

Instituer l'accès aux données des plateformes numériques à travers une convention entre toutes les parties prenantes

Pour AOC

Dominique Boullier

La requête des chercheurs ou des services publics d'accès aux données des plateformes numériques est tout-à-fait légitime tant elle correspond à un impératif de la nouvelle ère de quantification contemporaine : les traces de propagations accessibles après les sondages d'opinion et les recensements (voir mon livre « Propagations », Armand Colin, 2023).

Cependant, il serait vain d'attendre une coopération de la part de plateformes habituées pendant 15 ans à la prédation des données personnelles et au contournement de toutes leurs responsabilités. De ce fait, les demandes des chercheurs risquent de se réduire à quémander un accès provisoire et opaque comme un privilège qui peut à tout moment être révoqué comme on l'a vu avec les partenariats scientifiques noués par Facebook via Social Science One (NYU en 2021). L'enjeu est pourtant essentiel : il faut bâtir les conventions socio-économiques et techniques pour rendre opérationnelle et socialement utile la nouvelle ère de quantification qui est la nôtre, faite de Big Data, de machine Learning et d'IA et entièrement contrôlée par les plateformes du capitalisme financier numérique, les GAFAM. Il ne s'agit pas d'un pacte avec le diable, pas plus que ce fut le cas entre sciences sociales et dispositifs statistiques des états au début du XXe ou entre sciences sociales et sondages d'opinion gouvernés par les médias et les marques à partir des années 30. Dans les deux cas, les statistiques, comme le montre si bien Alain Desrosières, ont servi à prouver ET à gouverner, au prix d'alliances, de passerelles, de conventions réécrites et rendues à la fois efficaces et valides. La tâche actuelle est identique et il serait totalement naïf de passer sous silence le mécanisme de financement publicitaire qui gouverne toutes ces plateformes et qui justifie l'usage massif des traces. C'est plutôt sur la base de ce modèle économique qu'il convient de parvenir à intéresser les différentes parties prenantes, en premier lieu les marques, à une institution autonome et crédible de mesure de l'activité sur les réseaux sociaux.

Un idéal de refonte radicale

L'idéal serait d'obtenir de la part du législateur la prise en compte de ces plateformes au titre de médias à part entière, auxquels s'appliquerait entièrement le droit des médias, y compris dans ses pratiques des mesures d'audience, en tenant compte de leurs spécificités. C'est ce que je préconise depuis plusieurs années mais le lobbying des plateformes a été tel qu'elles peuvent continuer de s'abriter derrière l'article 230 du Decency Act de 1996, applicable à l'origine aux fournisseurs d'accès (technique) à Internet et aujourd'hui totalement obsolète. Cependant, le fait que Facebook et Twitter deviennent de plus en plus payants peut constituer une brèche pour leur faire admettre ce statut : ils ont des abonnés qui payent pour avoir accès à des contenus éditorialisés via les algorithmes, ils sont donc des médias.

Cependant les chances d'un changement de doctrine sur ce point sont très faibles, en raison de la puissance d'influence des plateformes et de l'absence de volonté politique de les contrôler de la part de gouvernements en Europe ancrés dans le dogme libéral du laisser-faire, malgré les tentatives des agences indépendantes, réussies comme pour le RGPD, ou très timides comme pour le DSA et le DMA . Il faut donc imaginer d'autres solutions pour aboutir à un accès réel aux données d'activité des plateformes.

Comprendre les ressorts d'un système de calcul

La question principale qui est trop vite oubliée est la suivante : pourquoi ces firmes collectent-elles toutes ces traces ? Non pas pour optimiser la performance de leurs systèmes de façon générale mais pour optimiser la performance *commerciale* de leurs plateformes. La clé de toute action en matière d'accès aux traces est bien celle du modèle économique qui est au cœur de la plateformesisation et par là de tout internet quasiment. Le modèle économique principal est le modèle publicitaire puisque Facebook surtout et Google en très grande partie ont gagné leur puissance grâce à ces revenus et captent à elles deux près de 75% des revenus publicitaires en ligne. Ces deux firmes (et les autres tout autant) prétendent attirer les placements publicitaires par leurs algorithmes de publicité programmatique, permettant d'exploiter toutes les traces (les plus infimes parfois comment des temps de présence sur une page) pour cibler des profils construits statistiquement grâce à la masse considérable de données collectées. Le Big Data et le Machine Learning ont été de point de vue des générateurs de revenus publicitaires considérables et ne doutons pas que c'est l'usage principal qui est déjà dans les tuyaux pour les IA génératives.

Peut-on dire qu'ils ont permis de mieux comprendre l'activité des utilisateurs, ce qui intéresserait évidemment au premier plan tous les chercheurs mais aussi les marques elles-mêmes qui investissent dans ces placements ? Chose curieuse, tout d'abord : on n'en sait rien ! Puisque tout le dispositif est totalement opaque, de plus en plus opaque avec le Deep Learning, et que seuls certains chercheurs de Facebook peuvent les exploiter pour publier (cf. l'équipe de L. Adamic par exemple), tous les autres accords avec des partenaires ayant capoté par obsession de contrôle de la part de la firme. Ensuite, parce qu'aucun travail académique indépendant n'a pu apporter le début d'une preuve de l'efficacité des placements ciblés sur le panier ou sur le taux de conversion pour les marques concernées (voir l'HDR de JS Beuscart à ce sujet).

La bulle publicitaire

Nous avons donc affaire en réalité à une bulle publicitaire sur ces plateformes, comme je l'ai montré dans un livre de 2020 (Comment sortir de l'emprise des réseaux sociaux) publié au même moment exactement que celui de Hwang (Subprime Attention Crisis, 2020) qui faisait la même analyse. Les firmes qui vivent de ces placements programmatiques comme Criteo en France n'ajoutent aucune intelligence à la compréhension de ces processus mais seulement de la complexité mathématique (compétence fort présente chez eux) : tout l'écosystème des agences de communication, de marketing, de publicité, de programmation, de placement, de mesures, a intérêt à maintenir cette opacité puisque, au bout du compte, les marques acceptent ce diktat des plateformes. Cela rappelle très nettement la configuration de l'écosystème de la finance spéculative qui à tous les niveaux (de la communication financière aux agences de notation) était complice d'un modèle de titrisation qui conduisait tout le système à sa perte et qui éclata brutalement en 2007-2008.

Pourquoi les marques acceptent-elles pour l'instant ce jeu de dupes ? Parce qu'en réalité les finalités du placement publicitaire sur les plateformes ne sont plus celui du panier ou du taux de conversion, c'est-à-dire le marché de la vente de produits ou services, mais celui du marché financier, celui des investisseurs. Il est très important de comprendre ce glissement financiarisé de toutes les marques si l'on veut obtenir un accès ouvert au calcul des traces. Car la finance est friande de données, de signaux quels qu'ils soient et d'indicateurs toujours plus sophistiqués que les plateformes peuvent leur fournir sous forme de taux d'engagement, qui indiquent avant tout des effets de réputation des marques, d'activité des communautés bien plus que de chiffres de vente effective. Dès lors, comme dans la finance en général, tout fait signal, y compris les données les plus fantaisistes du moment que cela altère la perception des autres investisseurs. C'est pourquoi les plateformes peuvent manipuler à volonté leurs algorithmes de traçage des activités, comme le faisait Facebook révélant en 2016 qu'il avait surestimé massivement le visionnage des vidéos (et donc leur facturation), sans pour autant

expliquer pourquoi son nouveau calcul serait plus fiable que le précédent puisqu'il restait encore totalement opaque !

Ce monde de la publicité en ligne est un jeu de dupes pseudoscientifique construit sur du fake et tout l'écosystème s'en satisfait, il faut en être conscient avant de prétendre rétablir un peu d'objectivité dans ce monde de spéculation.

Médias de masse et médias plateformes en ligne : des principes institutionnels radicalement différents

Certains pourraient dire qu'il en est de même pour les médias de masse. Ce qui plaiderait pour la fin de l'intoxication des médias par la publicité et donc pour son interdiction sur tous les contenus d'utilité publique et ce serait en effet une très bonne chose. Mais il s'agit là aussi d'un projet politique qui demanderait à la fois un autre courage que celui rencontré dans les gouvernements actuels, et un souci de la démocratie et de la santé mentale des citoyens qui n'est plus de mise, à l'heure où tous les politiques sont eux-mêmes pris dans cette course frénétique à la réputation en ligne, produits comme les autres, dans la compétition pour l'attention du public.

Cependant, la différence majeure entre le système publicitaire des mass médias et celui des plateformes internet réside dans l'existence d'une convention entre toutes les parties prenantes, explicite et contrôlable, d'où l'effet de stabilisation du marché comme le font toutes les conventions (voir la théorie des conventions sur le sujet, Eymard-Duvernay, Orléan, Thévenot etc.). La convention du monde des mass médias repose sur une division claire du travail entre l'institut qui effectue les mesures (Médiamétrie en France, Nielsen et d'autres ailleurs) et les intéressés à la mesure, les marques, les agences et les diffuseurs. Ce qui n'est pas le cas pour les plateformes puisqu'elles sont juges et parties, gardant tous leurs calculs opaques et vendant leurs placements via des enchères prétendument plus justes mais en réalité totalement opaques elles aussi. De plus, dans le système des mass médias, la mesure d'audience et sa valorisation en termes de prix de marché sont discutées dans un comité des parties prenantes qui s'impose à tous. Certes, il est très facile de démonter la fiction de la mesure d'audience, ses limites, ses points aveugles, etc. mais toutes les parties prenantes en sont averties, elles peuvent même contribuer à les réduire (ex. du passage de l'audimat au médiamat), et dans tous les cas, elles s'entendent pour accepter le côté arbitraire de la mesure. Ce n'est qu'ainsi que se construisent l'efficacité de la mesure et son acceptabilité, certes abusivement dès lors qu'un diffuseur en tire des conclusions du jour au lendemain pour une déprogrammation, mais c'est alors de sa responsabilité. Est-ce à dire que l'efficacité de ces placements publicitaires dans les mass media a été mieux démontrée en termes de chiffres de vente ? N'y aurait-il pas là aussi amplification abusive d'un score d'exposition vers un score de comportement ? Sans doute puisque toutes les tentatives d'objectiver ces corrélations restent très complexes en termes méthodologiques, comme cela est toujours fait à Angers par exemple en relation avec des centres commerciaux. Mais au moins, ce point aveugle est-il clair (!!!) et identifié.

Alors que pour les plateformes, aucun analytics ne vient confirmer les effets en termes de taux de conversion, qui, comme je l'ai dit, n'est pas l'objectif désormais, de toutes façons. Mais le problème pourrait être soulevé par une marque que cela ne changerait rien car les annonceurs n'ont rien à dire, ni aucun espace pour se faire entendre ni pour négocier, sauf cas de placement malheureux sur You Tube parfois, mais toujours trop tard. Mieux même, les plateformes vendent même leur savoir-faire, supposé issu de l'analyse opaque des données, aux marques, qui payent pour être éduquées à faire la bonne publicité selon Google par exemple (puisque la qualité de la publicité fait partie des critères des enchères !!). Les plateformes sont ainsi les opérateurs, les diffuseurs, l'instance de mesure mais aussi la régie publicitaire (jamais séparée de la plateforme) et l'agence de

communication potentiellement ! On atteint ainsi un dispositif idéal pour leur toute-puissance qui reste impénétrable pour les parties prenantes elles-mêmes et encore plus pour les chercheurs.

Construire des coalitions avec les marques pour contraindre les plateformes

Ce système autoréférentiel opaque et délirant fait la fortune des plateformes sur le dos des marques, il est donc fort peu probable qu'elles acceptent de dévoiler quelques sets de données pour les beaux yeux de chercheurs de service public. C'est pourquoi il faudra les y contraindre. Or, sans coalition forte qui peut attaquer leur modèle économique, les contraintes réglementaires ne seront d'aucun effet (même si je ne souhaiterais qu'être démenti).

Je propose donc que les chercheurs, les instituts de recherche et le régulateur (l'ARCOM) constituent une coalition avec les marques pour leur faire admettre qu'elles se font piller par les plateformes et qu'elles auraient tout intérêt à participer à la création d'une instance indépendante constituée par toutes les parties prenantes, chargée d'exploiter les données d'activité et de leur fournir une base, si ce n'est objective, tout au moins conventionnellement acceptable et explicite. Médiamétrie se positionne pour jouer ce rôle en prenant en compte dans ses panels (TV à domicile, TV en mobilité, internet) qu'il peut combiner désormais, l'activité des publics des plateformes à travers les boxes internet, pour des mesures élémentaires de connexion et de programmes visionnés (type Netflix) ce qui peut se faire sans l'accord des plateformes mais qui suppose cependant celui des panelistes évidemment, avec ajout de tags et de watermarks.

Pour convaincre ces plateformes, une campagne réputationnelle pour disqualifier leurs méthodes opaques serait très utile, puisque ce sont ces risques réputationnels seuls qu'elles craignent, elles aussi, auprès des investisseurs. On pourrait ensuite proposer aux marques de cesser à titre expérimental tout investissement dans telle ou telle plateforme pendant un certain temps et d'en mesurer les effets. Car, comme tout enjeu de réputation, ce point de départ est essentiel pour engendrer une contagion des comportements auprès des autres firmes, en montrant que c'est possible d'arrêter de placer de la publicité sur les plateformes avec un faible risque, et même avec un retour positif en termes de réputation. Il va de soi que les instances gouvernementales publiques, les services publics ou entreprises publiques devraient être les premières concernées par cette démarche et cesser tout placement à caractère publicitaire sur ces plateformes.

Les risques de cette menace pour les plateformes sont en fait les seuls qu'elles peuvent prendre en compte, dès lorsqu'ils portent sur leur chiffre d'affaires et sur leur réputation (financière par ricochet).

Cette stratégie ne règle pas les conditions précises de l'accès, qui devront respecter le RGPD par exemple (cf. le code de conduite de l'EDMO, European Digital Media Observatory), ni ne court-circuite les efforts du Règlement sur les Services Numériques (DSA européen) désormais opérationnel concernant les VLOPS (Very Large Online Platforms). Cette stratégie ne se contente pas de l'offre du statut de chercheur agréé prévu par le RSN, qui ne traite pas les conditions structurelles de l'accès aux données. Car ne pas traiter de la publicité et du modèle économique des plateformes qui vit précisément de ces données et qui les produit *pour* la publicité, c'est se retrouver pieds et poings liés par d'autres logiques qui dépassent de loin les exigences des chercheurs.

Or, il faut instituer cette coopération avec les plateformes durablement comme ce fut le cas avec les instituts nationaux de statistiques pour les recensements (qui servaient les gouvernements), avec les instituts de sondages pour les sondages d'opinion (qui servaient les médias et les marques) mais qui tous ont donné accès aux chercheurs et leur ont permis de produire des sciences sociales. Toutes les parties prenantes doivent être associées dans leur intérêt économique même à l'institution d'un tiers

de confiance capables de certifier les données et de produire une convention qui stabilise le marché pour le sortir de l'opacité, de la concurrence et de la prédation déloyales.

La stratégie proposée vise donc à pérenniser une instance tierce qui soit capable de valider cette nouvelle source de connaissance de l'activité sociale, réellement inédite et très riche, qui doit être contrôlée par les acteurs concernés eux-mêmes ou à défaut par une autorité publique. Nous devons être conscients de l'importance politique et cognitive de cette réflexivité d'une société sur elle-même et ne pas la traiter seulement comme une demande de privilège temporaire accordé aux chercheurs, restant dépendant du bon vouloir des plateformes.