

« Il faut évaluer l'IA au-delà de critères techniques »

Jean-Philippe Cointet

Professeur de sociologie au médialab de Sciences Po



Au-delà du mérite technologique, les usages de l'IA présentent-ils un réel intérêt démocratique ? « Il faut établir les critères pour mettre les outils à l'épreuve », fait valoir le sociologue Jean-Philippe Cointet associé au projet Communs démocratiques.

Que vous inspire la fulgurante popularité de ChatGPT ?

La vitesse d'adoption est effarante, le nombre d'utilisateurs déjà astronomique. Mais que font ces centaines de millions d'individus devant les interfaces intelligentes ? On ne sait pas grand-chose des usages. Nous sommes

tous dans une phase de découverte de ces outils révolutionnaires. Et tous abasourdis – y compris ceux qui développent la technologie - par les capacités nouvelles de traitement de l'information. Il est encore difficile d'en évaluer l'influence, les impacts sur la démocratie. En tant que chercheur au médialab de Sciences Po, ma collaboration au projet de recherche Communs démocratiques consiste à interroger les usages sociaux, au-delà des promesses technologistes.

Comment s'entendre sur la qualité démocratique des usages quand certains réfléchissent à sonder une machine à la place des citoyens ?

Avec les Grands modèles de langage (LLM), on dispose d'assistants qui peuvent, à toute vitesse, analyser, synthétiser, extraire des caractéristiques d'un grand volume de textes, avec une bonne qualité interprétative. À la place d'un questionnaire Google Form, on peut automatiser des entretiens afin de clarifier et reformuler la question, individualiser l'enquête. Certains vont effectivement jusqu'à imaginer se passer des enquêtés en les remplaçant par les LLM, car l'IA disposera des données pour dire ce que pense tel ou tel segment de population. Des recherches, aux États-Unis, entraînent ainsi la machine à s'exprimer à la place de groupes humains difficiles à mobiliser comme « les jeunes latinos gay au chômage qui vivent en Arkansas ».

On voit bien le saut technique, mais quel est l'avantage démocratique ?

C'est la question posée par le projet Communs démocratiques, qui s'intéresse surtout aux usages de l'IA dans les dispositifs délibératifs. Prenons un usage clé : le résumé automatique des débats. Les informaticiens qui développent ces systèmes disposent de critères reconnus pour mesurer et comparer les performances. Ces *benchmarks* techniques font autorité et, même s'ils sont peu interrogés sur les conditions de leur construction, ce sont aujourd'hui les seules grilles d'analyse dont on dispose. Pourtant, comment caractériser une synthèse automatique ou une traduction dans un espace social telle qu'une convention citoyenne où le pluralisme des opinions doit être respecté, où personne ne doit être invisibilisé ? Il faut établir de nouveaux *benchmarks*, non pas techniques mais conformes à des principes spécifiques. Deux entités de Science Po s'impliquent dans ces travaux : le Cevipof apporte un cadrage de philosophie politique, tandis que le médialab enquête auprès des citoyens utilisateurs. *In fine*, il faudra parvenir à transposer les bonnes pratiques dans des tests permettant aux informaticiens de mettre les LLM à l'épreuve. Cela suppose de traduire ces principes sous forme d'exemples pratiques, comme on le fait pour les biais de genre en vérifiant comment la machine traduit par exemple l'anglais « nurse » : infirmier ou infirmière ?

De quelles façons associez-vous les citoyens à vos travaux ?

Nous réalisons des ethnographies numériques en analysant les interactions des participants avec les LLM. À quel moment leur doigt reste-t-il suspendu au-dessus du clavier parce qu'ils sont surpris par la réponse du robot ou qu'ils sont en désaccord avec le traitement automatique de leur contribution ? Nous voulons savoir comment ils perçoivent ces usages, recueillir leur critique.

Le défi des dispositifs participatifs est de mobiliser les gens en nombre et en diversité. Les plateformes, aujourd'hui discriminantes, peuvent tirer parti des capacités de l'IA à assister les usagers en reformulant les contributions de ceux qui ne sont pas à l'aise avec l'écrit, ou avec le français. Mais comment cela est-il vécu ?

Par ailleurs, l'IA est capable de restituer en direct les principaux points d'entente dans un débat. Est-ce raisonnable de déléguer à la machine ce travail sur le consensus simplement parce qu'elle le fait très vite ? La machine doit aider les citoyens à agir, pas les priver du temps de mûrir leur réflexion.

La promesse de l'IA dans le champ délibératif est aussi de permettre aux dispositifs de passer à l'échelle. Encore faut-il s'assurer que les gens apprécient les outils. Ces derniers créent-ils de la confiance ou un sentiment de dépossession ?

La qualité démocratique d'une machine ne se définit-elle pas dans sa collaboration avec l'humain, notamment par le pouvoir des citoyens de valider toute production automatique ?

C'est certainement un impératif. De même que l'on peut affirmer que la machine, toujours un peu langue de bois et donneuse de leçons, ne sera jamais aussi fine qu'un humain pour modérer un débat. Le bon usage des IA ne vise pas à remplacer les gens. Mais gardons-nous aussi des fantasmes... Je me méfie des sciences sociales qui restent à distance des technologies pour les critiquer. Je crois qu'il faut aimer les outils, les ouvrir, les disséquer. C'est par cette expérience un peu intime que l'on parvient à une critique informée. C'est l'approche du médialab, avec une équipe pluridisciplinaire de sociologues, de designers et d'ingénieurs qui développent des outils numériques pour nous aider à enquêter sur le social.

Propos recueillis par Valérie Urman

Un jumeau numérique pour représenter les citoyens

Mobiliser massivement et faire gagner du temps : les promesses ne sont pas nouvelles mais la démocratie numérique compte cette fois sur l'IA. Au printemps 2025 débiteront les travaux du projet ADDI, un programme de recherche européen basé à Toulouse, doté d'un solide financement de 9,9 millions d'euros sur six ans.

« Notre premier terrain d'expérimentation recouvre les usages courants des plateformes numériques : s'exprimer lors d'une consultation en ligne, voter un budget participatif, liker des idées. Notre objectif est d'améliorer ces interfaces pour obtenir une participation plus massive, de 5 % à 30 % selon les cas d'usage », fait valoir le mathématicien Umberto Grandi, chercheur à l'université de Toulouse Capitole, coordonnateur du projet. Des essais randomisés et contrôlés doivent permettre de tester des hypothèses, par exemple l'anonymat des contributions.

L'ambition va au-delà de soigner l'expérience utilisateur des plateformes civiques. Coéquipier du programme, César Hidalgo, chercheur à la Toulouse School of Economics, défend, depuis des années, l'idée de jumeau numérique du citoyen, une intelligence artificielle capable de représenter une personne en s'exprimant ou en votant à sa place. Cet avatar trouverait des usages divers, l'IA générative pouvant agréger et synthétiser des masses d'informations pour restituer l'avis de cibles spécifiques, de minorités difficiles à mobiliser. « *Si on veut l'avis d'un panel de citoyens sur 150 projets, ils n'auront le temps d'en discuter que cinq ou six. On pourrait « augmenter » leurs préférences pour couvrir les 150 projets* », explicite Umberto Grandi. En partant d'avis connus, mais qui ne représentent que 5 % des contributions attendues, l'IA puise dans ses propres données pour prédire l'avis collectif sur d'autres sujets. Techniquement, cette technologie n'est pas aboutie. Mais l'idée agite la planète digitale, à défaut d'enthousiasmer les acteurs de la participation citoyenne.

En matière de délibération, le projet s'intéresse aux outils d'aide à la médiation et à la synthèse. La politiste Maija Setälä (Université de Turku, Finlande) a inspiré des objets de recherche. Par exemple : l'IA peut-elle favoriser la visibilité d'une convention citoyenne dans la population générale ? Peut-elle aider les participants à aligner leurs préférences ? Agit-elle sur la légitimité de l'assemblée citoyenne perçue à l'extérieur ? Le projet prévoit de réunir des panels délibératifs en présentiel et de s'introduire dans de futures conventions citoyennes comme terrains d'expérimentation.

V.U.

[1] Communs démocratiques est une initiative française lancée en mai 2024, associant Make.org, le CNRS, Sciences Po et Sorbonne université. C'est « *un programme mondial de recherche, d'expérimentation et de mise à disposition de solutions d'IA génératives open-source au service de la démocratie* ». Il s'appuie sur un conseil scientifique d'experts dont Yochai Benkler (Berkman Klein Center, États-Unis), Audrey Tang (ex-ministre des Affaires numériques de Taïwan), Hélène Landemore (Yale University, États-Unis), Karine Perset (OECD), Raja Chatila (Sorbonne Université, France), Michelle Barsa (Omidyar Network, États-Unis), Djamé Seddah (Inria, France), Constance de Leusse (Institut IA et Société, France). Communs démocratiques est l'un des lauréats de l'appel à projets Communs numériques pour l'IA générative financé par Bpifrance à hauteur de six millions d'euros. <https://newsroom.sciencespo.fr/>

[2] DeepMind (Google) développe la machine Habermas, une IA générative (LLM) entraînée pour favoriser le consensus en faisant apparaître les points d'accord au cours d'un débat. www.technologyreview.com/

[3] Le projet ADDI (Advancing digital democratic innovation) associe Umberto Grandi, Université Toulouse Capitole, Uille Endriss, Université d'Amsterdam (Pays-Bas), César Hidalgo, Toulouse School of Economics, Maija Setälä, Université de Turku (Finlande). Lauréat d'une bourse Synergy, le projet est financé à hauteur de 9,9 millions d'euros sur 72 mois par le Conseil européen de la recherche.

[4] Jairo F. Gudino, Umberto Grandi, César Hidalgo. *Large language models (LLMs) as agents for augmented democracy*. Royal Society, 2024. <https://royalsocietypublishing.org>



Jean-Philippe Cointet

Professeur de Sociologie au médialab de Sciences Po, Jean-Philippe Cointet travaille au développement de méthodes innovantes en sociologie computationnelle et particulièrement sur l'analyse de textes : publications scientifiques, posts, articles de presse, etc. Il dirige l'Institut libre des transformations numériques de Sciences Po.