

IA générative : une bifurcation radicale à reprendre en main par les sciences sociales

Dominique Boullier

Avril 2024

Est-ce que le monde de la recherche en SHS vous semble avoir pris la mesure de ce que l'irruption des IA va changer ?

L'IA n'a fait « irruption » que dans le monde des médias et de l'opinion publique mise en forme par les médias et par l'opération commerciale de Open IA, à savoir ChatGPT. Car cela fait des années que d'autres IA que l'IA générative (les IA symbolique et connexionniste) sont suivies, étudiées à divers titres (en linguistique, en sociologie des organisations ou des techniques), mais aussi mises en œuvre pour tout le traitement du langage notamment, et de façon plus réservée pour le traitement de données (même si certains classent l'utilisation d'une régression non linéaire dans le *machine learning*). Les sciences des données ont tendance à prendre la place vacante pour parler des thèmes des sciences sociales à notre place, ce qui n'est guère agréable, mais oblige à réagir en effet. Pour ce qui est de l'IA générative, les SHS n'ont pas encore pris la mesure de ce qu'elle peut changer parce qu'il faut du temps (et non des professions de foi ou des conversions ou aversions brutales et schématiques), qu'il faut les observer en situation, les tester, établir leurs limites et leurs atouts, avec cette contrainte particulière que les offres changent tous les 6 mois, voire continûment (ChatGPT 4 a des performances très supérieures à celles de ChatGPT 3.5 et l'offre de commande vocale est désormais disponible par exemple). Cela rend l'étalonnage compliqué, tant pour la recherche que pour la formation, car on ne peut se jeter sur le principe de la formation aux *prompts* sans avoir réfléchi un peu à ses conséquences et à son obsolescence. À cela, il faut ajouter que le mode de diffusion de l'innovation ChatGPT est beaucoup plus diffus et horizontal que d'autres innovations numériques dans les entreprises. Un ERP demandait une décision centrale, une vision à long terme et une longue *customization* et formation des personnels qui avaient rarement leur mot à dire (ce qui a conduit à 50 % d'échecs d'ailleurs). Mais avec ChatGPT, ce sont les employés, à des degrés de responsabilité très divers, qui ont commencé à tester l'offre gratuite et à imaginer des solutions pour leur propre poste de travail, sans le dire ouvertement à leur hiérarchie, car avouer qu'une partie du *reporting* que l'on fournit à son N+1 est désormais faite à 80 % par ChatGPT, ce n'est pas simple à défendre, notamment pour sa propre sécurité de l'emploi. Mais du point de vue des enquêtes sociologiques, c'est plus compliqué, car toute une partie du travail réel (et non de celui qui est annoncé dans les *workflows* et les organigrammes) se fait sous le manteau ou tout au moins avec un bricolage parfois inconnu de la hiérarchie. Il est donc impossible de savoir à quoi on peut accéder et quelle porte d'entrée il faut privilégier. En revanche, pour les chercheurs qui sont restés connectés aux entreprises et qui y interviennent, la demande est énorme de la part des directions et des départements, car c'est un peu la panique à bord et l'on voit que les sciences sociales sont sollicitées pour trouver des stratégies face à cette propagation d'une innovation hors des procédures habituelles et avec des conséquences totalement incertaines pour de nombreux métiers. Je dirais même que les

sciences sociales ont un *devoir* de mobilisation pour s'assurer que les collectifs de travail ou de citoyens ou les institutions reprennent la main sur cette diffusion virale de l'IA. C'est le moment où l'on peut encore « réencastrier » l'IA dans les organisations et non pas s'obliger à s'adapter à tout ce qui bouge et qui va bouger fréquemment dans les années à venir. Seules des approches familières des enjeux organisationnels peuvent aider à piloter cette « irruption » pour éviter des conséquences douloureuses.

Dans votre discipline, comment évaluez-vous la mobilisation des recherches sur ce sujet ? Quelles sont les hypothèses travaillées et les enjeux pour mieux comprendre ce phénomène de l'IA et ses conséquences sur les disciplines en SHS en général et dans la vôtre en particulier ?

Si l'on traite de l'IA en général et non de l'IA générative spécifiquement qui ne peut faire encore l'objet de recherches avancées, mais seulement de premières observations, l'axe de sociologie des algorithmes et plus généralement de sciences sociales des algorithmes (économie, histoire, SIC – sciences de l'information et de la communication –, etc.) paraît le plus original et le plus productif. Et cela en France même. Les chercheurs de SHS n'hésitent plus à entrer dans le bain technique pour suivre de près le travail des *data scientists* et pour observer comment se déroule l'implémentation de certaines IA dans certains contextes particuliers (police, justice, recommandations – commerciales ou non –, applications urbaines, médecine, etc.). Cette évolution est notoire par rapport aux études des usages de ce qu'on appelait les NTIC (nouvelles technologies de l'information et de la communication) auparavant où trois approches dominaient : l'analyse des usages, l'analyse des discours d'accompagnement et l'analyse des politiques publiques. Ces trois dimensions sont encore présentes et nécessaires (très peu les usages de l'IA finalement, car elle est extrêmement pervasive). Mais désormais, on entre dans la façon dont sont fabriquées les IA, dont elles sont paramétrées, dont les biais sont encapsulés, dont les courants de l'IA se sont concurrencés historiquement, dont les données sont prélevées, légalement ou non, dont les usagers eux-mêmes exploitent désormais ces données pour renverser l'asymétrie qui semble constitutive de cette puissance de ce qu'on appelle « la classe vectorialiste » (McKenzie Wark), celle qui maîtrise le code.

De plus, tout ce qui est fourni comme traces par les réseaux sociaux notamment (et donc par des IA) constitue une ressource exceptionnelle soit pour explorer leur dynamique, soit pour valider des hypothèses, soit même pour faire un nouveau type de sciences sociales quasi expérimentales qui s'inscrivent, comme je le propose, dans l'étude des propagations¹. Pour cela, il faut souvent soit une double compétence, soit une association avec des *data scientists* et souvent aussi un partenariat avec les détenteurs de ces données, parfois ouvertes, mais finalement de moins en moins. Ce serait pourtant une occasion pour les SHS en général de mobiliser certains types d'IA pour étudier des processus sociaux avec une robustesse quantitative inédite. Mais le coût d'entrée en matière de compétence en méthodes quantitatives reste élevé, ce qui ne peut qu'encourager les plateformes et les *data scientists* à occuper le terrain et à devenir les seuls capables de produire des analyses et même des prédictions (!) qui attireront l'attention (et les crédits) des médias, des institutions, et des entreprises. La tendance que l'on pouvait observer dans certaines disciplines où l'expertise en traitement de données et manipulation de dispositifs ou de modèles prenait toute la place dans la hiérarchie des recherches (en géographie avec les SIG – systèmes d'information géographique – ou en linguistique avec le TAL – traitement automatique des langues), cette tendance risque de devenir plus problématique, car les niveaux de compétence exigés pour piloter les IA vont

¹ Voir mon livre, Boullier Dominique, *Propagations. Un nouveau paradigme pour les sciences sociales*, Paris, Armand Colin, 2023.

devenir très difficiles à atteindre. On risque de voir encore apparaître une hiérarchie, mais en faveur des experts de *data sciences* qui ne sont pas issus des disciplines et qui peuvent coopérer avec les chercheurs de certaines disciplines. Or, cette division du travail, pour l'avoir expérimentée dans les meilleures écoles d'ingénieurs, n'est pas du tout simple à mettre en place et exige des décentrement réciproques. En sociologie, même quand on entre dans le suivi de la fabrication des algorithmes d'IA eux-mêmes, on est aisément décroché et cela devient un problème lorsque l'IA n'est pas seulement l'objet de la recherche, mais aussi son outil.

Quels sont les aspects positifs et négatifs de ces IA selon vous pour la recherche en général et/ou pour les chercheurs en particulier ?

Le problème majeur tient à l'opacité générale de tout le processus de traitement des données avec les IA connexionnistes et génératives. La collecte elle-même se restreint de plus en plus pour les chercheurs ou pour des tiers. Alors même que l'on pensait qu'un mouvement *open data* s'était enclenché, il s'est en réalité appliqué aux services publics, mais pas du tout à ces plateformes et grands acteurs d'intérêt général. Instituer l'obligation de faciliter l'accès aux données des plateformes pour les chercheurs devient un impératif de politique publique. En effet, ce sont toutes les activités humaines en société qui se trouvent désormais tracées et exploitées et qui servent à l'augmentation de l'expertise des plateformes. C'est ce que Shoshana Zuboff appelle la « *division of learning* » qui remplace la division du travail. Or cette réflexivité possible grâce à l'accès aux données devrait revenir à toutes les composantes de la société, et notamment aux chercheurs de la recherche publique. Sinon, la connaissance du social, mais aussi son gouvernement vont glisser vers une dépendance accrue aux plateformes ou aux sociétés expertes en IA.

L'opacité affecte encore plus la possibilité de vérifier la chaîne de traitement (les pondérations et leurs révisions), même pour les développeurs de ces algorithmes. C'est dire que l'on a confié désormais la plus grande partie de nos calculs, dans tous les éléments de nos existences et dans la recherche, à des systèmes qui gagnent en autonomie sans prévoir comment ouvrir les boîtes noires, comment les tester et comment contrôler ce que l'on appelle leur alignement (entre leurs façons de calculer et les principes, valeurs et visées de ceux qui les ont commandés). L'explicitation qui était la grande contrainte de l'informatique pour rendre tout calculable pouvait sembler abusive, mais le fait d'explicitier obligeait – en principe ! – à discuter (« *Calcuemus !* » disait Leibniz, scientifiquement et politiquement). Mais cela ne semble plus du tout une priorité, tout au moins pour les grandes plateformes qui exploitent les IA les plus puissantes. Si l'on n'inscrit pas l'explicitabilité dans le cahier des charges de toute IA mise sur le marché, c'est-à-dire l'obligation de rendre compte des classifications, des pondérations effectuées par les IA dans le cours même de leur entraînement puis de leur mise en place, on peut mettre la rationalité et la discussion scientifique aux orties, car elle ne sera plus possible. Et cela posera d'énormes problèmes juridiques, car le droit aussi exige de l'explicitation, et en cela il est parent de la science dans l'avènement des démocraties. Si les projets de voitures autonomes réduisent actuellement leurs ambitions les uns après les autres, c'est qu'il faut pouvoir distribuer les responsabilités juridiques et assurantielles en cas de problème. Or, le véhicule autonome est fait d'un grand nombre de composants, mais c'est encore plus vrai de l'environnement urbain qui atteint un degré d'incertitude extrêmement élevé et qui ne peut être traité par les IA si elles doivent intégrer la gestion des risques. Mais pour exiger cette explicitation dans tous les processus d'IA, cette auditableté, et cette responsabilité *in fine* de tous ceux qui ont contribué à paramétrer les IA (dont l'utilisateur !), il faut une vraie volonté politique et non un culte du business et de la disruption comme nous les imposent les libertariens californiens et tous ceux qui les suivent en Europe notamment.

Les IA constituent-elles un nouvel outil de travail, comme un autre (mais peut-être plus performant ?), ou un concurrent au chercheur ? Amènent-elles à redéfinir ce que sera demain l'activité scientifique ou n'est-ce qu'un outil au service d'une meilleure rapidité des analyses ?

L'exigence de répliquabilité des recherches qui était au cœur du processus de validation a déjà été quasiment abandonnée dans les sciences expérimentales, et c'est une catastrophe, au nom du « *publish or perish* ». Dans les SHS, on n'a quasiment jamais mis en place ce principe de répliquabilité puisqu'elles ne sont pas expérimentales, mais ont seulement un impératif de comparaison. Or, je prétends que précisément avec les capacités de calcul et de modélisation disponibles avec le *machine learning*, nous avons une chance de commencer à tester, à valider des énoncés sur des situations quasi équivalentes, pour vérifier certaines hypothèses. L'IA générative, l'absence d'éthique de ses vendeurs et son obligation de réponse à tout au prix d'hallucinations, sabotent cette possibilité. Un *data set* d'apprentissage évolue constamment et sans contrôle et il est donc impossible de répliquer des expériences. Cela permet d'annoncer des performances, des résultats, extrêmement précaires et éphémères, et donc si nécessaire de s'autoriser à laisser passer des hallucinations qui ne seraient rien d'autre que des hypothèses que l'on n'a même pas cherché à valider. Voire des résultats sous forme de corrélations qui n'avaient même pas fait l'objet d'hypothèses, puisqu'elles deviennent soi-disant inutiles dans un monde de l'induction généralisée capable de tester tout type de corrélations pour voir émerger des réponses significatives ! Le triomphe du *fake* est à l'agenda politique explicite des libertariens et est encapsulé dans les méthodes probabilistes opaques que l'on nous vend comme solution à toute question de recherche.

Alors que l'on pourrait pourtant explorer et tester des hypothèses de façon extraordinairement ouverte et de façon fiable si l'on y plaçait des contraintes élémentaires issues de l'éthique scientifique. Le délire du « *publish or perish* » était là bien avant l'IA générative, il a corrompu toute l'activité des chercheurs, mais l'IA générative va l'amplifier à un point tel que tout résultat scientifique sera désormais aussi suspect qu'une prédiction de Nostradamus, car les IA génératives produiront les articles, qui seront révisés par des IA génératives, lus par des publics fabriqués par des IA génératives qui produiront des classements de réputation produits par des IA génératives. Ce système que j'appelle auto-répliquant est déjà en place dans les médias sociaux et gagnera inévitablement les autres systèmes de publications déjà intoxiqués au *ranking* et à la réputation.

Ces IA vous semblent-elles propices à l'innovation, vecteurs d'innovation ? Permettront-elles aux chercheurs de demain de découvrir autrement ? Laissent-elles une place à l'invention ?

Oui, mais d'autres IA, pas celles mises sur le marché actuellement, des IA avec d'autres exigences et d'autres principes éthiques : *open source*, avec autorisation de mise sur le marché, explicabilité, limitation de ressources, etc. Dans ce domaine, beaucoup d'innovations restent à faire et elles sont en effet très prometteuses, à condition d'être sorties des logiques de compétition financière que les firmes qui les possèdent nous imposent, au grand public comme aux chercheurs.