

Pourquoi Sciences Po a eu raison d'interdire ChatGPT

Article pour AOC

Dominique Boullier

Professeur émérite de sociologie à Sciences Po

Le forçage du débat public sur l'IA par l'offre gratuite de conversation avec ChatGPT semble abaisser toute vigilance sur l'opération marketing ainsi en cours qui vise à promouvoir une adoption de fait au nom de l'Innovation indiscutable.

Sciences Po n'a pas besoin que l'on vole à son secours et je n'ai pas l'habitude de faire des plaidoyers pour défendre mon institution mais quand mon école prend une initiative justifiée comme l'interdiction de l'usage de ChatGPT par les élèves, je ne peux que m'en réjouir (sans avoir été pour quoi que ce soit dans cette décision) pour mieux débusquer les arguments des libertariens (qui s'ignorent parfois).

Chat GPT n'a rien d'une innovation radicale puisqu'elle prolonge tout ce que l'on sait sur les LLM (Large Language Models) et que ses applications ont été déjà développées chez Google (à partir de son architecture de transformeurs, LaMDA, GLaM, Pathways) et chez Meta (OPT-175B)....qui se sont abstenus de les mettre dans le public ([Meta a retiré Galactica](#) en trois jours en Novembre 2022). Voilà qui devrait nous intriguer. Certes, on peut dire qu'il s'agit seulement pour ces plateformes d'éviter des déboires réputationnels tant que l'on ne peut pas totalement contrôler la production de ces agents conversationnels fondés sur des algos dits transformeurs, toujours incapables de distinguer le vrai du faux par exemple. Mais pourtant, on ne va pas leur reprocher pour une fois de savoir se retenir, d'y réfléchir à deux fois, de différer, bref, tout ce qui construit du *désir* et qui permet de sortir de la dictature du *passage à l'acte*. Car c'est bien ce que fait Open AI, un passage à l'acte, en mettant son service à disposition du public pour prendre position de premier entrant, pour préempter l'avenir selon ses propres termes ou pour faire un coup réputationnel (très réussi). Or les firmes californiennes nous ont déjà fait systématiquement le coup du passage à l'acte, elles qui sont gouvernées, comme tout l'internet, par le *rough consensus* et le *running code*, devise de John Perry Barlow dans son manifeste pour l'indépendance du cyberspace (1996). Eh bien il faut le dire : non ! le cyberspace n'est en rien indépendant, et ne doit pas le devenir, même si les années 2010 du Far West des plateformes (de réseaux sociaux notamment) ont pu le faire croire. Non, la technique et l'innovation ne peuvent être indépendantes du droit et de la politique, elles doivent même être désormais réencastrées systématiquement dans nos normes sociales et discutées à ce titre. D'autant plus que leurs logiques ne sont pas en fait celles d'une supposée technique hors sol mais bien des choix techniques spécifiques au capitalisme financier numérique. Le passage à l'acte court-circuite toute étape de discussion politique, réglementaire ou seulement sociétale pour mieux attirer les investisseurs, disons-le clairement. Et la mutation en 2019 de Open AI, fondation devenue société capitaliste (« à but lucratif plafonné » avec un fort investissement de Microsoft), n'y est pas pour rien. La toute-puissance des développeurs qu'on appelle la *classe vectorialiste* (McKenzie Wark) est massivement soutenue et motivée par l'intérêt des investisseurs qui, suivant une dynamique spéculative qui gouverne toute cette économie, parient surtout sur ces effets d'anticipation et de création d'attentes. Bataille financière entre plateformes de l'IA que l'on semble vite oublier tant le

public s'est trouvé pris dans les phares de la voiture qu'est l'utilisation publique et gratuite d'une version d'agent conversationnel. Ce hobby d'un instant transforme cependant ce public en testeur et fournisseur bénévole de données pour l'apprentissage de ChatGPT et en agent marketing, VRP de Open AI, grâce au buzz ainsi généré. Ce buzz va des envolées philosophiques et visionnaires aux exemples les plus raffinés de mise à l'épreuve de l'algorithme (une des meilleures expériences étant celle d'Olivier Auber racontée sur son compte Facebook, où il parvient quasiment à faire émerger un inconscient de l'IA en lien avec ses concepteurs). Les agendas du débat public sur les choix socio-techniques semblent donc fixés par les plateformes, récemment Meta avec le métavers, désormais l'IA avec Chat-GPT, meilleurs moyens de mal poser les problèmes et d'installer une forme de servitude volontaire encouragée par tous les commentateurs qui doivent apparaître toujours favorables à l'innovation, sans dire qu'il existe toujours plusieurs versions et plusieurs solutions (dont l'abstinence !).

Si cette technique suscite tant de débats, c'est parce que la stratégie commerciale d'Open AI adopte la stratégie du choc (Naomi Klein), qui sidère et oblige à choisir son camp, en paralysant un vrai débat sur ce qu'on peut prendre et ce qu'on peut laisser dans de telles innovations au nom du bien commun et après discussion dans des arènes conçues pour cela. Dès lors, pour une bonne partie des débatteurs, « il est interdit d'interdire », et c'est le seul point de départ, le cadrage incontournable. Il ne s'agit en rien d'un retour de Mai 68 ou des hippies, mais d'une dérive des libertariens qui gouvernent les instances d'internet en général et de la plupart des firmes vers la version profitable du slogan, devenu machine de guerre contre les états et toute régulation. Si l'on part du principe qu'il est interdit d'interdire, on accepte de laisser les commandes de nos avenir à des firmes auto-référentes avec les conséquences que l'on a pourtant pu observer dans les réseaux sociaux. Le libertarianisme se confond ici avec le néo-libéralisme économique et les esprits ouverts qui prétendent qu'il sera toujours temps de réguler plus tard ont semble-t-il perdu une élémentaire capacité de vigilance que l'on trouve chez Google et chez Meta pour ce genre d'IA (ce qui est plutôt nouveau pour ces firmes, c'est vrai). D'autres, parmi ces « suprémacistes de la technique », comme les appelle Tariq Krim, sont encore plus cyniques et nous annoncent même qu'il ne faut pas interdire ni même réguler car ça ne marchera pas et qu'il est déjà trop tard. Argument qui devrait au contraire plaider pour une « régulation by design » dans la conception même, contre les boîtes noires que sont ces LLM (175 milliards de paramètres pour ChatGPT et 540 milliards pour Pathways), du point de vue même de ceux qui les programment, dont certains cherchent d'ailleurs à reprendre la main en plaidant pour un Machine Learning interprétable.

La démocratie technique, c'est pour quand ?

La crise écologique est pourtant là pour nous démontrer que le développement technique sans force de rappel pour compenser le délire du « progrès » se paye cher : les effets du tout automobile des années 60 auraient dû nous faire réfléchir à la nécessité d'anticiper les conséquences de nos décisions (ce que les firmes pétrolières savaient déjà très bien !), dans une démarche de modernisation réflexive comme la demandait Ulrich Beck. La sobriété n'est pas un slogan pour calculer nos économies d'énergie, c'est un principe éthique qui doit s'appliquer à TOUS nos choix, y compris numériques. Pourquoi est-ce si difficile de ralentir ces choix pour prendre le temps d'en discuter les conséquences ? Les gains immédiats de ChatGPT sont à peu près nuls sur le plan social, on est en phase de tests de services adaptés à ces capacités. Alors prenons le temps de lancer des enquêtes au cœur des algorithmes, de tester leurs biais, de vérifier les conséquences de chacune des applications possibles, de débattre démocratiquement de toutes les dimensions. Où est l'urgence si ce n'est celle de laisser s'établir un état de fait, une domination commerciale d'une firme, d'opérer un forçage de l'opinion qui rend les réticences ringardes au nom d'une *tyrannie du retard* que j'avais

déjà identifiée dans les années 80 pour le numérique. Bref, ces partisans du laisser-faire sont prêts à se transformer de fait en promoteurs d'un choix dicté par les intérêts d'une firme mais imaginent qu'on pourra toujours reprendre la main par de l'éducation ou par des régulations ad hoc. Mais transposons cela à la reconnaissance faciale et l'on se retrouve avec la même paralysie des états pris en otage par la classe vectorialiste, les firmes californiennes et leurs serviteurs libéraux qui gouvernent dans la plupart des pays. Et nous aurons une version (apparemment soft) du modèle chinois de surveillance totalitaire. Dès lors, tant que tous les garde-fous juridiques et techniques n'ont pas été installés, il faut interdire la reconnaissance faciale et ne pas se laisser embarquer dans l'urgence sécuritaire, si olympique soit-elle. Qui empêche d'interdire ces techniques ? Surtout lorsqu'on connaît leurs usages en Chine ou quand on sait leurs faiblesses techniques ? Et mieux même, seraient-elles parfaitement contrôlées juridiquement et indétournables politiquement, pourquoi serions-nous sommés de les mettre en œuvre ? Parce qu'il existe d'autres priorités sociales et d'investissement par exemple ? La même politique d'interdiction devrait s'appliquer aussi au Bitcoin (et non à toute la blockchain), qui est une catastrophe écologique (toute la version *proof of work* de la blockchain), une expansion infinie du domaine de la spéculation et un encouragement à toutes les opérations financières opaques, qui mènent à la faillite quantité d'investisseurs. Oui, on peut aussi interdire le Bitcoin et, face à cette proposition, j'ai déjà entendu les soupirs effondrés des ayatollahs du paradis de l'innovation pour l'innovation, de la pulsion startupiste du web 3.0, etc. Qui se plaindrait de cette interdiction si ce n'est une petite communauté de profiteurs puisque l'architecture supposée distribuée de cette blockchain s'est transformée en monopole de quelques fermes de minage. Alors que des exigences d'utilité sociale, de responsabilité environnementale et de conformité légale peuvent devenir des moteurs puissants d'innovation (et il existe plusieurs blockchains qui tentent de s'y conformer), si on sort du passage à l'acte qui n'est bien souvent qu'imitation.

Nous devons prendre soin de notre espace mental.

Imaginons un tel forçage pour d'autres technologies : une ligne TGV qui passe derrière chez vous, sans discussion, une usine chimique qui dégage des odeurs insupportables auprès de votre domicile, une éolienne que l'on implante à 100 mètres de votre jardin, tout cela sans débat, sans procédure contradictoire, sans enquête publique, sans possibilité d'amendement ou de rejet, parce qu'on assure le public que tout cela sera utile au bien commun (en fait à la « compétitivité » !) et en tous cas très innovant... Cette enveloppe que nous veillons à tisser autour de notre espace physique et personnel devrait aussi être notre souci lorsque la technique altère notre espace mental, pour produire ce que j'appelle une [habîtèle, pour rendre le numérique habitable](#). Le numérique semble invisible et pourtant il pénètre partout, d'autant plus avec tous les ressorts de captation de l'attention qui nous rend ouverts à toutes les propagations (ce que je traite dans mon prochain livre [« Propagations. Un nouveau paradigme pour les sciences sociales »](#), à paraître fin Février 2023 chez Armand Colin). Et nous cédonc notre pouvoir de contrôle, de décision, d'arbitrage par peur de manquer le train lancé à toute allure de ce qu'on appelait le Progrès. On sait désormais qu'il faut réguler voire interdire les pesticides ou les organismes génétiquement modifiés, pourquoi accepterait-on de nous trouver transformés sans discussion en Organismes Numériquement Modifiés ? Evidemment, il est aisé de qualifier de luddite cette exigence de ralentissement, de débat et de régulation ! Mais ce sont les attitudes de ces plateformes fondées sur la prédation de nos espaces mentaux et de nos capacités de décision qui imposent ces demandes de moratoire. Ces débats sur des choix aux lourdes conséquences anthropologiques devraient être le B-A BA de la démocratie dialogique comme le proposaient Callon, Lascoumes et Barthe (2000), démocratie totalement inintéressante pour les libertariens puisque procédurale et donc bureaucratique alors que le code ne doit pas cesser de courir...

Pour autant, cela ne veut en aucun cas dire qu'il ne faut pas s'y intéresser, étudier ces technologies et en faire aussi un objet de pédagogie. Mon enseignement à Sciences Po a toujours plaidé au contraire pour une obligation d'initiation technique sur les types de Machine Learning (et non d'IA), sur les types de blockchain, sur les choix possibles d'architectures de réseaux (et non pour subir une supposée fatalité de la 5G puis de la 6G), etc. Oui, il faut comprendre de l'intérieur ces dispositifs pour rendre les citoyens et les décideurs capables de contrôler les choix et de s'autocontrôler, sans fascination ni rejet. Mais pour cela, il faut du temps et non la dictature de l'urgence technique ou marchande. C'est possible de le faire depuis les petites classes, comme je l'ai fait dans le canton de Vaud en produisant avec tous les pédagogues de [Learn à l'EPFL](#) et les enseignants un cursus adapté à chaque niveau de formation à l'algorithmie associée immédiatement à une compréhension des enjeux sociaux, économiques, environnementaux et éthiques du numérique. En restituant toujours le pluralisme des politiques et des choix personnels.

Et nos experts qui s'alarment de l'interdiction faite par Sciences Po devraient consacrer leur énergie à nous montrer toutes les autres façons de faire des agents conversationnels, interprétables, contrôlables, avec des applications utiles et non seulement marchandes ou promotionnelles. En discutant toutes les variantes, on peut ainsi préparer les esprits à un véritable débat public européen puisque l'Europe partage des fondamentaux culturels et juridiques qui la distinguent radicalement des libertariens californiens comme du totalitarisme chinois. Car après tout, à Sciences Po comme dans toutes les écoles, il vaut mieux faire savoir que tricher n'est pas acceptable juridiquement ni éthiquement et que cela ne procure que des bénéfices très provisoires. La valeur des apprentissages personnalisés n'empêche pas des assistances selon les ressources de l'époque (un dictionnaire, une calculatrice, etc.) mais tout cela se prépare, se contrôle et surtout s'inscrit dans une stratégie éducative explicite et justifiée. Stratégie qui peut évoluer évidemment avec l'apparition de ces dispositifs mais qui ne DOIT pas s'y plier comme on le croit parfois lorsque les innovations pédagogiques sautent sur le premier gadget populaire (sinon nous fermerions les universités au profit de Tik Tok). Là encore, du temps est nécessaire, un cahier des charges doit être fixé à toutes les firmes qui veulent occuper le marché européen au moins (c'est ce qui se passe avec la préparation de l'IA Act européen), un dispositif de validation et de contrôle doit être mis en place et non une offre de détection automatique de textes générés par ChatGPT commercialisée par la même firme (sur le mode des firmes de sécurité vendeuses d'anti-virus couplées avec les pirates propagateurs des virus).

Comme la fascination publique le manifeste, c'est un enjeu anthropologique majeur qui se joue avec l'IA et non seulement un nouveau bouton sur l'interface du smartphone (même si je peux montrer que certains boutons ont entraîné des modifications profondes de nos façons d'être en société!). On peut en profiter pour montrer les limites de ces IA et sortir des fantasmes, pour montrer les enjeux socio-politiques et économiques majeurs à l'échelle internationale, pour exiger une IA interprétable. Mais pour cela, il faut d'abord refuser le cadrage de la controverse par une promotion sauvage d'un outil mal ficelé, attracteur de public conçu pour produire une adhésion de fait, boîte noire incontrôlable pour ses propres créateurs.