L’urgence d’exploiter les technologies émergentes

Entretien avec James Guszcza

par Jules Naudet

L’existence d’un panopticon numérique aux mains d’un petit nombre d’entreprises n’ayant aucun compte à rendre démocratiquement menace la dignité humaine et à l’autodétermination. Les risques et les avantages des nouvelles technologies doivent être maîtrisés par des technologies sociales adéquates.

Jim Guszcza est docteur en philosophie des sciences de l'université de Chicago ; il a enseigné à l'école de commerce de l'université du Wisconsin-Madison. Il a travaillé comme *data scientist* pendant deux décennies et est actuellement le *chief data scientist* de Deloitte Consulting, et membre de la pratique Advanced Analytics and Modeling de Deloitte. La création de systèmes hybrides homme-machine est un thème récurrent de son travail. Au cours des dernières années, il a appliqué des techniques d'incitation comportementale pour rendre les algorithmes d'apprentissage automatique plus éthiques et plus efficaces. Jim Guszcza siège au conseil consultatif scientifique de l'Institut de psychologie de la technologie.

La Vie des idées **: Le flux continu d'innovations technologiques dans le sillage de la révolution Internet a progressivement transformé la façon dont nous**

naviguons dans le monde d'aujourd'hui. De l'information circulant à grande vitesse à la surabondance de contenu, des *cookies* à la surveillance perpétuelle des comportements, de la banque en ligne aux *bitcoins*, du télétravail aux perspectives d'un monde de réalité virtuelle omniprésent, il semble que les cadres et les structures du monde dans lequel nous vivons aujourd'hui subissent des transformations radicales. Comment caractériseriez-vous ce moment précis de l'histoire dans lequel nous nous trouvons ?

Jim Guszcza : Notre époque est marquée par la technologie, de plus en plus d'aspects de notre vie sont médiatisés par le numérique. Je pense à ce moment historique en termes de forces déséquilibrées. Nos technologies numériques ont progressé à un rythme extraordinaire. Mais les « technologies sociales » – lois, politiques, dispositions institutionnelles, modèles commerciaux, normes sociales, stratégies éducatives, philosophies de conception – nécessaires pour les façonner et les limiter n'ont pas suivi le même rythme.

Cela est peut-être dû, du moins en partie, aux idéologies dominantes entourant les technologies numériques – techno-optimisme et déterminisme technologique – qui ont occulté la nécessité de les exploiter correctement par l'innovation sociale. Ces idéologies ont rendu difficile pour les décideurs, les technologues et le public de raisonner avec soin sur les risques, les avantages et le rôle des technologies émergentes dans la société. Elles ont en outre favorisé un sentiment collectif de complaisance.

Le techno-optimisme nous a invités à identifier le progrès technologique au progrès sociétal. Des titres comme « technologie exponentielle » et des slogans comme « l'information gratuite », « les DATA sont le nouveau pétrole » et « l'IA est la nouvelle électricité » sont tous porteurs d'idées utiles. Mais lorsqu'ils sont hissés au rang de mantras, ils encouragent le fantasme selon lequel les résultats technologiques produisent automatiquement des résultats désirables. En réalité, concevoir les bonnes pratiques sociales autour des technologies n'est généralement pas moins difficile que de développer les technologies elles-mêmes.

Nous avons négligé cette vérité à nos risques et périls. Jusqu'à une période récente, il était largement admis que le fait de connecter numériquement les gens, de leur donner accès à des informations illimitées et de créer de nouvelles formes d' « intelligence », grâce au pouvoir combiné du big data et de l'apprentissage automatique, aboutirait naturellement à des sociétés mieux connectées, mieux informées, plus intelligentes et plus productives. Sans doute les avantages de ces technologies sont-ils réels et substantiels. Mais leur développement anarchique a

conduit à des phénomènes qui menacent à la fois le bien-être individuel et le fonctionnement des sociétés démocratiques : addiction, dépression, polarisation, désinformation généralisée et accroissement des inégalités économiques.

L'idéologie techno-optimiste s'est mêlée à un esprit techno-déterministe qui considère la technologie comme une force quasi-autonome à laquelle les sociétés doivent s'adapter. Cet état d'esprit a prévalu jusqu'à très récemment, comme en témoignent les titres de livres tels que *The Singularity is Near*, *The Inevitable* et *Rise of the Robots*, ainsi que des slogans tels que « Les logiciels dévorent le monde », « Hâtez-vous de tout casser » et « Vous n'avez aucune vie privée ? Ce n'est pas grave ». Concevoir la technologie comme une force inévitable et autonome court-circuite la pensée critique et occulte la nécessité de concevoir les technologies – ainsi que les dispositions juridiques et institutionnelles qui les entourent – de manière à les conformer aux besoins et aux valeurs des individus.

Il ne s'agit pas d'un défi purement technique, mais plutôt d'un défi socio-technique. Et contrairement aux technologies développées en laboratoire, les systèmes socio-techniques sont intrinsèquement complexes. L'introduction de nouvelles technologies majeures peut donc avoir des conséquences imprévisibles. Nous devons concevoir nos systèmes sociotechniques en faisant preuve d'humilité intellectuelle et en reconnaissant la nécessité de tester, d'apprendre, d'innover et d'améliorer de manière progressive et itérative. Aller vite et casser les choses est une bonne façon d'innover dans des environnements de laboratoire contrôlés où les enjeux sont faibles, mais elle est perverse dans le contexte de la complexité sociotechnique. et des sociétés.

Reconsidéré sous cet angle, le mantra « l'IA est la nouvelle électricité » a une signification plus subtile. Pour exploiter l'électricité d'une manière qui a profité aux sociétés, il a fallu entourer la technologie de base de disciplines telles que la conception centrée sur l'homme, l'ingénierie, l'inspection de la sécurité, les politiques publiques et les systèmes réglementaires. Il en va de même pour les technologies fondées sur le big data, l'apprentissage automatique et les infrastructures numériques. Plutôt que de laisser ces forces se déchaîner, nous devons les exploiter de manière socialement bénéfique. Pour ce faire, il faut laisser tomber les idéologies et les mantras simplistes, et cultiver des stratégies de conception, de développement et de déploiement qui s'alignent sur les besoins et les valeurs humaines.

La vie des idées **: L'anthropologie structurelle a classiquement formulé l'hypothèse d'une homologie ou d'une correspondance entre, d'un côté, le monde physique construit dans lequel nous vivons et, de l'autre, la disposition des groupes sociaux et les « formes de classification » à travers lesquelles nous nous percevons et percevons le monde. Iriez-vous jusqu'à étendre cette analogie à la conception architecturale de nos structures numériques ? Dans quelle mesure diriez-vous que les systèmes informatiques, l'internet, les médias sociaux, les smartphones, etc. transforment la manière dont nous donnons un sens au monde dans lequel nous vivons et la manière dont nous essayons d'agir dans ce monde ?**

Jim Guszcza : Une petite anecdote personnelle peut illustrer le pouvoir qu'ont nos prothèses numériques de modifier la façon dont nous percevons, comprenons et nous comportons dans le monde. À Atlanta, lors d'un voyage d'affaires, j'ai hélé un taxi pour me rendre de Downtown à un restaurant dans le quartier de Midtown. Cela n'impliquait rien de plus qu'un simple plongeon dans une grande artère. Mais le chauffeur – qui était un habitant d'Atlanta – s'est fourvoyé dans un itinéraire sinueux et détourné. La raison : il suivait les instructions d'un appareil GPS dont les signaux étaient brouillés par les gratte-ciels environnants. Cette anecdote ne doit pas être interprétée comme une critique des appareils GPS. Au contraire, ces dispositifs peuvent être considérés comme un paradigme d'intelligence artificielle centrée sur l'homme. Ils exploitent des quantités massives d'informations de manière à améliorer l'autonomie humaine en nous aidant à surmonter nos limites cognitives. Au sens propre, ils rendent notre monde plus navigable. Néanmoins, dans ce cas, une bizarrerie de la psychologie humaine a interagi avec la technologie d'une manière qui a donné lieu à une « stupidité artificielle » plutôt qu'à une « intelligence artificielle ».

Une recherche sur le Web de l'expression « Death by GPS » (mort par GPS) révèle que cette histoire n'est pas un cas isolé, mais plutôt un exemple d'un phénomène répandu : les gens ont tendance à faire fi de leur bon sens, de leurs connaissances de base et de leur conscience de la situation au profit des indications faillibles du GPS, même dans des environnements éloignés et dangereux. L'une des maximes de base de la conception humano-centrée est que, si les erreurs ponctuelles peuvent être imputées à l'erreur humaine, les erreurs répétées doivent être imputées à une conception médiocre (ou complètement négligée).

Des arguments analogues peuvent être avancés à propos de la propagation de la désinformation dans les environnements en ligne, du malaise que produit la comparaison entre ce que nous pensons et ce que disent les amis et connaissances

d'après les médias sociaux, de la polarisation des groupes qui résulte du filtrage collaboratif des nouvelles et des contenus d'opinion, et des décideurs qui suppriment indûment leurs facultés de jugement scientifique et éthique au profit de résultats algorithmiques. Dans chaque cas, nous avons besoin d'intégrer la technologie dans un environnement décisionnel conçu pour s'adapter à la psychologie humaine.

Il est tout à fait pertinent de caractériser cette couche manquante – la conception d'environnements numériques centrés sur l'homme – comme une sorte d' « architecture ». Les racines grecques de ce mot nous aident à comprendre la couche manquante. Le génie logiciel et l'apprentissage automatique sont des types modernes de *technè* – « artisanat » ou « savoir-faire ». *Archè* signifie « chef » ou « maître ». Le maître d'œuvre – l'architecte – fournit les conceptions qui guident le développement technique. Tout comme il est peu probable que des villes vivables et humaines soient construites par des équipes qui ne connaissent que la science des matériaux et les principes de construction, il est peu probable que des environnements numériques vivables et humains soient créés par des équipes qui ne comprennent que le développement de logiciels et l'ingénierie de l'apprentissage automatique. Dans chaque cas, c'est la conception, ou « l'architecture », qui fait défaut.

Certes, les architectures de décision fondées sur des critères psychologiques ne sont pas totalement absentes de nos technologies numériques. Au contraire, les dark patterns (« interfaces truquées ») – des interfaces qui manipulent les utilisateurs de manière à entraver l'expression de leurs préférences et l'atteinte de leurs objectifs – sont aujourd'hui omniprésents dans les environnements en ligne. Les premiers "growth hackers" (« pirateurs de croissance ») de la Silicon Valley qui s'efforçaient de rendre addictifs les environnements numériques étaient imprégnés des théories de la « technologie de la persuasion ». Mais ces pathologies ne doivent pas nous amener à conclure que la prise en compte de la psychologie humaine dans la conception des technologies numériques est intrinsèquement dévoyée. Au contraire, la psychologie peut et doit être exploitée de manière éthique, de façon à renforcer (plutôt qu'à réduire) l'autonomie humaine. Pour y parvenir, il faudra plus qu'un appel candide à un comportement éthique de la part des grandes organisations. Les technologies numériques doivent être développées et déployées dans le cadre de dispositions réglementaires appropriées et de modèles commerciaux qui alignent les incitations économiques sur les besoins de la société. Se contenter de réclamer un développement éthique de la technologie a peu de chances d'être efficace. Le développement éthique doit être encouragé.

La Vie des idées **: La matérialité du « vieux » monde devient-elle obsolète en raison de nos nouvelles façons de vivre le monde ? Comment abordez-vous les craintes de ceux qui voient un danger dans le fait de passer au tout virtuel et de devenir étranger à la réalité ?**

Jim Guszcza : Les craintes de voir la réalité virtuelle éclipser la réalité physique vécue me rappellent les craintes de voir émerger la version « Singularité » de l'IA, appelée à dominer l'intelligence humaine et à rendre le travail humain obsolète. Dans chaque cas, je pense que la question doit être retournée : pourquoi devrions-nous croire que ce sont plus que de simples scénarios de science-fiction ?

Dans le cas de l'IA, les titres des journaux regorgent d'exemples d'algorithmes capables de réaliser des exploits surhumains : rechercher des modèles subtils dans des bases de données massives, prouver des théorèmes mathématiques, battre les champions du monde d'échecs et de go, peser les facteurs de risque de manière statistiquement optimale. Mais un titre tout aussi important est généralement passé sous silence : les tâches les plus faciles pour les humains – reconnaître des objets ou des voix, se déplacer dans l'espace, juger des motivations humaines, comprendre des récits simples – sont généralement les plus difficiles à mettre en œuvre sous forme de machine. Le philosophe Andy Clark a fait remarquer un jour que les humains sont « bons au frisbee, mauvais en logique ». Les algorithmes d'IA sont tout le contraire. Ils ont tendance à briller dans les tâches impliquant les facultés de raisonnement de la cognition humaine qui ont évolué le plus récemment. Mais leur avantage est nettement moindre sur les capacités perceptives et motrices qui ont évolué pendant des millions d'années. C'est l'une des principales raisons pour lesquelles il est plus judicieux de concevoir des technologies algorithmiques pour compléter – et non remplacer – la cognition humaine.

De même, il serait vraisemblablement très difficile de construire des technologies de réalité virtuelle capables de se substituer de manière robuste à la réalité physique dans laquelle nous avons évolué pendant des millions d'années. De même que l'IA est mieux conçue comme un moyen d'étendre – plutôt que d'imiter ou de remplacer – la cognition humaine, la RV est mieux conçue comme un moyen d'améliorer – et non de remplacer – notre expérience de la réalité.

La Vie des idées **: Comment vos recherches aident-elles à comprendre ou à gérer les conséquences de ces transformations ? Que nous apprennent-elles des effets que ces changements causent à notre vie quotidienne ?**

Jim Guszcza : Je co-dirige un projet au *Center for Advanced Study in the Behavioral Sciences* de l'Université de Stanford, financé par la Fondation Rockefeller, dont l'objectif est d'élaborer l'exigence d'une nouvelle pratique de l'IA. Ce nouveau domaine s'attaquerait directement à certains des défis évoqués ci-dessus. Il existe des méthodologies bien établies qui permettent aux ingénieurs en apprentissage automatique d'optimiser des fonctions objectives, pré-spécifiées sur des échantillons de données (souvent à l'échelle du web). Mais l'apprentissage automatique n'offre ni les outils scientifiques ni les ressources conceptuelles nécessaires pour évaluer quels objectifs nous devrions optimiser, ou comment construire des échantillons de données d'entraînement appropriés à la situation en question. De telles décisions nécessitent des jugements scientifiques et éthiques, éclairés par des connaissances spécifiques au contexte. Le fait qu'elles tendent à sortir du cadre de la pratique courante de l'IA est mis en évidence par les nombreuses histoires de partialité algorithmique de ces dernières années. On pourrait appeler cela le « problème du premier kilomètre » de l'apprentissage automatique : avant de former les algorithmes, nous devons d'abord construire un échantillon de données qui enregistre correctement le monde de manière scientifique et éthique.

Le domaine envisagé s'attaquerait également au « problème du dernier kilomètre » de l'apprentissage automatique : notre objectif n'est généralement pas simplement d'optimiser les résultats algorithmiques, mais plutôt les résultats dans le monde réel. Cela implique d'intégrer les algorithmes dans des environnements de décision et des flux de travail qui s'alignent bien sur des caractéristiques de la psychologie humaine telles que la cognition limitée, le contrôle de soi limité, l'intérêt personnel limité, l'éthique limitée et le rôle des émotions dans le raisonnement humain. L'apprentissage automatique nous permet d'optimiser les algorithmes, mais le véritable objectif est généralement d'optimiser les systèmes humains qui travaillent avec des technologies algorithmiques. Pour ce faire, il faut une base scientifique qui va au-delà de l'apprentissage automatique et englobe également l'éthique et les sciences du comportement.

Prenons l'exemple de la prise de décision médicale. Indépendamment des révolutions du *big data* et de l'apprentissage automatique, les psychologues savent depuis longtemps que même des algorithmes simples surpassent en performance le

jugement spontané des cliniciens experts pour un large éventail de décisions. Pourtant, les bons scientifiques des *data* savent aussi que les algorithmes ne peuvent pas distinguer les apparences (les données) de la réalité (les processus qui ont généré les données) ; et qu'ils sont dépourvus du type de jugement nécessaire pour évaluer si une indication est appropriée dans une situation spécifique. Par conséquent, ils ne peuvent pas remplacer les experts humains.

L'enjeu va donc au-delà de l'utilisation de l'apprentissage automatique pour optimiser les algorithmes. Il est nécessaire de concevoir des processus de collaboration homme-algorithme dans lesquels les forces et les limites relatives de l'un contrebalancent celles de l'autre. Contrairement au domaine bien établi de l'ingénierie de l'apprentissage automatique, la conception de tels processus relève aujourd'hui davantage de l'art que de la science. D'où la nécessité, formulée dans notre proposition, de créer un domaine de l'ingénierie qui soit celui de « l'intelligence hybride » homme-machine. Nous soutenons qu'un tel domaine ne peut être fondé sur les seules sciences informatiques et de l'information. Il doit également intégrer des principes tirés de l'éthique et des sciences comportementales. En outre, il doit incorporer des approches participatives afin de s'assurer que le savoir local, ainsi que les besoins et valeurs des multiples parties prenantes, soient tous intégrés dans la conception du système homme-machine.

***La Vie des Idées* : Le fait que les grandes entreprises technologiques et les États aient accès à une sorte de panopticon crée-t-il une réelle menace pour la démocratie ? Voyez-vous des moyens par lesquels ces nouvelles technologies pourraient plutôt donner du pouvoir aux citoyens et consolider la démocratie ?**

Jim Guszcza : La consolidation des données comportementales granulaires par les grandes entreprises technologiques constitue une menace pour la démocratie d'au moins deux façons : par leur taille même, et par leur capacité croissante à médiatiser le discours politique et à manipuler de manière algorithmique la diffusion de l'opinion, de l'information et de la désinformation.

En ce qui concerne le premier point, les entreprises numériques peuvent facilement passer à une échelle supérieure, grâce à leurs coûts fixes élevés et à leurs faibles coûts marginaux. En outre, nombre d'entre elles bénéficient également de puissants effets de réseau : plus elles attirent d'utilisateurs et de développeurs, plus elles deviennent précieuses pour les futurs utilisateurs et développeurs. Il en résulte

un cycle vertueux de croissance. Le big data permet aux entreprises numériques d'amplifier ces avantages inhérents en utilisant l'apprentissage automatique pour améliorer et personnaliser en permanence leurs services.

Une telle dynamique produit des organisations extrêmement riches et puissantes. L'existence de ces entreprises dominantes et indéboulonnables éloigne les sociétés de la démocratie et les oriente vers la ploutocratie. Leur poids économique leur confère un pouvoir excessif qui leur permet d'écraser ou de racheter des concurrents plus petits, d'opprimer leurs employés et les travailleurs indépendants, et de faire main basse sur la réglementation gouvernementale. En bref, la domination d'un petit nombre d'entreprises technologiques très puissantes éloigne les sociétés de l'idéal démocratique pour les rapprocher de la ploutocratie.

Outre les problèmes liés à la taille, les grandes entreprises technologiques riches en données menacent encore la démocratie en classant les utilisateurs dans des groupes d'affinité partageant les mêmes idées, et en formant des algorithmes sur des données comportementales granulaires pour recommander à des utilisateurs ciblés différentes nouvelles, opinions et informations erronées. Ces algorithmes exploitent souvent les tendances psychologiques naturelles de l'homme, telles que le biais de confirmation, le biais de groupe et le rôle des émotions dans le jugement politique. Ces tactiques profitent aux entreprises en augmentant le temps d'écran des utilisateurs. Mais elles créent également de graves externalités négatives menaçant la démocratie, sous la forme d'un grand nombre de citoyens mal informés et polarisés. L'attentat du 6 janvier 2021 contre le capitole des États-Unis illustre de manière frappante la gravité de cette menace.

Au-delà des questions de taille, de pouvoir et de polarisation, l'existence d'un « panopticon numérique » entre les mains d'un petit nombre d'entreprises qui n'ont aucun compte à rendre à la démocratie est tout simplement un affront à la dignité humaine et à l'autodétermination. Dans sa configuration actuelle, la big tech sape directement certaines des valeurs humaines fondamentales que la démocratie est censée soutenir.

Sur une note plus encourageante, le travail de la ministre taïwanaise du numérique, Audrey Tang, montre que les technologies numériques et le big data peuvent être exploités de manière à rendre les démocraties plus, et non moins, démocratiques. La communauté g0v ("gov-zero") de Taiwan, composée de « hackers civiques » décentralisés, a créé des plates-formes ouvertes et populaires sur lesquelles diverses parties prenantes peuvent auto-organiser diverses utilisations des données,

débattre de questions politiques et s'associer à des fonctionnaires pour améliorer les services publics. Cette approche « ascendante » est intrinsèquement démocratique : elle utilise des plateformes numériques pour permettre l'organisation collective, renforcer l'action des citoyens sur leurs données et leurs technologies, et accroître leur capacité à orienter les activités de leur gouvernement.

Traduit de l'anglais par A. Suhamy

Publié dans laviedesidees.fr, le 8 juin 2022.