PROTECTION DE DONNÉES PERSONNELLES

LES CONSOMMATEURS PRIS AU PIÈGE DU BIG BROTHER COMMERCIAL

janvier 2024

UFC-QUE CHOISIR • *Service des études* • *http://www.quechoisir.org*

RÉSUMÉ

Ces dernières années, la prise de conscience des consommateurs quant à la collecte et au traitement massifs de leurs données personnelles par des entreprises du quotidien a augmenté. Cependant, la plupart ignorent encore largement ce qui se passe réellement avec leurs données une fois collectées. Il est donc impératif de rendre plus transparentes les activités de commercialisation des données personnelles, actuellement très opaques.

Aujourd'hui, des entreprises de toutes tailles, dans presque tous les secteurs de la consommation hors ligne et en ligne, sont impliquées dans l'économie des données personnelles. Outre les géants du Net et les plateformes en ligne, cela concerne le commerce traditionnel, les médias, les fabricants d'objets connectés, les opérateurs de communications électroniques, les banques, les assureurs et les services publics. En plus, de nombreuses entreprises largement inconnues des consommateurs collectent et traitent également des données personnelles à grande échelle, agissant dans l'ombre. Il s'agit notamment de courtiers en données et d'intermédiaires de la publicité ciblée en ligne.

Les données personnelles des consommateurs sont collectées par l'ensemble de ces acteurs en quasi-permanence grâce à des technologies de pistage, dont les cookies tiers sont les plus connus. Ainsi, le consommateur partage ses données en un simple clic avec des centaines d'entreprises dont il n'a pour la plupart jamais entendu parler. Les tests de l'UFC-Que Choisir démontrent qu'en consultant à peine une dizaine de sites parmi les plus fréquentés en France, les données personnelles collectées sont partagées plus de 4 000 fois avec plus de 1 000 tiers. Au cœur de cette surveillance commerciale est la publicité ciblée en ligne, alimentée par le système d'enchères en temps réel, ou *real-time bidding* (RTB) en anglais, une technologie publicitaire présente sur quasiment tous les sites web et applications.

Grâce à ce pistage sophistiqué et omniprésent, les marchands de données personnelles peuvent créer un profil publicitaire précis de chaque consommateur, exposant des détails allant de ses habitudes de consommation à sa situation financière ou son état de santé mentale. Par exemple, la filiale de Microsoft spécialisée dans la publicité classe les consommateurs en fonction de plus de 650 000 traits de personnalité et situations personnelles. Nombre de ces traits sont extrêmement intimes, tels que « dysfonction érectile », « dépression », « gros acheteur de test de grossesse », « dépendance aux opioïdes », « sympathisant de syndicats », « réceptif aux messages émotionnels », « dépendance au jeu de hasard », ou encore « problèmes d'argent ».

Ainsi, les algorithmes analysent minutieusement les comportements de navigation, les préférences et les historiques d'achats pour créer des annonces sur mesure, de nature à inciter les consommateurs à succomber à des achats impulsifs. Cette pratique d'exploiter les vulnérabilités psychologiques et de créer un sentiment de besoin immédiat et de satisfaction instantanée conduit à une surstimulation constante des consommateurs, les poussant à acheter des produits dont ils n'avaient peut-être même pas connaissance auparavant. La publicité ciblée en ligne se révèle ainsi être une force motrice majeure d'une consommation déraisonnée. S'y ajoute le risque concret de piratage des données et d'actes cybercriminels au détriment de la vie privée des consommateurs, qui est multiplié par la circulation constante de données personnelles entre des milliers d'entreprises.

Malgré l'opposition de 84 % des consommateurs au pistage et à la monétisation de leurs comportements en ligne, les entreprises utilisent des pratiques trompeuses pour obtenir leur

consentement, contournant ainsi l'obligation légale du règlement général sur la protection des données (RGPD). Elles ont recours, entre autres, aux *dark patterns*, des interfaces conçues pour manipuler le libre choix des consommateurs, ou à des conditions générales interminables et inintelligibles. Et bien que les consommateurs aient théoriquement le droit de demander l'effacement de leurs informations, la réalité est tout autre. L'enquête de l'UFC-Que Choisir révèle que sur 1 040 tiers examinés, plus de la moitié (54 %) ne fournissent aucun moyen de contact ou ignorent simplement les demandes d'effacement qui leur sont adressées. Ceux qui répondent dissuadent fréquemment les consommateurs en exigeant des démarches supplémentaires fastidieuses.

Au vu de ces constats, l'UFC-Que Choisir, soucieuse de garantir aux consommateurs une réelle maîtrise de leurs données personnelles, exige des sites internet et applications qui les collectent une véritable transparence sur l'utilisation qui en est faite, et, de garantir aux consommateurs un accès et un contrôle sur les données personnelles qu'ils ont transmis à des tiers.

Par ailleurs, l'association rappelle que dans le cadre de sa campagne « Je ne suis pas une data », elle met à disposition des consommateurs sur le site respectemesdatas.fr un outil gratuit leur permettant de découvrir ce que les plateformes en ligne savent sur eux et de reprendre la main sur leurs données personnelles.

TABLE DES MATIÈRES

I. UN CADRE RÉGLEMENTAIRE SYSTÉMATIQUEMENT BAFOUÉ PAR LES ENTREPRISES

1. La collecte de données entre consentement forcé et intérêt « légitime »

Depuis 2018, la collecte et le traitement des données à caractère personnel sont encadré par le règlement général sur la protection des données (RGPD)[1], qui a remplacé la directive sur la protection des données personnelles[2] adopté en 1995.

Le RGPD établit les grands principes relatifs au traitement des données personnelles. Avant tout, il stipule que tout traitement de données doit être fait sur une des six bases légales définies dans l'article 6 :

- L'exécution d'un contrat auquel le consommateur est partie,
- Des intérêts légitimes poursuivis par l'entreprise, tant qu'ils ne prévalent pas sur les intérêts ou les libertés et droits fondamentaux du consommateur concerné,
- Le consentement par le consommateur concerné,
- Une obligation légale à laquelle le responsable du traitement est soumis,
- La sauvegarde des intérêts vitaux de la personne concernée,
- L'exécution d'une mission d'intérêt public.

Les trois dernières ne sont en général pas applicable au traitement par des entreprises à des fins publicitaires et commerciales. Les bases légales utilisées par les entreprises afin de justifier la collecte et le traitement de données personnelles sont donc principalement l'exécution d'un contrat, l'intérêt légitime, et/ou le consentement.

Les mauvaises pratiques des entreprises commencent à cette étape primordiale. En effet, les bases légales souvent mises en avant dans les conditions générales et les politiques de protection des données personnelles des entreprises ne permettent pas la collecte et le traitement intrusif de données personnelles de millions de consommateurs chaque jour.

Deux décisions de la Commission de protection des données irlandaise (DPC) visant Meta et adoptées en décembre 2022[3] ont clarifié que l'« exécution du contrat » n'était pas de base légale adaptée au traitement de données personnelles pratiqué par l'entreprise. Plus précisément, selon les décisions l'objet principal du contrat entre Meta et les consommateurs utilisant Facebook et/ou Instagram est un service de communication en ligne et non l'affichage de publicités ciblées. Le traitement de données personnelles n'est donc pas strictement nécessaire pour l'exécution du contrat.

Meta, qui disposait d'un délai de trois mois pour modifier sa politique en matière de protection des données personnelles, a annoncé peu après la décision qu'elle remplacerait la base juridique par l'« intérêt légitime »[4]. En pratique, cela signifie que l'entreprise n'a

[1] Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE.

[2] Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

[3] Data Protection Commission announces conclusion of two inquiries into Meta Ireland, DPC, 2023.

[4] How Meta Uses Legal Bases for Processing Ads in the EU, Meta, 2023.

modifié que la formulation de ses conditions d'utilisation, sans réellement toucher à ses pratiques de collecte de données.

La CJUE a donc dû clarifier, dans une décision adoptée en juillet 2023[5], que cette base légale ne pouvait pas non plus justifier le traitement excessif des données des consommateurs par le géant américain. La Cour a beau reconnaitre l'intérêt légitime de Meta d'afficher de la publicité ciblée, elle précise que celui-ci ne prévale pas sur les droits et libertés fondamentales des consommateurs, qui ne sauraient raisonnablement pas s'attendre à un tel traitement de leurs données personnelles. Par la suite, Meta n'avait plus le choix que de changer encore la base légale et de désormais demander aux consommateurs leur consentement[6].

Néanmoins, le fait de s'appuyer sur le consentement comme base juridique ne garantit toujours pas que les consommateurs soient traités de manière équitable dans l'environnement numérique. Le RGPD définit le consentement comme « toute manifestation de volonté, libre, spécifique, éclairée et univoque par laquelle la personne concernée accepte, par une déclaration ou par un acte positif clair, que des données à caractère personnel la concernant fassent l'objet d'un traitement » (article 4 paragraphe 11). Tout en prétendant adhérer à cette base juridique en demandant aux consommateurs leur consentement sous une forme ou une autre, les entreprises ne manquent pas de stratagèmes pour rendre cette décision aussi compliquée que possible, interférant ainsi avec le choix libre et éclairé des consommateurs.

Cela commence par la pratique de submerger le consommateur dans l'information à un point tel qu'il lui est impossible de s'éclairer sur les conséquences de son consentement. Une étude de l'UFC-Que Choisir montre qu'il faut en moyenne 34 minutes pour lire les conditions d'utilisation d'un site web[7]. Lire l'intégralité des conditions de Meta, un des mastodontes de la publicité en ligne, prendrait en effet pas moins de 2 heures et 45 minutes. Une autre étude publiée en 2008 (une époque où les conditions générales étaient encore plus succinctes qu'aujourd'hui) estime que si le consommateur lisait les politiques de protection des données personnelles de tous les sites web consultés au fil d'un an, il y consacrerait 35 jours de travail (244 heures)[8].

Une simple lecture des conditions générales et politiques de protection des données ne garantit pourtant en aucun cas que le consommateur lambda est en mesure de comprendre leur contenu et les conséquences. Ces textes sont généralement formulés dans un jargon juridique et gardent délibérément vagues les passages où l'utilisation des données personnelles du consommateur est expliquée. Il est par exemple courant de présenter une pratique de collecte de données comme quelque chose d'avantageux pour le consommateur, sans entrer dans les détails. Ainsi, les données personnelles seraient traitées afin de « vous présenter des publicités personnalisées susceptibles de vous intéresser » ou d'« améliorer les systèmes et logiciels existants et pour développer de nouveaux produits ».

Ensuite, les entreprises se servent des *dark patterns*, ces fameuses interfaces conçues pour manipuler le libre choix des consommateurs. Le consommateur les rencontre le plus souvent sur les bannières cookies. Certaines pratiques auparavant omniprésentes ont été identifiées par le Comité européen de la protection des données (CEPD) comme non conformes au RGPD : il s'agit notamment des cases pré-cochées, de l'absence d'un bouton de refus dès l'apparition de la bannière, et de l'usage de boutons incitatifs (par exemple, un petit bouton

[5] Arrêt de la Cour (grande chambre) du 4 juillet 2023, Meta Platforms Inc. e.a. contre Bundeskartellamt, Affaire C-252/21.

[6] How Meta Uses Legal Bases for Processing Ads in the EU, Meta, 2023.

[7] Conditions générales : À l'épreuve du chrono, UFC-Que Choisir, 2022.

[8] The Cost of Reading Privacy Policies, Aleecia M. McDonald, Lorrie F. Cranor, 2008.

« refuser » gris et en contraste faible à côté d'un grand bouton vert « accepter »)[9]. Alors que ces pratiques les plus impudentes sont désormais de plus en plus rares, de nombreux autres *dark patterns* sont pratiqués par les entreprises.

Malgré le fait que Google ait déjà été condamné par CNIL en 2019 d'avoir obfusqué les informations nécessaires pour un consentement éclairé[10], une pratique toujours courante consiste à présenter une option facile pour accepter le traitement des données, tout en obligeant les consommateurs à cliquer sur diverses sous-pages pour obtenir des informations supplémentaires et refuser le traitement des données.

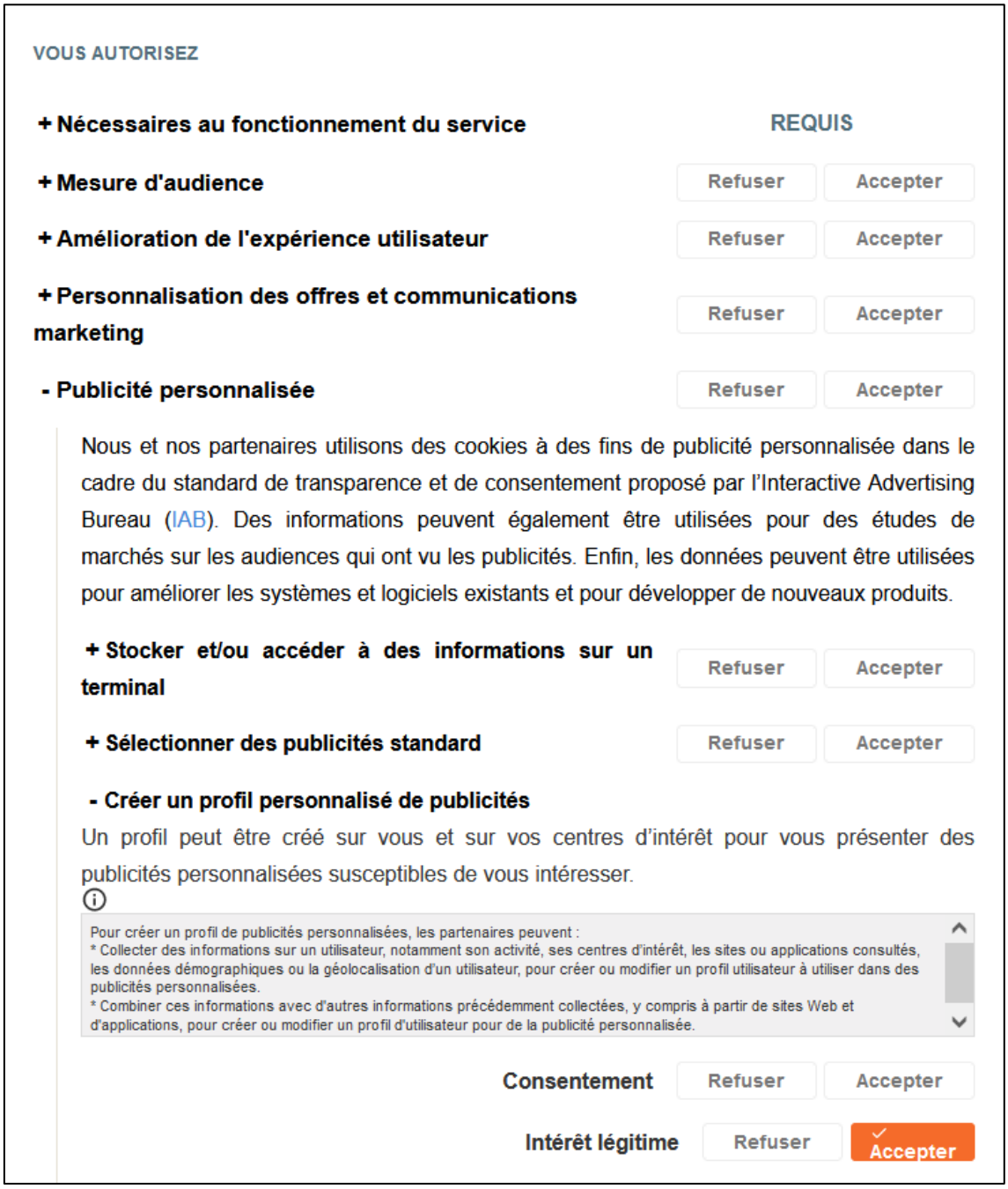
A titre d'illustration, la capture d'écran suivante montre la bannière cookies qui s'affiche en consultant le site leboncoin.fr.

La bannière présente le consommateur avec plusieurs options. Deux boutons proéminents attirent d'abord l'attention : un pour accepter tout traitement, et un autre pour « personnaliser » son choix. L'option de refuser tous les cookies tiers est également présente, mais en forme d'un lien « Continuer sans accepter » beaucoup moins visible que les boutons.

[9] Report of the work undertaken by the Cookie Banner Taskforce, CEPD, 2023.

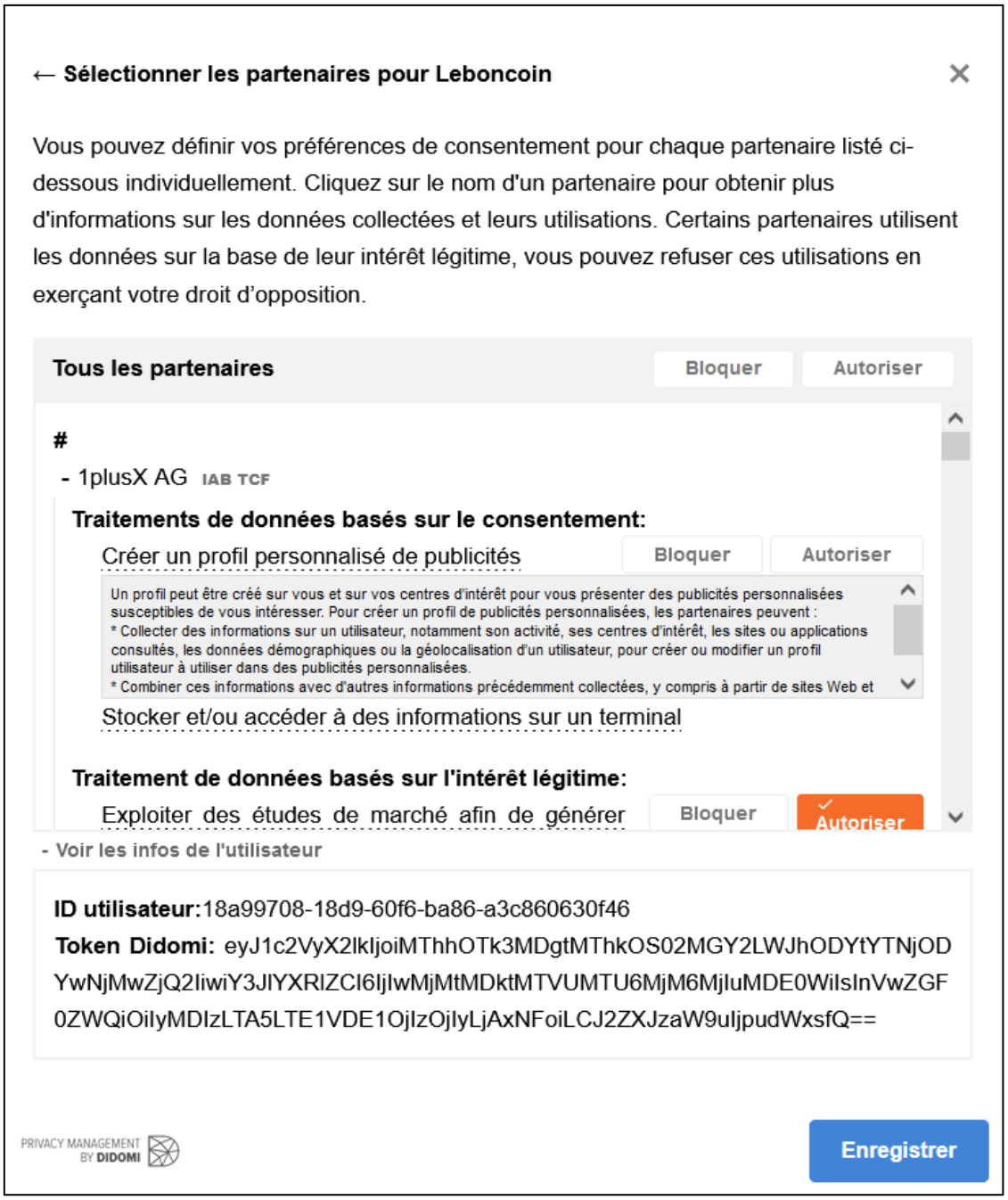
[10] The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC, CEPD, 2019.

Quand le consommateur clique sur « personnaliser », le site ouvre une autre fenêtre sur laquelle il faut cliquer plusieurs dizaines de fois pour pouvoir voir toutes les informations. La capture suivante ne présente qu'un détail de cette fenêtre, qui est dans son intégralité, avec toutes les explications affichées, six fois plus longue.



A noter qu'il faut faire défiler jusqu'à la fin de la fenêtre pour trouver l'option de tout refuser.

A nouveau sur la bannière initiale, lorsque le consommateur clique sur le lien « partenaires », il tombe sur encore une autre fenêtre, encore plus longue que la précédente.



Le principe est le même, sauf qu'ici, il faut cliquer des centaines de fois pour afficher toutes les informations concernant les 266 partenaires de leboncoin.fr. Un exercice qui serait pourtant bien inutile, car le consommateur lambda n'a jamais entendu parler de la quasi-totalité de ces entreprises, et les explications du traitement précis qu'elles effectuent restent entièrement vagues.

L'exemple des bannières cookies montre que les outils de consentement mis en place par les entreprises vont, sous le couvert de la transparence, à l'encontre de ce qu'exige le GDPR, à savoir une décision libre et informée de la part du consommateur.

Par conséquent, en 2022, l'Autorité belge de protection des données (APD) a estimé que le *Transparency and Consent Framework* (TCF) n'était pas conforme à de nombreuses dispositions du RGPD. Le TCF est la technologie qui sous-tend le partage et traitement des

données personnelles dans le cadre des enchères en temps réel (RTB). Il a été développé par le *Interactive Advertising Bureau* (IAB), qui est une organisation de lobbying regroupant les acteurs de la publicité en ligne. La décision a fait l'objet d'un appel de la part du IAB devant la CJUE[11].

Il convient de souligner que certains sites, bien qu'ils affichent une bannière cookies, ignorent complètement le choix du consommateur. Soit ils placent des cookies sur l'appareil du consommateur avant que celui-ci donne ou refuse son consentement, soit ils en placent malgré le fait qu'il l'a refusé. Amazon a par exemple été condamné de cette pratique par la CNIL en 2020[12].

Enfin, Meta a commencé en novembre 2023 de proposer aux utilisateurs de Facebook et d'Instagram de souscrire un abonnement payant s'ils souhaitent refuser le traitement de leurs données personnelles. Ainsi, les utilisateurs refusant d'être pistés doivent souscrire un abonnement débutant à 9,99 € par mois, et qui atteindra jusqu'à 20,99 € par mois en 2024[13]. Cette pratique est susceptible de ne pas être conforme avec le RGPD, qui exige un consentement libre et univoque[14]. En outre, l'abonnement proposé par Meta repose sur plusieurs pratiques commerciales trompeuses et agressives[15].

Il s'ensuit que le cadre légal actuel, aussi ambitieux qu'il ait pu être au départ, est trop inefficace pour garantir un environnement numérique respectueux de la vie privée des consommateurs. Il a fallu plus de cinq ans, après le dépôt initial de la plainte en 2018, pour que soit adoptée la décision précisant que Meta avait collecté et traité les données des consommateurs sans base juridique valable. Et l'idée que les consommateurs peuvent consentir librement et en connaissance de cause au traitement de leurs données personnelles reste une fiction juridique. En effet, à ce jour peu d'éléments indiquent que les mesures réglementaires décrites ci-dessus ont entraîné des changements significatifs et structurels en ce qui concerne la collecte et le traitement des données personnelles des consommateurs par des entreprises[16].

2. L'exercice de leurs droits rendu chimérique pour les consommateurs

Le RGPD ne se contente par ailleurs pas d'établir le cadre légal pour le traitement des données, mais il accorde également aux consommateurs un ensemble de droits de contrôle de leurs données personnelles : il s'agit notamment du droit d'opposition à un traitement de ses données (article 21), du droit d'accès à ses données (article 15), et du droit à l'effacement (article 17). En pratique, il est cependant extrêmement difficile voire impraticable pour les consommateurs d'exercer ces droits, en partie à cause de la complexité de l'écosystème de l'industrie des données et en partie à cause des pratiques ressemblant à celles que les entreprises utilisent pour inciter les consommateurs à partager leurs données.

---

[11] L'APD remet de l'ordre dans l'industrie de la publicité en ligne : IAB Europe est tenue responsable d'un mécanisme qui viole le RGPD, Autorité de protection des données, 2022.

[12] Cookies : le Conseil d'État valide la sanction de 2020 prononcée par la CNIL contre Amazon, CNIL, 2022.

[13] Le tarif de base débute à 9,99 € pour l'utilisation d'un seul compte souscrit sur la version web de la plateforme, et atteindra, à partir de mars 2024, 20,99 € pour un utilisateur faisant usage simultané de Facebook et Instagram, et souscrivant via l'application mobile.

[14] Article 4 du

[15] Facebook et Instagram : Plainte européenne contre Meta pour pratiques commerciales trompeuses et agressives, UFC-Que Choisir, 2023.

[16] Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

Il est pourtant évident que la majorité des consommateurs ne sont pas d'accord avec la collecte et le partage de leurs données personnelles à des fins publicitaires. Une enquête exclusive du Bureau européen des unions de consommateurs (BEUC)[17], dont l'UFC-Que Choisir est membre, révèle que seulement 14 % des Français trouvent acceptable que les applications et les sites web ont le droit de pister et monétiser leur comportement. En outre, seulement 15 % trouvent acceptable que les entreprises peuvent les cibler avec des publicités basées sur des informations concernant leur situation de vie, leurs vulnérabilités et leurs faiblesses.

D'autres enquêtes confirment ces résultats. Selon un sondage effectué par YouGov en 2021, 89 % des Français ne sont pas d'accord que leurs données sont utilisées afin de leur afficher des publicités personnalisées[18]. En même temps, alors que 90 % des Français estiment qu'il est nécessaire de connaître l'identité des entreprises susceptibles de suivre leur navigation sur le web, 23 % n'ont jamais eu d'information sur ces entreprises et 45 % trouvent que les informations qu'ils ont sont insuffisantes[19]. Et depuis qu'Apple a soumis l'accès à l'IDFA (*Identifier for advertisers*), un identifiant des consommateurs permettant aux annonceurs de les cibler, au consentement explicite de chaque consommateur, 75 % l'ont refusé[20].

Dans la vie quotidienne des consommateurs, il est toutefois souvent impossible de protéger leurs données. Par exemple, le droit de s'opposer au traitement de ses données est traduit par les entreprises en une case à cocher, avant l'utilisation de leur service, pour accepter l'intégralité de leurs pratiques en matière de données personnelles. Sans cet accord, il est impossible pour le consommateur d'utiliser le service. En réalité, le droit d'opposition n'est donc pas plus qu'un droit de ne pas utiliser la plupart des services numériques, ce qui correspond au « droit » de ne pas participer à l'économie et la société numérique du 21ème siècle. Dans ces conditions, le consentement donné par les consommateurs ne peut guère être considéré comme libre.

Si le consommateur change d'avis et souhaite mettre fin au traitement de ses données qu'il a accepté, il peut tout d'abord retirer son consentement. En théorie, cette étape est encore relativement simple du point de vue technologique : il suffirait de supprimer les cookies ou de désinstaller les applications concernées. En pratique, certains cookies sont difficiles à supprimer ou sont placés à nouveau sur l'appareil du consommateur lors d'une prochaine consultation d'un site web. Quant aux applications, certaines ne peuvent pas être désinstallées, comme le système d'exploitation ou des logiciels préinstallés par le fabricant.

En tout état de cause, retirer le consentement n'implique pas que les données déjà collectées soient automatiquement effacées. En effet, obtenir cet effacement s'avère souvent impossible. Le consommateur a beau pu consentir au traitement de ses données en un clic, il est obligé de faire une demande individuelle auprès de chaque responsable du traitement des données afin d'avoir accès aux données détenues et de demander leur effacement.

Tout d'abord, le consommateur doit connaître les noms des entreprises détenant ses données. Comme il sera démontré dans les parties suivantes de cette étude, la quasi-totalité des tierces parties collectant les données sont pourtant inconnus par les consommateurs. Mais supposons qu'un consommateur ait une vue d'ensemble parfaite de toutes les entreprises concernées. Pour la présente étude, nous avons utilisé la liste de 1 040 tiers

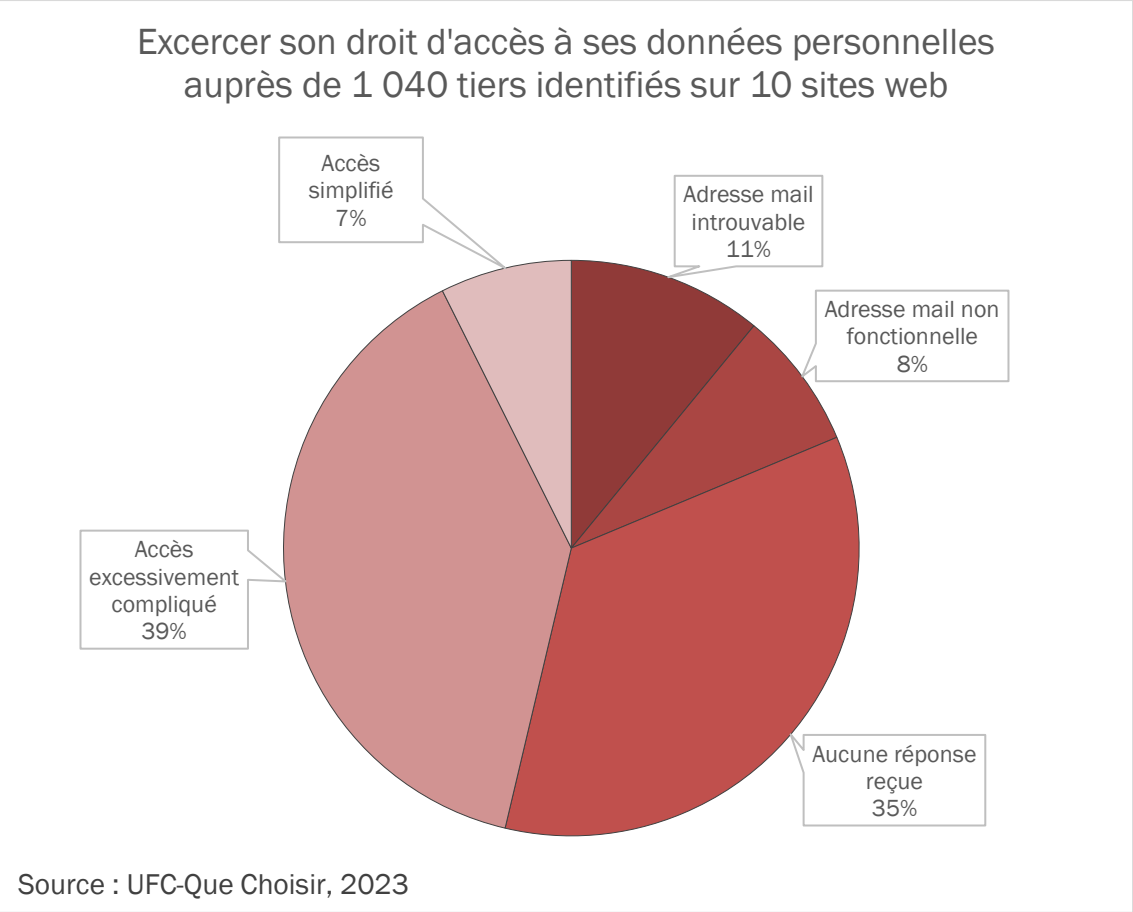
[17] Enquête réalisée en 2023 par le BEUC auprès de 615 consommateurs français. Connected, but unfairly treated: Consumer survey results on the fairness of the online environment, BEUC, 2023.

[18] Do people really want personalised ads online?, Global Witness, 2021.

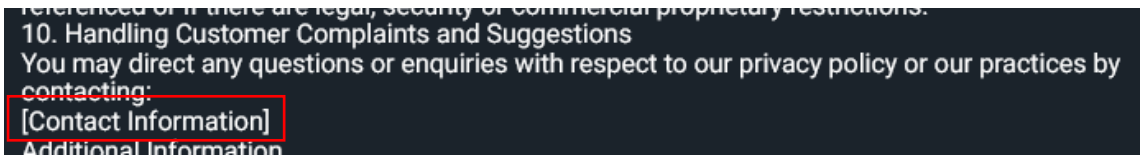
[19] Les Français et la réglementation en matière de cookies, Ifop pour la CNIL, 2019-2022.

[20] App Tracking Transparency Opt-In Rate – Monthly Updates, Flurry, 2022.

collectant les données personnelles sur un échantillon de 10 sites web[21] pour simuler le parcours d'un consommateur demandant l'effacement de ses données auprès de toutes ces entreprises.



Dans une première étape, il faut trouver le contact de tous les délégués à la protection des données (DPD) désignés par les entreprises. Alors que des liens vers leurs sites web peuvent être retrouvés dans la liste de « partenaires » affichée, 26 de ces sites ne fonctionnait pas et 715, donc plus de deux tiers, ne proposaient pas leur politique de protection des données en français. Sur 117 sites (11 %), l'adresse électronique ou un autre moyen de contact était introuvable. Sur certains de ces sites, la politique de protection des données a visiblement été copié-collé sans prendre la peine de remplir les espaces réservés avec les informations correspondantes, comme l'illustre la capture d'écran suivantes.



[21] Sur un échantillon de 10 sites web, l'UFC-Que Choisir a répertorié la présence de 1 040 tiers, cumulant un total de 4 332 occurrences (certains d'entre eux étant présents sur plusieurs sites). Les tests ont été effectués en juin 2023 sur les 10 sites suivants : 20minutes.fr, allocine.fr, cdiscount.fr, leboncoin.fr, lemonde.fr, marmiton.org, meteofrance.com, orange.fr, programme-tv.net, yahoo.com. Ces sites sont classés parmi les 50 sites les plus consultés en France, selon les mesures de Similarweb. Voir détails dans le chapitre III.1.b.

En total, sur les 1040 entreprises, 923 adresses mails ont pu être identifiées et contactées avec une demande d'accès et d'effacement de données personnelles. Pourtant, 83 (8 %) de ces adresses électroniques ne fonctionnaient pas et 374 (35 %) des mails envoyés sont restés sans aucune réponse au bout du délai de 30 jours fixé par le RGPD. Moins de la moitié des demandes (46 %) ont donc déclenché une réaction de la part des responsables du traitement des données.

Sur les 465 réponses que nous avons reçues, 189 sollicitaient des informations supplémentaires du consommateur, telles que l'identifiant publicitaire du navigateur ou de l'appareil mobile utilisé. En raison de la pseudonymisation de données[22] et de l'utilisation d'identifiants publicitaire, un simple nom ou une adresse électronique ne suffisait souvent pas aux entreprises pour identifier un consommateur. Ainsi, la demande d'un identifiant technique était une étape logique pour répondre aux requêtes. Cependant, la complexité de cette démarche variait. Dans 79 cas, le DPD fournissait un lien vers un site permettant au consommateur d'afficher de manière simplifiée l'identifiant du navigateur, ce qui peut être considéré comme la méthode la plus claire et facile pour le consommateur. Dans les autres réponses, les démarches demandées au consommateur étaient plus complexes : 11 réponses incluaient un lien défectueux, 87 demandaient au consommateur de trouver lui-même l'identifiant publicitaire dans les paramètres du navigateur ou du système d'exploitation, et 11 exigeaient même des justificatifs excessifs tels qu'une pièce d'identité ou une déclaration sur l'honneur.

Ensuite, 227 réponses du DPD confirmaient simplement que l'adresse électronique de l'expéditeur ne pouvait pas être trouvée dans les bases de données de l'entreprise. En raison de l'opacité de leurs modèles d'entreprise, il est difficile de déterminer avec certitude comment ces entreprises traitent les données du consommateur. Cependant, il est probable qu'elles utilisent des identifiants techniques de la même manière que les autres entreprises et tentent simplement d'exploiter l'ignorance du consommateur en ne mentionnant pas ce point dans leur réponse.

Enfin, 49 réponses se contentaient d'envoyer un message générique renvoyant vers la politique de protection des données de l'entreprise, sans tenir compte de la demande spécifique.

Parmi les 647 entreprises potentiellement détenant des données personnelles, seulement 79 (7 %) proposent donc une démarche relativement simple et convaincante.

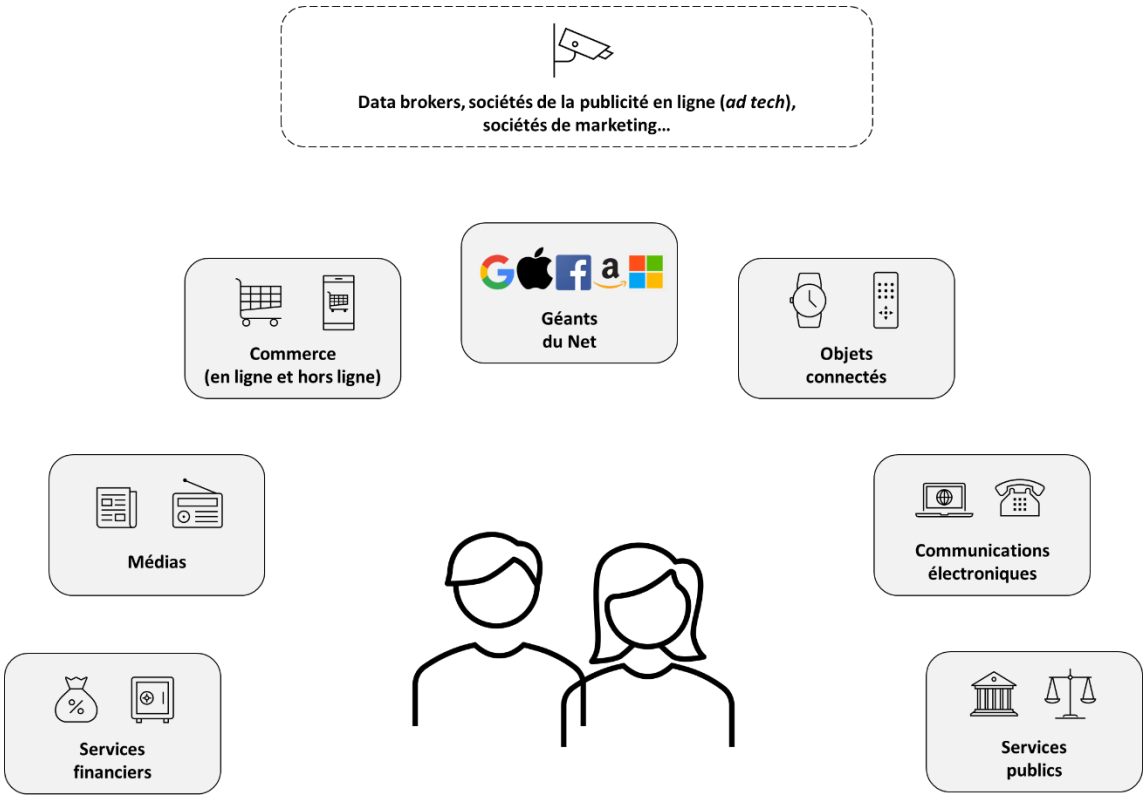
[22] Voir chapitre III.2.b pour plus de détails.

II. L'ÉCOSYSTÈME DE L'INDUSTRIE DES DONNÉES PERSONNELLES

Ces dernières années, les consommateurs ont de plus en plus pris connaissance du fait que leurs données personnelles sont collectées et traitées à grande échelle par les entreprises avec lesquelles ils interagissent au quotidien, comme les réseaux sociaux, les moteurs de recherche ou les plateformes en ligne. En revanche, la plupart des consommateurs ignorent, et donc sous-estiment encore largement, ce qu'il se passe exactement avec leurs données après qu'elles ont été collectées. Il est donc opportun de rendre plus visibles les activités de commercialisation de données personnelles[23], qui sont très opaques.

Il ne s'agit ici pas uniquement d'activités des géants du Net. En effet, des milliers d'entreprises, à travers pratiquement tous les domaines économiques, collectent, traitent, partagent et commercialisent les données personnelles. Et elles n'agissent pas indépendamment les unes des autres. Au contraire, c'est l'interconnexion d'acteurs traditionnels et émergents de l'industrie qui a créé un tout nouvel écosystème permettant le suivi de milliards de consommateurs à des fins commerciaux.

Le graphique suivant cartographie cet écosystème de l'industrie des données personnelles.



[23] Cette partie est principalement basée sur les études suivantes : Corporate surveillance in everyday life, Cracked Labs, 2017 ; Out of Control, Forbrukerrådet, 2020 ; Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

1. Les acteurs numériques et traditionnels avec qui les consommateurs partagent des données personnelles

Avec certaines des entreprises collectant des données personnelles, les consommateurs ont des relations commerciales plus ou moins directes, car ils achètent leurs produits et utilisent leurs services. Les consommateurs leur fournissent donc directement des données (par exemple leur nom et adresse dans le cadre d'un contrat) ou en génèrent lors de l'utilisation des services. Alors que certaines de ces entreprises, comme les réseaux sociaux, sont de plus en plus sous les feux des projecteurs, d'autres sont beaucoup moins sur le radar des consommateurs en ce qui concerne la protection de leurs données personnelles.

a. Les géants du Net

Quand on pense à la collecte de données, on pense en premier aux géants du Net. Il s'agit des opérateurs de réseaux sociaux comme Meta (Facebook, Instagram, WhatsApp) ByteDance (TikTok), Snap (SnapChat) ou X (ex-Twitter), de vendeurs en ligne (Amazon, AliExpress), de services et matériel informatiques comme Alphabet (Google, Android), Microsoft, Apple ou Samsung, ou de services de mobilité et de voyage (Uber, Booking, Airbnb).

Ces grandes plateformes offrent souvent une multitude de services et produits interconnectés, et ont établi une position dominante dans l'économie numérique. Par exemple, Amazon est, *inter alia*, un site de vente en ligne, une place de marché, un hébergeur (Amazon Web Services), et un fournisseur de contenu en ligne (Amazon Prime, Twitch). Il est le leader du marché du e-commerce en France[24] et du marché global de l'hébergement Internet[25].

Leurs services peuvent jouer un rôle de contrôleur d'accès à d'autres services (comme l'Apple App Store et le Google Play Store dont on a besoin pour installer des applications sur le smartphone) ou peuvent être intégrés dans un grand nombre de services tiers (comme le bouton j'aime de Facebook qu'on trouve sur des milliers de sites web), ce qui leur donne un accès privilégié à des données personnelles de centaines de millions, voire de milliards de consommateurs dans le monde entier.

Beaucoup de consommateurs s'imaginent que les géants du Net vendent leurs données personnelles de manière très classique : la remise d'un dossier de données en échange d'argent. Cette image ne correspond pourtant pas exactement au fonctionnement du marché numérique. En effet, les données étant leur principal atout, les géants du Net n'ont aucun intérêt de les céder à leurs concurrents. Cependant, cela ne les empêche pas de les commercialiser en vendant l'accès à leur exploitation dans des conditions contrôlées par eux-mêmes. Ainsi, les données retenues par Google ou Facebook peuvent par exemple être utilisées par des marques à des fins publicitaires, sans que celles-ci y aient un accès direct.

b. Le commerce traditionnel (en ligne et hors ligne)

Les acteurs de l'économie numérique ont beau être les plus avancés en matière de collecte de données, ils ne sont pas les seuls ni les premiers.

Aujourd'hui, les entreprises, des plus petites aux plus grandes, dans pratiquement tous les domaines de la consommation sont impliquées dans l'économie des données personnelles.

---

[24] Baromètre de l'audience du e-commerce : 1er trimestre 2023, Fevad, 2023.
https://www.fevad.com/1er-trimestre-2023-barometre-de-laudience-du-e-commerce/

[25] Q1 Cloud Spending Grows by Over $10 Billion from 2022; the Big Three Account for 65% of the Total, Synergy Research Group, 2023.
https://www.srgresearch.com/articles/q1-cloud-spending-grows-by-over-10-billion-from-2022-the-big-three-account-for-65-of-the-total

Bien avant la disponibilité généralisée d'outils numériques, des entreprises dans des domaines traditionnels, comme les compagnies aériennes ou les chaines de commerçants en alimentation ou de vêtement, ont commencé à mettre en place des programmes de fidélité leur permettant de collecter des informations sur les transactions de leurs clients. D'autres secteurs économiques ont suivi cette tendance, et des programmes de fidélité sont désormais omniprésents. Bien entendu, il n'y pas de séparation stricte entre distribution classique et en ligne : un consommateur peut utiliser sa carte de fidélité de la même enseigne dans les magasins physiques et sur le site web de celle-ci.

Les données collectées comprennent les données fournies directement par les consommateurs (leur nom, adresse, adresse électronique, numéro de téléphone, date de naissance etc.), mais également les données transactionnelles recensées à travers les achats effectués : l'historique des produits achetés, l'heure et le lieu des achats (magasin physique ou en ligne), mode de paiement utilisé etc. Grace à l'analyse de ces données, il est possible de disposer d'informations sur les préférences et le comportement des consommateurs. Ces analyses sont complétées par des données supplémentaires collectées via des technologies de pistage en ligne, qui ne sont désormais plus réservées aux géants du Net mais, au contraire, accessible à toutes les entreprises.

Les bases de données ainsi établies sont souvent croisées et agrégées au sein d'un groupement où d'une alliance d'entreprises. Ces données sont également partagées avec des entreprises spécialisées dans les études de marché, afin d'aider les commerçants à optimiser leurs pratiques commerciales.

c. Les médias

Les médias comprennent les éditeurs de médias traditionnels (livres, journaux et magazines, radio, télévision etc.), les fournisseurs de contenu vidéo (Netflix, Amazon Prime, YouTube etc.) et audio (Spotify, Deezer etc.) en ligne, les distributeurs de jeux vidéo (Ubisoft, Sony, Microsoft, Tencent, Activision Blizzard etc.), ou encore les entreprises de services web (Yahoo!, Google etc.).

Les modèles économiques dans le secteur s'appuient depuis longtemps non seulement sur la vente directe, mais aussi sur la publicité. La mesure d'audience (via les abonnements ou des sondages, par exemple) est donc antérieure à l'essor des technologies numériques, mais a largement été amplifiée par celles-ci. Aujourd'hui, les éditeurs de contenu en ligne, qu'il s'agisse de contenus sur des sites web ou dans des applications, ont à leur disposition des outils pour collecter en temps réel des données démographiques et comportementales détaillées sur leur audience.

d. Les objets connectés

Les objets connectés sont de plus en plus courants dans les foyers. Avant tout, le smartphone s'est établi comme compagnon incontournable des consommateurs : 87 % des Français en possédait un en 2022[26]. Il permet une collecte de données exhaustive au niveau du système d'exploitation (typiquement, le consommateur doit se connecter avec son Apple ID ou son compte Gmail avant de pouvoir utiliser toutes les fonctionnalités), des applications et de la navigation en ligne.

Au-delà du smartphone, 40 % des Français sont équipés d'au moins un objet connecté[27]. Il s'agit par exemple de montres connectées, de jouets, de liseuses, et de l'ensemble des

[26] Baromètre du numérique 2022, CRÉDOC, 2023.

[27] Ibid.

objets de la maison connecté (téléviseurs, ampoules, prises, aspirateurs robotisés, réfrigérateurs, enceintes, thermostats, vidéosurveillance etc.).

S'y ajoutent notamment les moyens de transport individuel comme la voiture, devenue de plus en plus connectée ces dernières années, mais également des objets dont le consommateur n'est pas le propriétaire et qu'il utilise en fonction de ses besoins dans le cadre d'un service de location (les vélos et trottinettes en libre-service, par exemple).

Le fait que certains de ces objets peuvent communiquer entre eux ouvre la voie au partage des données personnelles collectées entre les différents fabricants, ainsi qu'avec des prestataires tiers qui fournissent par exemple les logiciels ou l'infrastructure permettant l'interopérabilité entre objets connectés. A nouveau, les géants du Net y jouent un rôle important : pour gérer ce réseau d'objets, les consommateurs ont typiquement recours soit à leur smartphone, soit à un assistant virtuel (comme Siri d'Apple, Alexa d'Amazon ou l'Assistant Google), qui interagissent donc avec, et ont par conséquent accès aux données de, tous ces objets.

e. Les opérateurs de communications électroniques

Les opérateurs de communications électroniques sont des acteurs de l'écosystème de la collecte de données que les consommateurs ont moins en vue. Pourtant, en tant que fournisseur d'accès à l'infrastructure d'Internet fixe et mobile, ils ont accès et contrôlent les flux de données de tous les consommateurs. Cette position cruciale leur permet de collecter, par exemple, des métadonnées concernant leurs connexions (temps, lieu et durée des connexions), les appareils utilisés, et l'historique de leurs communications (appels et messages).

f. Les fournisseurs de services financiers

Vu l'importance que leurs décisions peuvent avoir dans la vie des consommateurs, les banques, assurances et fournisseurs de cartes de crédit méritent une mention spéciale en ce qui concerne la collecte de données personnelles. Elles détiennent des données sensibles concernant l'identité des consommateurs, qu'elles collectent dans le cadre de la prévention de la fraude, ainsi que leurs transactions et leur situation financière (biens et revenus, paiements, habitudes de consommation etc.). Ces données sont par exemple utilisées afin d'effectuer des évaluations des risques-clients, avec des conséquences importantes pour les consommateurs (octroi d'un prêt bancaire, tarif et conditions d'une assurance etc.).

Ces acteurs étant traditionnellement plutôt conservateurs et lourdement réglementés, ils ont été comparativement lents à adopter des nouvelles technologies numériques de collecte et traitement de données, mais l'émergence de nouveaux acteurs, comme les fintechs, les ont poussés à rattraper leur retard et s'ouvrir à d'autres acteurs de l'économie des données personnelles.

g. Les services publics

Enfin, il est essentiel de souligner que les données personnelles des consommateurs sont également collectées par les services publics : le système de santé, le système éducatif, les services sociaux etc. Bien que le caractère public de ces services limite considérablement la probabilité que des données sensibles soient partagées avec des acteurs commerciaux, il n'y a pas de garantie absolue qu'elles ne seront jamais utilisées à des fins commerciales. Par exemple, dans le cadre des négociations européennes concernant la création d'un espace de données de santé européen qui sont actuellement en cours, l'option de permettre le partage de données avec des entreprises développant des applications de bien-être est à l'étude.

2. Les tierces parties largement inconnues par les consommateurs

Le consommateur partage ainsi ses données personnelles avec en moyenne au moins des dizaines, voire des centaines d'entreprises. Il peut effectivement être choquant de se rendre compte du simple nombre de ces entreprises, mais au moins les consommateurs sont en principe conscients de la relation commerciale qu'ils entretiennent avec elles (dans le cadre de l'achat de leurs produits ou de l'utilisation de leurs services).

Toutefois, il existe un nombre important d'entreprises qui collectent et traitent également des données personnelles à grande échelle mais qui agissent dans l'ombre, et restent donc inconnues par la plupart des consommateurs mais aussi des décideurs.

Au cœur de leur activité : la publicité ciblée fondée sur des technologies publicitaires numériques. Celle-ci constitue la base du modèle économique derrière la plupart des services non-payants comme les réseaux sociaux, les journaux gratuits en ligne, les applications gratuites etc. Toutefois, il faut noter que le fait qu'un service est payant ne garantit pas l'absence de pistage. Au contraire, la majorité des services et applications payants ont également recours à des technologies de traçage.

a. Les acteurs de la publicité ciblée

La relation entre annonceurs et éditeurs constitue le cadre général de la publicité tant classique que ciblée.

D'un côté, les annonceurs sont les entreprises qui cherchent à faire de la publicité pour leur produits et services afin d'attirer les consommateurs. Plus de 44 % des dépenses globales dans la publicité au niveau global (152 milliards d'euros en 2021) proviennent des trente annonceurs les plus importants au monde (dont LVMH, L'Oréal, Stellantis et Renault-Nissan-Mitsubishi), qui sont principalement actifs dans les domaines de l'alimentation, de produits d'hygiène et de beauté, et de voitures[28].

De l'autre côté, les éditeurs sont les entreprises proposant des contenus aux consommateurs et générant une partie ou l'intégralité de leurs revenus grâce à l'espace publicitaire présent sur leur site ou dans leur application, et dont se servent les annonceurs. Dans l'espace numérique, les éditeurs sont par exemple les réseaux sociaux, les journaux et magazines en ligne, les plateformes de streaming (YouTube, Twitch etc.), ou les jeux vidéo et d'autres applications de divertissement (en particulier sur smartphones)[29]. Dans la mesure où les éditeurs cherchent à accroître leur propre base d'utilisateurs, ils peuvent également jouer le rôle d'annonceurs.

Les relations commerciales entre ces deux parties sont facilitées par différents intermédiaires, qui ensemble forment la filière de la technologie publicitaire, surnommé adtech (pour *advertising technology* en anglais). Ces intermédiaires fournissent l'infrastructure technologique nécessaire pour la publicité ciblée. C'est à ce point focal entre annonceurs et éditeurs que l'agrégation et l'analyse des données personnelles et le profilage des consommateurs a lieu, afin de déterminer à qui, quand, et comment une publicité est affichée. Les données personnelles utilisées proviennent en partie des éditeurs et des annonceurs et en partie des intermédiaires adtech eux-mêmes.

Cela se produit sans aucune connaissance ou compréhension de la part des consommateurs. En effet, de nombreux acteurs complètement inconnus par les consommateurs sont impliqués dans ce processus. Le système est tellement complexe que

[28] Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

[29] Ibid.

les intermédiaires sont organisés en plusieurs rangs, c'est-à-dire que certains ne proposent pas leurs services directement aux éditeurs ou annonceurs, mais à d'autres intermédiaires.

En quelques mots, les principaux types d'entreprises dans l'écosystème de l'adtech sont les suivants :

- Les *supply-side platforms* (SSP) permettent aux éditeurs d'organiser, d'optimiser et de vendre de l'espace publicitaire sur leurs sites ou dans leurs applications.
- Les *demand-side platforms* (DSP) proposent aux annonceurs des services d'achat automatisé d'espace publicitaire afin d'afficher leurs publicités de manière ciblée.
- Les *ad exchanges* (bourses publicitaires en français) facilitent l'entremise de transactions entre DSP (annonceurs) et SSP (éditeurs) ; leur fonction est aujourd'hui souvent exercée directement par ces derniers.
- Les *data management platforms* (DMP), ou plateformes de gestion de données, permettent à tous les acteurs de l'adtech, et notamment aux SSP et DSP, l'analyse algorithmique de données personnelles qui est nécessaire pour un bon ciblage de la publicité.
- Les entreprises qui mesurent le rendement des publicités ciblées, par exemple pour voir quels consommateurs ont cliqué sur la publicité ou acheté le produit.
- Les entreprises qui effectuent, typiquement au nom des éditeurs ou des annonceurs, une vérification du processus publicitaire afin de détecter les escroqueries (comme des mesures manipulées) et de protéger la marque (par exemple, éviter que la publicité soit affichée sur un site qui ne correspond pas à l'image que la marque souhaite communiquer).

Toutes ces acteurs fournissent des solutions technologiques pour mettre en œuvre le système de la publicité ciblée, mais il existe un autre type d'acteurs dont le rôle est d'alimenter le système en données personnelles : les *data brokers*, ou courtiers en données en français. Leur activité principale est la collecte et l'agrégation de données personnelles de différentes sources hors ligne et en ligne (sources publiques, pistage web, achat direct de données etc.), leur analyse, et la création de segments publicitaires et de profils individuels, qui permettent de cibler la publicité affichée selon les comportements, préférences et caractéristiques socio-démographiques des consommateurs. Les données traitées proviennent de l'ensemble d'entreprises de tous les secteurs économiques présentées dans la partie précédente, et sont mis à la disposition de tous les autres acteurs impliqués dans la publicité ciblée.

Certains data brokers existent depuis des décennies et ont donc pu accumuler des bases de données historiques importantes. Les technologies numériques de pistage en ligne et d'analyse algorithmique n'ont fait qu'exploser le nombre et l'ampleur des données personnelles auxquelles ils ont accès.

Ce qu'ils ont tous en commun, c'est que les données qu'ils détiennent ne leur ont pas été fournies par les consommateurs mais ont été collectées de manière clandestine. Par conséquent, les consommateurs ne sont pas du tout conscients de cette collecte massive et ce traitement invasif de données personnelles concernant leur vie privée. En effet, qui a déjà entendu parler d'entreprises comme Axciom, Experian or Oracle ? Il s'agit pourtant d'entreprises avec des milliers de collaborateurs, actifs partout dans le monde, et possédant les données personnelles de milliards de consommateurs[30].

[30] A titre d'exemple, selon les informations officielles publiées sur son site web ou dans ses rapports annuels, Axciom emploie 6 200 personnes et détient des données personnelles de plus d'un milliard de personnes.

Les différents rôles dans cet écosystème de la publicité ciblée ne sont pas strictement séparés entre des entreprises distinctes. En effet, de nombreuses entreprises, et notamment les grandes plateformes, y jouent plusieurs rôles. En particulier, la position de Google et Meta ne peut être sous-estimée : non seulement sont-ils à la fois éditeurs et annonceurs, mais ils pratiquent également les principales activités intermédiaires. Grâce à leur position dominante, ils contrôlent ainsi une partie importante de toute la chaine de valeur de l'adtech. Dans une moindre mesure, TikTok, Twitter, Apple, Amazon, Netflix et Spotify jouent également un rôle important.

b. Le système des enchères en temps réel

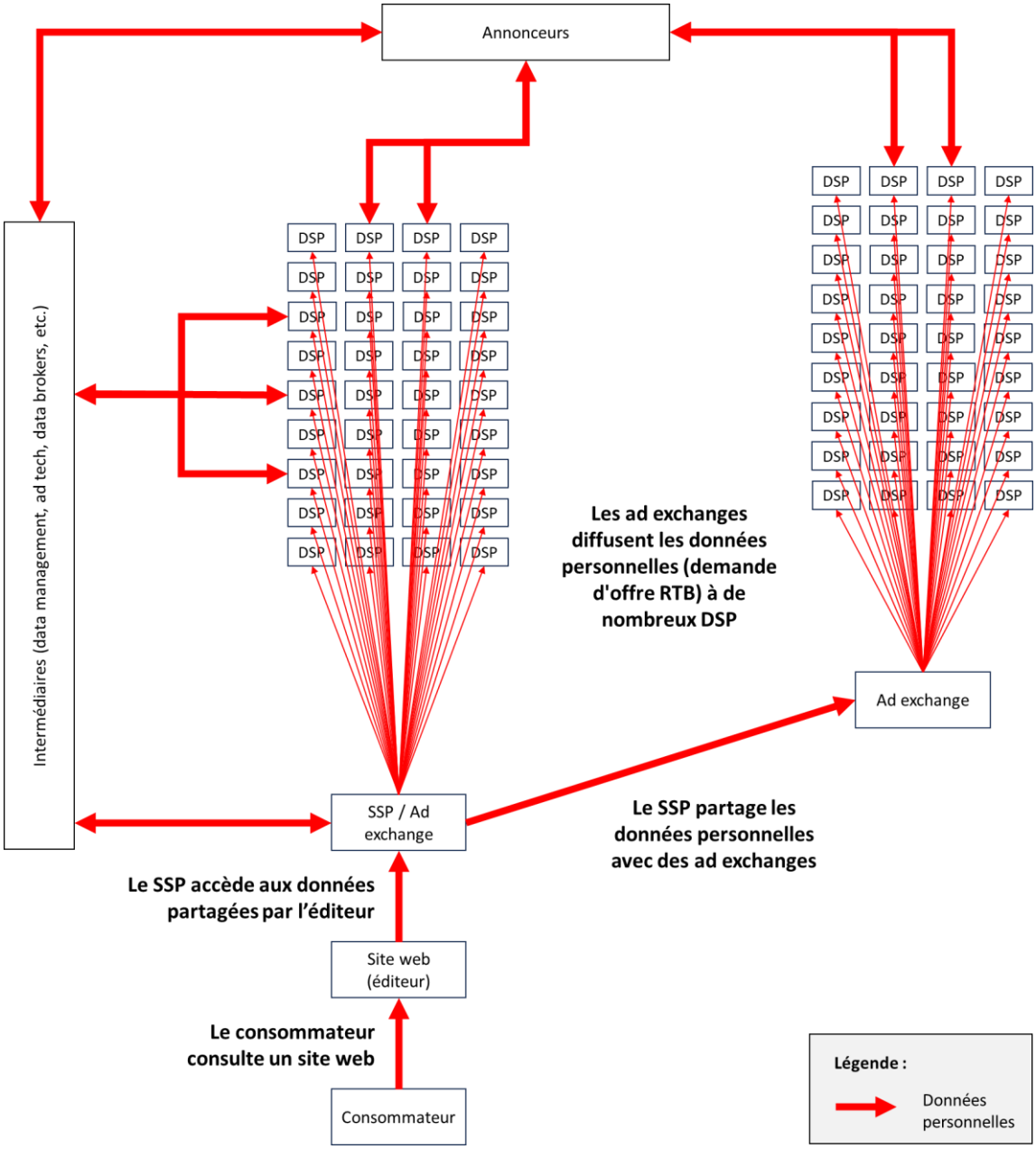
Une pierre angulaire de cet écosystème de la publicité en ligne, qui met en relation tous ces acteurs, est le système des enchères en temps réel, ou *real-time bidding* (RTB) en anglais. Ce système est au cœur du pistage massif des consommateurs. En effet, chaque jour les données personnelles d'un consommateur européen sont en moyenne exposées 376 fois par le système du RTB[31].

L'image que certains consommateurs peuvent avoir d'une transaction commerciale traditionnelle lors de laquelle un commissaire-priseur vend aux enchères des fichiers de données personnelles à une salle remplie d'hommes d'affaires ne reflète pas la réalité technologique du RTB. Chaque fois qu'un consommateur voit une publicité sur un site web ou dans une application, c'est très probablement le résultat de ce processus d'enchères complexe, qui se déroule de manière automatisée en quelques millisecondes et impliquent de nombreux intermédiaires ainsi que des dizaines, voire des centaines d'annonceurs souhaitant diffuser leurs publicités[32].

Le graphique suivant illustre, de manière simplifiée, le fonctionnement de ce système.

[31] The Biggest Data Breach: ICCL report on the scale of Real-Time Bidding data broadcasts in the U.S. and Europe, ICCL, 2022.

[32] Les enchères en temps réel (RTB) : un système complexe, LINC, 2020.



Lorsqu'un consommateur se rend sur un site web ou utilise une application affichant des publicités, qu'il s'agisse d'un site web ou d'une application, des données personnelles circulent de l'appareil du consommateur vers une multitude d'acteurs le long de la chaîne de valeur.

Le site visité ou l'application utilisée par le consommateur est l'éditeur, qui propose de l'espace publicitaire à des annonceurs souhaitant afficher de la publicité pour leurs produits ou services. Le RTB constitue un processus automatique au cours duquel la négociation, la conclusion et l'exécution de ce contrat entre éditeur et annonceur se déroulent en quelques millisecondes, grâce à la chaîne d'intermédiaires présentées dans la section précédente.

Le processus est déclenché par le consommateur qui consulte un site web ou utilise une application de l'éditeur. Le site ou l'application envoie ensuite des données personnelles (un identifiant technique, la géolocalisation, l'historique de sites web visités etc.) à un SSP (*supply-side platform*) qui gère l'espace publicitaire de l'éditeur. Via une bourse publicitaire

(*ad exchange*), qui peut être fournie par le SSP ou par une entreprise distincte, les données sont diffusées à jusqu'à des centaines de DSP (*demand-side platform*) représentant les annonceurs. Selon l'attractivité du profil publicitaire du consommatoire, chaque DSP fait une offre pour l'affichage des publicités de leur client, en tenant compte des préférences de celui-ci. Le DSP gagnant est choisi au sein du *ad exchange* sur la base des offres de tous les DSP, mais aussi d'autres critères qui ne sont pas toujours transparents[33].

Tout au long de ce processus, afin d'enrichir l'ensemble de données en circulation et d'effectuer des analyses de profilage, des *data management platforms* (DMP), des *data brokers* ou d'autres intermédiaires peuvent également être impliqués et ainsi accéder au flux de données personnelles.

Une fois la publicité a été placée et vue par le consommateur, le DSP ayant gagné la transaction paie le SSP. Celui-ci rémunère ensuite l'éditeur dont l'espace publicitaire a été utilisé, alors que le DSP est rémunéré par l'annonceur dont la publicité a été affichée. Tous les acteurs paient également les autres intermédiaires (DMP, *data brokers* etc.) dont ils ont éventuellement utilisé les services.

[33] Certains acteurs peuvent par exemple négocier un accès privilégié, comme en témoigne un accord entre Google et Meta, connu par son nom de code *Jedi Blue*. En échange de la promesse de la part de Meta de réduire certaines de ses propres activités publicitaires, Meta a obtenu la garantie de gagner 90 % des enchères quelles que soit son offre, d'avoir 300 millisecondes pour enchérir contre 160 millisecondes proposées à la concurrence, et d'avoir le droit d'identifier 80 % des utilisateurs de smartphones et 60 % des internautes. Inquiète par cet accord susceptible de fausser la concurrence, la Commission européenne a lancé une enquête. Source : La Commission ouvre une enquête sur un possible comportement anticoncurrentiel de Google et de Meta dans le secteur de l'affichage publicitaire en ligne, Représentation de la Commission européenne en France, 2022.

III. LE PARCOURS DES DONNÉES PERSONNELLES

Les données personnelles des consommateurs sont donc collectées, traitées et partagées par un nombre important d'entreprises. Selon leur activité et leur relation avec le consommateur, elles se servent de différentes technologies et bases légales afin d'avoir accès aux données.

1. Les techniques de collecte et de partage de données

La plupart des entreprises avec lesquelles le consommateur conclut un contrat collectent les données personnelles sur la base de celui-ci. En revanche, d'autres entreprises n'ont pas cette relation commerciale directe avec le consommateur, ce qui ne les empêche pourtant pas de collecter ses données en s'appuyant sur des « partenariats » avec les premières parties et en utilisant des technologies avancées de pistage en ligne, comme les cookies.

a. La collecte de données par les entreprises dont les consommateurs utilisent les services

Certaines données (le nom, la date de naissance, l'adresse électronique etc.) sont typiquement fournies par le consommateur au moment de la conclusion d'un contrat avec une entreprise. Il peut s'agir d'un contrat dans le sens classique (c'est-à-dire un document papier signé), par exemple lors d'une souscription d'un programme de fidélité dans un magasin physique, mais aussi d'une action plus implicite, par exemple la création d'un compte sur un réseau social. Peu importe le type de contrat, celui-ci est typiquement complété par des conditions générales de vente précisant sur une dizaine de pages en tout petits caractères les détails de la collecte et du traitement de données personnelles. Ces documents sont formulés dans un langage juridique inaccessible au consommateur lambda, et ne sont par conséquence pratiquement jamais lu ni compris par celui-ci.

Cependant, la plupart des données personnelles ne sont pas collectées à ce moment précis, mais sont générées par le consommateur tout au long de son quotidien lors de l'utilisation des services fournis par les entreprises : les achats qui sont effectués dans des magasins physiques et enregistrés sur la carte de fidélité, les interactions sociales qui ont lieu et les contenus (photos, posts etc.) qui sont publiés sur les réseaux sociaux, etc.

Pourtant, les entreprises ne s'arrêtent pas à la collecte de ces données directes, pour lesquelles les consommateurs peuvent avoir du moins une vague idée qu'elles sont collectées. Elles utilisent notamment des technologies de pistage en ligne pour traquer jusqu'au moindre détail les activités et comportements des consommateurs.

La technologie la plus connue, grâce aux fameuses bannières que le consommateur voit à chaque fois qu'il visite un site web pour la première fois, sont les cookies.

Les éditeurs de site web utilisent notamment des cookies propriétaires, ou *first party cookies* en anglais. Le fonctionnement est simple : lorsque le consommateur entre l'adresse d'un site web dans son navigateur, ce dernier envoie la demande d'afficher le site au serveur de l'éditeur. Celui-ci envoie le contenu demandé, accompagné d'un petit fichier contenant des informations supplémentaires, le cookie. Le cookie est sauvegardé sur l'appareil du consommateur et y reste même après la fermeture du site. Si le consommateur revisite le site à un moment ultérieur, et renvoie donc une nouvelle demande d'affichage du site, il envoie aussi les informations enregistrées dans le cookie, à partir desquelles le site peut désormais être adapté.

Certains cookies propriétaires peuvent être nécessaire pour assurer le fonctionnement correct d'un site. Par exemple, quand le consommateur met un article dans le panier sur un

site de e-commerce, cette information est enregistrée dans un cookie, qui permet d'afficher le panier avec l'article choisi la prochaine fois que le consommateur se rend sur le site.

Les éditeurs peuvent toutefois placer aussi des cookies optionnels sur l'appareil du consommateur. Ces cookies, souvent masqués par des formulations nébuleuses comme « amélioration de l'expérience utilisateur » ou « personnalisation des offres et communications marketing », permettent de mesurer le comportement du consommateur : combien de temps il passe sur le site, quels contenus il regarde, les informations qu'il entre dans un formulaire etc.

Afin de collecter les données personnelles des consommateurs sur les sites web, les entreprises ont besoin de technologies de pistage comme les cookies car le consommateur accède au site via son navigateur, que les éditeurs de sites web ne contrôlent pas. En revanche, la collecte de données s'avère beaucoup plus simple quand le consommateur utilise les applications propriétaires des éditeurs (par exemple utiliser l'application d'Amazon pour smartphones pour fait des achats, au lieu de visiter le site amazon.fr via le navigateur). Comme ce sont les éditeurs eux-mêmes qui développent leurs applications, ils peuvent directement y intégrer des fonctionnalités de pistage[34].

b. La collecte de données par les tierces parties

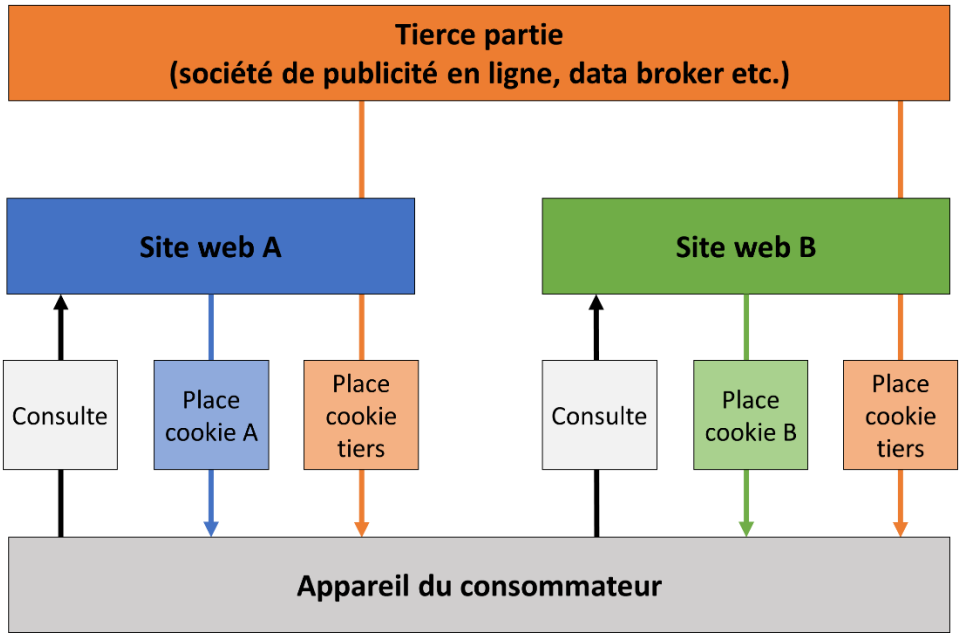
Les entreprises dont le consommateur utilise les produits et services ne sont pourtant pas les seules à le traquer. En effet, de nombreuses tierces parties sont également impliquées dans la collecte et l'analyse de données personnelles. En établissant un partenariat avec les premières parties, elles peuvent se servir des mêmes technologies de pistage en ligne.

Ainsi, de la même façon que les cookies propriétaires, les cookies tiers sont des fichiers contenant des informations sur le consommateur. Comme le consommateur n'utilise pas les services des parties tierces, elles ont besoin des premières parties (les éditeurs de sites web) pour placer leurs cookies. La majorité des sites, même s'ils semblent être entièrement fournis par leur éditeur, sont donc en réalité composés d'éléments de plusieurs entreprises. En échange d'un service (l'utilisation d'une police ou d'un élément visuel sur le site, la mesure d'audience ou un service publicitaire par exemple), les éditeurs permettent ainsi aux tierces parties d'intégrer leurs trackers dans le site web des premiers.

Certains de ces trackers sont visibles sur le site : le bouton j'aime de Facebook, une vidéo YouTube intégrée, l'option de se connecter avec son compte Google ou Facebook etc. La plupart sont pourtant masqués, et peuvent même prendre la forme d'un pixel espion transparent dont la seule fonction est d'aspirer et de transmettre des données personnelles.

Le fonctionnement des cookies tiers est le même que pour les cookies propriétaires : lorsque le consommateur visite un site web, il envoie la demande d'afficher le contenu. Comme certains contenus ne sont pas fourni par l'éditeur du site, le site transfère la demande aux parties tierces. Celles-ci envoient le contenu au consommateur en plaçant leur cookie tiers.

[34] Il faut souligner que les navigateurs sont également des applications, et les plus utilisés représentant ensemble presque 90 % des utilisateurs sont tous développés par les géants du Net (Chrome par Google, Safari par Apple et Edge par Microsoft).



La plus importante différence entre cookies propriétaires et cookies tiers est le fait que les parties tierces sont en général présentes sur un grand nombre de sites différents. Ainsi, lorsque le consommateur visite le site d'un deuxième éditeur, celui-ci n'a pas accès aux informations stockées dans les cookies propriétaires du premier éditeur. En revanche, les tierces parties ayant intégré leur trackers dans les deux sites peuvent relier l'activité du consommateur sur les deux sites afin de créer un profil détaillé de ses préférences et comportements.

Concrètement, quand un consommateur consulte d'abord un site web d'un hôtel dans un domaine skiable et ensuite un magasin en ligne d'équipement sportif, ce dernier ne peut pas savoir que le consommateur a visité le premier – contrairement à une tierce partie présentes sur les deux sites qui peut donc déduire que le consommateur est potentiellement intéressé par l'équipement de ski.

Paradoxalement, alors que les consommateurs ne sont pas conscients de l'existence de ces tierces parties, celles-ci sont en mesure de collecter et analyser beaucoup plus de données personnelles que les sites que les consommateurs consultent.

Dans les applications, les tierces parties peuvent également intégrer leur tracker grâce à des kits de développement logiciel (SDK, pour *software development kit* en anglais) et des bibliothèques de logiciels. En termes simples, ces kits sont comme une boîte à outil numérique contenant du code pré-écrit pour des fonctionnalités basiques et courantes. Au lieu d'écrire tout le code eux-mêmes, les développeurs peuvent ainsi utiliser un SDK pour gagner du temps et des efforts. Cette pratique permet aux tierces parties de proposer des solutions publicitaires prêtes à l'emploi que les développeurs peuvent intégrer dans leurs applications, mais qui collecteront également des données personnelles pour le compte de la tierce partie ayant mis à disposition l'outil.

En moyenne, un site web envoie des données à 34 tierces parties[35], et une application partage des données avec une dizaine de tierces parties (sachant que le consommateur moyen a installé une trentaine d'applications sur son smartphone)[36].

[35] Online Tracking: A 1-million-site Measurement and Analysis, Steven Englehardt et Arvind Narayanan, 2016.

[36] Before and after GDPR: tracking in mobile apps, Konrad Kollnig et al, 2021.

Les valeurs moyennes ne doivent pourtant pas occulter le fait que certains sites communiquent avec beaucoup plus de tierces parties. Le tableau suivant présente le nombre de « partenaires » listés sur un échantillon de 10 sites web parmi les plus fréquentés en France[37].

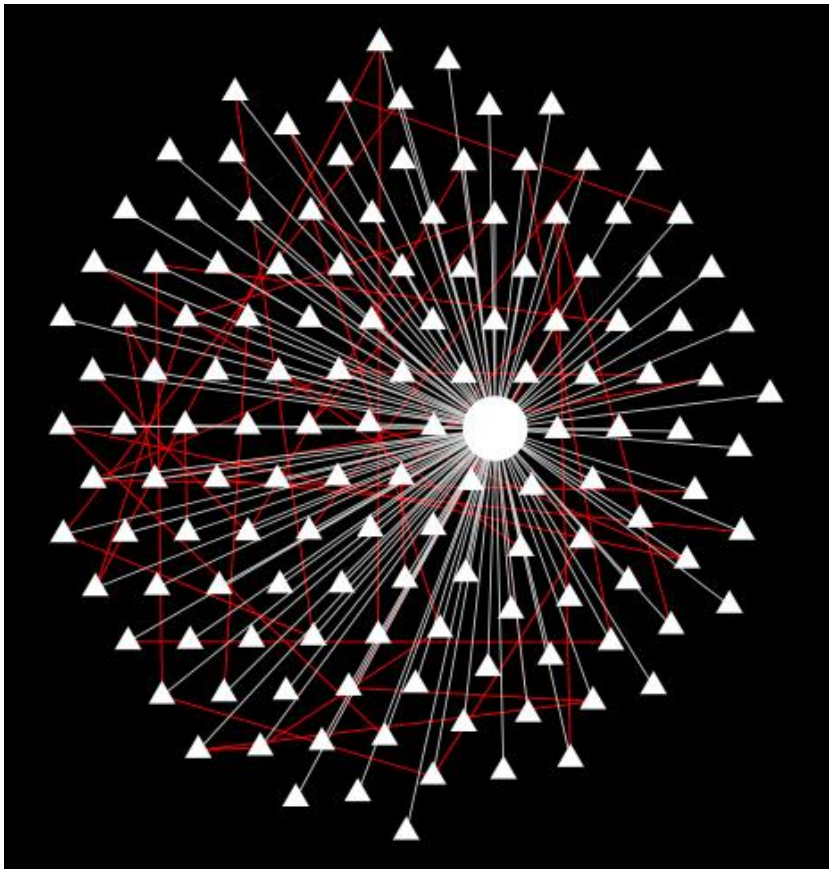
Nom du site	Nombre de « partenaires
20minutes.fr	609
allocine.fr	786
cdiscount.com	578
leboncoin.fr	266
lemonde.fr	169
marmiton.org	237
meteofrance.com	131
orange.fr	778
programme-tv.net	537
yahoo.com	241

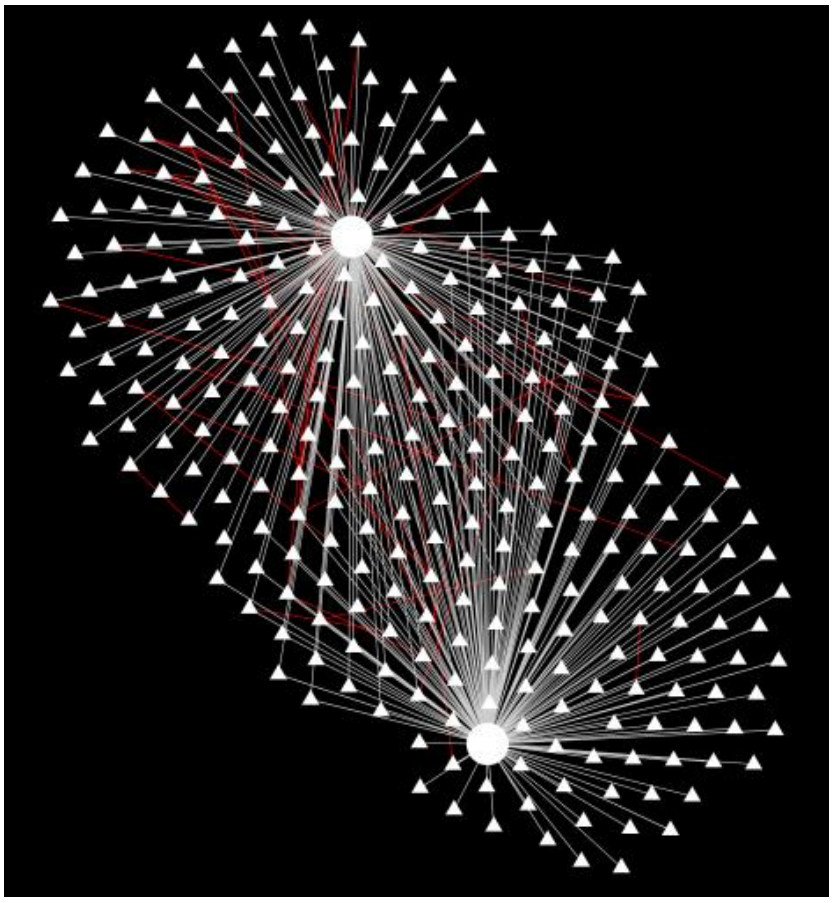
D'entre ces 4 332 « partenaires », des centaines sont présents sur plusieurs sites et donc en mesure de combiner les données collectées à travers le parcours en ligne du consommateur. En ne consultant pas plus que ces 10 sites, celui-ci se fait donc pister par jusqu'à 1040 tierces parties individuelles.

A titre illustratif, les graphiques suivants[38] illustrent le nombre de cookies placés, et reliés, par les tierces parties au cours de la consultation d'abord de leboncoin.fr et ensuite de cdiscount.com. Les cercles représentent les deux sites visités, et les triangles les cookies tiers. Les lignes blanches indiquent sur quel site un cookie est présent (un nombre important est donc présent sur les deux sites). Les lignes rouges relient les tierces parties qui échangent des données personnelles grâce à la synchronisation de leurs cookies.

[37] L'UFC-Que Choisir a recensé en juin 2023 le nombre de « partenaires » affichés par la bannière cookie sur les 10 sites web suivants : 20minutes.fr, allocine.fr, cdiscount.fr, leboncoin.fr, lemonde.fr, marmiton.org, meteofrance.com, orange.fr, programme-tv.net, yahoo.com. Ces sites sont classés parmi les 50 sites les plus consultés en France, selon les mesures de Similarweb.

[38] Les visualisations ont été réalisées avec l'extension de navigateur Thunderbeam-Lightbeam.





Qui sont donc toutes ces tierces parties ? Bien que la liste des « partenaires » puisse être consultée par le consommateur, il est impossible de vraiment comprendre qui ils sont, et à quelle fin ils utilisent les données partagées. Sur la liste des 1 040 « partenaires » de notre échantillon, le consommateur ne reconnaitra pas plus qu'une vingtaine, notamment les plateformes en ligne : Alphabet et ses services (Google, YouTube), Amazon, Meta et ses services (Facebook, Instagram), Microsoft et ses services (Bing, LinkedIn), Twitter, Amazon etc. Ces entreprises dominent l'écosystème de la publicité en ligne et ont placé leurs outils de pistage dans un nombre important d'applications et de sites web. Par exemple, sur un échantillon de 2 millions d'applications Android, les trackers d'Alphabet était présent dans 89 % et ceux de Meta dans 38 %[39].

Les autres entreprises « partenaires » recensés, qui représentent donc la quasi-totalité de la liste, sont les intermédiaires de la publicité ciblée présentés ci-dessus dont les consommateurs n'ont jamais entendu leurs noms : AdGear, Comscore, Criteo, Equativ, Impact, Kwanko, LiveRamp, Magnite, OpenX, Quantcast, RTB House, Vectaury… Bien que leur position sur le marché de la publicité en ligne soit moins importante que celle des géants du Net, il s'agit malgré tout d'entreprises avec des centaines, voire des milliers de collaborateurs et dont les chiffres d'affaires peuvent atteindre des centaines de millions d'euros.

c. La fin des cookies tiers ne mettra pas fin au pistage en ligne

Il est opportun de noter que les cookies tiers, un des outils de pistage en ligne les plus importants depuis des décennies, sont susceptible de perdre en importance. Leur usage a déjà été considérablement limité par défaut dans les navigateurs Firefox de Mozilla en 2019[40], et Safari d'Apple en 2020[41].

Google prévoit également d'abandonner progressivement les cookies tiers dans son navigateur Chrome et de les remplacer par une nouvelle technologie, nommé *Privacy Sandbox*. La fin des cookies de pistage a initialement été annoncée pour début 2022[42], mais a été reportée plusieurs fois et était, au moment de la publication de la présente étude, prévue pour la fin de 2024[43].

Cependant, le fait que les cookies tiers deviendront donc potentiellement obsolètes dans les prochaines années ne signifie pas la fin de l'intrusion massive dans la vie privée des consommateurs. Alors que Mozilla, en tant qu'organisation à but non lucratif, peut encore être considéré comme ayant des objectifs nobles, les géants du Net ne renonceront pas à une source de revenus importante, voire majoritaire dans le cas de Google[44], sans la remplacer par une autre plus rentable.

En effet, les nouvelles « solutions publicitaires » de Google et Apple ont en commun qu'elles renforcent la position de ces entreprises vis-à-vis leurs concurrents. La collecte et le traitement des données personnelles se déroule ainsi principalement dans les applications de Google (Chrome, Android)[45] et Apple (Safari, MacOS), qui sont ainsi en mesure d'imposer

---

[39] Before and after GDPR: tracking in mobile apps, Konrad Kollnig et al, 2021.

[40] Today's Firefox Blocks Third-Party Tracking Cookies and Cryptomining by Default, Mozilla, 2019.

[41] Full Third-Party Cookie Blocking and More, WebKit, 2020.

[42] Building a more private web: A path towards making third party cookies obsolete, Chromium Blog, 2020.

[43] The next stages of Privacy Sandbox: General availability and supporting scaled testing, Privacy Sandbox, 2023.

[44] En 2022, 79 % (soit 224 milliards d'euros) des revenus d'Alphabet provenaient de ses activités de publicité en ligne.

[45] Ciblage publicitaire : pour remplacer les cookies, Google Chrome exploite l'historique de navigation, La Tribune, 2023.

leurs conditions d'accès à ces données à leurs concurrents, c'est-à-dire les intermédiaires de la publicité[46]. La commercialisation des données personnelles n'est donc pas arrêtée.

Enfin, les cookies tiers ont beau être un outil de pistage important, ils ne sont pas le seul. En effet, depuis plusieurs années déjà, les consommateurs se connectent le plus souvent à Internet en utilisant leur smartphone et non leur ordinateur[47]. Les applications mobiles, qui utilisent d'autres technologies de pistage que les cookies, sont donc devenues un point d'entrée de plus en plus important pour les collecteurs de données personnelles.

Une technique alternative est par exemple l'utilisation de l'empreinte digitale d'appareil, ou *device fingerprinting* en anglais, qui permet d'identifier l'appareil d'un consommateur à partir des informations qu'il transmet. Des données comme la langue du système, la version du navigateur et les paramètres choisis, les extensions installées, les polices utilisées, la résolution de l'écran ou la géolocalisation sont typiquement communiquées afin d'afficher correctement le contenu consulté. La combinaison d'une dizaine de ces paramètres individuels est souvent suffisante pour créer une empreinte unique de l'appareil.

A noter également que les cookies propriétaires ne sont pas touchés par ces développements. Les éditeurs de sites (les premières parties) peuvent donc continuer à placer des cookies et pister les consommateurs.

La fin des cookies de pistage ne mettra donc pas fin à la collecte généralisée de données personnelles, mais ne fera que déplacer le rapport de force au sein de l'écosystème de la publicité en ligne, des intermédiaires vers les grandes plateformes (qui sont déjà dominantes) et, dans une moindre mesure, les éditeurs. Tant que la publicité ciblée s'appuie sur des pratiques de personnalisation basées sur des données personnelles intimes des consommateurs, les acteurs de l'industrie trouveront toujours des solutions techniques pour collecter et partager les données des consommateurs à grande échelle.

2. Les types de données collectées, et ce qu'elles permettent de savoir sur le consommateur

a. Le profilage des consommateurs sur la base des données collectées

Le fait qu'un nombre important d'entreprises dans tous les secteurs de l'économie sont impliquées dans la collecte et le traitement de données personnelles, et la disponibilité de technologies de pistage et d'analyse algorithmique de plus en plus avancées, permettent aux entreprises d'exploiter un réservoir de données d'une ampleur sans précédent afin de créer des profils individuels détaillés des consommateurs.

Ces pratiques ont existé bien avant la généralisation d'Internet. Parmi les sources « traditionnelles » de données personnelles, qui sont exploitées depuis des décennies, on trouve notamment[48] :

- Les données disponibles publiquement, par exemple dans des annuaires ou des registres publics,

[46] En 2023, L'Autorité de la concurrence a notifié un grief à Apple pour abus de position dominante. Plusieurs acteurs de la publicité en ligne avaient déposé une plainte auprès de l'autorité au sujet de l'IDFA (*Identifier for advertisers*), un identifiant des consommateurs permettant aux annonceurs de les cibler et auquel Apple a conditionné l'accès au consentement individuel de chaque consommateur. Les plaignants accusent Apple de fausser la concurrence, sous couvert de protection des données personnelles. Publicité sur applications mobiles iOS : le rapporteur général indique avoir notifié un grief au groupe Apple, Autorité de la concurrence, 2023.

[47] Baromètre du numérique 2022, CRÉDOC, 2023.

[48] Corporate surveillance in everyday life, Cracked Labs, 2017.

- L'historique d'achats de consommateurs disponible via les programmes de fidélité,
- Les abonnements médiatiques (journaux, magazines, télévision payante etc.) souscrits par les consommateurs,
- L'historique de transactions bancaires,
- Des données collectées au moyen de jeux-concours ou de sondages.

Les données collectées grâce à ces sources permettent notamment d'identifier un consommateur (nom, adresse, numéro de téléphone, genre, âge), de déduire sa position socio-économique (le quartier dans lequel il vit, la voiture immatriculée à son nom), de connaitre son pouvoir d'achat et ses biens et revenus (son historique de transaction financières), et de faire des prévisions concernant ses préférences et orientations politiques (les produits qu'il a achetés, les médias qu'il consomme).

Si la collecte massive de données personnelles a donc déjà eu lieu à l'époque pré-Internet, les bases de données ainsi créées sont souvent restées sporadiques, incomplètes, incohérentes et réparties entre diverses entreprises. Depuis les années 2000, la généralisation d'Internet et de smartphones et l'émergence des réseaux sociaux ont pourtant propulsé ces pratiques vers de nouveaux niveaux. Non seulement les nouvelles technologies ont facilité la collecte de données via les sources traditionnelles, mais elles ont conduit à la création d'une multitude de nouvelles sources de données personnelles prêtes à être exploitées[49] :

- Les données disponibles publiquement en ligne (par exemple, des informations publiées par les consommateurs sur les réseaux sociaux),
- Les données demandées lors de la création d'un compte en ligne,
- Le pistage des activités en ligne (sites web consultés etc.),
- Les interactions sociales sur les réseaux sociaux,
- L'historique d'achats et de souscriptions en ligne,
- La mesure d'utilisation d'applications de smartphone et de services numériques,
- La mesure d'utilisation d'objets connectés (voitures, téléviseurs, maison connectée etc.).

Grâce à l'accès à ces nouvelles sources, en combinaison avec le développement de nouveaux outils d'analyse algorithmique de données, l'échelle et la profondeur de l'exploitation des données personnelles des consommateurs ont changé de manière significative. Les entreprises peuvent désormais connaître les centres d'intérêt d'un consommateur de façon beaucoup plus granulaires : les sites qu'il consulte, les applications qu'il utilise, sa consommation précise de médias (streaming vidéo et audio, jeux vidéo, presse en ligne etc.), ses recherches en ligne, son historique d'achats, etc. Contrairement aux pratiques pré-Internet, l'accès à des informations particulièrement intimes (par exemple, les recherches sur les maladies, la consommation de sites pornographiques, les habitudes de jeux de hasard et d'argent) est notamment devenu beaucoup plus facile.

En outre, les métadonnées sont une source d'informations particulièrement intéressantes pour les entreprises. En termes simples, les métadonnées sont des données qui contiennent des informations concernant d'autres données. Par exemple, la géolocalisation permet de

[49] Corporate surveillance in everyday life, Cracked Labs, 2017 ; Out of Control, Forbrukerrådet, 2020.

savoir dans quels magasins le consommateur est allé, mais pas quels produits il a acheté. Concernant la communication électronique, les métadonnées indiquent avec qui le consommateur a communiqué, par quel moyen (appel vocal, messagerie instantanée), à quel moment, et pendant quelle durée. En revanche, elles ne permettent pas de savoir de quoi il a parlé.

Ces informations peuvent néanmoins être très utiles, surtout lorsqu'elles sont combinées à d'autres données. Ainsi, grâce à la géolocalisation une entreprise peut par exemple mesurer si, après avoir vu une publicité en ligne, le consommateur s'est rendu dans un magasin physique de la marque. Même son parcours dans le magasin peut être traqué avec précision. Désactiver la géolocalisation sur son smartphone ne suffit pas forcément pour limiter ce pistage, car l'appareil peut aussi se localiser grâce aux réseaux wifi disponibles qu'il détecte autour de lui, ou aux autres appareils (smartphones ou autres) qu'il identifie via Bluetooth.

L'ignorance des consommateurs quant à l'importance des métadonnées permet aux entreprises de s'ériger en protectrices de la vie privée, comme le fait par exemple WhatsApp qui chiffre le contenu des conversations de ses utilisateurs mais ne se prive pas de collecter les métadonnées.

De manière plus générale, les données collectées grâce à des technologies numériques enrichissent les données collectées à partir de sources « traditionnelles » et permettent aux entreprises d'établir des profils individuels de chaque consommateur. Ce jumeau numérique informe les entreprises d'un grand nombre de caractéristiques du consommateur :

- Ses habitudes et comportements : où il se déplace et quand, son domicile et son lieu de travail, quand et pour combien de temps il utilise ses appareils, pour quoi il s'en sert (jouer aux jeux, communiquer sur les réseaux sociaux, suivre les actualités, travailler etc.), quand et où il pratique des activités culturelles ou sportives, son orientation sexuelle, etc.,
- Son réseau social : qui il fréquente, qui sont les membres de sa famille, ses amis et ses collègues, ses relations intimes (grâce aux données collectées par les applications de rencontre, par exemple), etc.,
- Son état de santé : ses maladies, sa condition physique, son cycle menstruel (grâce aux données collectées par les applications de « bien-être »), etc.,
- Son orientation politique, sa religion, son origine ethnique, etc.

De nombreux études ont également montré que l'analyse algorithmique des données personnelles permet de déduire, avec une grande précision, des caractéristiques concernant la personnalité du consommateur qui ne sont jamais exprimé explicitement par celui-ci, comme son impulsivité, sa stabilité émotionnelle, ou son état de dépression – des caractéristiques qui peuvent ensuite être exploitées afin d'influer sur son comportement et ses décisions.

b. L'usage d'identifiants

Comme indiqué ci-dessus, les données personnelles sont collectées par un grand nombre d'entreprises. Afin de maximiser leur utilisation à des fins d'exploitation commerciale, il est nécessaire de fusionner les données pour créer des profils les plus complets possible.

Les identifiants jouent un rôle clé dans cet établissement de liens entre les bases de données des différentes entreprises. En effet, afin de pouvoir utiliser une donnée pour afficher une publicité ciblée à un consommateur, elle doit être liées à un identifiant qui est associé à celui-ci.

Il existe de nombreux identifiants de ce type[50]. Certains, comme le nom, le numéro de téléphone ou l'adresse électronique permettent de directement identifier la personne. Le nom n'est pourtant pas forcément l'information la plus intéressante pour les entreprises. L'adresse électronique ou le numéro de téléphone peuvent s'avérer plus utile, car ils sont souvent directement liés aux appareils du consommateur. Tel est le cas des comptes identifiants des grandes plateformes : Google, Microsoft et Apple tous forcent le consommateur à s'authentifier dans leur système d'exploitation respectif (Android, Windows et macOS/iOS). Google et Facebook proposent également l'option d'utiliser leur compte pour se connecter à un service tiers.

D'autres identifiants sont plus techniques : l'adresse IP, les cookies, l'adresse MAC qui est un identifiant physique et unique attribué à chaque appareil connectable, le numéro IMEI qui permet d'identifier de manière unique chacun des terminaux de téléphonie mobile, l'empreinte digital d'appareil, ou encore les identifiants publicitaires qu'Android, Windows et macOS/iOS attribuent à chaque utilisateur. En outre, les data brokers et autres intermédiaires de l'écosystème utilisent leurs propres identifiants.

La multiplicité des identifiants ne constitue pourtant pas un obstacle au fusionnement de données personnelles de différentes sources dans un seul profil. La synchronisation des identifiants se fait grâce à des solutions technologiques. Elle peut s'effectuer de manière automatisée, par exemple entre cookies tiers (dit *cookie matching* ou *cookie synching*), ou être proposée comme service par un intermédiaire.

Les identifiants utilisés sont en général pseudonymisés, c'est-à-dire qu'ils ne contiennent pas de référence directe au consommateur qu'ils identifient[51]. Les noms ou adresses électroniques sont convertis en identifiants alphanumériques randomisés. Jean Dupont devient ainsi par exemple « a12b3cd4–5678–9e01-fa23–456-b78c9d0e ».

Selon les entreprises, les données d'un jumeau numérique pseudonymisé ne doivent plus être considérées comme données personnelles protégées, car la personne derrière est masquée par l'identifiant. Pourtant, pseudonymisation n'égale pas anonymisation. Comme chaque identifiant pseudonymisé est unique et lié à un seul consommateur spécifique, l'identification de la personne est toujours possible[52].

Certaines entreprises vont jusqu'à tenir deux registres distincts : un premier avec les noms, adresses, numéros de téléphone et identifiants pseudonymisés, et un deuxième avec les identifiants et les jumeaux numériques. Grâce aux mêmes identifiants présents dans les deux registres, ceux-ci peuvent être fusionnés sans aucun problème.

Et même si toutes les entreprises n'agissent pas de manière aussi audacieuse, il reste très simple de réidentifier la personne réelle sur la base de son jumeau numérique. En fait, il n'est même pas nécessaire de connaître son nom pour le réidentifier.

En effet, une célèbre étude réalisée en 2000[53] a montré que seulement trois informations personnelles (date de naissance, genre et code postale) sont suffisantes pour identifier près de 90 % d'individus. Les entreprises détiennent bien plus que ces simples informations démographiques : les résultats de recherches en ligne qui sont extrêmement individuels, la géolocalisation qui permet d'identifier un consommateur en combinant son domicile avec son lieu de travail, les identifiants techniques comme l'adresse IP ou les cookies, etc.

---

[50] Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation, Frederik Zuiderveen Borgesius, 2016.

[51] Ibid.

[52] Ibid.

[53] Simple Demographics Often Identify People Uniquely, Latanya Sweeney, 2000.

Dans la pratique, cette réidentification est peut-être rarement effectuée par les entreprises, mais cela ne change rien au fait que les données personnelles collectées et le profil comportemental créé restent liés à un consommateur individuel. Si, par exemple, le jumeau numérique d'un consommateur porte le trait de caractère « ne paie pas ses factures », les services numériques qu'il utilise ont accès à cette information et vont l'attribuer au consommateur (ou, plus précisément, à son appareil) lorsqu'il navigue sur Internet. Ils ne connaissent peut-être pas son nom, mais les conséquences sont les mêmes. Par exemple, le consommateur ne verra pas tous les modes de paiement disponibles aux autres consommateurs.

La CNIL[54] a résumé ce constat dans son analyse des pratiques de Google : « le seul et unique objectif poursuivi par la société consiste à recueillir un maximum d'éléments sur des personnes singularisées afin d'optimiser la valorisation de leurs profils sur le plan publicitaire. Son modèle économique ne requiert donc pas de connaître les nom, prénom, adresse ou autres éléments directement identifiants sur les personnes, qui ne lui sont pas nécessaires pour les reconnaître lors de chaque nouvel usage qu'elles feront de ses services. Cependant, les éléments qu'elle collecte à cette fin, qui peuvent être combinés entre eux [...], lui permettent de cerner avec une précision extrême le comportement d'une machine et, derrière celle-ci, de son utilisateur, à laquelle elle est capable in fine d'affecter les caractéristiques de son activité quotidienne, ses interactions avec autrui, ses centres d'intérêt, des éléments liés à sa personnalité, ses choix de vie, etc. En d'autres termes, l'accumulation de données qu'elle détient sur une seule et même personne lui permet de la singulariser à partir d'un ou de plusieurs éléments qui lui sont propres. Ces données doivent, en tant que telles, être considérées comme identifiantes et non comme anonymes ».

La fuite en 2023 d'une base de données de segments publicitaires maintenue par Xandr, filiale de Microsoft spécialisée dans la publicité, donne un aperçu unique du rôle du profilage dans la publicité ciblée[55].

La base de données comprend plus de 650 000 traits de personnalité et situations personnelles en fonction desquels les consommateurs sont classés. Nombre de ces traits sont extrêmement intimes, tels que « dysfonction érectile », « dépression », « gros acheteur de test de grossesse », « dépendance aux opioïdes », « sympathisant de syndicats », « réceptif aux messages émotionnels », « dépendance au jeu de hasard », « ratio revenu/dette inférieur à la moyenne », ou encore « problèmes d'argent ». Les données personnelles utilisées pour attribuer ces caractéristiques aux consommateurs proviennent de plus de 90 fournisseurs de données, c'est-à-dire des *data brokers* et d'autres entreprises ayant collecté les données personnelles des consommateurs, y compris des géants du courtage en données comme Oracle et Axciom.

Ces segments publicitaires, en combinaison avec les identifiants publicitaires, jouent un rôle essentiel dans le cadre des enchères en temps réel (RTB). C'est sur cette base qu'une *demand-side platform* (DSP) décide si elle veut faire une offre pour l'affichage des publicités de leur client, et si oui de quel montant.

c. Les impacts nocifs sur la vie des consommateurs

La plupart des activités en ligne des consommateurs déclenchent donc un flux de données personnelles se déroulant en arrière-plan et dans lequel un nombre important d'entreprises sont impliquées. Cela ne nécessite pas forcément une interaction consciemment initiée par le consommateur (ouvrir une application, consulter un site web, saisir une requête dans un

[54] Délibération 2013-420 du 3 janvier 2014, CNIL, 2014.

[55] From "Heavy Purchasers" of Pregnancy Tests to the Depression-Prone: We Found 650,000 Ways Advertisers Label You, The Markup, 2023.

moteur de recherche, etc.) : grâce aux objets connectés, avant tout le smartphone, le flux de données peut être passif et quasi-permanent même durant les moments pendant lesquels le consommateur n'utilise pas ses appareils. Les données personnelles peuvent ainsi être collectées, partagées, fusionnées et analysées en permanence. Le résultat de ce flux de données dynamique sont des décisions prises de manière automatisée qui impactent l'expérience du consommateur dans l'espace numérique mais aussi dans le monde réel[56].

Ces décisions ne sont pas forcément négatives. Le fait qu'un moteur de recherche affiche des résultats personnalisés, qu'un service de streaming musical propose des nouveaux chanteurs adaptés au goût musical du consommateur, ou qu'un aspirateur robot ajuste son parcours en fonction de la disposition de la maison de l'utilisateur, sont tous possible grâce à l'analyse de données personnelles. Il s'agit de pratiques généralement acceptées, voire réclamées par les consommateurs.

Par ailleurs, les entreprises utilisent les données personnelles à leurs propres fins. S'il n'est en principe pas critiquable que les entreprises cherchent à réaliser des bénéfices, elles dupent les consommateurs en faisant tout pour dissimuler à quel point le profilage et le partage de données impactent leurs décisions.

La publicité ciblée est l'utilisation la plus évidente des données personnelles susceptible de nuire aux consommateurs. La promesse des entreprises est qu'elle leur permet « d'afficher des annonces plus pertinentes pour le consommateur ». En réalité, il est toutefois dans leur intérêt économique d'afficher en priorité des annonces qui leur rapportent le plus d'argent. Celles-ci ne sont pas nécessairement celles qui intéresseraient le consommateur le plus mais plutôt celles qui le poussent à l'achat[57]. Les algorithmes déployés sont en effet capables d'analyser minutieusement les comportements de navigation, les préférences et les historiques d'achats pour créer des annonces sur mesure, incitant ainsi les consommateurs à succomber à des achats impulsifs. Cette pratique d'exploiter les vulnérabilités psychologiques et de créer un sentiment de besoin immédiat et de satisfaction instantanée conduit à une surstimulation constante des consommateurs, les incitant à acheter des produits dont ils n'avaient peut-être même pas connaissance auparavant. Ainsi, la publicité ciblée en ligne se révèle être une force motrice majeure d'une consommation déraisonnée[58].

La publicité ciblée a beau être au cœur de la collecte de données, elle n'est pas la seule pratique susceptible d'avoir un impact négatif sur les consommateurs. Par exemple, la personnalisation automatisée et dynamique de prix est également basée sur le profilage des consommateurs. Il s'agit encore une fois d'une pratique très opaque et difficile à reconnaitre, mais l'exemple de Tinder montre à quel point les consommateurs peuvent subir des préjudices financiers à cause de la collecte et du partage de données personnelles qu'ils n'auraient très probablement jamais acceptés si les conséquences ne leur auraient pas été masquées. Il a effectivement été constaté à plusieurs reprises que l'application de rencontre appliquait une tarification discriminatoire sur la base de l'âge, la géolocalisation, le genre et l'orientation sexuelle du consommateur. En fonction de ces caractéristiques, certains consommateurs payaient cinq fois plus que d'autres[59].

---

[56] Corporate surveillance in everyday life, Cracked Labs, 2017.

[57] Online advertising: the impact of targeted advertising on advertisers, market access and consumer choice, Parlement européen, 2021.

[58] Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

[59] Tinder charges older people more, CHOICE, 2020 ; Tinder's unfair pricing algorithm exposed, Which?, 2022.

Concernant les recommandations et la personnalisation de contenus, elles risquent d'impacter la santé mentale des consommateurs, en particulier des mineurs. En effet, alors que l'efficacité des algorithmes pour attirer l'attention des consommateurs et faire en sorte qu'ils passent le plus de temps possible sur une application est à l'origine du succès des plateformes, et notamment des réseaux sociaux, leur potentiel de causer des dépendances à l'écran inquiète chercheurs et décideurs[60]. À ce jour, le lien de causalité entre l'utilisation des réseaux sociaux et les problèmes de santé mentale n'a pas pu être prouvé, notamment parce que les données nécessaires pour effectuer ce type d'analyse sont détenues par les plateformes de réseaux sociaux elles-mêmes. Néanmoins, des documents internes fuités de Meta (Facebook, Instagram) et ByteDance (TikTok) ont révélé que des impacts négatifs sur la santé mentale de ses utilisateurs existent, et que les ingénieurs et les manageurs de ces entreprises en ont bien conscience[61].

Enfin, le profilage des consommateurs est aussi utilisé pour orienter leurs opinions et leur comportement politiques notamment dans le contexte des élections. Le scandale de Facebook / Cambridge Analytica en est la plus célèbre illustration : cette affaire, qui a été révélée en 2018 par un lanceur d'alerte, a concerné environ 87 millions d'utilisateurs de Facebook dont les informations ont été collectées et utilisées sans leur consentement par la société Cambridge Analytica, spécialisée dans l'analyse comportementale à des fins politiques. Ces données ont servi à influencer les intentions de vote (ou provoqué l'abstention) en faveur de Donald Trump lors de l'élection présidentielle américaine de 2016, ainsi que d'autres campagnes politiques dans le monde[62].

En résumé, la collecte de données personnelles et le profilage permettent aux entreprises de comprendre le comportement des consommateurs et de le manipuler en les poussant à utiliser certains services, à regarder certains contenus, à acheter certains produits ou à prendre certaines décisions de vote. Certains consommateurs peuvent éventuellement considérer qu'ils sont immunisés contre ce type d'influence comportementale, mais ils risquent de sous-estimer les effets cognitifs qui ont été prouvés par la recherche. Le fait que les grandes entreprises technologiques disposent d'importants départements de recherche comportementale ne trompe pas. Leur base d'utilisateurs massive, qui atteint des centaines de millions, voire des milliards de consommateurs, est constamment utilisée, pour des tests comportementaux à grande échelle, en affichant des versions légèrement différentes de contenus et d'interfaces à différents groupes d'utilisateurs[63]. La plupart de ces recherches ne sont pas publiques, mais un exemple notoire est la recherche menée par Facebook en 2014 sur les émotions des utilisateurs. Dans cette expérience qui a concerné près de 690 000 consommateurs, le caractère positif ou négatif des messages dans les fils d'actualité des utilisateurs a été manipulé, ce qui a influencé le nombre de messages portant des émotions positives ou négatives que les utilisateurs ont eux-mêmes publiés par la suite[64].

Il faut souligner que les technologies d'analyse utilisées s'appuient sur des prédictions probabilistes dont le niveau de précision n'atteint pas 100 %. En d'autres termes, le profil numérique d'un consommateur peut être partiellement erroné. Dans la pratique, cela ne fait toutefois peu de différence : les décisions automatisées décrites dans ce chapitre, qui sont basées sur la collecte massive de données à caractère personnel, impactent la vie des consommateurs en tout état de cause, que l'information soit correcte ou non.

---

[60] Addiction, abrutissement, souffrance : TikTok et les risques de troubles psychologiques, EURACTIV, 2023.

[61] Mark Zuckerberg Was Warned on Social Media Addiction, Filing Shows, Bloomberg, 2023.

[62] Cambridge Analytica : 87 millions de comptes Facebook concernés, Le Monde, 2018.

[63] Corporate surveillance in everyday life, Cracked Labs, 2017.

[64] Experimental evidence of massive-scale emotional contagion through social networks, Kramer et al., PNAS vol. 111 no. 24, 8788–8790, 2014.

A part l'utilisation des données personnelles par les entreprises, les fuites posent un énorme risque de sécurité pour les consommateurs. Des cybercriminels peuvent par exemple s'en servir pour commettre des fraudes bancaires ou usurper l'identité d'un consommateur.

Les fuites massives de données personnelles sont effectivement très fréquentes et ne concernent pas uniquement les petites entreprises. Au contraire, beaucoup de grandes plateformes ont déjà été frappées, et dans ces cas le nombre de consommateurs dont les données sont compromises peut atteindre plusieurs centaines de millions au niveau global. Sur la liste des grandes entreprises touchées figurent, *inter alia*, Facebook (avec 533 millions de dossiers fuités)[65], Microsoft (250 millions)[66], TikTok (42 millions)[67], Uber (57 millions)[68], eBay (145 millions)[69], Yahoo (3 milliards)[70], Instagram (200 millions)[71], ou encore PlayStation Network (25 millions)[72]. En France, 4 731 violations de données à caractère personnel ont été signalées à la CNIL en 2022, deux fois plus qu'en 2019. Il est estimé que les données d'environ 5 millions de Français ont été impactés en 2022[73].

Sans oublier les fuites malveillantes en interne, comme le cas récent de Tesla où pendant plusieurs années, les salariés ont partagé dans des messageries internes des enregistrements souvent intimes réalisés par les caméras intégrées dans les voitures de la marque, y compris des enregistrements réalisés pendant que le véhicule était stationné sur la propriété ou au garage des consommateurs.[74]

Les risques liés à la collecte de données personnelles sont donc bien réels. Alors que certains usages de données personnelles sont tout à fait justifiables, notamment quand ils améliorent vraiment l'expérience des consommateurs, c'est la pratique du profilage au service de la publicité ciblée qui est particulièrement nocive.

Des études récentes soulèvent pourtant des doutes quant aux avantages du (re)ciblage basé sur les données personnelles des consommateurs[75]. Cette forme de publicité est avant tout dans l'intérêt de l'industrie de la technologie publicitaire elle-même, mais il n'a jusqu'à présent jamais été démontré qu'elle apporte des gains significatifs aux éditeurs ou aux annonceurs, par rapport à la publicité non ciblée. En effet, les revenus des éditeurs stagnent depuis des années alors que ceux des intermédiaires de la publicité ciblée ne cessent de croitre. En Espagne, l'association nationale des médias, qui réunit plus de 80 éditeurs de presse, a même porté plainte contre Meta pour concurrence déloyale, en soulignant l'aspect anticoncurrentiel de la collecte des données à des fins publicitaires par la plateforme[76]. En outre, en France l'Union des marques estime que « la domination du marché par les plateformes implique la fixation de prix sans négociation », ce qui crée « un environnement

---

[65] 533 millions de données utilisateurs de Facebook compromises, Le Monde Informatique, 2021.

[66] 250 millions de données de Microsoft ont fuité : quelles conséquences ?, Numerama, 2020.

[67] 235 Million Instagram, TikTok And YouTube User Profiles Exposed In Massive Data Leak, Forbes, 2020.

[68] 2016 Data Security Incident, Uber, 2017.

[69] eBay asks 145 million users to change passwords after data breach, The Washington Post, 2014.

[70] All 3 Billion Yahoo Accounts Were Affected by 2013 Attack, The New York Times, 2017.

[71] 235 Million Instagram, TikTok And YouTube User Profiles Exposed In Massive Data Leak, Forbes, 2020.

[72] Le piratage du PSN, JV, 2011.

[73] 4ème édition du Baromètre Data Breach, Bessé, Almond, 2023.

[74] Tesla workers shared 'intimate' car camera images, ex-employees allege: 'Massive invasion of privacy', The Guardian, 2023.

[75] Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

[76] Facebook owner Meta faces $600 mln lawsuit from Spanish media, Reuters, 2023.

dans lequel les marques ne peuvent pas opérer sans elles »[77]. Mettre fin à la collecte et le partage excessifs de données par l'écosystème de la publicité en ligne protège donc non seulement la vie privée des consommateurs, mais est également dans l'intérêt des petites et moyennes entreprises, qui sont actuellement plus ou moins à la merci du pouvoir des géants de la publicité, Alphabet et Meta en tête.

Toutefois, il ne faut en aucun cas opposer le bon fonctionnement de la concurrence des marchés numériques et la protection des données personnelles. La garantie de la concurrence, qui est sans aucun doute importante, ne doit pas être utilisée comme argument fallacieux pour défendre et maintenir des pratiques contraires à la protection des données. Au lieu de cela, l'accent devrait être mis sur la mise en place de technologies publicitaires qui allient protection des données et concurrence loyale. C'est par exemple le cas de la publicité contextuelle, c'est-à-dire une publicité qui s'adapte au contexte du site sur lequel elle est affichée, sans avoir besoin de connaitre l'identité ou les caractéristiques du consommateur qui la regarde[78].

En conclusion, bien que la majorité des consommateurs ne souhaitent pas être suivis, ils se trouvent dans une asymétrie par rapport aux entreprises dans l'environnement en ligne. Ils sont délibérément maintenus dans l'ignorance de ce qui advient de leurs données personnelles et n'ont guère d'autre choix que de renoncer à utiliser la plupart des services en ligne, bien que ceux-ci jouent un rôle essentiel dans la société actuelle[79].

Malgré les bonnes intentions, le RGPD a effectivement contribué à une individualisation du droit à la vie privée. La protection des données personnelles repose donc sur un contrat individuel entre l'utilisateur et l'entreprise, mettant toute la responsabilité sur le consommateur qui doit lire les conditions d'utilisation d'un service numérique, en comprendre tous les aspects et faire un choix éclairé[80].

L'autorité de protection des données norvégienne (Datatilsynet) a en tiré les conséquences en 2023 dans le cadre d'une action visant Meta, en tout simplement interdisant la publicité ciblée basée sur le ciblage des utilisateurs sur Facebook et Instagram sans leur consentement[81]. La décision a été confirmé par le tribunal d'instance, à la suite d'un appel de Meta, et validé par le Comité européen de la protection des données (CEPD) dans une décision contraignante d'urgence le 27 octobre 2023[82]. Meta est désormais obligé de demander leur consentement aux consommateurs avant de collecter et traiter leurs données personnelles.

[77] Study on the impact of recent developments in digital advertising on privacy, publishers and advertisers, Commission européenne, 2023.

[78] Ibid.

[79] EU consumer protection 2.0: Protecting fairness and consumer choice in a digital economy, BEUC, 2022.

[80] Quand le décideur européen joue le jeu des Big techs… Engager une transition technologique pour sortir des dépendances numériques, Ophélie Coelho, 2021.

[81] Midlertidig forbud mot adferdsbasert markedsføring på Facebook og Instagram, Datatilsynet, 2023.

[82] EDPB Urgent Binding Decision on processing of personal data for behavioural advertising by Meta, CEPD, 2023.

DEMANDES DE L'UFC-QUE CHOISIR

Au vu de ces constats, l'UFC-Que Choisir, soucieuse de garantir aux consommateurs une réelle maîtrise de leurs données personnelles, exige des sites internet et applications qui les collectent une véritable transparence sur l'utilisation qui en est faite, et, de garantir aux consommateurs un accès et un contrôle sur les données personnelles qu'ils ont transmis à des tiers.

Par ailleurs, l'association rappelle que dans le cadre de sa campagne « Je ne suis pas une data », elle met à disposition des consommateurs sur le site respectemesdatas.fr un outil gratuit leur permettant de découvrir ce que les plateformes en ligne savent sur eux et de reprendre la main sur leurs données personnelles.