

# Ouvrir le code des algorithmes ? — Oui, mais... (1/2)

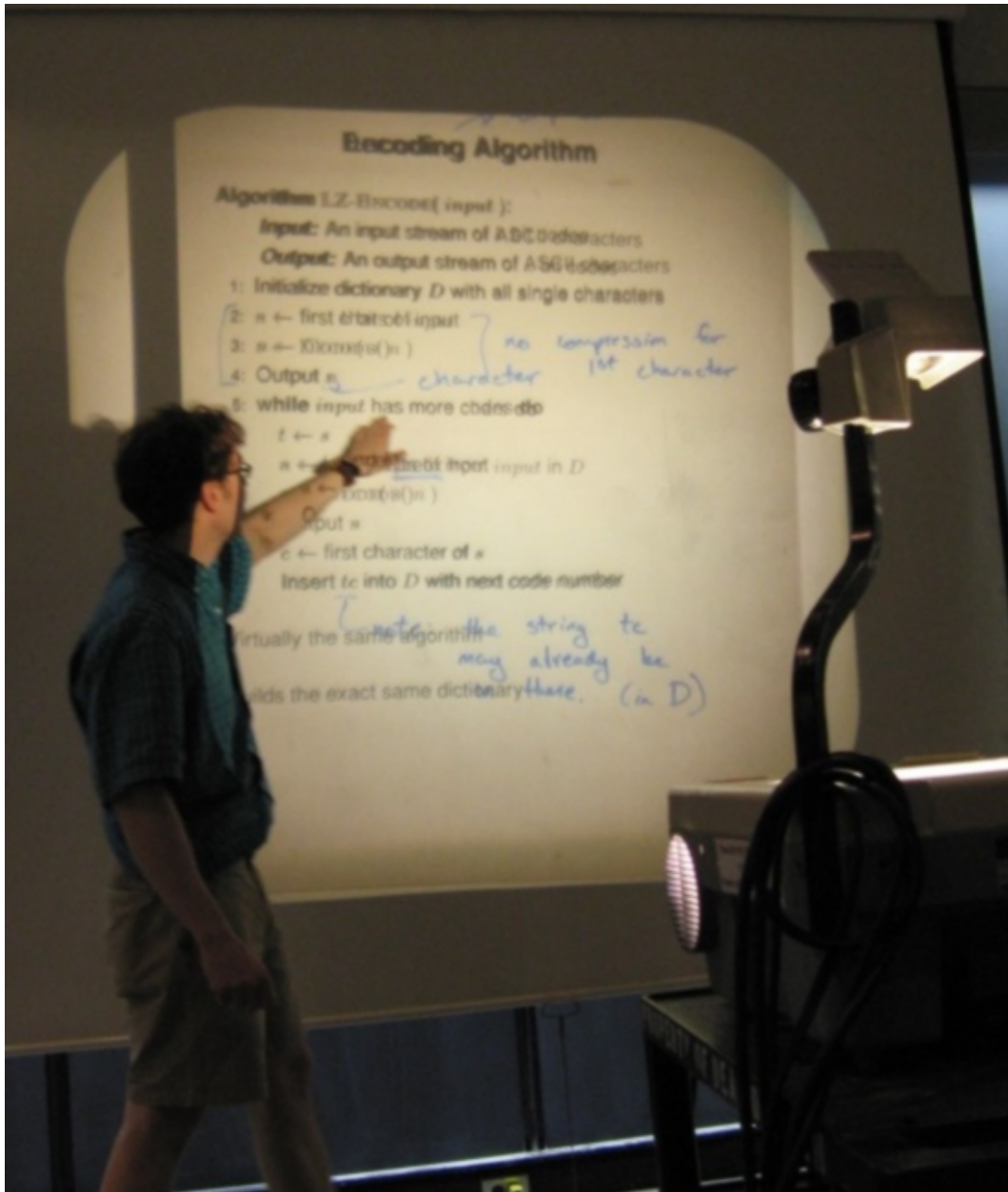
Voici le premier des deux articles qu'Hubert Guillaud nous fait le plaisir de partager. Sans s'arrêter à la surface de l'actualité, il aborde la transparence du code des algorithmes, qui entraîne un grand nombre de questions épineuses sur lesquelles il s'est documenté pour nous faire part de ses réflexions.

---

## **Dans le code source de l'amplification algorithmique : publier le code ne suffit pas !**

**par Hubert GUILLAUD**

Le 31 mars, Twitter a publié une partie du code source qui alimente son fil d'actualité, comme l'a expliqué l'équipe elle-même dans un billet. Ces dizaines de milliers de lignes de code contiennent pourtant peu d'informations nouvelles. Depuis le rachat de l'oiseau bleu par Musk, Twitter a beaucoup changé et ne cesse de se modifier sous les yeux des utilisateurs. La publication du code source d'un système, même partiel, qui a longtemps été l'un des grands enjeux de la transparence, montre ses limites.



« LZW encoding and decoding algorithms overlapped » par nayukim, licence CC BY 2.0.

## Publier le code ne suffit pas

Dans un excellent billet de blog, le chercheur Arvind Narayan (sa newsletter mérite également de s'y abonner) explique ce qu'il faut en retenir. Comme ailleurs, les règles ne sont pas claires. Les algorithmes de recommandation utilisent l'apprentissage automatique ce qui fait que la manière de classer les tweets n'est pas directement spécifiée dans le code, mais apprise par des modèles à partir de données de Twitter sur la manière dont les utilisateurs ont réagi aux tweets dans le passé. Twitter ne divulgue ni ces modèles ni les données d'apprentissages, ce qui signifie qu'il n'est pas possible d'exécuter

ces modèles. Le code ne permet pas de comprendre pourquoi un tweet est ou n'est pas recommandé à un utilisateur, ni pourquoi certains contenus sont amplifiés ou invisibilisés. C'est toute la limite de la transparence. Ce que résume très bien le journaliste Nicolas Kayser-Bril pour AlgorithmWatch (pertinemment traduit par le framablog) : « *Vous ne pouvez pas auditer un code seulement en le lisant. Il faut l'exécuter sur un ordinateur.* »

« *Ce que Twitter a publié, c'est le code utilisé pour entraîner les modèles, à partir de données appropriées* », explique Narayan, ce qui ne permet pas de comprendre les propagations, notamment du fait de l'absence des données. De plus, les modèles pour détecter les tweets qui violent les politiques de Twitter et qui leur donnent des notes de confiance en fonction de ces politiques sont également absentes (afin que les usagers ne puissent pas déjouer le système, comme nous le répètent trop de systèmes rétifs à l'ouverture). Or, ces classements ont des effets de rétrogradation très importants sur la visibilité de ces tweets, sans qu'on puisse savoir quels tweets sont ainsi classés, selon quelles méthodes et surtout avec quelles limites.

La chose la plus importante que Twitter a révélée en publiant son code, c'est la formule qui spécifie comment les différents types d'engagement (likes, retweets, réponses, etc.) sont pondérés les uns par rapport aux autres... Mais cette formule n'est pas vraiment dans le code. Elle est publiée séparément, notamment parce qu'elle n'est pas statique, mais qu'elle doit être modifiée fréquemment.

Sans surprise, le code révèle ainsi que les abonnés à Twitter Blue, ceux qui payent leur abonnement, bénéficient d'une augmentation de leur portée (ce qui n'est pas sans poser un problème de fond, comme le remarque pertinemment sur Twitter, Guillaume Champeau, car cette préférence pourrait mettre ces utilisateurs dans la position d'être annonceurs, puisqu'ils payent pour être mis en avant, sans que l'interface ne le signale clairement, autrement que par la pastille bleue). Reste que le code n'est pas clair sur l'ampleur de cette accélération. Les notes attribuées aux tweets des abonnés Blue sont multipliées par 2 ou 4, mais cela ne signifie pas que leur portée est pareillement multipliée. « *Une fois encore, le code ne nous dit pas le genre de choses que nous voudrions savoir* », explique Narayan.

Reste que la publication de la formule d'engagement est un événement majeur. Elle permet de saisir le poids des réactions sur un tweet. On constate que la

réponse à tweet est bien plus forte que le like ou que le RT. Et la re-réponse de l'utilisateur originel est prédominante, puisque c'est le signe d'une conversation forte. À l'inverse, le fait qu'un lecteur bloque, mute ou se désabonne d'un utilisateur suite à un tweet est un facteur extrêmement pénalisant pour la propagation du tweet.

Type of engagement	Weight
Probability the user will like the tweet	0.5
Probability the user will retweet the tweet	1.0
Probability the user replies to the tweet	13.5
Probability the user opens the tweet author profile and likes or replies to a tweet	12.0
Probability (for a video tweet) that the user will watch at least half of the video	0.005
Probability the user replies to the tweet and this reply is engaged by the tweet author	75.0
Probability the user will click into the conversation of this tweet and reply or like a tweet	11.0
Probability the user will click into the conversation and stay there for at least 2 minutes	10.0
Probability the user will react negatively (requesting "show less often" on the tweet or author, block or mute the tweet author)	-74.0
Probability the user will click report tweet	-369.0

Tableau du poids attribué en fonction des types d'engagement possibles sur Twitter.

Ces quelques indications permettent néanmoins d'apprendre certaines choses. Par exemple que Twitter ne semble pas utiliser de prédictions d'actions implicites (comme lorsqu'on s'arrête de faire défiler son fil), ce qui permet d'éviter l'amplification du contenu trash que les gens ne peuvent s'empêcher de regarder, même s'ils ne s'y engagent pas. La formule nous apprend que les retours négatifs ont un poids très élevé, ce qui permet d'améliorer son flux en montrant à l'algorithme ce dont vous ne voulez pas - même si les plateformes devraient permettre des contrôles plus explicites pour les utilisateurs. Enfin, ces poids ont des valeurs souvent précises, ce qui signifie que ce tableau n'est valable qu'à l'instant de la publication et qu'il ne sera utile que si Twitter le met à jour.

Les algorithmes de recommandation qui optimisent l'engagement suivent des modèles assez proches. La publication du code n'est donc pas très révélatrice.

Trois éléments sont surtout importants, insiste le chercheur :

*« Le premier est la manière dont les algorithmes sont configurés : les signaux utilisés comme entrée, la manière dont l'engagement est défini, etc. Ces informations doivent être considérées comme un élément essentiel de la transparence et peuvent être publiées indépendamment du code. La seconde concerne les modèles d'apprentissage automatique qui, malheureusement, ne peuvent généralement pas être divulgués pour des raisons de protection de la vie privée. La troisième est la boucle de rétroaction entre les utilisateurs et l'algorithme ».*

Autant d'éléments qui demandent des recherches, des expériences et du temps pour en comprendre les limites.

Si la transparence n'est pas une fin en soi, elle reste un moyen de construire un meilleur internet en améliorant la responsabilité envers les utilisateurs, rappelle l'ingénieur Gabriel Nicholas pour le Center for Democracy & Technology. Il souligne néanmoins que la publication d'une partie du code source de Twitter ne contrebalance pas la fermeture du Consortium de recherche sur la modération, ni celle des rapports de transparence relatives aux demandes de retraits des autorités ni celle de l'accès à son API pour chercheurs, devenue extrêmement coûteuse.

*« Twitter n'a pas exactement 'ouvert son algorithme' comme certains l'ont dit. Le code est lourdement expurgé et il manque plusieurs fichiers de configuration, ce qui signifie qu'il est pratiquement impossible pour un chercheur indépendant d'exécuter l'algorithme sur des échantillons ou de le tester d'une autre manière. Le code publié n'est en outre qu'un instantané du système de recommandation de Twitter et n'est pas réellement connecté au code en cours d'exécution sur ses serveurs. Cela signifie que Twitter peut apporter des modifications à son code de production et ne pas l'inclure dans son référentiel public, ou apporter des modifications au référentiel public qui ne sont pas reflétées dans son code de production. »*

L'algorithme publié par Twitter est principalement son système de recommandation. Il se décompose en 3 parties, explique encore Nicholas :

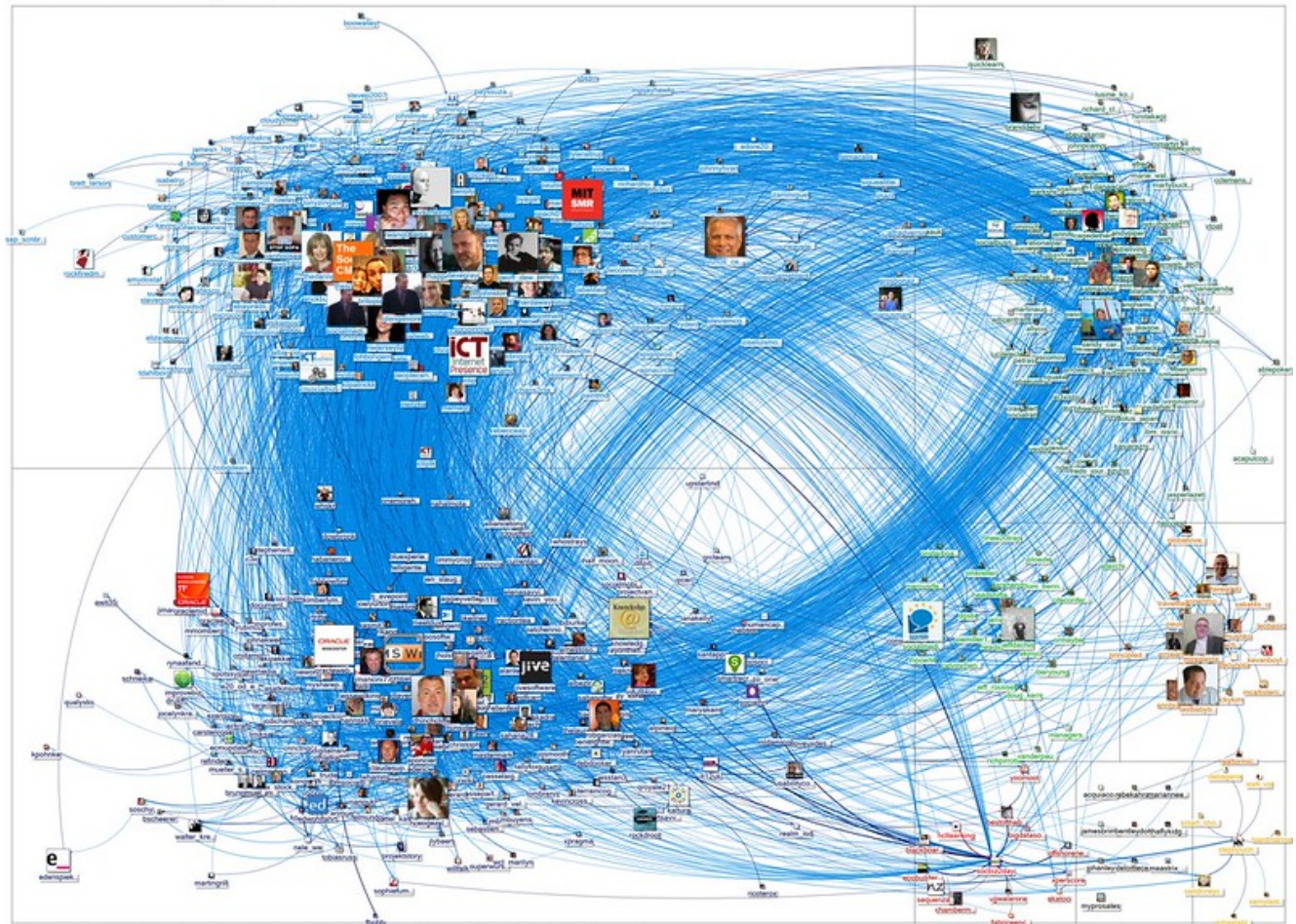
- Un système de génération de contenus candidats. Ici, Twitter sélectionne 1500 tweets susceptibles d'intéresser un utilisateur en prédisant la probabilité que l'utilisateur s'engage dans certaines actions pour chaque tweet (c'est-à-dire qu'il RT ou like par exemple).
- Un système de classement. Une fois que les 1 500 tweets susceptibles d'être servis sont sélectionnés, ils sont notés en fonction de la probabilité des actions d'engagement, certaines actions étant pondérées plus fortement que d'autres. Les tweets les mieux notés apparaîtront généralement plus haut dans le fil d'actualité de l'utilisateur.
- Un système de filtrage. Les tweets ne sont pas classés strictement en fonction de leur score. Des heuristiques et des filtres sont appliqués pour, par exemple, éviter d'afficher plusieurs tweets du même auteur ou pour déclasser les tweets d'auteurs que l'utilisateur a déjà signalés pour violation de la politique du site.

Le score final est calculé en additionnant la probabilité de chaque action multipliée par son poids (en prenant certainement en compte la rareté ou la fréquence d'action, le fait de répondre à un tweet étant moins fréquent que de lui attribuer un like). Mais Twitter n'a pas publié la probabilité de base de chacune de ces actions ce qui rend impossible de déterminer l'importance de chacune d'elles dans les recommandations qui lui sont servies.

Twitter a également révélé quelques informations sur les autres facteurs qu'il prend en compte en plus du classement total d'un tweet. Par exemple, en équilibrant les recommandations des personnes que vous suivez avec celles que vous ne suivez pas, en évitant de recommander les tweets d'un même auteur ou en donnant une forte prime aux utilisateurs payants de Twitter Blue.

Il y a aussi beaucoup de code que Twitter n'a pas partagé. Il n'a pas divulgué beaucoup d'informations sur l'algorithme de génération des tweets candidats au classement ni sur ses paramètres et ses données d'entraînement. Twitter n'a pas non plus explicitement partagé ses algorithmes de confiance et de sécurité pour détecter des éléments tels que les abus, la toxicité ou les contenus pour adultes, afin d'empêcher les gens de trouver des solutions de contournement, bien qu'il ait publié certaines des catégories de contenu qu'il signale.





Created with NodeXL (<http://nodexl.codeplex.com>) from the Social Media Research Foundation (<http://www.smrfoundation.org>)

« 20120212-NodeXL-Twitter-socbiz network graph » par Marc\_Smith; licence CC BY 2.0.

Pour Gabriel Nicholas, la transparence de Twitter serait plus utile si Twitter avait maintenu ouverts ses outils aux chercheurs. Ce n'est pas le cas.

Il y a plein d'autres points que l'ouverture de l'algorithme de Twitter a documentés. Par exemple, l'existence d'un Tweepcred, un score qui classe les utilisateurs et qui permet de voir ses publications boostées si votre score est bon, comme l'expliquait Numerama. Ou encore le fait que chaque compte est clustérisé dans un groupe aux profils similaires dans lequel les tweets sont d'abord diffusés avant d'être envoyés plus largement s'ils rencontrent un premier succès... De même, il semblerait qu'il y ait certaines catégories d'utilisateurs spéciaux (dont une catégorie relative à Elon Musk) mais qui servent peut-être plus certaines statistiques qu'à doper la portée de certains comptes comme on l'a entendu (même s'il semble bien y avoir une catégorie VIP sur Twitter - comme il y a sur Facebook un statut d'exception à la modération)...

# Ouvrir, mais ouvrir quoi ?

En conclusion de son article, Narayan pointe vers un très intéressant article qui dresse une liste d'options de transparence pour ceux qui produisent des systèmes de recommandation, publiée par les chercheurs Priyanjana Bengani, Jonathan Stray et Luke Thorburn. Ils rappellent que les plateformes ont mis en place des mesures de transparence, allant de publications statistiques à des interfaces de programmation, en passant par des outils et des ensembles de données protégés. Mais ces mesures, très techniques, restent insuffisantes pour comprendre les algorithmes de recommandation et leur influence sur la société. Une grande partie de cette résistance à la transparence ne tient pas tant aux risques commerciaux qui pourraient être révélés qu'à éviter l'embarras d'avoir à se justifier de choix qui ne le sont pas toujours. D'une manière très pragmatique, les trois chercheurs proposent un menu d'actions pour améliorer la transparence et l'explicabilité des systèmes.

## Documenter

L'un des premiers outils, et le plus simple, reste la documentation qui consiste à expliquer en termes clairs - selon différentes échelles et niveaux, me semble-t-il - ce qui est activé par une fonction. Pour les utilisateurs, c'est le cas du bouton « *Pourquoi je vois ce message* » de Facebook ou du panneau « *Fréquemment achetés ensemble* » d'Amazon. L'idée ici est de fourbir un « *compte rendu honnête* ». Pour les plus évoluées de ces interfaces, elles devraient permettre non seulement d'informer et d'expliquer pourquoi on nous recommande ce contenu, mais également, permettre de rectifier et mieux contrôler son expérience en ligne, c'est-à-dire d'avoir des leviers d'actions sur la recommandation.

Une autre forme de documentation est celle sur le fonctionnement général du système et ses décisions de classement, à l'image des rapports de transparence sur les questions de sécurité et d'intégrité que doivent produire la plupart des plateformes (voir celui de Google, par exemple). Cette documentation devrait intégrer des informations sur la conception des algorithmes, ce que les plateformes priorisent, minimisent et retirent, si elles donnent des priorités et à qui, tenir le journal des modifications, des nouvelles fonctionnalités, des changements de politiques. La documentation doit apporter une information solide et loyale, mais elle reste souvent insuffisante.

## Les données



Pour comprendre ce qu'il se passe sur une plateforme, il est nécessaire d'obtenir des données. Twitter ou Facebook en ont publié (accessibles sous condition de recherche, ici pour Twitter, là pour Facebook). Une autre approche consiste à ouvrir des interfaces de programmation, à l'image de CrowdTangle de Facebook ou de l'API de Twitter. Depuis le scandale Cambridge Analytica, l'accès aux données est souvent devenu plus difficile, la protection de la vie privée servant parfois d'excuse aux plateformes pour éviter d'avoir à divulguer leurs pratiques. L'accès aux données, même pour la recherche, s'est beaucoup refermé ces dernières années. Les plateformes publient moins de données et CrowdTangle propose des accès toujours plus sélectifs. Chercheurs et journalistes ont été contraints de développer leurs propres outils, comme des extensions de navigateurs permettant aux utilisateurs de faire don de leurs données (à l'image du Citizen Browser de The Markup) ou des simulations automatisées (à l'image de l'analyse robotique de TikTok produite par le *Wall Street Journal*), que les plateformes ont plutôt eu tendance à bloquer en déniaient les résultats obtenus sous prétexte d'incomplétude - ce qui est justement le problème que l'ouverture de données cherche à adresser.

### **Le code**

L'ouverture du code des systèmes de recommandation pourrait être utile, mais elle ne suffit pas, d'abord parce que dans les systèmes de recommandation, il n'y a pas un algorithme unique. Nous sommes face à des ensembles complexes et enchevêtrés où « *différents modèles d'apprentissage automatique formés sur différents ensembles de données remplissent diverses fonctions* ». Même le classement ou le modèle de valeur pour déterminer le score n'explique pas tout. Ainsi, « *le poids élevé sur un contenu d'un type particulier ne signifie pas nécessairement qu'un utilisateur le verra beaucoup, car l'exposition dépend de nombreux autres facteurs, notamment la quantité de ce type de contenu produite par d'autres utilisateurs.* »

Peu de plateformes offrent une grande transparence au niveau du code source. Reddit a publié en 2008 son code source, mais a cessé de le mettre à jour. En l'absence de mesures de transparence, comprendre les systèmes nécessite d'écluser le travail des journalistes, des militants et des chercheurs pour tenter d'en obtenir un aperçu toujours incomplet.

### **La recherche**

Les plateformes mènent en permanence une multitude de projets de recherche

internes voire externes et testent différentes approches pour leurs systèmes de recommandation. Certains des résultats finissent par être accessibles dans des revues ou des articles soumis à des conférences ou via des fuites d'informations. Quelques efforts de partenariats entre la recherche et les plateformes ont été faits, qui restent embryonnaires et ne visent pas la transparence, mais qui offrent la possibilité à des chercheurs de mener des expériences et donc permettent de répondre à des questions de nature causale, qui ne peuvent pas être résolues uniquement par l'accès aux données.

Enfin, les audits peuvent être considérés comme un type particulier de recherche. À l'heure actuelle, il n'existe pas de bons exemples d'audits de systèmes de recommandation menés à bien. Reste que le *Digital Service Act* (DSA) européen autorise les audits externes, qu'ils soient lancés par l'entreprise ou dans le cadre d'une surveillance réglementaire, avec des accès élargis par rapport à ceux autorisés pour l'instant. Le DSA exige des évaluations sur le public mineur, sur la sécurité, la santé, les processus électoraux... mais ne précise ni comment ces audits doivent être réalisés ni selon quelles normes. Des méthodes spécifiques ont été avancées pour contrôler la discrimination, la polarisation et l'amplification dans les systèmes de recommandation.

En principe, on pourrait évaluer n'importe quel préjudice par des audits. Ceux-ci visent à vérifier si « *la conception et le fonctionnement d'un système de recommandation respectent les meilleures pratiques et si l'entreprise fait ce qu'elle dit qu'elle fait. S'ils sont bien réalisés, les audits pourraient offrir la plupart des avantages d'un code source ouvert et d'un accès aux données des utilisateurs, sans qu'il soit nécessaire de les rendre publics.* » Reste qu'il est peu probable que les audits imposés par la surveillance réglementaire couvrent tous les domaines qui préoccupent ceux qui sont confrontés aux effets des outils de recommandations.

### **Autres moteurs de transparence : la gouvernance et les calculs**

Les chercheurs concluent en soulignant qu'il existe donc une gamme d'outils à disposition, mais qu'elle manque de règles et de bonnes pratiques partagées. Face aux obligations de transparence et de contrôles qui arrivent (pour les plus gros acteurs d'abord, mais parions que demain, elles concerneront bien d'autres acteurs), les entreprises peinent à se mettre en ordre de marche pour proposer des outillages et des productions dans ces différents secteurs qui leur permettent

à la fois de se mettre en conformité et de faire progresser leurs outils. Ainsi, par exemple, dans le domaine des données, documenter les jeux et les champs de données, à défaut de publier les jeux de données, pourrait déjà permettre un net progrès. Dans le domaine de la documentation, les cartes et les registres permettent également d'expliquer ce que les calculs opèrent (en documentant par exemple leurs marges d'erreurs).

Reste que l'approche très technique que mobilisent les chercheurs oublie quelques leviers supplémentaires. Je pense notamment aux conseils de surveillance, aux conseils éthiques, aux conseils scientifiques, en passant par les organismes de contrôle indépendants, aux comités participatifs ou consultatifs d'utilisateurs... à tous les outils institutionnels, participatifs ou militants qui permettent de remettre les parties prenantes dans le contrôle des décisions que les systèmes prennent. Dans la lutte contre l'opacité des décisions, tous les leviers de gouvernance sont bons à prendre. Et ceux-ci sont de très bons moyens pour faire pression sur la transparence, comme l'expliquait très pertinemment David Robinson dans son livre *Voices in the Code*.

Un autre levier me semble absent de nombre de propositions... Alors qu'on ne parle que de rendre les calculs transparents, ceux-ci sont toujours absents des discussions. Or, les règles de traitements sont souvent particulièrement efficaces pour améliorer les choses. Il me semble qu'on peut esquisser au moins deux moyens pour rendre les calculs plus transparents et responsables : la minimisation et les interdictions.

La minimisation vise à rappeler qu'un bon calcul ne démultiplie pas nécessairement les critères pris en compte. Quand on regarde les calculs, bien souvent, on est stupéfait d'y trouver des critères qui ne devraient pas être pris en compte, qui n'ont pas de fondements autres que d'être rendus possibles par le calcul. Du risque de récidive au score de risque de fraude à la CAF, en passant par l'attribution de greffes ou aux systèmes de calculs des droits sociaux, on trouve toujours des éléments qui apprécient le calcul alors qu'ils n'ont aucune justification ou pertinence autres que d'être rendu possibles par le calcul ou les données. C'est le cas par exemple du questionnaire qui alimente le calcul de risque de récidive aux Etats-Unis, qui repose sur beaucoup de questions problématiques. Ou de celui du risque de fraude à la CAF, dont les anciennes versions au moins (on ne sait pas pour la plus récente) prenaient en compte par exemple le nombre de fois où les bénéficiaires se connectaient à leur espace en

ligne (sur cette question, suivez les travaux de la Quadrature et de Changer de Cap). La minimisation, c'est aussi, comme l'explique l'ex-chercheur de chez Google, El Mahdi El Mhamdi, dans une excellente interview, limiter le nombre de paramètres pris en compte par les calculs et limiter l'hétérogénéité des données.

L'interdiction, elle, vise à déterminer que certains croisements ne devraient pas être autorisés, par exemple, la prise en compte des primes dans les logiciels qui calculent les données d'agenda du personnel, comme semble le faire le logiciel Orion mis en place par la SnCF, ou Isabel, le logiciel RH que Bol.com utilise pour gérer la main-d'œuvre étrangère dans ses entrepôts de logistique néerlandais. Ou encore, comme le soulignait Narayan, le temps passé sur les contenus sur un réseau social par exemple, ou l'analyse de l'émotion dans les systèmes de recrutement (et ailleurs, tant cette technologie pose problème). A l'heure où tous les calculs sont possibles, il va être pertinent de rappeler que selon les secteurs, certains croisements doivent rester interdits parce qu'ils sont trop à risque pour être mobilisés dans le calcul ou que certains calculs ne peuvent être autorisés.

Priyanjana Bengani, Jonathan Stray et Luke Thorburn, pour en revenir à eux, notent enfin que l'exigence de transparence reste formulée en termes très généraux par les autorités réglementaires. Dans des systèmes vastes et complexes, il est difficile de savoir ce que doit signifier réellement la transparence. Pour ma part, je milite pour une transparence "projective", active, qui permette de se projeter dans les explications, c'est-à-dire de saisir ses effets et dépasser le simple caractère narratif d'une explication loyale, mais bien de pouvoir agir et reprendre la main sur les calculs.

## **Coincés dans les boucles de l'amplification**

Plus récemment, les trois mêmes chercheurs, passé leur article séminal, ont continué à documenter leur réflexion. Ainsi, dans « Rendre l'amplification mesurable », ils expliquent que l'amplification est souvent bien mal définie (notamment juridiquement, ils ont consacré un article entier à la question)... mais proposent d'améliorer les propriétés permettant de la définir. Ils rappellent d'abord que l'amplification est relative, elle consiste à introduire un changement par rapport à un calcul alternatif ou précédent qui va avoir un effet sans que le comportement de l'utilisateur n'ait été, lui, modifié.

L'amplification agit d'abord sur un contenu et nécessite de répondre à la question

de savoir ce qui a été amplifié. Mais même dire que les *fake news* sont amplifiées n'est pas si simple, à défaut d'avoir une définition précise et commune des *fake news* qui nécessite de comprendre les classifications opérées. Ensuite, l'amplification se mesure par rapport à un point de référence précédent qui est rarement précisé. Enfin, quand l'amplification atteint son but, elle produit un résultat qui se voit dans les résultats liés à l'engagement (le nombre de fois où le contenu a été apprécié ou partagé) mais surtout ceux liés aux impressions (le nombre de fois où le contenu a été vu). Enfin, il faut saisir ce qui relève de l'algorithme et du comportement de l'utilisateur. Si les messages d'un parti politique reçoivent un nombre relativement important d'impressions, est-ce parce que l'algorithme est biaisé en faveur du parti politique en question ou parce que les gens ont tendance à s'engager davantage avec le contenu de ce parti ? Le problème, bien sûr, est de distinguer l'un de l'autre d'une manière claire, alors qu'une modification de l'algorithme entraîne également une modification du comportement de l'utilisateur. En fait, cela ne signifie pas que c'est impossible, mais que c'est difficile, expliquent les chercheurs. Cela nécessite un système d'évaluation de l'efficacité de l'algorithme et beaucoup de tests A/B pour comparer les effets des évolutions du calcul. Enfin, estiment-ils, il faut regarder les effets à long terme, car les changements dans le calcul prennent du temps à se diffuser et impliquent en retour des réactions des utilisateurs à ces changements, qui s'adaptent et réagissent aux transformations.

Dans un autre article, ils reviennent sur la difficulté à caractériser l'effet *bulle de filtre* des médias sociaux, notamment du fait de conceptions élastiques du phénomène. S'il y a bien des boucles de rétroaction, leur ampleur est très discutée et dépend beaucoup du contexte. Ils en appellent là encore à des mesures plus précises des phénomènes. Certes, ce que l'on fait sur les réseaux sociaux influe sur ce qui est montré, mais il est plus difficile de démontrer que ce qui est montré affecte ce que l'on pense. Il est probable que les effets médiatiques des recommandations soient faibles pour la plupart des gens et la plupart du temps, mais beaucoup plus importants pour quelques individus ou sous-groupes relativement à certaines questions ou enjeux. De plus, il est probable que changer nos façons de penser ne résulte pas d'une exposition ponctuelle, mais d'une exposition à des récits et des thèmes récurrents, cumulatifs et à long terme. Enfin, si les gens ont tendance à s'intéresser davantage à l'information si elle est cohérente avec leur pensée existante, il reste à savoir si ce que l'on pense affecte ce à quoi l'on s'engage. Mais cela est plus difficile à mesurer car cela suppose de



savoir ce que les gens pensent et pas seulement constater leurs comportements en ligne. En général, les études montrent plutôt que l'exposition sélective a peu d'effets. Il est probable cependant que là encore, l'exposition sélective soit faible en moyenne, mais plus forte pour certains sous-groupes de personnes en fonction des contextes, des types d'informations.

Bref, là encore, les effets des réseaux sociaux sont difficiles à percer.

*Pour comprendre les effets de l'amplification algorithmique, peut-être faut-il aller plus avant dans la compréhension que nous avons des évolutions de celle-ci, afin de mieux saisir ce que nous voulons vraiment savoir. C'est ce que nous tenterons de faire dans la suite de cet article...*