

« L'éducation au numérique est la clé pour éviter les effets pervers des algorithmes »

Entretien avec M. Tristan Nitot, VP Advocacy chez Qwant

Third (T) : Pourriez-vous présenter les grandes étapes de votre parcours dans le numérique ?

Tristan Nitot (TN) : Je suis tombé très jeune dans le numérique, à l'époque on disait l'informatique. À 14 ans, j'ai commencé à programmer et j'ai voulu en faire mon métier.

En 1997, je suis rentré chez Netscape et, en 1998, le projet Mozilla a été annoncé. Ce projet consistait à ouvrir le code source du navigateur pour permettre son amélioration par la communauté des internautes. Ce fut un véritable bouleversement de l'ordre établi car cela donnait un accès libre à l'actif essentiel (le code source) et qu'on mettait en avant la collaboration, ces deux dimensions étant contraires à l'approche utilisée dans l'ensemble de l'industrie et de en particulier chez le leader du marché, Microsoft.

Si beaucoup de gens n'ont pas compris la démarche (y compris chez Netscape), cette approche m'a tout de suite séduit car j'utilisais déjà des logiciels libres et je côtoyais cette communauté. L'idée de Mozilla m'est apparue extraordinaire et, avec le recul, je dirais que c'était révolutionnaire. C'était une nouvelle manière de créer de la valeur que je qualifierais « d'utopie en marche » : chaque personne qui améliore le logiciel contribue au bien commun de la communauté des utilisateurs. Ce projet était d'autant plus important qu'il s'agissait du navigateur, c'est-à-dire de la pierre angulaire permettant d'accéder au web.

J'ai trouvé cela fascinant et j'ai décidé de mettre mon énergie au service de ce projet. En 2003, le projet Mozilla s'est autonomisé de Netscape et le 9 novembre 2004, le navigateur Firefox est sorti. En quelques années, il y a eu plus de 500 millions d'utilisateurs et cela a validé le modèle du logiciel libre tout en donnant de nouvelles perspectives à la navigation sur le web.

T : Internet est une zone de liberté et d'épanouissement. Pourtant, vous avez récemment dénoncé la surveillance de masse et insisté sur la nécessité de protéger la vie privée des utilisateurs : quel regard portez-vous sur les récentes évolutions d'internet ?

TN : Avant toute chose, je suis un enthousiaste d'internet car c'est une zone de liberté inouïe. J'aime beaucoup la notion de « cyber-espace » car elle transmet l'idée d'une zone infinie dont on peut faire la découverte.

C'est d'ailleurs ancré aux origines car Tim Berners-Lee, l'inventeur du web, a voulu créer un endroit où les gens pouvaient partager et publier leurs idées. Cependant, il n'avait pas anticipé le mouvement de monétisation des contenus et la nécessité, pour les entreprises, de financer le développement des applications et leur fonctionnement.

Aujourd'hui, les revenus proviennent essentiellement de la publicité ciblée qui implique une collecte massive de données personnelles.

Avec les révélations d'Edward Snowden, on s'est rendu compte que tous les utilisateurs laissent des traces sur internet et qu'elles pouvaient être exploitées pour surveiller les citoyens. La publication des documents de la *National Security Agency* (NSA) a mis à jour une forme de surveillance généralisée des citoyens basée sur la collecte de leurs données. De la même manière, les GAFAs participent à la mise en place de cette surveillance et assurent l'hégémonie commerciale des États-Unis.

À mon sens, le problème majeur provient du fait que (i) les utilisateurs ne sont pas éduqués aux dangers du numérique et (ii) qu'il y a une injonction permanente d'avoir recours aux outils numériques. On encourage des usages massifs sans que les consommateurs ne soient en mesure de comprendre que leurs données sont collectées et utilisées. Par exemple, on vante le prix des téléphones Samsung par rapport à ceux d'Apple mais personne n'insiste sur la pré-installation de toutes les applications de Google sur les premiers, qui sont de véritables mouchards.

Les utilisateurs ont de plus en plus recours à des produits gratuits (ou moins chers) car la vraie valeur réside dans les données personnelles qui sont collectées à des fins publicitaires.

T : Ce que vous expliquez fait écho à l'adage « si c'est gratuit, c'est vous le produit ». Si tel est le cas, Qwant étant gratuit, doit-on en déduire que les utilisateurs sont le produit ?

TN : Étant donné la concurrence, Qwant fournit un service gratuit car les utilisateurs ne comprendraient pas de devoir payer pour un moteur de recherche. Par conséquent, il a fallu trouver un mode de financement car il s'agit d'une entreprise privée qui doit développer son infrastructure informatique et gagner de l'argent. Dans cette perspective, il a été décidé d'avoir de la publicité contextuelle mais pas de la publicité ciblée. Chez Qwant, la publicité dépend de la requête (le mot clé saisi) et pas de l'utilisateur, de sorte que ses données personnelles et sa vie privée sont protégées.

Par exemple, toute personne qui cherche le mot « vélo » sur Qwant verra la même publicité. Certes les revenus sont moins importants pour Qwant mais la vie privée des utilisateurs est protégée.

Cette approche est l'expression du principe de « neutralité des résultats », c'est-à-dire que les résultats et les publicités sont les mêmes pour tous les utilisateurs effectuant une requête similaire. Il ne faut pas le confondre avec la « neutralité du net » qui se situe au niveau des fournisseurs d'accès à internet. Ce principe fondateur signifie que toute personne peut avoir accès au contenu sans être discriminé et sans demander l'autorisation en amont.

L'abolition de ce principe aux États-Unis me fait froid dans le dos car il n'y a plus d'égalité d'accès au réseau entre les utilisateurs et les libertés s'estompent. Je prends un exemple pour qu'on se rende compte de ce que cela signifie en pratique. En août dernier, alors que d'immenses incendies ravageaient la Californie, la société Verizon (un opérateur téléphonique) a réduit le débit internet des pompiers car ils utilisaient plus de data que leur forfait ne le permettait. Or la limitation de la bande passante a eu un impact direct sur les outils informatiques des soldats du feu, ce qui les a rendus moins performants dans la gestion de la crise.

La vigilance s'impose face aux nouvelles technologies car les dérives sont vite arrivées.

T : La vigilance incombe à chacun mais comment faire cette vérification sans se contenter de l'apparence ? Est-ce en explorant et contrôlant l'algorithme pour savoir exactement ce que font les opérateurs ?

TN : Effectivement, il me semble important d'être attentif à cela.

En informatique, un algorithme est une méthode, une séquence d'actions qui donne un résultat. Analyser l'algorithme d'un produit revient à prendre connaissance du mode d'emploi et à comprendre son fonctionnement.

Cependant, aujourd'hui, dans le langage commun, lorsqu'on utilise le terme « algorithme » j'ai l'impression qu'on fait plutôt référence au *machine learning*. Il s'agit du mécanisme par lequel la machine prend des décisions à partir d'un protocole pré-déterminé et d'un apprentissage à partir d'un ensemble de données. Dans cette configuration, il n'est pas possible d'expliquer comment la décision a été prise car on passe tout de suite au résultat sans que la machine ne détaille les étapes du raisonnement.

C'est un véritable sujet car on ne peut pas expliquer pourquoi ou comment les décisions sont prises. C'est d'autant plus problématique que ces algorithmes complexes peuvent être porteurs de biais dans leur conception ou leur fonctionnement.

Prenons l'exemple d'un algorithme qui étudie les dossiers de patients atteints d'un cancer du poumon et qui conclut que les asthmatiques ont de meilleures chances de guérison que les autres. C'est très bien mais l'algorithme qui analyse le critère « asthme » par rapport au résultat « guérison/non guérison » ignore que les asthmatiques sont structurellement plus suivis que le reste de la population, de sorte que les cancers sont statistiquement détectés plus tôt. Il pourrait donc y avoir un effet pervers à l'utilisation de cet algorithme car une mutuelle pourrait être vendue plus chère à des non asthmatiques.

Les algorithmes sont créés par des entreprises qui ont leurs propres intérêts, lesquels divergent avec ceux du public.

Pour éviter des situations non désirées, il me semble impératif d'introduire le maximum de transparence dans les algorithmes, en particulier le code source, tout en investissant dans la recherche visant à faire du *machine learning* explicable.

T : Vous évoquez la transparence : que mettez-vous derrière ce terme ? Est-ce toujours désirable d'être transparent ? Pourquoi pourrait-on l'exiger des entreprises (transparence de leurs codes informatiques) et pas des individus (droit à la dissimulation) ?

TN : Tout d'abord, je pense que plus il y a de transparence, mieux c'est. Ensuite, il faut bien distinguer le domaine étatique (régalien) du domaine privé (les entreprises).

En ce qui concerne le premier, le citoyen a le droit de savoir comment sont prises les décisions le concernant (en particulier lorsqu'elles concernent ses droits et libertés). Autrement, cela signifie qu'on est dans l'arbitraire et que l'État n'est plus redevable vis-à-vis des citoyens, ce qui revient à dire qu'on n'est plus en démocratie. La publication du code source de « Parcoursup » est un excellent exemple de la nécessité d'expliquer les décisions étatiques prises à l'aide d'algorithmes.

Pour la sphère privée, il faut faire la distinction entre « transparence » et « rendre public ». Il est impératif que les fournisseurs de technologies soient en mesure de détailler la logique et le fonctionnement de leurs algorithmes, notamment pour pouvoir être audités et contrôlés.

Cependant, cela ne veut pas dire qu'il faut que ce soit public car on pourrait arriver à un résultat différent du but recherché.

À titre d'illustration, le code du moteur de recherche de Qwant n'est pas *open source*. Ce n'est pas une contradiction mais une précaution. En effet, si tout le monde connaissait les critères de classement alors il serait possible de contourner les règles. En d'autres termes, en rendant public le code source nous échouons dans la délivrance du service de qualité que nous voulons pour nos utilisateurs (en particulier, des listes de résultats pertinents et une recherche respectueuse des données personnelles). Cette part d'obscurité des algorithmes nous sert chez Qwant mais nous sommes capables de répondre du fonctionnement de notre produit si une autorité nous le demande.

T : Ces décisions sont prises par des opérateurs privés et le citoyen est souvent désemparé face aux puissances algorithmiques. Peut-on réellement faire la balance des intérêts lorsqu'on est utilisateur ?

TN : L'éducation au numérique est la clé pour éviter les effets pervers des algorithmes.

Aujourd'hui les utilisateurs ne se rendent pas compte des conséquences de leurs actions.

Il faut leur apprendre (en particulier, les enfants) à faire la part des choses concernant les outils numériques. Il faut prendre du recul afin de prendre conscience que les intérêts des entreprises éditant ces solutions ne sont pas forcément alignés avec ceux des utilisateurs.

Par conséquent, je crois qu'il faut instaurer une véritable « hygiène » pour le numérique, comme cela existe à des fins sanitaires. On peut faire le parallèle avec les microbes : il y a plusieurs siècles on ne les voyait pas et ils n'existaient pas pour les hommes. Avec le développement de la science et de la médecine, on s'est rendu compte qu'il était bon de faire certains gestes (par exemple, se laver les mains).

C'est pareil pour le numérique et je pense que l'humanité va intégrer des réflexes d'hygiène. On pourrait imaginer qu'un jour, les internautes ne se servent plus de Google car c'est « sale » en raison de la collecte massive de données personnelles.

Les utilisateurs doivent être formés en vue d'être responsabilisés.

T : Mais les utilisateurs sont intégrés à un monde global qui a sa propre inertie, de sorte qu'il peut être vain d'espérer que les usages vont permettre d'« assainir » l'utilisation d'internet. Voyez-vous d'autres actions possibles ?

TN : Bien sûr, je vais en citer deux.

La première est de développer des algorithmes performants et respectueux des intérêts des utilisateurs. On peut donc développer des alternatives comme Qwant (qui ne capturent pas les données personnelles) ou Firefox (qui fonctionnent sur le modèle du logiciel libre). Cependant, ces modèles sont moins complets pour l'instant car ils sont moins rentables et qu'ils coûtent cher à construire.

La seconde réside dans le choix que les ingénieurs peuvent faire dans leurs carrières. Puisqu'ils sont la cheville ouvrière du numérique, les ingénieurs doivent avoir une conscience et participer à des initiatives éthiques. Il est possible de créer des outils respectueux et il faut être acteur de la construction d'un web meilleur qui met au cœur de son projet les précautions de respect des droits des utilisateurs.

Chacun a un rôle à jouer et internet est un actif commun qu'il faut construire ensemble.

T : Dans cette perspective, la régulation a un rôle à jouer. Quel regard portez-vous sur la régulation du numérique ? Les choses vont-elles dans le bon sens ?

TN : Voilà une question difficile ! À vrai dire, je n'ai pas de réponse précise.

Réguler le numérique n'est pas simple car les opérateurs sont beaucoup plus rapides que le législateur. Il a fallu 10 ans pour que le RGPD entre en application alors que les smartphones existent déjà depuis 11 ans... le monde change tellement vite qu'il est difficile d'imaginer des règles et, lorsqu'on le fait, il est parfois vain de penser que ces normes seront adaptées à la réalité opérationnelle.

Je pense que le régulateur part avec des boulets aux pieds vis-à-vis des industriels du numérique et que la dimension mondiale de ces sociétés introduit d'importantes disparités qui rendent les choses plus compliquées encore.

Malgré ce handicap structurel difficile à compenser, je pense que le niveau pertinent pour la régulation est celui de l'Union Européenne, et le RGPD est un bon début. Cela prendra du temps mais on y arrivera. Dans l'intervalle, c'est l'éducation qui doit prendre le relais. ■■■