Le Monde
22 mars 2019

article suivant
Maternités Les trajets s'allongent,...

Grand débat : le travail très politique des algorithmes

Différents logiciels ont commencé à décrypter la masse des contributions des citoyens pour aboutir à des synthèses début avril

Faut-il confier à des algorithmes les clés du débat démocratique ? La question s'invite dans le débat politique alors que les logiciels des prestataires choisis par le gouvernement sont en train de mouliner les millions de mots des contributions au grand débat pour en établir une synthèse, prévue début avril. En fait, il y aura plusieurs synthèses. En fonction des formats, les organisateurs ont réparti le travail entre différentes entreprises spécialisées dans le traitement automatique de textes.

Les contributions en ligne de la plate-forme *Legranddébat.fr*, attribuées à OpinionWay et la société Qwam, sont traitées par un algorithme qui s'appuie sur un -référentiel de mots et de concepts déjà existant. Il s'agit, selon le directeur général de Qwam, Christian Langevin, d'une *" solution de traitement sémantique associée à l'intelligence artificielle et capable de reconnaître le sens des mots dans une phrase pour classer les contributions en fonction des thématiques abordées : développement durable, augmentation du pouvoir d'achat... ".*

Les cahiers citoyens, courriers manuscrits et synthèses des réunions locales, de leur côté, numérisés par la Bibliothèque nationale de France et son prestataire Numen, sont traités avec une autre méthode. Le logiciel de la société Cognito repère les mots-clés et leur contexte, puis crée un lexique *" directement à partir des textes eux-mêmes, au fil de l'eau, ce qui nous semble plus respectueux de la manière dont les -citoyens s'expriment,* assure son fondateur, Gilles Proriol. *Sur les centaines de milliers de contributions déjà analysées, nous avons listé pour le moment quelque 600 propositions différentes, regroupées autour de trente-huit sous-thèmes et huit thèmes ".*

Chacun des dispositifs prévoit des interactions fréquentes entre

l'homme et la machine. Une part essentielle du travail reste néanmoins automatisée, ce qui, dans un processus démocratique, soulève de nombreuses questions.

" Les risques d'erreurs et les difficultés d'interprétation sont réels, notent un ingénieur et un chercheur du laboratoire pluridisciplinaire Triangle, spécialisé dans l'analyse de l'action et du discours politique. *Pour qu'un traitement automatisé fonctionne, il faut que les contributeurs utilisent les bons termes et le bon ordre, ce qui n'est évidemment pas toujours le cas. Il faut compter avec les fautes de frappe ou d'orthographe, dont il reste un nombre incompressible, même en supposant un important travail de nettoyage des données, par exemple “fiancé” au lieu de “financé”. Il est aussi difficile d'extraire le sens lorsque la phrase est ironique ou que la syntaxe est erronée ".*

" Travailler en transparence "

Pour le mathématicien David Chavalarias, qui pilote le projet Politoscope de l'Institut des systèmes complexes de Paris Ile-de-France (ISC-PIF), *" le recours à la technologie peut être intéressant, mais tout dépend ce que l'on -recherche. Si le politique veut comprendre les grandes tendances ou trouver quelques bonnes idées parmi une foule de contributions, cette consultation peut être utile. Mais ce serait une erreur d'en déduire ce que veut le peuple, car les biais méthodologiques sont trop nombreux ".*

Le recours aux traitements automatisés soulève aussi des enjeux de transparence. Lors d'une journée organisée au Conseil économique, social et environnemental par le think tank Décider ensemble, lundi 18 mars, Isabelle Falque-Pierrotin, membre du collège des garants, a prévenu : *" Nous sommes en train de travailler avec les prestataires pour ouvrir la boîte noire et que la restitution soit incontestable. Les citoyens doivent être en capacité de savoir comment elle a été -construite. Nous devons comprendre comment les prestataires -travaillent. "*

Dans une démarche d'" open data ", une partie des contributions est déjà mise en ligne et le reste devrait suivre. Les prestataires proposent aussi de publier les liens entre chaque contribution et la catégorie dans laquelle elle a été classée. *" On veut travailler en transparence totale sur ce que fait notre logiciel "*, affirme Gilles -Proriol, de la société Cognito.

Pour autant, aucun des outils des prestataires n'étant sous licence " open source ", il n'est donc pas possible d'avoir accès aux détails de l'algorithme. Pour David Chavalarias, *" ne pas avoir accès au code du logiciel, c'est-à-dire aux détails de la méthodologie et des paramètres choisis, pose problème. Les thèmes retenus seront-ils représentatifs ? Comment être sûr qu'il n'en reste pas d'autres que le logiciel n'aura pas trouvés ? Il suffit souvent de modifier un peu les curseurs pour que les résultats soient très différents. Il faut pouvoir justifier ces choix et rendre publics les biais qu'ils introduisent, ce qui n'est pas possible sans un accès direct au code et aux -paramètres utilisés ".*

Analyses parallèles

L'ouverture des données suscite des initiatives parallèles à l'analyse officielle. A l'Assemblée nationale, un hackathon réunira samedi des développeurs volontaires sur des projets collaboratifs. Plusieurs unités de recherche universitaire travaillent aussi sur les contributions des différentes plates-formes.

Au sein du laboratoire Triangle, qui explore les échanges publiés sur le site du vrai débat, lancé par un groupe de " gilets jaunes ", les ingénieurs en sont aux premiers constats : *" Nous avons repéré une grosse présence de la sensibilité sur l'écologie et l'urgence climatique, ou bien encore une forte demande de service public (proximité, demandes massives de -nationalisation ou de renationalisations...). Les revendications identitaires et réactionnaires existent (Frexit, immigrés, etc.) mais sont très minoritaires, et les propositions visant à rétablir la peine de mort ou à revenir sur la loi Taubira sur le mariage pour tous sont massivement rejetées. "*

Pour la première fois, une -consultation citoyenne suscite aussi l'intérêt d'entreprises et d'experts spécialisés dans l'analyse et la sécurité des données. Un *data scientist* a ainsi montré que des informations person-nelles publiées sur le site du grand débat n'avaient pas été anonymisées. Avec un simple filtre, il a -retrouvé près de 200 adresses électroniques et une quarantaine de numéros de téléphone de -personnes dont *" une partie des contributions montrent clairement quel est le bord politique de l'auteur ".*

Claire Legros