démocratieS
le média critique de la participation

Faire délibérer les foules avec l'IA ? Pas encore...

Aurélien Bellet
Chercheur à l'Institut national de recherche en sciences et technologies du numérique

L'IA peut-elle permettre la délibération citoyenne à grande échelle ? Ce n'est pas pour tout de suite, tempère le chercheur Aurélien Bellet, qui pointe des enjeux immédiats de conception, de fiabilité et de transparence.

Avec des collègues chercheurs de l'Inria, vous avez évalué la synthèse officielle de la consultation numérique du Grand débat national. Les résultats officiels, produits par l'IA, étaient-ils fiables ?

La synthèse officielle du Grand débat national est un échec, de notre point de vue. La consultation citoyenne en

ligne comprenait un questionnaire de type sondage, dont les résultats sont faciles à restituer, mais aussi des questions ouvertes. Les Français ont répondu à ces dernières en rédigeant des textes dont beaucoup expriment des opinions détaillées et argumentées : 5 millions de réponses à comprendre, classer et analyser pour rendre compte des avis. Pour ce travail, l'État a fait appel à des entreprises privées, non aux chercheurs des institutions publiques. Toutefois, nous avons entrepris, *a posteriori*, une étude de cette analyse officielle.

Ce qui frappe d'emblée, c'est le défaut de transparence. Les prestataires disent employer leurs propres algorithmes, sans autre précision, sans donner accès au code ni indiquer les paramétrages utilisés ; ils ne détaillent pas, non plus, la construction des catégories dans lesquelles ils classent les contributions ; enfin, ils ne décrivent pas clairement les étapes de l'analyse. Tout cela rend difficile l'appréciation de leur travail.

Nous avons donc essayé de répliquer leurs résultats, en utilisant plusieurs approches standard dans le domaine de l'analyse textuelle et en publiant le code nécessaire à la reproduction de nos expériences. Nous avons conservé leurs catégories d'analyse aux fins de comparaison. Les résultats prennent ainsi la forme de pourcentages, comme dans un sondage : tant de citoyens parlent de ceci, tant de cela. L'objectif est de trouver des résultats proches de la synthèse officielle.

Premier enseignement : il est impossible de répliquer la synthèse officielle. Aucune des approches n'a permis de retomber sur leurs chiffres. Les résultats sont même, parfois, sensiblement différents.

Est-ce surprenant d'obtenir des synthèses différentes ?

C'est inhérent à l'IA car il existe différents types d'algorithmes et différentes manières de les paramétrer. Dans le cas du Grand débat national, nos résultats sont assez proches de la synthèse officielle lorsque les questions sont peu ouvertes, mais nous constatons des différences importantes, et même des aberrations, sur des questions très ouvertes.

L'autre problème de la synthèse officielle, c'est le nombre élevé de réponses qui n'entrent dans aucune des catégorie établies, ces réponses étant alors reversées dans le groupe « autres contributions » : selon les questions, de 15 % à 30 % des points de vue sont ainsi déclassés. Il s'agit en particulier des réponses exprimant des émotions, comme la colère ou la frustration, par exemple « C'est de la foutaise, ces questions sont orientées ! » ou des idées originales.

D'une façon générale, il serait donc plus honnête de présenter les résultats d'une concertation non pas comme la synthèse vraie mais comme une synthèse possible ?

Effectivement, il est important de reconnaître les limites de ces technologies basées sur l'apprentissage statistique et dont les prédictions ne sont pas la vérité. S'ajoute, dans le cas de l'analyse d'écrits, l'ambiguïté

propre au texte, pas simple à analyser, même par un humain. C'est pourquoi le manque de transparence est un problème majeur, particulièrement dans le champ démocratique. Il semble légitime de s'appuyer sur des outils automatiques d'IA pour analyser les avis recueillis à grande échelle mais la contrepartie est de publier rigoureusement la méthode et les paramétrages, ainsi que le code, idéalement.

Les avancées techniques permettraient-elles d'obtenir aujourd'hui une meilleure synthèse qu'en 2020 ?

Pas sûr. Les IA génératives, les chatbots type ChatGPT basés sur de Grands Modèles de Langage, ont apporté des améliorations notables à l'analyse de texte : le résumé, le compte-rendu. Les technologies sont meilleures mais cela ne change pas vraiment le constat de fond : on peut utiliser des modèles différents et les paramétrer de différentes façons. Si l'on ne documente pas clairement la manière dont on s'y est pris, cela ne permet pas de répliquer l'analyse ni de valider les résultats. Avec le risque qu'ils soient contestés.

Comment utiliser l'IA sans la boite noire, en préservant l'apport critique des experts mais aussi le pouvoir de contrôle des profanes, des participants ?

Nous en sommes encore au temps du développement, il faut donc confronter différentes méthodes et différentes analyses pour expérimenter le traitement automatique de très grand corpus qui, effectivement, dépassent la capacité de lecture des personnes. Si l'on trouve de nombreuses convergences entre les approches, on peut leur accorder un degré de fiabilité plus important. Pour l'instant, j'inviterais donc toutes les parties prenantes d'une démarche participative à prendre les résultats d'une analyse automatique avec prudence. À coup sûr, la promesse de fournir une « synthèse officielle » est abusive si elle ne s'accompagne pas de moyens de validation et de vérification.

La manière de représenter les résultats compte aussi. Dans le cas du Grand débat national, la méthode est proche du sondage, donc elle résume les textes à des catégories et des pourcentages. On perd potentiellement beaucoup d'informations. Une des possibilités est d'associer les citoyens à l'analyse des contributions, en initiant une campagne d'annotation de l'analyse. On peut combiner IA et participation humaine pour vérifier, par exemple, l'affectation de certains textes vers certaines catégories. Ensuite, on peut reprendre ces corrections pour ré-entrainer l'IA et renforcer ainsi ses performances, dans un processus itératif. La synthèse participative devient alors une modalité intéressante pour concevoir une IA de confiance.

Un des reproches adressés à l'IA est d'écraser les signaux faibles, les contributions marginales. Est-ce justifié ?

En effet, les technologies d'IA sont basées sur l'apprentissage statistique, le fait de trouver des régularités dans les données. Le défaut commun à toutes les approches est d'avoir tendance à sur-présenter et mieux capturer les

éléments majoritaires ou qui apparaissent de manière suffisamment répétée. Elles risquent de rater ce qui est moins fréquent, donc ce qui est émergent, innovant.

C'est un point de vigilance dans beaucoup de contextes. Quand on interroge des modèles actuels, comme ChatGPT, ils reproduisent et amplifient les biais sociaux des données d'entrainement, comme les stéréotypes de genre dans les métiers : dans une phrase comportant les termes docteur et secrétaire, l'IA aura tendance à considérer que « docteur » se rapporte à un homme et « secrétaire » à une femme.

Si la qualité de l'IA dépend des données d'apprentissage, est-il plus sûr de l'entrainer sur un corpus spécialement produit lors d'un processus participatif ?

Pas nécessairement. Les approches basées sur l'IA sont beaucoup plus performantes lorsqu'elles sont entrainées sur de grandes masses de données. De manière générale, l'état de l'art actuel consiste à pré-entrainer le modèle d'IA sur de grands volumes de données variées (typiquement, issues d'Internet) puis à raffiner le modèle sur un corpus spécifique à la tâche que l'on souhaite réaliser.

L'engouement pour l' IA générative ne doit pas conduire à l'utiliser pour tout et n'importe quoi. Des approches plus simples et moins gourmandes en données et en ressources de calcul sont parfois plus efficaces.

Peut-on envisager un standard d'IA démocratique ?

Une réflexion dans ce sens me semble possible et souhaitable. Dans le champ de la participation citoyenne, il faut poser les attentes, définir des critères de fiabilité, pour faire émerger une conception partagée d'une IA "démocratique". On pensait tenir l'occasion de le faire au moment du Grand débat national, le gouvernement avait même lancé un appel à manifestation d'intérêt pour financer des projets de recherche, forcément pluridisciplinaires. Passée la crise politique, ce programme est resté lettre morte, hélas.

Beaucoup espèrent que l'IA générative permettra une délibération de qualité à l'échelle de milliers, voire de millions de participants. Est-ce crédible ?

Ce n'est pas un fantasme mais cela va nécessiter de la réflexion et du travail. Dans ce cas, il ne s'agit plus seulement de synthétiser des données en grand nombre, mais de faciliter l'expression, modérer les discussions, contrôler l'équité de la parole, partager en temps réel un état de la délibération ; traduire et transcrire un débat possiblement multilingue, mais aussi produire ou vérifier des informations utiles aux participants... L'IA présente tout le potentiel pour rendre possible une démarche de délibération citoyenne à grande échelle, mais il me semble totalement illusoire de croire que l'on pourra déléguer un tel dispositif à un système automatisé.

À mon sens, une question importante est de concevoir d'emblée la combinaison des tâches remplies par les outils et par l'humain. Quel rôle exactement doit tenir la machine, sur quoi intervient l'humain ? À cette condition, l'IA peut faire suffisamment bien une série de tâches pour permettre aux démarches participatives de passer à une large échelle. Même alors, les dispositifs ne seront pas parfaits, ils nécessiteront aux humains de continuer à y passer beaucoup de temps !

L'IA peut faire plus vite ou à la place des humains, mais que fait-elle mieux ?

Je travaille sur les applications de l'IA en santé. Dans ce domaine aussi, les systèmes automatiques fonctionnent beaucoup mieux si on les combine intelligemment avec l'humain. Les algorithmes peuvent être meilleurs que les médecins sur certains types de diagnostics, mais ils sont bien plus mauvais sur d'autres. Pour les scientifiques qui conçoivent des modèles d'IA, il est essentiel d'intégrer à la source le partage des rôles. L'IA est bonne à ce pour quoi on l'entraine. Plutôt que lui assigner la résolution complète d'un problème, on peut l'entrainer à tenir un rôle complémentaire à l'humain.

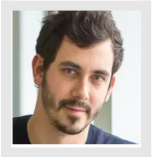
Propos recueillis par Valérie Urman

[1] Le Grand Débat national (GDN) s'est tenu du 15 janvier au 15 mars 2019. Environ 1,5 million de Français se sont exprimés sur la fiscalité, la transition climatique, les services publics, la démocratie. Quelques repères chiffrés : https://granddebat.fr

[2] L'État a fait appel à des prestataires privés pour produire la synthèse officielle des différents formats de contribution (cahiers de doléances, comptes-rendus des réunions publiques locales, consultation numérique nationale, conférences citoyennes). Chargé d'analyser la consultation numérique, le sondeur Opinion Way a livré les résultats des 52 questions fermées et a délégué à la société Qwam la synthèse des 5 millions de réponses aux questions ouvertes.

[3] Aurélien Bellet, Pascal Denis, Rémi Gilleron, Mikaela Keller, Nathalie Vauquier. Pour plus de transparence dans l'analyse automatique des consultations ouvertes : leçons de la synthèse du Grand Débat National. 2021. *Statistique et Société*, vol. 9, n° 1 et 2. Un résumé vulgarisant cette étude a été publié conjointement dans Le Monde et The Conversation.

[4] Voir notre article IA générative, LLM... De quoi parle-t-on ?



Aurélien Bellet

Aurélien Bellet est directeur de recherche à l'Institut national de recherche en sciences et technologies du numérique (Inria). Spécialiste de l'apprentissage statistique et de l'intelligence artificielle, ses travaux se concentrent sur le développement d'approches d'IA de confiance répondant aux enjeux cruciaux de protection de la vie privée, d'équité et de transparence.