts invariants revient à placer tous les problèmes rencontrés dans un cadre de raisonnement général. Plusieurs problématiques signifient aussi plusieurs manières d'exploiter un concept. Par extension, la représentation du concept va s'enrichir et devenir progressivement plus universelle. Quasiment tous les problèmes d'apprentissage automatique reviennent à chercher la réponse à une question particulière : Y a-t-il un avion sur la photo ? Cette revue est-elle positive ? Ce signal présente-t-il une anomalie ?... Le fait de posséder une chaîne de prétraite-

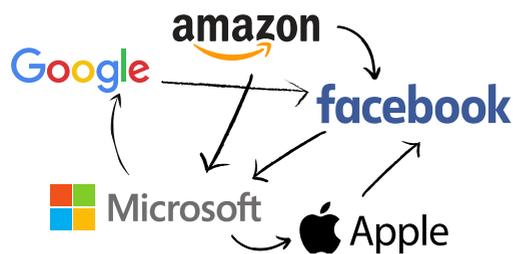
ments aboutissant à une représentation sémantique des concepts - photo, revue, signal... - permet de répondre à de nouvelles questions en ayant peu - voire pas - de données étiquetées spécifiques (few/0-shot learning).

Plusieurs exemples applicatifs illustrent aussi cette volonté de convergence. Sur le plan architectural, les systèmes de traduction ont changé de paradigme récemment. Les représentations de phrases encodées à l'aide de réseaux récurrents ou transformés dans le langage d'origine, peuvent ensuite être décodées vers n'importe quelle langue. Ces systèmes, comme d'autres, s'appuient sur des représentations de mots riches apprises sur des corpus gigantesques à l'aide d'algorithmes plus efficaces. Dans le domaine de l'image et de la vision, il devient maintenant courant d'utiliser des systèmes appris sur des tâches spécifiques par certains, pour traiter des tâches de plus haut niveau par d'autres ou sur des petits jeux de données qu'il n'aurait pas été possible de traiter efficacement sans transfert de connaissances. Cette exploitation des représentations, non seulement dans plusieurs tâches, mais aussi dans plusieurs applications impliquant diffé-

ÉTAT DE L'ART

rents auteurs, est un élément nouveau et structurant pour la communauté de l'apprentissage automatique.

La convergence concerne les implémentations et le matériel utilisé. Quelques bibliothèques de référence open-source, à l'image de SciKit-Learn, TensorFlow, ou Torch (et sa récente version Python), servent de base à la plupart des implémentations, l'exécution reposant sur des GPU toujours plus puissants. Cette concentration de plateformes, supportée par les GAFAM, aboutit à des cadres logiciels de très grande qualité qui favorisent la



productivité, la reproductibilité des expériences, les échanges et, plus généralement, l'accès à des architectures très complexes. Enfin, la convergence s'illustre au niveau des plateformes de données et des groupes industriels. La communauté informatique partage de longue date des jeux de données permettant une comparaison équitable des modèles, cependant un nouveau

palier a été franchi récemment avec la multiplication des compétitions ouvertes de grande envergure, l'initiative CLEF ou le site Kaggle.

Autour de 2015, les outils de type machine learning atteignent un niveau de maturité et de performances en NLP tel qu'ils deviennent incontournables en matière de sémantique puis de traduction et enfin d'extraction d'informations. La communauté industrielle s'adapte alors à cette nouvelle génération d'outils malgré leur côté boîte noire et le manque de garanties sur les sorties. Coté algorithmique, le changement est brutal : l'ensemble des technologies évoluent très vite, de nouveaux paradigmes apparaissent, d'autres à l'inverse s'amenuisent voire disparaissent. En parallèle, l'industrialisation rapide de ces nouvelles stratégies par les GAFAM et les start-up force les acteurs industriels à des arbitrages fréquents. De plus, dans de nombreuses applications, certains aspects du machine learning sont toujours inacceptables pour le client final. Il convient alors de trouver une nouvelle interaction entre différentes familles d'outils.

Sur le plan industriel, l'adaptabilité, la réactivité et l'agilité des organisations deviennent impératives. Les conséquences pour les principaux acteurs sont profondes. Les rachats se multiplient (DeepMind par Google, KXEN par SAP) pour aboutir à la création de quelques laboratoires mondiaux privés dans le domaine de l'apprentissage automatique (DeepMind, Microsoft Research, OpenAI, Google Brain, FAIR).

QWAM ET L'IA



Ivan Monnier
Directeur Technique

Sundar Pichai, CEO de Google, l'affirmait en 2016 : « nous allons vers un monde où l'Intelligence artificielle va prédominer. » Cette (R)évolution de l'IA est portée par le Web et la quantité astronomique de données qui y sont stockées et y circulent. Si les algorithmes sur lesquels reposent les applications actuelles liées au traitement du langage étaient pour la plupart connus il y a vingt ans (voire plus), le big data permet d'obtenir des résultats d'ores et déjà très impressionnants. Et ce n'est que le début... L'intelligence artificielle associée au big data constitue une nouvelle étape dont les applications, notamment pour ce qui concerne le positionnement et l'offre de QWAM, sont particulièrement prometteuses.

Dés sa création en 2007, QWAM, en tant que nouvel entrant sur le marché du traitement automatique du langage, s'est appuyée sur la R&D pour développer son savoir-faire technologique et son offre commerciale. Depuis 2017, QWAM participe à des programmes de recherche ambitieux en intelligence artificielle avec des équipes de renommée mondiale dont celle du laboratoire machine learning & Information Access - LIP6 de l'Université Pierre et Marie Cu-

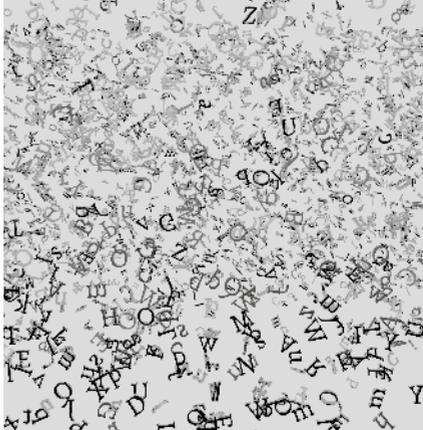
rie (Paris 6). Pour rappel, notre métier consiste à développer des applications permettant de valoriser des données textuelles. Or, en l'absence de structuration, ces données sont difficilement exploitables notamment dans une perspective d'analyse.

Les pionniers du Traitement automatique du langage (TAL ou NPL en anglais) ont développé l'élaboration de référentiels, souvent limités à un métier, en utilisant des techniques de modélisation décrivant un univers sémantique donné en relation le plus souvent avec un secteur d'activités ou une thématique. Les linguistes spécialisés ont alors construit à partir de règles, de dictionnaires, grammaires et réseaux sémantiques, des ontologies, permettant d'obtenir des résultats plus ou moins probants. Un tel procédé nécessite toutefois de faire appel à une armée de linguistes, ce qui engendre des coûts de mise en œuvre élevés et parfois rédhibitoires du point de vue du client final.

Ainsi, la principale difficulté pour la plupart des acteurs du marché, a longtemps porté sur l'adaptabilité des dictionnaires d'un domaine à un autre afin

d'être en capacité de répondre aussi finement et rapidement que possible au contexte du client. Cette exigence a poussé QWAM à investiguer le champ très particulier de la génération automatique d'ontologie en s'appuyant sur des outils de TALN (Traitement Automatique du Langage Naturel) renforcés par l'apport de l'IA. Cette approche innovante et spécifique permet de générer automatiquement une ontologie, et cela pour un coût nettement moindre que les procédés classiques. Nous avons inversé l'approche : plutôt que de créer à l'avance et mot à mot une ontologie en fonction de l'univers du client, ce dernier fournit son propre corpus à partir duquel des listes candidates revues par l'IA sont extraites. Le linguiste, toujours indispensable, intervient en fin de process pour reprendre la liste de quelques termes revus par l'IA et enlever les scories avant d'intégrer le tout dans une ontologie adaptée au client final.

La prochaine étape se dessine d'ores et déjà pour QWAM : il s'agit d'intégrer et d'appliquer dans nos solutions les nouveaux langages bi-directionnels BERT (Bidirectional Encoder Representations from Transformers) développés et pré-entraînés (en anglais pour l'essentiel) par Google. Basé sur des réseaux de neurones, BERT permet de comprendre des requêtes (groupe de mots, phrases...) en tenant compte du contexte. Le traitement du langage BERT tient compte de l'ensemble des éléments du langage naturel (prépositions, expressions, nuances polysémiques, pronoms...). L'analyse des mots



s'effectue en relation avec l'ensemble des éléments de la phrase, plutôt que sur l'analyse de chaque mot, l'un à la suite de l'autre. Les modèles tiennent donc compte du contexte complet ce qui est particulièrement utile et efficace pour comprendre l'intention derrière les requêtes de recherche et pour traiter les expressions naturelles et conversationnelles. BERT constitue une véritable rupture que nous devons être en mesure de maîtriser notamment dans nos corpus. Reste au préalable à adapter BERT à la langue française, passer de BERT à CamemBERT, version française développée par l'INRIA. La communauté des experts du langage avec laquelle nous collaborons régulièrement s'y emploie.

Enfin, en complément de BERT et toujours porté par la montée en puissance de l'IA, nous nous intéressons également au Natural Language Understanding (NLU) qui s'attache à analyser dans un texte des intentions, des avis, des opinions, des sondages ou encore des sentiments et à détecter le sens de ce qui est dit ou écrit. Les applications sont multiples, notamment pour ce qui est de l'analyse automatisée et ultra-rapide de tendances, le retour clientèle, que ce soit dans le cadre de relations de suivi marketing, de l'information (réseaux sociaux notamment) ou dans le monde des RH.

CAS APPLICATIFS

CRM-Ventes

Études Marketing

Ressources Humaines

Analyse de risques et d'incidents

Compliance

IA, Défense et Sécurité

Moteur de recherche,
Information & Media

Analyse de l'information web

L'IA au service de la démocratie



Loïc Deo Van

Fondateur et CEO des start-up ByPath en 2013 puis d'EverReady.ai en 2019, Loïc Deo Van est un expert de la vente BtoB et des solutions CRM innovantes : les Sales Tech.

QWAM : de quand datent les premiers outils CRM ? Que représente ce marché ?

Loïc Deo Van : Les outils CRM sont apparus à la fin des années 90 sous la forme, à l'époque, d'un progiciel métier vertical parmi d'autres (finance, comptabilité, RH, supply chain...) et intégré comme sous module des ERP. En l'espace de vingt ans, le CRM s'est imposé comme un système d'information majeur remplaçant même l'ERP notamment dans les sociétés de services. Selon les dernières études, le marché du CRM devrait dépasser celui de l'ERP d'ici 2025 et peser près de 80 milliards.

Les premiers CRM étaient assez simples, alimentés essentiellement par les équipes de vente, leur but était

avant tout de constituer une base de données clients/prospects, d'historiser les transactions et de les partager avec l'ensemble des équipes. Ils disposaient également de quelques fonctionnalités permettant d'améliorer la productivité et de générer des rapports d'activité. L'explosion du web, des réseaux sociaux, du e-commerce ou encore du marketing digital, a généré des flux d'informations et un nombre d'interactions gigantesques. Le CRM n'a pas pu exploiter toutes les possibilités offertes par ce big data et a permis aux Sales & Marketing Tech de se développer sur des niches à valeur ajoutée : Marketing automation, Lead generation, predictive, retargeting... Ces Sales & Marketing Tech sont venues se greffer sur le CRM sous forme de briques com-

plémentaires ou Add on. Certaines ont d'ailleurs été rachetées et intégrées par les CRM.

QWAM : comment s'est insérée l'IA dans ces outils ?

Il faut faire attention : l'IA est à la mode et fait vendre. Du coup, le terme est souvent galvaudé pour simplifier et vulgariser l'ensemble des technologies (algorithmes, automatisation, programmes...) nécessaires pour traiter le Big Data et y extraire de la valeur métier. Mais au sens strict du terme il n'y a que très peu d'applications CRM qui utilisent vraiment la machine learning ou les réseaux neuronaux qui constituent véritablement l'IA.

L'IA demandait jusqu'il y a très peu de temps de lourds investissements R&D, IT, ressources humaines... pour des résultats souvent peu probants. Depuis peu, on assiste à une démocratisation des plateformes clé en main de machine learning, de traitement des data qui rendent maintenant beaucoup plus accessible l'expérimentation et l'utilisation de l'IA sur des problématiques liées à l'activité commerciale.

QWAM : quels types d'informations sont utiles pour le commercial ?

Ces dernières années, Google, LinkedIn, Twitter, ont fait que chacun d'entre nous s'est mis à générer de l'information parfois même sans en être conscient. Bien sûr, l'écrasante majorité de ces informations n'ont pas de valeur d'un point de vue commercial mais certaines d'entre elles peuvent être cruciales dans un cycle de vente : un article de presse qui mentionne le

nom de votre interlocuteur, un tweet qui montre que votre prospect a assisté à une présentation de votre concurrent, un like sur LinkedIn sur un sujet qui vous concerne, ...

C'est en partant de ce constat que j'ai créé en 2013 la start-up ByPath avec pour but l'exploitation des signaux faibles. Il s'agissait de trouver des aiguilles dans une énorme botte de foin : internet ! ByPath est devenue rapidement une solution d'intelligence commerciale incontournable sur le marché, véritable « GPS du commercial », capable d'indiquer quel prospect contacter, à quel moment et avec quel discours !

Comme ByPath beaucoup d'autres SmarTech (Sales & Marketing Tech) ont su exploiter les nouvelles briques technologiques : algorithmes, indexation, NLP (traitement automatique du langage), etc. afin de répondre de façon plus efficace aux problématiques commerciales. La plupart de ces techs sont désormais bien maîtrisées. Ce qui les différencie porte sur leur capacité respective d'assemblage et de « fine tuning ». De ce point de vue, ma rencontre avec QWAM, experte sur le NLP, a été décisive dans le succès de ByPath. En effet, de prime abord, collecter, trier et extraire d'un texte les données avec pertinence (entités nommées, thématiques, organigrammes, CV, annonces...) paraît simple. En réalité c'est vraiment complexe, surtout quand on traite au quotidien des millions d'articles provenant de centaines de milliers de sources. Par exemple : que signifie orange ? La couleur, le fruit,



Retrouvez le cas client



ÉTUDES MARKETING



François Laurent

Fondateur du cabinet de conseil marketing « Consumer Insight », Past Président de l'« ADETEM », premier regroupement associatif français de marketeurs.

la société, la ville... Nous nous sommes appuyés sur des spécialistes comme QWAM dans chaque domaine afin de gagner du temps tout en garantissant un bon niveau de qualité.

QWAM : quels sont selon vous les principaux atouts de l'IA en matière d'aide à la vente ?

Au-delà des sujets précédemment évoqués : prospection, marketing, je pense que l'IA va permettre d'assister le commercial dans son quotidien. Elle sera demain la nouvelle interface entre le commercial et ses outils CRM, mail, téléphone, agenda... Elle fera la synthèse de l'ensemble des data disponibles dans les différents systèmes d'informations de l'entreprise afin de créer une véritable intelligence collective. Cette intelligence permettra d'améliorer les performances de chaque commercial notamment grâce au partage des meilleures pratiques.

C'est sur la base de cette conviction que j'ai créé en 2019 la société EverReady.ai. Notre objectif est de mesurer, comprendre et d'améliorer la performance des équipes commerciales à partir de

l'analyse de leurs interactions clients et prospects. En effet, les entreprises génèrent aujourd'hui des tonnes d'informations dans leurs différents outils CRM, messagerie, téléphones, agenda... qui restent souvent inexploitées et non corrélées les unes avec les autres. Pour faire simple, on reproche à son CRM d'être mal rempli alors même que l'information se trouve dans les mails ou le journal d'appels des commerciaux. Ce type d'interface intelligente existe déjà dans le domaine grand public, à l'image de Siri (Apple) ou d'Alexa (Amazon), de Waze. Notre ambition est mettre cette intelligence collective au service des équipes commerciales pour leur donner du feed-back sur leurs activités, les assister dans leur stratégie, améliorer leurs prévisions de vente, leurs recommander les meilleures actions à réaliser, capitaliser sur l'expérience de chaque commercial... L'approche que nous proposons va fluidifier le pilotage du business, répondre aux nouvelles formes d'organisation et de management (télétravail notamment). L'instan-tanéité collective engendre une performance décisionnelle accrue et inédite. Seule l'IA le permet.

François Laurent possède une vaste expérience dans le domaine du marketing et de la communication. Cet expert qui a notamment travaillé chez Eurocom Havas, Idémédia Publicis, Millward Brown ou Thomson Multimédia... s'implique désormais à travers ses activités de conseil dans la mise en place de méthodes de marketing plus performantes en adéquation avec les mutations de la nouvelle société numérique (empowered consumer, Web 2.0, médias sociaux, économie circulaire, consommation collaborative...).

QWAM : comment l'IA est-elle impliquée dans le domaine du marketing ?

François Laurent : je précise tout d'abord que je ne suis pas, ou plutôt plus, en situation opérationnelle en entreprise. Ma position est celle d'un expert indépendant spécialisé dans le marketing en général. Pour répondre à votre question, je distingue l'intelligence artificielle dans son état actuel de ce qu'elle devrait être demain selon moi. Aujourd'hui, l'IA est un terme à la mode dont tout le monde parle, et pas seulement dans le monde du marketing. On « colle » de l'IA partout, par exemple pour dialoguer avec des consommateurs sur un site. En réalité, ce n'est

qu'un simple chatbot (agent conversationnel), plus ou moins efficace, basé sur un algorithme plus ou moins bien conçu et qui répond à peu près n'importe quoi... Mais le premier algorithme venu n'est pas forcément intelligent.

QWAM : l'IA n'en est-elle encore qu'aux balbutiements ? Qu'en est-il de l'état de l'art ?

Il est extrêmement limité, mais cela n'empêche pas ces spécialistes de fonder beaucoup d'espoir dans l'IA. La plupart des gens qui disent développer ou utiliser de l'IA à des fins de marketing se contentent de bricoler. Rien ne fonctionne vraiment et les applications dans le domaine marketing et des ressources humaines sont encore balbutiantes, à quelques rares exceptions près. Pour autant, elles sont nettement plus développées dans les domaines scientifiques.

QWAM : quel est selon vous son apport, potentiel ou avéré ?

Il faut tout d'abord se poser la question de savoir ce que nous pouvons en espérer. L'IA associée à un CRM va-t-elle permettre par exemple de regrouper les clients au sein d'une même base de données référentielles, tout banaliser



et faire du marketing de façon plus ou moins automatisée ? Dans cet esprit, la plupart des grandes entreprises ont développé leur propre « Data Lake » et embauché des data-scientists. Je ne constate toutefois pas d'avancée majeure en termes de performance ou d'efficacité. Et à la limite, si ça marchait, je pense que ce serait une catastrophe. Cela voudrait dire en effet que la différenciation, l'un des fondements du marketing, disparaîtrait, tout le monde utilisant des données comparables et effectuant les mêmes traitements, tous barbotant aux mêmes endroits. Entre le non-existant d'aujourd'hui et l'idéal de demain, il faut trouver une juste mesure, des outils qui permettent de faire des choses intelligentes tout en laissant à l'expert la possibilité de prendre les

décisions qui vont bien aussi bien du point de vue de l'action que de l'analyse.

Dans le cas du grand débat, l'intelligence artificielle développée par QWAM en complément de ces outils habituels de data et text-mining ne se substitue pas à l'analyse de l'expert. L'IA est ici un outil qui permet de prendre des décisions mais ce n'est pas à l'outil de prendre des décisions. Ou alors, cela équivaudrait à passer du marketing à la science-fiction. On se rapprocherait des romans d'anticipation d'Isaac Asimov.

QWAM : comment voyez-vous l'avenir de l'IA en tant qu'outil marketing ?

Je ne suis ni opérateur ni ingénieur. Mais je suis les gens qui s'y intéressent

de près, à l'image par exemple des assureurs. C'est vrai que dans ce domaine de l'assurance, l'IA peut répondre à de véritables challenges. Aujourd'hui l'offre des assureurs est bâtie sur des probabilités du passé. Personne n'a été véritablement capable de travailler sur une nouvelle approche de l'assurance. Des systèmes experts à base d'IA permettraient d'inventer de nouveaux produits d'assurance. Mais cette démarche se heurte à la remise en cause de la notion même d'assurance, qui est basée sur la notion de mutualisation des risques. Mais si l'IA permet de concevoir et de mettre sur le marché des offres de moins en moins chères et de plus en plus individualisées, cela veut dire que, pour compenser, il faut qu'il y ait des produits plus chers visant des populations à risque. L'IA permet donc certes de créer de nouveaux produits innovants mais risque aussi de mettre à terre tout un secteur d'activité.

QWAM : la crise liée au Covid 19 que nous traversons depuis mars dernier peut-elle avoir une incidence sur l'approche du marketing 2.0 à base d'IA ?

La crise du Covid19 a changé beaucoup de choses et cela en un temps record. On s'aperçoit que les entreprises ne cherchent plus uniquement à maximiser l'efficacité et le profit comme l'IA pousse à le faire. Mais l'IA va-t-elle pouvoir intégrer la dimension du « développement soutenable » ou la notion d'éthique ? Et d'ailleurs, peut-on modéliser l'éthique ? Nous sommes confrontés à des enjeux de société

gigantesques. Le risque de l'IA serait de conforter un retour vers le « business as usual », sans que soient tirées les leçons de la situation que nous venons de traverser : par exemple la nécessité de relocaliser, de refaire confiance aux salariés, etc. Les experts de l'ADETEM et son comité scientifique réfléchissent à ce que doit être la nouvelle réalité du marketing et ses responsabilités. En quoi l'IA est-elle compatible avec ces notions ? La société et le monde de l'entreprise ne reposent pas sur des machines. Il faut se donner des garde-fous. À défaut, sous prétexte que la machine puisse prendre des décisions à notre place, on risque de passer à côté de choix importants. Il ne faut pas que ces nouveaux outils deviennent une facilité pour les décideurs et les patrons. Voici le principal enjeu de l'IA !

RESSOURCES HUMAINES

ANALYSER, PRÉVOIR ET MIEUX GÉRER LES RISQUES EN ENTREPRISE AVEC QWAM TEXT ANALYTICS

Considérées comme l'un des actifs les plus importants d'une entreprise, les ressources humaines participent directement à son développement au même titre que ses capacités d'innovation et de production ou encore sa valorisation financière. Dans ces conditions, la gestion RH constitue un enjeu social et économique majeur. Sous-évalués ou mal gérés, les risques psycho-sociaux ont un impact sur le fonctionnement et l'image même de l'entreprise. Ils se traduisent par une hausse du turnover, une mauvaise ambiance parmi les salariés, une augmentation des arrêts maladie et de l'absentéisme, ou encore des conflits internes qui tendent les relations sociales.

Face à ces risques aux conséquences parfois dramatiques, les responsables des ressources humaines cherchent donc à anticiper les conflits avant qu'ils ne se déclenchent, en « captant » notamment quelques signes avant-coureurs souvent ténus. L'objectif est d'améliorer la politique RH en améliorant la capacité d'écoute des personnels. En analysant et en synthétisant les avis et les contributions des collaborateurs d'une entreprise, QWAM Text Analytics participe à l'amélioration de la politique RH. Basée sur l'analyse sémantique, cette solution

se fonde sur l'analyse et le croisement de sources contributives diversifiées : entretiens annuels, enquêtes de satisfaction, rapports de groupes de travail thématiques, informations recueillies sur le web, les réseaux sociaux...

Grâce à son moteur sémantique, cet outil permet également de gérer et de cartographier les compétences des collaborateurs ou encore de cribler les bases de CV et les profils des candidats. Il intègre enfin un module de recueil et d'analyse du sentiment d'un groupe de personnes donné selon le contexte métier.

La solution QWAM Text Analytics a notamment été déployée auprès de la Direction du Personnel Militaire de la Marine Nationale (DPMM) avec pour objectif de sonder au plus près le moral des 39 000 marins, leurs préoccupations et leurs suggestions, et d'être donc en capacité d'améliorer rapidement et de façon concrète les conditions de travail. La solution déployée par QWAM permet d'obtenir une évaluation fiable de la tonalité des expressions recueillies auprès des personnels concernés en alliant la détection des occurrences pré-définies à l'apprentissage automatique (machine learning).



À cet effet, les commentaires issus de diverses enquêtes internes, (indicateur de mesure du moral, sondages...) sont analysés et leurs ressentis objectivés avant de bâtir une cartographie des attentes. Cette cartographie permet de quantifier au plus près la force des préoccupations qui remontent afin que les instances RH soient en mesure d'améliorer les conditions d'exercice de leur métier. La Marine Nationale dispose pour cela de données statistiques mais

également textuelles, d'autant plus utiles qu'elles permettent à chacun de décrire précisément attentes et ressentis. Ces commentaires écrits, qui représentent chaque jour plusieurs dizaines de milliers de lignes, sont impossibles à analyser en dehors d'outils faisant appel à l'Intelligence Artificielle appliquée à la sémantique, à l'image de la solution QWAM Text Analytics sélectionnée après un appel d'offres.

Retrouvez le cas client

ANALYSE DU RISQUE

PROJET EDF SIGMA IA & TLN : QUAND LE PASSÉ PERMET D'ANALYSER ET D'ANTICIPER LES RISQUES.

Inondations, raz de marée, secousses sismiques, l'histoire est jalonnée de catastrophes naturelles aux conséquences humaines, économiques et environnementales parfois dramatiques et qui ont marqué leur temps. Le souvenir de ces événements s'estompe cependant rapidement jusqu'à s'effacer de la mémoire collective au fil des siècles. Des archives écrites subsistent néanmoins, dont le décodage peut se révéler riche d'enseignements pour les experts des phénomènes naturels, sismologues, météorologistes, climatologues... Ces traces écrites forment un témoignage précieux qui permet de dater un événement, d'en préciser les circonstances, l'ampleur et les conséquences sur l'environnement de leur époque.

Sismologue au Département Technique d'Essais en Géologie, Géotechnique et de Génie civil (EDF), Emmanuelle Nayman revient sur les objectifs et les résultats du projet d'amélioration de la connaissance des séismes historiques en France métropolitaine dans le cadre du programme de recherches SIGMA v2 (Seismic Ground Motion Assessment). En synthèse, ce projet vise à collecter de nouvelles informations sur la sismicité historique sur internet en utilisant des techniques de

data mining et de text mining, à l'aide de la solution QWAM Text Analytics.

QWAM : quel est l'objet du projet Sigma ?

Emmanuelle Nayman : SIGMA 2 (pour Seismic Ground Motion Assessment,) est un grand programme international de recherche et développement lancé en 2017 dont l'objectif est de faire évoluer les connaissances en matière de caractérisation des mouvements sismiques en cas de tremblement de terre. Le consortium regroupe des experts dans le domaine de l'aléa sismique tant du monde industriel (EDF, CEZ, PGE&E, AREVA) que du monde académique (CEREMA, CEA, Universités, etc.) Le projet SIGMA 2 est structuré en différents groupes de travail et s'organise en différentes thématiques, allant des caractéristiques du sous-sol jusqu'aux domaines des sciences de l'ingénieur. Le second axe, intitulé « Paramètres des tremblements de terre », étudie les caractéristiques des séismes (magnitude, profondeur, localisation). Les deux objectifs principaux sont d'améliorer la connaissance de la sismicité dite historique et de mieux contraindre les caractéristiques des séismes dits instrumentaux (enregistrés par des capteurs), dans



des régions de faible sismicité, comme c'est le cas en France.

QWAM : pouvez-vous nous décrire la méthodologie appliquée à ce projet ?

Nos recherches en matière de sismicité historique se concentrent sur la France métropolitaine, une région du monde caractérisée par une faible sismicité. La période de récurrence d'un séisme est très longue, de l'ordre de 2 000 à 3 000 ans. Il s'agit d'une sismicité intraplaque, autrement dit, il n'existe pas en France de failles tectoniques majeures comme celles au Japon ou en Californie (Faille de San Andreas) pouvant générer des tremblements de terre aussi importants.

La sismicité dite instrumentale, enregistrée par des capteurs permet d'obtenir directement la magnitude du tremblement de terre par exemple. Les capteurs n'ayant été déployés qu'à partir des années 1960, la période est beaucoup trop courte pour être représentative de la sismicité en France, caractérisée par des périodes de retour bien plus longues. Il est donc nécessaire de remonter à des périodes beaucoup plus anciennes : il s'agit de l'étude de la sismicité dite historique qui se base sur des études macrosismiques, c'est-à-dire sur le ressenti de la population et

des dégâts constatés lors des séismes. Ces informations, ces témoignages sont conservés dans le patrimoine littéraire, au travers des archives textuelles (témoignages retranscrits, journaux, patrimoine littéraire, registres paroissiaux, actes notariés...) et iconographiques (tableaux, gravures, plus récemment photos et films).

L'étude de la sismicité sous un angle historique, sans capteur donc, repose sur des informations souvent parcelaires, un simple « ressenti » issu d'un témoignage, description rapide et partielle rapportée par un chroniqueur des dégâts occasionnés dans son village suite à un tremblement de terre (l'écroulement du clocher, l'effondrement d'un bloc de maison, le nombre de victimes...). L'assemblage des témoignages permet d'obtenir un champ macrosismique en croisant la localisation et les effets ressentis.

En France, nous disposons de la base de données macrosismiques SisFrance. Cette institution tripartite (BRGM, EDF et IRSN) collecte systématiquement depuis 1975 toutes les informations écrites sur les séismes. À ce jour, la base en répertorie plus de 6 000 ressentis en France métropolitaine, de 463 à 2007, et décrits par plus de 10 000 références bibliographiques. Il s'avère

que 80% des séismes présents dans cette base sont décrits par moins de 10 documents. Bien que les tremblements les plus significatifs soient désormais répertoriés, il est néanmoins nécessaire d'en améliorer la connaissance en trouvant des archives dans de nouvelles sources patrimoniales littéraires non exploitées à ce jour.

QWAM : est-il encore possible de découvrir de nouveaux documents ?

Nous recherchons des éléments à ce jour non encore versés dans SisFrance. Les historiens enrichissant la base de données ont sillonné les archives départementales de France et d'autres institutions : le travail d'une vie ! Il faut trouver d'autres sources. De ce point de vue, Internet est une source incontournable et permet de récupérer pas mal de choses. Il nous a fallu cependant cibler les sources à crawler sous peine d'être submergés sous un flot d'informations sans réel intérêt. Notre choix s'est alors porté sur le site documentaire Gallica développé par la Bibliothèque Nationale de France (BNF) et dont nous savions qu'il contenait de nombreuses sources intéressantes en matière de sismologie. Ce portail met à disposition plus de quatre millions de documents dont une grande partie d'entre eux d'ores et déjà transformés en fichier texte, ou « océrésés ». Pour des raisons de temps, il était évidemment impossible de les parcourir un par un, d'où la nécessité et l'intérêt de faire appel à un outil automatisé de data et text-mining à l'image de ceux

développés par QWAM. Le pré-traitement sous forme de fichiers texte de ces sources permet d'obtenir des informations rapidement analysables, quantifiables et exploitables dans le cadre de nos recherches. Le fonds Gallica est loin d'être épuisé. Je suis persuadée qu'il nous reste beaucoup de choses à découvrir.

QWAM : quel est l'apport des outils de datamining ?

Compte tenu de la multitude de textes, un simple moteur de recherche n'aurait pas suffi pour filtrer de façon efficace les résultats. Les techniques de datamining se sont révélées indispensables, même si, au départ, je n'avais qu'une intuition quant à l'intérêt d'un outil sémantique de ce type associé à de l'IA. Le pari est réussi : la méthode est efficace en termes de résultats et la démarche fonctionne sur le plan technique. Dans un premier temps, nous avons travaillé avec les experts de QWAM sur l'indexation de l'intégralité de la base de données Gallica, sans pré-filtre. Afin d'affiner nos requêtes et d'améliorer les performances du text mining, nous avons développé un dictionnaire (ontologies) dédié au tremblement de terre et à ses effets, regroupé en thématiques. Pour l'établir, nous sommes partis de la base de SisFrance afin de constituer un corpus multithématique et une ontologie en rapport avec le contexte de la sismologie historique (comportement des personnes au cours et après les tremblements de terre, description des dommages sur



le bâti, expressions en rapport avec le ressenti sonore ou le descriptif visuel des séismes...). La superposition d'éléments contextuels combinés avec une ontologie dédiée permet de récupérer et de sélectionner de façon plus fine les documents. Par ailleurs, cette démarche réduit de façon significative la remontée d'informations parasites, le « bruit », issues de la base Gallica. Les informations recueillies sont alors confiées à des historiens et des experts chargés de mettre à jour la base de données SisFrance et les caractéristiques des séismes.

QWAM : cette approche est-elle selon vous applicable à d'autres types de phénomènes et catastrophes naturelles (pandémie, raz de marée, éruption volcanique...)?

J'en suis convaincue, à condition de déterminer des dictionnaires adaptés

à chaque type de phénomènes. Pour autant, le monde de la sismologie est en avance par rapport à d'autres phénomènes et autres catastrophes naturelles dont les experts ne disposent pas de bases de données historiques équivalentes à celle de SisFrance.

QWAM : sur quels critères avez-vous décidé de retenir la solution de QWAM ?

En tant que sismologue, je ne dispose d'aucune expertise ni compétence en matière de datamining. Nous nous sommes appuyés sur les compétences de la R&D de EDF qui dispose d'un département dédié au Traitement Automatisé de Langue (TAL). Nous avons fait un benchmark pour sélectionner un candidat avec qui travailler. Le POC réalisé avec la startup s'est avéré infructueux, ne pouvant répondre techniquement à nos attentes. J'ai alors

ANALYSE DU RISQUE

«crawlé» de façon systématique tous les sites d'entreprises spécialisées dans le datamining pour sélectionner au final QWAM dont j'ai apprécié la culture R&D. Une expérimentation sous la forme d'un POC a permis de confirmer cette première impression avant de commencer le projet proprement dit.

QWAM : quelles sont les prochaines étapes ?

Nous travaillons en étroite collaboration avec QWAM pour faire évoluer les

outils, les techniques, etc. La méthode est efficace, la démarche projet est stabilisée et pérenne suite au processus d'optimisation. À ce jour, 1000 récits inédits ont été récupérés. Nous allons poursuivre l'exploration systématique de la base Gallica, en incluant le latin, l'anglais et le vieil anglais, et investiguer d'autres bases de données. Nous sommes loin d'avoir épuisé les sources historiques et de nombreuses découvertes issues du passé attendent les

COMPLIANCE



Michel Meyer

Président de Data Integrity et du Cercle des DPO, des CDO, de la conformité et de la gouvernance

Michel Meyer est un spécialiste de la compliance (data integrity), activité qu'il développe au sein de son cabinet Michel Meyer Data Integrity. Expert depuis 45 ans dans le marketing direct, le data mining et la gestion des connaissances, il a notamment travaillé chez EY en tant que CIO ainsi que chez Arthur Andersen en charge du département Knowledge. Michel Meyer préside également le Cercle des DPO, des CDO, de la conformité et de la gouvernance qu'il a créé en 2019. Son parcours professionnel est marqué par son intérêt pour l'innovation technologique et la mise en œuvre d'outils naturellement intuitifs dans la gestion et la diffusion de l'information.

QWAM : quand avez-vous entendu parler d'IA pour la première fois ?

Michel Meyer: Il y a près de 35 ans même si on parlait alors d'analyse cognitive et de recherche opérationnelle. J'ai pratiqué dans les années 90, notamment pour Bouygues, les premières analyses de scoring destinées à envoyer des mailings marketing à des clients potentiellement intéressés par l'achat d'une maison. Nous partions d'annuaires accessibles via le Minitel et sous forme papier. Une fois océrisés, nous en extrayions les bases nominatives avant d'en croiser les données

avec différents critères. J'ai appliqué la même méthode pour le CNIT et Air France avec d'excellents taux de retour. Nous étions alors des précurseurs et rapidement pratiquement tous les secteurs d'activités se sont mis à croiser leurs bases de données dans une démarche de type B to B puis B to C.

QWAM : quelles sont selon vous les limites de l'IA ?

On parle d'analyse et on est capable de détecter des modifications, des « signaux faibles » à partir desquels sont établies des interprétations à caractère le plus souvent statistique. Cependant, ces informations proviennent par définition du passé et c'est sur ce passé que se fondent les projections vers le futur. Quand on analyse le passé, on analyse le comportement du passé mais ce comportement est loin de la réalité d'aujourd'hui et encore plus loin de celle de demain. Si on ignore ou si on néglige cet aspect cognitif, les projections seront fausses et dans l'incapacité d'être exploitées pour bâtir des recommandations dans le contexte d'aujourd'hui ou de demain. C'est un paradoxe... Toutes les sociétés qui ont tenté de développer des outils décisionnels à partir des signaux faibles ont



QWAM Text Analytics
Quand le passé éclaire le présent et anticipe le futur
Cas pratique - Amélioration de la connaissance des séismes historiques

La France Métropolitaine est caractérisée par une activité sismique faible à modérée, et donc une période de récurrence longue des événements sismiques majeurs. Il s'est produit, en moyenne, un séisme fortement destructeur et quatre séismes responsables de dommages mineurs par siècle sur mille ans d'histoire. Prenons l'exemple du séisme de Lille au XIIème siècle ou plus récemment, celui de Lambès en Juin 1929.

La sismicité historique est basée sur la recherche et l'analyse des témoignages sur le comportement de terre, construits dans le patrimoine historique. Ces témoignages constituent la base de la macro-sismicité, c'est à dire la sismicité dont les effets peuvent être décrits et quantifiés via des intensités macro-sismiques.

En France, la base de données Sifrance (consortium IGN, BRN et BRGM) <http://www.sifrance.org/> dispose de données macro-sismiques.

Les séismes historiques sont plus ou moins bien connus, et la robustesse de l'estimation de leurs paramètres sismologiques (moment angulaire, localisation) dépend fortement (i) du nombre (N) de la qualité et (ii) du niveau de détails de la transcription des effets des séismes des documents trouvés en utilisant leur géo-graphique où le séisme a été ressenti. Pour améliorer la connaissance des séismes

39 rue Talbot - 75009 PARIS
www.qwam.com - info@qwam.com
01 44 69 59 30

Retrouvez le cas client



d'ailleurs échoué dans leur démarche. Il n'y a là rien d'étonnant quand on considère, ce qui est mon cas, que les technologies et l'informatique ne sont qu'un moyen, un outil, et en aucune façon une solution. On a essayé de faire croire qu'on vendait une solution. Mais si on n'a pas les hommes, les experts, ces solutions ne servent à rien. Le B to B pour moi c'est le Brain to Brain. On cherche le savoir pour le mettre à la disposition des autres.

QWAM : l'IA doit-elle toujours être associée à l'expertise humaine ?

J'ai toujours cherché à récupérer les informations fiables qualifiées par de vrais experts et de m'assurer que leurs procédures de mises à jour étaient optimales. Prenons l'exemple de la santé : l'IA est incapable de dire si tel ou tel médicament est utile ou pas pour la Covid. En revanche, l'IA va pouvoir dire que

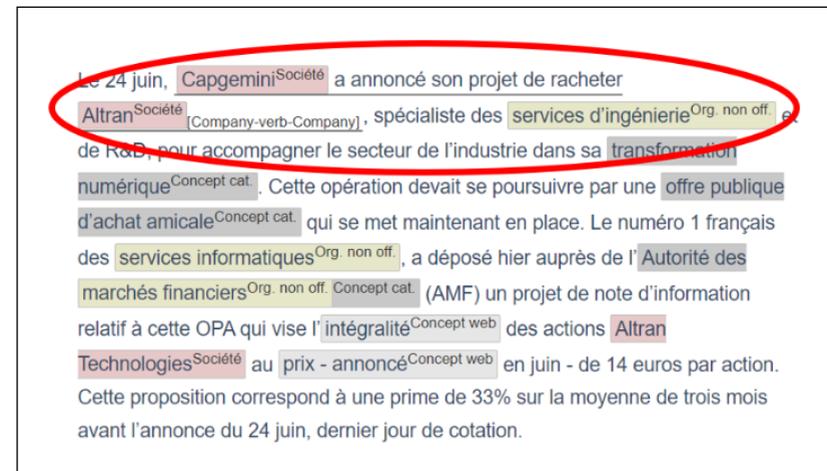
tel traitement a x% de chances d'être efficace. La masse d'informations sur laquelle se fonde l'analyse, aussi vaste soit-elle, ne peut permettre d'obtenir que des recommandations d'ordre prédictif, des projections. Mais rien ne garantit que ces projections soient fiables à partir du moment où l'on peut garantir que l'information de base, qui sert de pivot, est elle-même fiable. L'analyse via l'IA de la masse d'informations va générer des alertes et des règles qui vont elles-mêmes permettre à des « ultra-sachants », des scientifiques de déterminer quels peuvent être les écarts et d'établir avec pertinence des probabilités. Le pire danger est de croire le résultat d'analyses statistiques. Sans experts, c'est la démocratie des non-sachants qui prévaut. Regarder les réseaux sociaux et les théories des complots... Maintenant, nous avons tous vu les « experts » de la Covid s'écharper

sur le sujet. Choisir les bons experts est donc indispensable pour décider à partir des résultats de l'IA.

QWAM : à quel niveau et sur quelles bases l'IA peut-elle s'appliquer à la compliance ?

La compliance se fonde sur des textes réglementaires, des déclarations, des audits et des certifications. Dans ce contexte à la fois spécifique et basé sur la confiance, l'IA va permettre par exemple de repérer dans les déclarations les éléments en fonction des différentes réglementations en vigueur (et variables) d'un pays à l'autre. L'analyse issue de l'IA va aboutir à l'élaboration de recommandations prédictives et à la création de règles disposant de plusieurs prismes (juridiques, législatifs, culturels...), auto-améliorables et corrigées par des experts. Cette pre-

mière étape devrait suffire pour ce qui touche à la compliance déclarative. Mais la compliance nécessite d'obtenir un taux de 100% en matière de certitude des informations données et cela à un instant T. Ce sont les 2 ou 3 % qui manquent qui posent problème et que l'IA est en capacité de repérer. Les modules à base d'IA, à l'image de Data discovery sont capables de détecter des cas de non-conformité, de générer des alertes prédictives et d'engager des actions de remédiation. Mais il en va de la compliance comme des autres domaines applicatifs, même si l'IA est capable de préconiser des recommandations ou de générer des alertes, ce sont les décideurs qui auront le dernier mot !. Avec une dernière évidence : si les experts sont mauvais, les décisions le seront également...



Sémantique et finance... une approche opérationnelle
Retrouvez le cas client

IA, DÉFENSE ET SÉCURITÉ

UN DOUBLE ENJEU TECHNOLOGIQUE ET DE SOUVERAINETÉ



Christian Langevin
Directeur Général QWAM

Depuis plus de 20 ans, l'intérêt pour l'IA chez les militaires n'a cessé de croître jusqu'à devenir une priorité en termes de R&D. Aux USA, la CIA finance et mène près de 200 projets basés sur l'IA. Les domaines applicatifs de l'IA sont d'ores et déjà nombreux : aide à la décision et au commandement, planification, robotisation, armement, maintien en condition opérationnelle, systèmes de géolocalisation... Mais c'est dans le renseignement et la cybersécurité que l'IA est aujourd'hui la plus prometteuse et la plus attendue. Ainsi, la plupart des analyses et interprétations d'images satellites reposent sur des logiciels ayant recours à l'IA facilitant le travail de renseignement. De même l'IA est utilisée pour traiter les données issues du big data, recouper les informations, les corrélées, leur donner un sens tactique, voire prédictif. Enfin, en matière de cyber-guerre, la contribution de l'IA porte notamment sur la résilience des réseaux, la détection et la prévention des cyber-attaques.

En matière de sécurité intérieure, l'apport de l'IA s'inscrit dans un contexte différent de celui de la Défense. L'environnement technologique n'est pas le même tout comme le montant des budgets alloués aux équipements informatiques, plus modeste. Pourtant, les enjeux et les objectifs sont relativement proches. Dans les deux cas, à l'heure de la transformation numérique, l'automatisation des tâches les plus simples et les plus répétitives, est une priorité partagée. L'objectif est de gagner en rapidité et en performance sachant que l'IA peut y contribuer. Les applications métiers les plus répandues portent sur l'analyse décisionnelle et prédictive, la reconnaissance faciale et la biométrie, la criminalistique orientée image, les outils de prévention et de lutte contre la criminalité, les processus d'authentification dynamique, ou encore la détection et le croisement de données. Par ailleurs, tout comme à la Défense, l'exploitation du big-data et le renforcement du renseignement sont deux priorités où

les technologies de machine learning et de deep learning sont particulièrement attendues. Le potentiel de l'IA repose ici sur son aptitude en tant qu'outil et multiplicateur d'efficacité opérationnelle, décisionnelle et informationnelle.

Au-delà des opportunités capacitaires de l'IA, reste à surmonter plusieurs défis. Politique et industriel tout d'abord. En matière de Défense et de sécurité intérieure, tous deux de niveau régalién, la capacité de la France à garantir sa souveraineté en matière IT (et particulièrement la confidentialité de nos données les plus sensibles vis-à-vis d'états tiers) est une priorité absolue. Cette exigence passe par la protection de l'ensemble des maillons de la chaîne de gestion des données (hébergement via un cloud souverain et/ou sur des serveurs de préférence physique, logiciels de traitement, API dédiées...). Dans un contexte où les ruptures technologiques sont fréquentes, cette souveraineté nécessite des investissements conséquents en matière de R&D que ni l'État ni le secteur industriel ne peuvent porter seuls. Une coopération étroite entre l'État, les États si l'on se situe dans une vision de défense européenne, et le monde de la R&D, les développeurs et les porteurs de solutions est donc une nécessité, sachant par ailleurs que les grands groupes industriels peinent à attirer les talents de la tech, rares et convoités. Ces derniers optent désormais souvent pour une forme d'indépendance via la création de leur propre entreprise. Par ailleurs, ces mêmes industriels s'inscrivent dans une logique de redéveloppement ou d'utilisation de l'open source afin de maximiser leur CA et rechignent à utiliser les PME à qui elles devraient reverser une partie de leurs revenus.

Sur le plan technique, l'alimentation en données constitue un défi majeur. La masse disponible doit être suffisamment importante et représentative pour être en capacité de se rapprocher au plus près du monde réel. À défaut, le résultat final issu de l'IA et du deep learning apparaît comme faussé et l'analyse automatique qui en découle distordue par rapport à la réalité. L'accès à des bases de données aussi vastes que possible est donc une nécessité, sachant par ailleurs que ces bases sont particulièrement sensibles et que l'anonymisation réduit la pertinence des résultats obtenus. Enfin, la fiabilité des outils et la protection des données vis-à-vis de toute tentative d'intrusion ou de détournement sont autant de défis supplémentaires : la moindre vulnérabilité et/ou interruption de l'alimentation en données faussent voire bloquent tout calcul et donc toute interprétation basée sur l'analyse. On citera enfin parmi les enjeux techniques de l'IA, la puissance de calcul et l'interopérabilité des outils (dans un contexte éminemment évolutif).

L'analyse des données textuelles, domaine d'expertise de QWAM, est l'un des enjeux majeurs de la Défense et de la sécurité intérieure. On citera à titre d'exemple les innombrables comptes-rendus et documents de toute nature produits ou analysés par les opérateurs de ces secteurs, les transcriptions d'écoute, l'analyse d'informations web ou réseaux sociaux à grande échelle, etc. L'IA pour l'analyse des données textuelles peut être au cœur de l'amélioration significative des applications à venir pour une meilleure efficacité opérationnelle.

MOTEUR DE RECHERCHE, INFORMATION & MÉDIA



Sophie Guerrier

Responsable du service Documentation
et Archives numériques
Groupe Le Figaro



Eric Roustit

Responsable informatique éditoriale
Groupe Le Figaro

QWAM : depuis quand utilisez-vous les outils de QWAM ?

Eric Roustit : nous travaillons avec QWAM depuis onze ans. On a commencé avec l'indexation automatique des articles destinés à alimenter la base documentaire du Groupe Le Figaro. En 2014, nous avons intégré la solution Ask'n'Read pour enrichir à la volée, directement depuis le CMS et avant publication, l'ensemble de nos articles pour le web à l'aide de mots-clefs et de rebonds vers les articles similaires. Plus récemment, nous avons amélioré notre dispositif de requête via la dernière génération de leur moteur d'indexation. En ouvrant son portail documentaire, le matin, chaque journaliste dispose de ses cinq articles les plus récents (print et web) ainsi que ceux qui sont publiés par le service auquel il appartient. Cette

sélection est complétée par une mise en avant visuelle des principales tendances de l'actu du jour traitées par tous les titres print de la PQN et PQHN ainsi que ceux issus des agences AFP et Reuters.

QWAM : quels sont les principaux critères qui vous ont amené à choisir les solutions de QWAM ?

Sophie Guerrier : nous visions un outil unique d'accès à l'ensemble des flux nécessaires au travail de la rédaction et capable d'effectuer des recherches sur de multiples bases de données, incluant nos propres productions ainsi que la presse spécialisée et internationale issue de l'intégrateur Factiva développé par Dow-Jones. Notre objectif était de simplifier au maximum le processus de requête et d'aboutir à



une démarche aussi transparente que possible vis-à-vis de l'utilisateur final. Le journaliste tape sa question sans se soucier de savoir s'il fait sa recherche via un système de crawling ou un autre outil. C'est vraiment cela qui a fait la différence avec les solutions concurrentes ainsi que le fait que le service proposé par QWAM soit accessible sur l'ensemble des écrans, y compris les tablettes et les smartphones. En temps de confinement, c'est très important !

QWAM : vous avez déployé une nouvelle version de QWAM. Pour quelle raison ?

La nouvelle version est un outil de gestion d'accès au flux et de recherche documentaire qui conserve cette logique multibase simplifiée à l'extrême. Elle s'adresse à la rédaction mais également à tous les services administratifs du groupe, dont le service juridique, celui du personnel et le service data du marketing. Une présentation sous forme de tags vient enrichir les résultats de recherche du moteur qui permet d'accéder très rapidement à l'information en cliquant sur des sources de façon très intuitive. Pour affiner le sujet, il suf-

fit de cliquer sur un mot (ou tag) et on accède alors à l'ensemble des articles répondant à cette recherche. C'est un système simplissime qui, en outre, offre d'autres critères de recherche que l'habituel aspect chronologique. Je précise que les archives du Figaro depuis 1997 sont consultables en ligne et en pdf. Nous remonterons prochainement jusqu'en 1944 puis jusqu'en 1854, date de création du journal.

QWAM : quel a été l'accueil des journalistes ?

Sophie Guerrier : dans un premier temps, le changement n'est pas forcément perçu comme un ami. Il faut apprivoiser l'outil, acquérir de nouveaux automatismes et se donner le temps de la découverte. Une phase d'apprentissage et d'accompagnement est nécessaire.

Eric Roustit : les journalistes veulent accéder rapidement à l'information sans vraiment structurer leur recherche. On essaie de pousser QWAM dans ses retranchements et de challenger son équipe technique sur des solutions permettant de faciliter

les recherches des journalistes et de simplifier au maximum le processus de requête... Il faut faire en sorte que les utilisateurs trouvent la démarche vraiment facile et qu'elle devienne incontournable. Mais ce qui paraît simple demande en fait au préalable beaucoup de réflexion et de tests pour proposer quelque chose qui soit accessible et compréhensible par tous.

QWAM : avez-vous suivi les démarches de QWAM en matière d'IA, de NLP ou encore d'ontologie ? Apport des technos limites ?

Sophie Guerrier : la dernière version du moteur et de la surcouche applicative développée par QWAM permet de diminuer le temps de travail des documentalistes dévolu à la correction et à la qualification des productions éditoriales. Cette nouvelle version est beaucoup plus riche en termes de résultats de recherche et de pertinence des infos. C'est vraiment efficace. Le fait que le critère de parution simultanée, et donc l'aspect chronologique, a sauté permet de faire remonter des affaires similaires mais qui se sont produites 15 ou 20 ans avant. Cette capacité de rapprochement résulte de l'apport de l'ontologie développée par les équipes de QWAM.

QWAM : qu'attendez-vous des innovations en matière de NLP et d'IA ?

Sophie Guerrier : QWAM nous propose des solutions innovantes que l'on n'a pas l'habitude de voir dans des systèmes documentaires certes très efficaces mais souvent figés et austères pour les utilisateurs finaux. La nouvelle

version de QWAM représente un vrai progrès de ce point de vue, tant sur la forme que sur le fond. On a envie de faire plein d'autres choses avec les experts de QWAM. Le champ des possibles est large. Grâce aux tests, nous avons déjà identifié des besoins, pour l'essentiel liés au cycle de publication et de republication de nos productions éditoriales. Mais aussi à la pertinence du moteur par rapport à nos contenus et à un outil pour suivre et enrichir ces mises à jour. La réactivité est notre impératif, les productions web devenant de plus en plus stratégiques.

Par ailleurs, nous avons un certain nombre de chantiers en cours à l'image des outils de Speech to Text implémentés sur les vidéos produites par le journal. L'objectif de cette application consiste à générer et à indexer de façon automatique des textes à partir des vidéos produites par Le Figaro. Il devient donc possible de chercher et de visionner un extrait précis dans une vidéo, sur la base d'une phrase ou d'un mot. Nous attendons également beaucoup de l'amélioration de la pertinence de l'indexation automatique pour laquelle la marge de progrès est importante. Seuls des moteurs particulièrement puissants, à base d'IA, pourront permettre de développer ce type d'applications.

Eric Roustit : la data-visualisation est également une piste très prometteuse avec le développement de modules métiers qui devrait permettre à nos journalistes de concevoir eux-mêmes et de façon rapide la scénarisation de leurs recherches et alimenter la production de leurs textes sous forme visuelle.

ANALYSE DE L'INFORMATION WEB



Alain Beauvieux
Senior advisor, PremièresVentes

Docteur en informatique, Alain Beauvieux est un expert de la commercialisation de logiciels innovants, notamment pour la reconnaissance et le traitement de la parole, l'aide à la traduction, le traitement d'images numériques ou encore la valorisation des informations circulant sur le Net. Après plusieurs années chez IBM, il a présidé la société AMI Software, leader des solutions d'Intelligence économique. En 2017, il fonde « PremièresVentes », une société d'aide aux start-ups dans l'IT.

L'industrie informatique est coutumière des modes et sa capacité à se réinventer est assez surprenante ; les mêmes concepts sont souvent présentés comme des nouveautés là où il n'y a que redite... Il en est ainsi de l'Intelligence Artificielle, et notamment de l'apprentissage automatique, dont la majeure partie de l'algorithmique exploitée aujourd'hui a été inventée au cours des années 1980. La véritable nouveauté vient plutôt des données dont la volumétrie a littéralement explosé : le Zeta octet (10^{22}) est devenu l'unité de mesure du web bien qu'il soit difficile de savoir ce que l'on mesure exactement.

Ce gigantesque gisement de données, le « big data », a changé la donne. Il y a 30 ans, la difficulté était plutôt de pouvoir faire fonctionner les algorithmes sur un périmètre significatif. Ainsi, la reconnaissance de la parole et la construction d'un « modèle de langage » nécessitent une véritable « chasse à la donnée » qui pouvait prendre plusieurs mois. L'application de reconnaissance vocale SIRI d'Apple n'a plus vraiment ce problème et les applications de cette nature ne semblent pas plus handicapées.

D'une façon générale, ce changement de paradigme a eu un impact significatif sur les technologies mises en œuvre pour l'analyse sémantique et plus généralement le traitement automatique des langues (Natural Language Processings - NLP en anglais). Jusqu'à récemment, deux écoles se faisaient face : d'une part, les tenants d'une sémantique formalisée via des modèles de type réseaux sémantiques, cartouches de connaissances, graphes conceptuels... D'autre part, les chercheurs et ingénieurs prônant une sémantique calculatoire, c'est-à-dire reposant sur une approche probabiliste du NLP.

ANALYSE DE L'INFORMATION WEB

Cette dichotomie a été au cœur des approches des outils de traduction automatique qui appartenait à l'une ou l'autre des écoles ; Google et son formidable gisement de données indexées voire alignées a mis fin à la polémique dans le domaine avec la performance que l'on connaît.

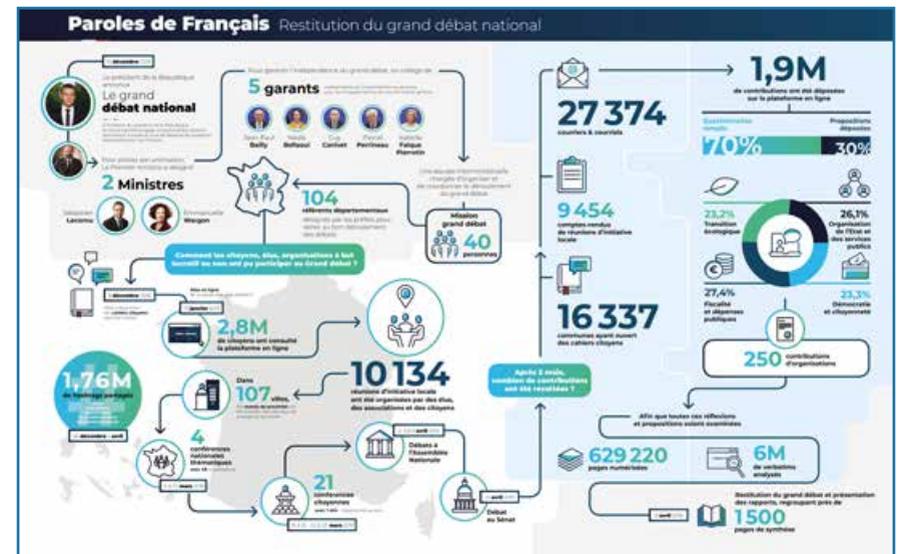
Le même chemin a été poursuivi par les analystes et notamment les personnes en charge de la veille, de la Market Intelligence et plus largement de l'analyse de données. Il y a encore une dizaine d'années, « transmettre la bonne information à la bonne personne au bon moment » était une promesse enthousiasmante. Tout simplement, parce que trouver de l'information était encore un challenge. Aujourd'hui, le problème est plutôt de « séparer le bon grain de l'ivraie » avec une dimension nouvelle qui est la vitesse de propagation des informations, un des 3 V qui caractérisent le big data. Une information chasse l'autre et les auteurs de Fake News s'en régalaient sachant pertinemment qu'il devient difficile de contrôler ces flux d'informations.

Alors que faire devant ces gigantesques flux d'informations ? Il s'agit précisément de les utiliser à bon escient pour permettre aux algorithmes d'apprentissage automatique issus de l'IA de travailler de façon performante en utilisant toutes les technologies disponibles comme la reconnaissance vocale ou la traduction automatique. Un second enjeu porte sur la mise en évidence des informations-clés par des techniques d'analyse telles que celles proposées par la société QWAM avec le logiciel QWAM Text Analytics. La promesse n'est plus tellement de faire de la veille (la bonne information au bon moment...) mais plutôt d'extraire de ces gigantesques flux d'informations les éléments qui permettront de mettre en évidence des tendances, de dégager les points importants, de corréliser afin de confirmer.

Cette profusion de données ne « tue pas l'information ». Elle permet au contraire de mettre en œuvre des techniques d'analyse et de réflexion extrêmement performantes. Promesse qui ne peut être tenue que par des outils performants comme QWAM Text Analytics.

L'IA AU SERVICE DE LA DÉMOCRATIE

L'IA, OUTIL DE LA DÉMOCRATIE PARTICIPATIVE



L'intelligence Artificielle est arrivée sur le devant de la scène politique à l'occasion du Grand Débat National lancé par le Président de la République en janvier 2019. Dès le démarrage de la consultation et anticipant un volume important de réponses, le Gouvernement s'est mis en quête de prestataires pour le dépouillement. Le site <https://granddebat.fr/>, plateforme de contribution en ligne fournie par la société Cap Collectif, a accueilli le questionnaire conçu

par le Gouvernement et composé de plus de 50 questions fermées et de 76 questions ouvertes (voir rapport méthodologique <https://granddebat.fr/pages/syntheses-dugrand-debat>). Le dépouillement et la synthèse globale de ces questions ont été confiés à l'institut OpinionWay qui a fait appel à QWAM pour l'analyse des questions ouvertes. Au regard du nombre de contributions aux 76 questions ouvertes (5,5 millions de réponses) et de l'exigence



**Intelligence artificielle :
outil incontournable du
traitement de vos textes.
En savoir plus sur le blog**

L'IA AU SERVICE DE LA DÉMOCRATIE

par le Gouvernement de délais de restitution très court (quelques semaines), un traitement manuel n'était pas envisageable. Le recours à des outils de traitement en masse s'est donc imposé. Ces outils complétés de solutions d'IA développées par QWAM et couplés à des validations humaines ont permis de superviser le travail des algorithmes et de s'assurer de la cohérence des résultats. L'apport réussi de l'IA dans le Grand Débat démontre son intérêt en matière de dépouillement et d'analyse fine d'enquêtes à très grande échelle. L'IA permet de prendre en compte des expressions, suggestions ou propositions ciblées afin de faire émerger des thématiques multiples (environnement, santé, sécurité, transport...) et de devenir un levier de la démocratie

participative. À l'ère de la data, la prise en compte à grande échelle de données émanant des citoyens permet le développement d'une démocratie participative. L'IA peut s'avérer extrêmement utile dans d'autres domaines liés à l'expression démocratique. Au-delà, l'IA peut également constituer un outil efficace de lutte contre les "fake news", une vraie menace pour la démocratie, en limitant sensiblement leur propagation sur les réseaux sociaux. En conclusion, quand elle est maîtrisée et utilisée à bon escient, l'IA peut être une opportunité de diversification des outils de pratique de la démocratie, sachant qu'il existe, à l'heure du numérique, une forte demande d'expression directe des citoyens.



>> À lire l'article complet paru dans la revue de l'ENA

CONCLUSION



Fanny Maffre
Responsable clientèle QWAM

Les différents témoignages d'experts et de clients qui jalonnent ce livre blanc montrent que l'Intelligence Artificielle constitue une solution innovante et particulièrement bien adaptée aux métiers intégrant des données textuelles. Et même si le terme d'IA appliquée au domaine du Text Mining peut apparaître comme galvaudé dans l'état actuel des technologies, il ne reste pas moins que l'attente est très forte de la part des utilisateurs et les visées applicatives prometteuses.

Dans les débats entre pro IA, attentistes et sceptiques, QWAM a choisi la voie de l'innovation et de l'IA pour le traitement des données textuelles tout en intégrant le pilotage humain. Plus que tout discours, l'expérience que nous avons accumulée, notamment à travers l'analyse des contributions du Grand Débat National, nous a confirmé dans l'idée qu'une méthodologie spécifique et des outils de pilotage dédiés sont nécessaires afin de garantir la pertinence d'une solution et d'une offre intégrant l'IA.

L'IA doit permettre aux analystes de réaliser à la fois plus facilement et avec plus de pertinence des tâches ou des analyses qui seraient trop chronophages ou impossibles à réaliser à cause, par exemple, du nombre et de

la pluralité des données à traiter. Ainsi, chaque analyste est formé et accompagné avec pour objectif d'être autonome sur son outil de travail, de maîtriser les paramètres et les restitutions et de comprendre la forme du traitement qui aura été appliqué dans le processus d'analyse. Pour éviter les biais cognitifs, il nous semble indispensable que les utilisateurs de nos solutions d'Intelligence Artificielle appréhendent au mieux les différents traitements réalisés sur leurs données textuelles afin qu'il n'y ait pas d'effet « boîte noire ».

L'Intelligence Artificielle appliquée au NLP suppose de relever de nombreux défis, tant sur le plan technique que sur le plan financier. En la matière, la R&D coûte cher et les compétences sont très recherchées. QWAM s'emploie depuis sa création à attirer les meilleurs experts en développement, afin de demeurer à la pointe des avancées technologiques et d'appréhender au mieux les besoins des utilisateurs ainsi que les nouveaux usages auxquels nous devons répondre. Quels que soient l'environnement technique et les innovations à venir, chez QWAM nous restons persuadés que l'intelligence artificielle prend tout son sens tant qu'elle est et reste au service de l'humain.



Ask'Read

Solution de détection et d'analyse des informations web



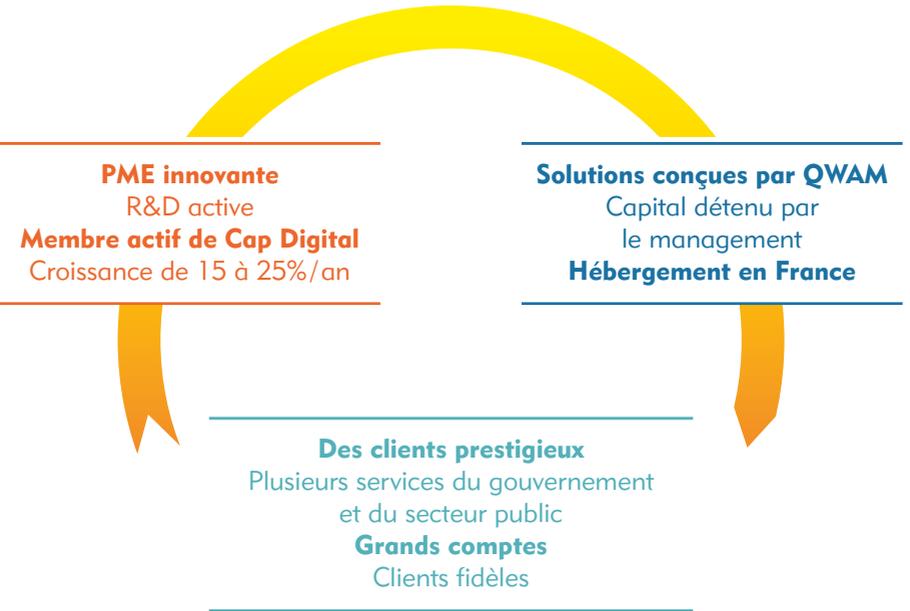
QWAM Text Analytics

Solution d'Intelligence Artificielle pour l'analyse et la valorisation des données textuelles



QES InfoMedia Suite

Solution de recherche, gestion et diffusion des flux numériques



NOS CLIENTS



CONTACTS / CONTRIBUTIONS

Merci à celles et ceux qui ont participé à l'écriture de ce Livre Blanc.

Françoise Soulié-Fogelman
Conseiller scientifique du Hub France IA

Vincent Guigue
Professeur associé, Maître de conférences, UPMC LIP 6

Loïc Deo Van
Fondateur / CEO EverReady.ai

François Laurent
Fondateur Consumer Insight, co-président ADETEM

Emmanuelle Nayman
Sismologue, EDF SIGMA

Michel Meyer
Président de Data Integrity et du Cercle des DPO, des CDO, de la conformité et de la gouvernance

Sophie Guerrier
Groupe Le Figaro, Responsable Documentation et Archives numériques

Eric Roustit
Groupe Le Figaro, Responsable informatique éditoriale

Alain Beauvieux
Senior advisor, PremièresVentes

Ainsi qu'aux contributeurs de QWAM.

Christian Langevin
Directeur Général

Ivan Monnier
Directeur Technique

Fanny Maffre
Responsable clientèle

Jordan Saltiel
Assistant ingénieur d'affaire

Conception, interviews, rédaction, coordination éditoriale

Denis Fortier

Conception graphique

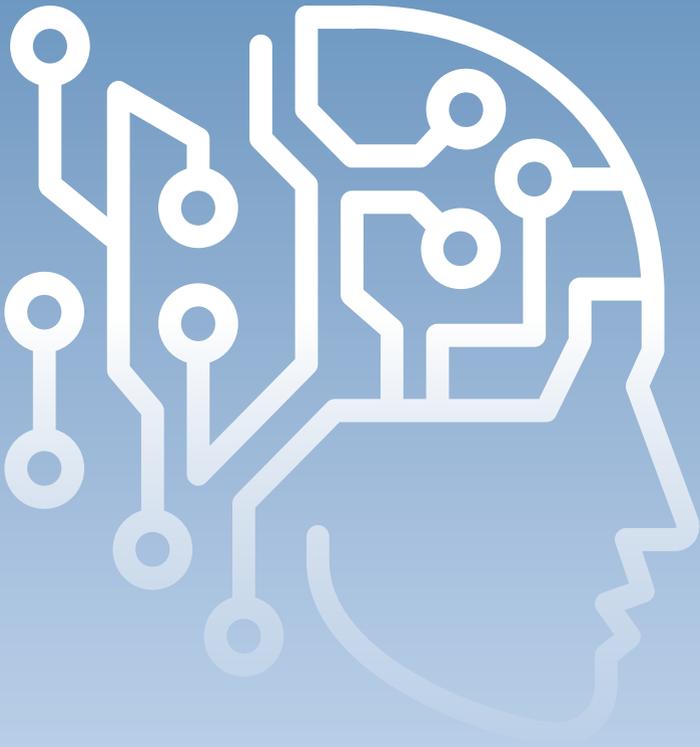
Alexis Jacov

QWAM CONTENT INTELLIGENCE
29 rue Taitbout, 75009 Paris, France
+33 1 44 69 59 20 • info@qwamci.com • www.qwamci.com

Mentions légales

Copyright © 2020 QWAM CI. Document non contractuel. QWAM CONTENT INTELLIGENCE, immatriculée au registre du commerce et des sociétés : RCS 493 317 374 Paris, et dont le siège est situé 29 rue Taitbout 75009 Paris. Téléphone: +33 (0)144695920. L'ensemble des éléments figurant dans ce Livre Blanc sont protégés par les dispositions du Code de la Propriété Intellectuelle. En conséquence, toute reproduction de ceux-ci, totale ou partielle, ou imitation, sans notre accord exprès, préalable et écrit, est interdite. Cette interdiction s'étend notamment, sans que cette liste ne soit limitative, à tout élément rédactionnel figurant dans le présent Livre Blanc, à la présentation des logos, images, photos, graphiques, de quelque nature qu'ils soient.

QWAM
CONTENT INTELLIGENCE



www.qwamci.com