CSA

CONSEIL SUPÉRIEUR DE L'AUDIOVISUEL

FOCUS

CAPACITÉ À INFORMER DES ALGORITHMES DE RECOMMANDATION

Une expérience sur le service YouTube

Novembre 2019

© Conseil supérieur de l’audiovisuel

Résumé

Une étude empirique visant à mesurer l'impact de l'algorithme de recommandation du service YouTube sur la diversité et le pluralisme de l'information

Cette étude restitue les résultats d'une étude empirique menée par le Conseil visant à appréhender le fonctionnement d'un algorithme de recommandation d'une plateforme et son impact en matière de pluralisme de l'information et de diversité des opinions présentée aux utilisateurs.

Les plateformes, au premier rang desquelles se trouvent les réseaux sociaux, jouent un rôle de plus en plus important dans l'accès à l'information, en particulier chez les jeunes générations : 62 % des internautes déclarent utiliser les réseaux sociaux pour s'informer[1]. L'immédiateté, la fraîcheur des informations disponibles et la personnalisation des contenus selon les centres d'intérêts expliquent la place toujours croissante des réseaux sociaux dans l'accès à l'information[2].

Les algorithmes de recommandation, qui opèrent une sélection des informations mises à disposition de l'utilisateur, jouent un rôle central dans la consommation de contenus d'information sur ces plateformes. Les principaux critères régissant ces algorithmes sont connus : les préférences et historique de consommation de l'utilisateur, les caractéristiques des contenus (source, durée), leur succès (nombre de vues, partages, commentaires), etc. Toutefois, l'importance accordée à chacun d'entre eux et leur articulation sont plus floues et plus difficiles à observer. La question se pose alors de la capacité des recommandations d'un algorithme à offrir une variété de points de vue sur un sujet donné, permettant de s'informer de manière complète et diversifiée.

L'étude se concentre sur la plateforme de partage de vidéos YouTube qui occupe une place majeure dans la consommation de contenus vidéo sur Internet et propose une offre très diversifiée dans les thématiques abordées, les types d'éditeurs présents (chaînes de télévision, professionnels d'Internet, vidéastes amateurs, etc.), les formats utilisés ou encore la durée des contenus. Par ailleurs, YouTube rend publiques un grand nombre de données qui facilitent la caractérisation des contenus. Enfin, YouTube dispose d'une fonctionnalité de lecture automatique, qui s'appuie sur un algorithme de recommandation enchaînant les vidéos sans que le vidéonaute n'ait à intervenir. Elle se rapproche ainsi à certains égards à la diffusion linéaire en télévision. Ce sont les recommandations faites à partir de cette fonctionnalité qui ont été étudiées ici.

[1] Kantar pour La Croix. *Baromètre de la confiance dans les médias.* 01/2019.
[2] Médiamétrie. *La place croissante des réseaux sociaux dans l'accès à l'information*. 29/11/2016.

Principe de l'étude et méthodologie : l'examen des recommandations du mode lecture automatique de YouTube faites à chacun des participants

Le principe de l'expérience a consisté à lancer, à partir du compte Google de chaque participant, des vidéos sur divers sujets. Ces vidéos étaient les mêmes pour tous les participants. À partir de chaque vidéo de départ, les 10 vidéos successivement recommandées par la lecture automatique ont été analysées selon les prismes de la diversité des points de vue et de la proximité avec les centres d'intérêt des participants.

Les vidéos de départ portaient sur 23 thèmes présélectionnés, très divers et susceptibles de susciter des clivages d'opinions marqués, tels que la laïcité, le réchauffement climatique ou la vaccination[1]. Entre 3 et 5 vidéos présentant une diversité de points de vue ont été sélectionnées pour chacun de ces thèmes, soit un total de 101 vidéos de départ.

L'expérience a été réalisée à partir de 42 profils, dont ceux de 38 agents du Conseil[2] et 4 profils fictifs créés spécialement pour l'étude. Les participants ont rempli un questionnaire sur leurs habitudes et préférences de consommation sur YouTube, dans le respect du cadre juridique relatif à la protection des données personnelles[3]. Ces informations fournies par les participants et les vidéos suggérées par la lecture automatique de YouTube (plus de 39 000 au total) constituent la matière première de l'étude.

Résultats et enseignements

Les résultats de l'expérience menée par le Conseil permettent de dégager quelques enseignements sur le fonctionnement de l'algorithme de recommandation de YouTube s'agissant de la fonctionnalité de lecture automatique.

Dans ses premières recommandations, l'algorithme semble accorder moins d'importance au nombre de vues, à la date de publication ou encore au nombre de réactions qu'aux mots-clés associés au thème de départ des vidéos.

L'analyse des vidéos recommandées aux participants selon leur nombre de vues et leur date de publication fait apparaître que celles-ci ne sont ni les plus vues ni les plus récentes sur la plateforme. La moitié des vidéos recommandées comptent moins de 150 000 vues et 80 % en totalisent moins de 500 000. De même, seules 22 % des vidéos recommandées ont moins de 6 mois. L'algorithme ne semble pas non plus recommander en priorité les vidéos récentes qui ont fait un « *buzz* » : les deux vidéos de moins de deux mois les plus recommandées au cours de l'expérience comptaient respectivement 757 et 1 215 vues.

[1] Liste exhaustive des 23 thèmes : « Le grand remplacement », Bitcoin et cryptomonnaies, Corrida, Drague dans la rue, Droits sexuels et reproductifs, Éducation sexuelle à l'école, Épilation, Euro, Homosexualité, Intelligence artificielle, Laïcité, Machisme vs. Féminisme, Missions sur la lune, Mort de Michael Jackson, Réchauffement climatique, Réforme de la SNCF, Sectes, SIDA, Téléphone à l'école, Territoires palestiniens occupés, Vaccins, Véganisme, Vie extraterrestre.
[2] Le nombre de participants retenu répond avant tout à des préoccupations de nature pratique.
[3] RGPD, Règlement Général sur la Protection des Données.

Les vidéos les plus recommandées (plus de 40 fois au cours de l'expérience) portent généralement sur le thème de départ et sont majoritairement recommandées juste après la vidéo de départ. L'analyse des caractéristiques de ces vidéos n'a pas montré de trait commun : le nombre de vues, le nombre de « j'aime » / « je n'aime pas », et le nombre d'abonnés de la chaîne de publication varient fortement.

Dans ses premières recommandations, l'algorithme semble dès lors se focaliser sur le thème de départ et proposer un même corpus de vidéos aux utilisateurs, indépendamment des caractéristiques des vidéos. Pour proposer du contenu supposé pertinent aux utilisateurs, l'algorithme s'appuie sur les mots-clés présents dans le nom des chaînes, les titres et les *tags*[1] afin de lier les vidéos entre elles par un dénominateur commun.

La recommandation algorithmique pourrait produire des phénomènes dits de chambre d'écho à l'échelle du groupe de participants pouvant affecter la diversité des contenus proposés et le pluralisme des opinions exprimées dans les vidéos recommandées.

Sur l'ensemble des vidéos recommandées, 18 % concourent au débat associé au thème de départ (27 % lorsqu'on prend en compte les vidéos se rapportant à ce thème de manière plus large). Ce chiffre varie cependant considérablement selon les thèmes et peut aller de 2 % à 33 %. Parmi les facteurs pouvant expliquer ces différences figurent le nombre de vidéos disponibles sur le sujet sur YouTube et le dynamisme de la communauté d'utilisateurs portant le sujet en question et qui se matérialise notamment par le nombre de commentaires, partages ou réactions.

De plus, les résultats de l'étude mettent en exergue que la puissance des chaînes, c'est-à-dire la probabilité pour qu'un utilisateur qui vient de regarder une vidéo d'une chaîne donnée se voie ensuite proposer une vidéo de la même chaîne, est un facteur important de recommandation : sur l'ensemble des vidéos recommandées, 41 % des vidéos lues proviennent de la même chaîne que la vidéo directement précédente. Un phénomène dit de chambre d'écho[2], consistant à se voir recommander des vidéos de même type ou d'une même source pendant une longue période de temps, pourrait donc apparaître ici aussi.

S'agissant de la diversité des points de vue exprimés dans les vidéos recommandées sur un thème donné, plus d'un tiers des vidéos recommandées portant sur le thème de départ expriment le même point de vue que la vidéo de départ. Au fil des recommandations, 44 % des vidéos proposés par la lecture automatique vont dans le même sens que la vidéo directement précédente. Ce phénomène pourrait provenir des *tags* qui peuvent porter la marque de l'opinion exprimée dans la vidéo et prioriser ainsi dans les recommandations les vidéos exprimant le même point de vue. Le dynamisme de la communauté semble également jouer un rôle central.

[1] Un *tag* (ou étiquette, marqueur, libellé) est un mot-clé associé à de l'information (par exemple une vidéo), qui décrit une caractéristique de l'objet et permet un regroupement facile des informations contenant les mêmes mots-clés. Les *tags* sont habituellement choisis de façon personnelle par l'auteur ou l'utilisateur ; ils ne font souvent pas partie d'un schéma de classification prédéfini.

[2] Une chambre d'écho est une expression généralement retenue pour décrire une situation dans laquelle un individu se voit proposer de l'information, des idées, ou des croyances amplifiées ou renforcées par la répétition.

L'algorithme paraît individualiser les recommandations sans toutefois nécessairement tenir compte des profils des utilisateurs.

L'expérience a mis en lumière un « pivot » autour duquel l'algorithme semble définitivement dévier du thème de départ et individualiser les recommandations.

Après de la troisième vidéo proposée en lecture automatique, il recommande des vidéos de plus en plus populaires (nombre de vues) et de plus en plus récentes : près de 80 % des vidéos de plus d'un million de vues et 75 % des vidéos de moins de 2 mois ont été recommandées après la troisième itération.

L'analyse des parcours de consommation montre que lors de la première recommandation, 7 des 42 participants regardent la même vidéo. Ce chiffre baisse ensuite constamment au fil des itérations pour tomber à 1,6 à la dixième recommandation, ce qui laisse entrevoir une individualisation des contenus proposés aux participants.

Toutefois, le nombre important de vidéos recommandées plusieurs fois (6 000 vidéos sur les 16 700 vidéos différentes recommandées) indique que les participants se sont vus recommander les mêmes vidéos, indépendamment de leur profil. La personnalisation, c'est-à-dire la correspondance entre les recommandations et les centres d'intérêts de participants, semble relativement limitée. Il ressort d'une analyse sur un échantillon certes réduit que seules 11 % des vidéos recommandées aux participants correspondent effectivement à leurs habitudes ou à leurs préférences déclarées.

Les vidéos s'individualisent sans pour autant tout à fait coller aux préférences et habitudes des utilisateurs. Que ce soit pour les thématiques les plus consultées par les participants sur YouTube ou pour les recherches déjà effectuées par les participants sur les thèmes de départ, l'algorithme ne semble pas aller au-delà de l'individualisation en personnalisant les recommandations.

Sommaire

Introduction 9
Observations de YouTube 12
Accès à l'information 13
Médias utilisés pour s'informer 13
Place des algorithmes dans l'accès à l'information sur Internet 15
Principes de l'étude 22
Choix de YouTube 22
Échantillon de population 24
Choix des thèmes et vidéos de départ 25
Questionnaire sur les habitudes et préférences de consommation sur YouTube 27
Conformité avec le Règlement général sur la protection des données (RGPD) 28
Éléments d'attention sur le mode opératoire de l'étude 29
Résultats 31
Typologie des vidéos recommandées 31
Offre de vidéos portant sur le thème de départ 41
Diversité des opinions exprimées dans les vidéos recommandées 48
Influence des profils personnels sur les vidéos recommandées 52
Ingénierie inversée (*reverse engineering*) 58
Conclusion 64
Bibliographie 65
Annexes 66
Liste des vidéos de départ sélectionnées 66
Questionnaire sur les habitudes et préférences de consommation sur YouTube 71

Introduction

Le Conseil supérieur de l'audiovisuel (CSA) est le garant du respect du caractère pluraliste de l'expression des courants de pensée et d'opinion dans les médias audiovisuels[1]. Pour satisfaire à cet objectif, le Conseil s'assure que le principe déontologique de diversité de l'information, dans les opinions exposées, dans ses sources, dans son traitement et dans sa présentation, est au cœur des processus d'éditorialisation des services qu'il régule.

S'informer sur Internet est devenu une pratique courante, en particulier chez les jeunes générations, qui ont parfois tendance à délaisser les médias « traditionnels »[2]. La tendance à la baisse, encore légère, du temps passé à regarder la télévision (-5 minutes en quatre ans jusqu'à décembre 2018 pour les individus de 4 ans et plus) ou à écouter la radio (-4 minutes pour les 13 ans et plus) témoigne d'un glissement des habitudes vers les plateformes numériques, pour tout type de contenus, aussi bien de divertissement que d'information.

Sur Internet, de nombreuses plateformes jouent un rôle d'agrégateur de contenus d'information pour les utilisateurs. On y retrouve les médias traditionnels, de nouveaux médias disponibles uniquement sur Internet et des contenus créés par les utilisateurs eux-mêmes. Ces plateformes numériques, et notamment les réseaux sociaux, suggèrent des contenus en recourant à des algorithmes de recommandation, de classement ou de référencement, dont la transparence du fonctionnement fait débat. La capacité de ces algorithmes à informer les citoyens de façon pluraliste pose donc question.

Cet enjeu est au cœur des préoccupations des pouvoirs publics depuis plusieurs années.

Au niveau national, le Conseil national du numérique (CNNum) a introduit dans un avis rendu en 2014 sur la neutralité des plateformes[3] le principe général de loyauté des plateformes vis-à-vis des utilisateurs, qui implique une plus grande transparence du fonctionnement de leurs algorithmes. Les conclusions du CNNum ont été reprises par le Conseil d'État dans son étude annuelle de 2014[4], puis par le législateur dans la loi n° 2016-1321 pour une République numérique[5], qui

impose aux « *opérateurs de plateformes* » un certain nombre d'obligations relatives à la transparence des modalités de référencement, de classement et de déréférencement des biens

[1] L'article 13 de la loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication indique que « *le CSA assure le respect de l'expression des courants de pensée et d'opinion dans les programmes des services de radio et de télévision [...].* »
[2] CSA. *Vidéos en ligne ou télévision chez les jeunes publics : étude économétrique*. 01/2018.
[3] Le Conseil national du numérique dans son avis n° 2014-2 estime que « *l'impératif de loyauté implique aussi que les critères de pertinence et les principes régissant le fonctionnement des algorithmes soient explicités aux utilisateurs dans une démarche de littératie numérique* », page 33.
[4] Conseil d'État. *Étude annuelle 2014. Le numérique et les droits fondamentaux*. 09/09/2014.
[5] Loi n°2016-1321 du 7 octobre 2016 pour une République numérique.

ou des services qu'ils proposent[1]. De son côté, la Commission nationale de l'informatique et des libertés (CNIL) s'est interrogée en 2017 sur l'effet de la personnalisation algorithmique sur le pluralisme démocratique et culturel[2].

En 2016 puis en 2017, le CSA s'est quant à lui intéressé au rôle des algorithmes dans le modèle économique des plateformes numériques[3] et dans l'accès aux contenus audiovisuels[4]. En outre, parmi ses 20 propositions pour une régulation audiovisuelle refondée, rendues publiques en septembre 2018, le Conseil a préconisé une transparence et une loyauté accrues des algorithmes afin de favoriser la diversité culturelle et de diminuer le risque d'enfermement de l'utilisateur dans des choix prédéfinis[5].

Enfin, le Parlement a adopté la loi du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, qui impose des obligations aux opérateurs de plateforme en ligne, en particulier en matière de transparence des algorithmes, et confie au CSA le contrôle de leur bonne application.

Au niveau européen, le rapport d'experts sur la désinformation mandatés par la Commission Européenne considère que les plateformes doivent fournir aux utilisateurs toutes informations utiles sur le fonctionnement des algorithmes impliqués dans l'accès à l'information[6].

Dans ce contexte, le Conseil a souhaité engager une démarche novatrice pour les pouvoirs publics, tant dans sa méthode que dans ses objectifs, en menant une étude visant à mieux appréhender la capacité des algorithmes à informer. L'approche empirique retenue par le Conseil dans cette étude apporte une « vision du réel » qui complète des analyses théoriques que peut développer le monde de la recherche.

Cette étude tente de déterminer si les internautes ont accès à une information aussi pluraliste que les téléspectateurs qui s'informent dans un environnement régulé. Sur un sujet donné, les recommandations d'un algorithme offrent-elles une variété de points de vue permettant de s'informer de manière diversifiée et honnête ?

[1] Articles L.111-7 à L.111-8 du Code de la consommation.
[2] Commission Nationale Informatique & Libertés. *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*. 12/2017.
[3] CSA. *Plateformes et accès aux contenus audiovisuels.* 23/09/2016.
[4] CSA Lab. *Le rôle des données et des algorithmes dans l'accès aux contenus*. 01/2017.
[5] CSA. *Refonder la régulation audiovisuelle*. 09/2018.
[6] European Commission. *A multi-dimensional approach to disinformation*. 03/2018.

L'étude s'intéresse dans un premier temps aux différents moyens d'accès à l'information en France et au rôle joué dans ce contexte par les algorithmes. Les deux autres parties de l'étude sont consacrées à l'analyse des vidéos recommandées sur la plateforme de partage de vidéos YouTube afin de disposer d'une première appréhension du fonctionnement de son algorithme. La deuxième partie explique les principes de cette analyse ; la troisième en présente les résultats.

Dans une démarche de coopération avec les plateformes numériques, le Conseil a partagé avec YouTube et sa maison-mère Google les objectifs et résultats de ce rapport avant sa publication. Les représentants de la plateforme ont pu soumettre leurs remarques au Conseil. Celui-ci a pris en compte celles d'entre elles qu'il jugeait pertinentes dans la rédaction finale du rapport. Il a par ailleurs proposé à YouTube de disposer d'un espace d'expression dans ce rapport afin d'apporter toutes précisions que la société jugeait utile, notamment sur les efforts qu'elle met en œuvre en matière de régulation des contenus.

Observations de YouTube

Avec plus de cinq cents heures de vidéos mises en ligne chaque minute, et plus d'un milliard d'heures de vidéos visionnées chaque jour, YouTube attache une grande importance à ce que nos utilisateurs puissent trouver sur sa plateforme des vidéos de qualité, pertinentes et diverses. Nos utilisateurs disposent de plusieurs moyens pour accéder à cette diversité de vidéos, qu'il s'agisse de la page d'accueil de la plateforme, de son moteur de recherche ou des recommandations de vidéos.

La présente étude s'intéresse à l'une de ces fonctionnalités, la lecture automatique des vidéos recommandées par la fonctionnalité "à suivre", en particulier sur des sujets d'informations pour lesquels il existe une grande diversité de points de vue sur la plateforme, tant du côté des créateurs que des utilisateurs. À travers différents exemples, cette étude tend à démontrer que cette diversité se reflète dans les recommandations qui sont faites, notamment lorsqu'elle établit que 56 % des vidéos montrées expriment un point de vue différent de la vidéo précédente ou que les recommandations de vidéos divergent rapidement du thème de départ.

YouTube travaille constamment dans le but d'améliorer la qualité des recommandations faites à nos utilisateurs. À titre d'illustration, plusieurs centaines de modifications visant à améliorer les algorithmes de recommandations ont été déployées entre la réalisation de cette étude et sa publication.

Parmi ces mises à jour, nous avons notamment annoncé en janvier 2019[1] plusieurs améliorations visant à mettre en avant des informations de qualité qui font autorité, en particulier sur les sujets d'intérêt public, et à réduire les recommandations relatives aux vidéos qui se situent à la limite du règlement de la communauté YouTube ou qui sont susceptibles de désinformer les utilisateurs et peuvent être nuisibles (des vidéos vantant par exemple les mérites d'un remède miracle censé guérir une maladie grave, ou affirmant que la Terre est plate ou encore ayant manifestement pour but de désinformer les utilisateurs à propos de certains événements historiques).

Nous avons également annoncé en juin 2019[2] l'ajout de fonctionnalités donnant plus de contrôle à nos utilisateurs sur les recommandations qui leur sont faites sur la page d'accueil et dans les recommandations de vidéos "à suivre" concernées par la présente étude. Ainsi, nos utilisateurs peuvent désormais explorer plus facilement des sujets et des vidéos liés à ceux qu'ils regardent habituellement directement depuis leur page d'accueil, supprimer les suggestions de chaînes qu'ils ne souhaiteraient pas regarder ou encore mieux comprendre pourquoi une vidéo leur est suggérée dans les recommandations.

L'amélioration des recommandations sur la plateforme YouTube demeure un objectif constant pour l'ensemble de nos équipes. À ce titre, les différents retours de la communauté et des autorités nous sont précieux.

[1] YouTube Official Blog. *Continuing our work to improve recommendations on YouTube*. 25/01/2019.
[2] YouTube Official Blog. *Giving you more control over your Homepage and Up Next videos*. 26/06/2019.

Accès à l'information

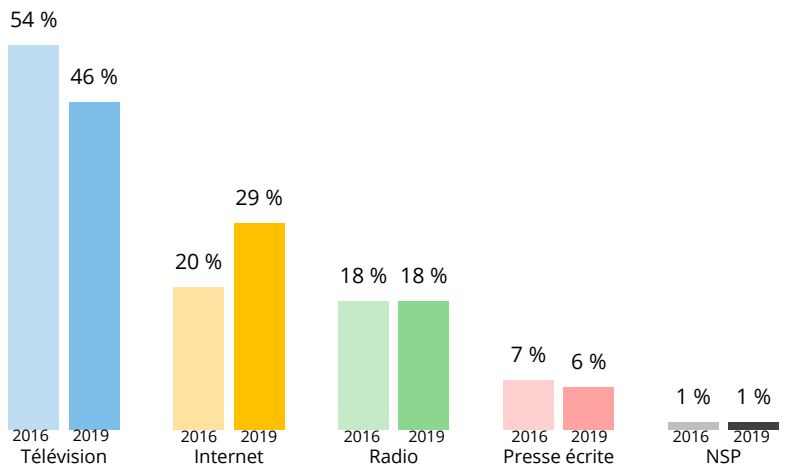
Médias utilisés pour s'informer

Dans cette étude, le terme « information » est utilisé dans une acception large, prenant en compte l'actualité et tout contenu apportant de la connaissance sur un sujet donné.

Les citoyens disposent de multiples canaux d'information : les canaux « traditionnels » tels que la presse imprimée, la télévision et la radio et les canaux apparus avec Internet tels que les *pure players* de l'information en ligne (Slate, Rue 89, Médiapart, Huffington Post…etc.), les pages d'actualité[1] et les réseaux sociaux, ces deux derniers agrégeant une part importante de l'information disponible sur Internet.

La dernière édition du baromètre de la confiance dans les médias, réalisé par Kantar[2] pour le journal La Croix, montre qu'en France, 46 % des individus s'informent principalement en regardant la télévision (-8 points depuis 2016). Internet arrive en deuxième position avec 29 % des individus l'utilisant comme première source d'information (+9 points depuis 2016). La radio et la presse imprimée sont respectivement utilisées en priorité par 18 % (stable depuis 2016) et 6 % des répondants (-1 point depuis 2016). Ces résultats montrent notamment que les utilisateurs se détournent de la télévision en faveur du numérique.

Les différents moyens d'accès à l'information en France en 2016 et en 2019

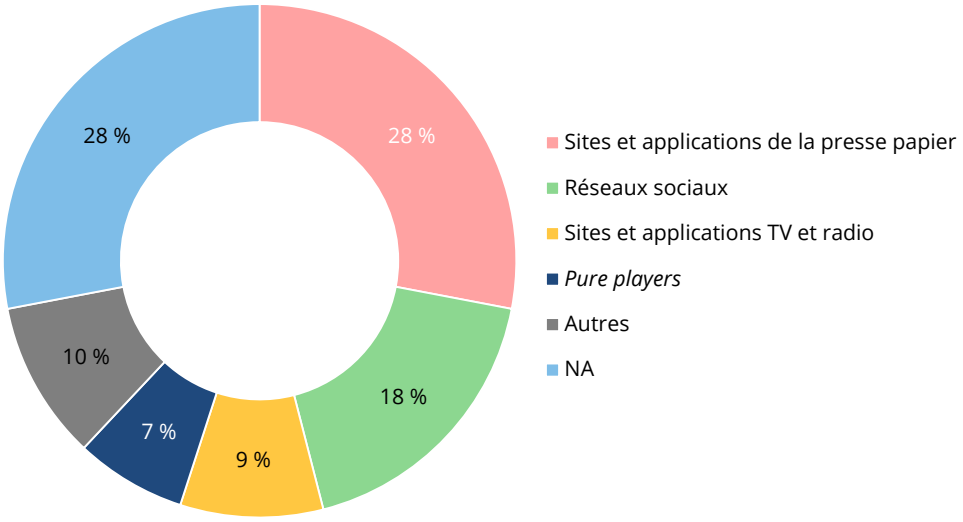


Source : Kantar pour La Croix. Baromètre de la confiance dans les médias. 01/2016 et 01/2019.
Base : ensemble de la population de 18 ans et plus.
Note : NSP = ne sait pas.

[1] Il peut s'agir de la page d'actualité des moteurs de recherche (exemple : https://fr.yahoo.com/) ou d'un fournisseur d'accès à Internet (exemple : https://www.orange.fr/portail).
[2] Kantar pour La Croix. *Baromètre de la confiance dans les médias*. 01/2019.

Sur Internet, les acteurs de la presse imprimée constituent la première source d'information à laquelle ont recours 28 % des individus déclarant s'informer en ligne. Les réseaux sociaux arrivent en deuxième position avec 18 % des répondants les présentant comme leur première source d'information sur Internet. Au total, 62 % des répondants affirment consulter les réseaux sociaux pour s'informer. Une étude de Médiamétrie menée début 2016 précisait que 51 % des personnes s'informant sur les réseaux sociaux lisaient en fait des articles provenant de pages des médias traditionnels[1]. Utilisés par 9 % des 18 ans et plus, les sites et applications des acteurs du secteur audiovisuel arrivent en cinquième position. La consommation de contenus d'information sur Internet passe donc bien souvent par les déclinaisons numériques des médias traditionnels. Toutefois, il convient de souligner que ces chiffres ne prennent pas en compte les jeunes de moins de 18 ans, qui sont des utilisateurs coutumiers d'Internet et des réseaux sociaux.

L'accès à l'information sur Internet en France en 2019



Source : Kantar pour La Croix. Baromètre de la confiance dans les médias. 01/2019.
Base : ensemble de la population de 18 ans et plus.
Note : NA = non applicable

Une étude de la Direction générale des médias et des industries culturelles (DGMIC) s'intéressant notamment à la consommation d'information des moins de 18 ans montre que 79 % des 15-34 ans utilisent Internet tous les jours ou presque pour s'informer[2]. Les réseaux sociaux sont pour eux le premier moyen d'accès à l'information sur Internet (71 %), sans toutefois être exclusifs de moyens plus traditionnels : les répondants déclarent utiliser en moyenne quatre sources d'information, parmi lesquelles les journaux télévisés (49 %) et les sites de vidéos en ligne (46 %).

[1] Médiamétrie. *La place croissante des réseaux sociaux dans l'accès à l'information*. 29/11/2016.
[2] Médiamétrie pour la DGMIC. *Les jeunes et l'information*. 07/2018.

Les caractéristiques sociodémographiques influent sur le choix des médias utilisés pour s'informer, mais également sur le degré d'analyse critique des informations proposées. Ainsi, la deuxième vague de l'enquête sur le complotisme menée par l'Ifop pour la Fondation Jean Jaurès et Conspiracy Watch[1] révèle que le degré de perméabilité aux théories du complot diminue avec l'âge (28 % des 18-24 ans croient à 5 théories ou plus contre 18 % des 50-64 ans). De même, les individus qui s'informent prioritairement sur Internet (plutôt les jeunes, donc) croient plus souvent à des théories du complot (27 % contre 15 % pour les personnes qui s'informent en priorité par la radio).

Pour autant, les réseaux sociaux sont perçus comme le moyen d'accès à l'information le moins crédible par les répondants du baromètre de la confiance dans les médias (25 % de taux de crédibilité). La radio est le média dans lequel les Français ont le plus confiance (50 %), suivie par la presse écrite (44 %) et la télévision (38 %). Toutefois, alors que ce niveau de confiance est resté stable pour Internet entre 2018 et 2019, il a fortement baissé pour les médias traditionnels (-6 points pour la radio, -8 pour les journaux et -10 pour la télévision).

Selon l'étude de Médiamétrie précitée[2], les réseaux sociaux répondent à deux attentes essentielles de la part des utilisateurs : des informations immédiates d'une part et présélectionnées pour eux d'autre part. Pour la moitié des répondants, « *la mise à jour rapide des informations* » est évoquée comme une raison de passer par les réseaux sociaux pour s'informer. La capacité des réseaux sociaux à proposer des contenus personnalisés, à agréger, trier et choisir les informations qui intéressent les utilisateurs reste un des principaux facteurs expliquant le recours à Internet et aux réseaux sociaux.

Les algorithmes de classement, de référencement et de recommandation jouent donc un rôle central dans les usages et dans l'attractivité des réseaux sociaux.

Place des algorithmes dans l'accès à l'information sur Internet

Parmi les agrégateurs d'informations, les portails d'actualité (tels que Google Actualités ou Yahoo Actualités) utilisent notamment des algorithmes de classement ou de référencement qui attribuent aux contenus des caractéristiques puis les trient en fonction de ces dernières[3]. L'approche algorithmique est alors centrée sur le contenu lui-même. Les réseaux sociaux utilisent quant à eux des algorithmes qui vont au-delà de ces fonctions de tri en personnalisant la mise en avant des contenus d'information. Le rôle de l'algorithme consiste alors à mettre en correspondance les attributs des contenus et ceux de l'utilisateur.

[1] Ifop pour la Fondation Jean Jaurès et Conspiracy Watch. *Enquête sur le complotisme*. 01/2019.
[2] Médiamétrie. *La place croissante des réseaux sociaux dans l'accès à l'information*. 29/11/2016.
[3] Certains portails d'actualité utilisent également des algorithmes de recommandation et de personnalisation.

Dans une note du CSA Lab publiée en 2017[1], le Conseil distinguait trois types d'algorithmes de mise en correspondance (« *matching* ») :

- les algorithmes-menus : les contenus proposés sont classés par genres et sous genres, selon une typologie prédéfinie, et l'utilisateur choisit au sein de ce menu ;
- les algorithmes statistiques : les contenus sont « poussés » vers l'utilisateur en fonction de ses propres consommations passées ainsi que des contenus « aimés » par les autres utilisateurs de la plateforme (cas de YouTube) ;
- les algorithmes sémantiques : les contenus sont analysés et indexés par la plateforme à l'aide de mots-clés. De son côté, l'utilisateur saisit des mots-clés caractérisant ses goûts. L'algorithme établit alors une correspondance entre les mots-clés descriptifs des contenus et ceux de l'utilisateur.

Les deux dernières catégories font explicitement appel aux goûts et préférences des consommateurs. L'algorithme statistique révèle par construction les préférences avérées des utilisateurs pour tel ou tel type de contenus et exploite ensuite ces informations pour recommander ce même type de contenus à l'utilisateur. L'algorithme sémantique y ajoute une part d'exploration en proposant des contenus qui sont conformes aux goûts déclarés par l'utilisateur mais qui ne sont pas nécessairement semblables à ses consommations passées.

Le cas spécifique de l'algorithme de YouTube

L'algorithme de YouTube a récemment fait l'objet de plusieurs enquêtes, notamment empiriques, visant à observer les mécanismes de recommandations de vidéos ou leurs effets[2]. Un article du *Guardian*[3] a ainsi mis en exergue que, dans les mois qui ont précédé l'élection présidentielle américaine et à partir d'un nombre égal de vidéos pro-Trump et pro-Clinton, l'algorithme avait recommandé six fois plus de vidéos en faveur du candidat devenu président des États-Unis. Le projet *Algo Transparency*[4] (qui a par ailleurs fourni les ressources de données nécessaires à l'enquête du *Guardian*) a montré que l'algorithme recommandait souvent des vidéos à caractère sensationnel, conspirationniste ou incitant à la haine.

Ces enquêtes ont été critiquées par YouTube qui réfute en particulier les méthodes et *in fine* les résultats. Par ailleurs, dans un objectif d'amélioration de ses recommandations, YouTube a depuis modifié son algorithme à de multiples reprises afin de garantir la transparence de l'actualité sur la plateforme[5]. YouTube a notamment annoncé le 26 juin 2019 avoir créé des fonctions additionnelles afin de donner plus de contrôle aux utilisateurs sur la page d'accueil et

[1] CSA Lab. *Le rôle des données et des algorithmes dans l'accès aux contenus*. 01/2017.
[2] Des témoignages d'utilisateurs ont également pu être mis en avant. Voir par exemple : Fondation Mozilla. *Vous avez des regrets sur YouTube ? C'est le cas de milliers de personnes*. 10/2019. ou The New York Times. *The making of YouTube radical*. 08/06/2019.
[3] The Guardian. *Fiction is outperforming reality*. 02/02/2018.
[4] https://algotransparency.org/
[5] YouTube Official Blog. *Greater transparency for users around news broadcasters*. 02/02/2018.

les recommandations automatiques[1]. Cette évolution permet à l'utilisateur de supprimer des suggestions de chaînes qu'il ne souhaite plus voir dans ses recommandations. Elle consiste également en l'ajout d'un encart en-dessous des vidéos recommandées aux utilisateurs sur la base des vidéos appréciées et vues par d'autres utilisateurs aux intérêts similaires, afin de mieux expliquer pourquoi certaines vidéos leur sont recommandées.

Ces enquêtes descriptives n'expliquent toutefois pas pourquoi l'algorithme de YouTube recommande certains contenus plus que d'autres. L'algorithme statistique de YouTube prend en compte de nombreux critères pour déterminer les contenus à proposer à l'utilisateur. Une partie importante des critères est documentée par YouTube[2] et par les utilisateurs eux-mêmes. Ces derniers s'appuient sur leur propre expérience et sur les précisions apportées par YouTube[3] pour augmenter la visibilité des vidéos. En revanche, les règles qui régissent l'articulation entre les différents critères et ainsi la capacité et la logique de recommandations de l'algorithme restent méconnues.

Les critères de recommandation utilisés par YouTube sont de différentes natures. En premier lieu, pour être répertoriées, les vidéos sont classées par grandes catégories (humour, musique, cuisine, auto…etc.). Les *tags* (ou étiquettes) des vidéos, renseignés par l'auteur de la publication, mais non visibles par l'utilisateur, sont ensuite utilisés par l'algorithme pour répertorier les vidéos plus finement et lier les vidéos entre elles. L'algorithme est alors capable d'enchaîner des vidéos présentant un certain niveau de proximité.

Une autre catégorie de critères repose sur la consommation de vidéos sur YouTube par les autres utilisateurs. Le nombre de vues et le temps de lecture (en proportion de la durée totale de la vidéo) permettent de mesurer la popularité de la vidéo. De même, le nombre de « j'aime » (et de « je n'aime pas ») est supposé refléter la satisfaction (l'insatisfaction) et l'approbation (la désapprobation) des utilisateurs. Le taux de clics (ou « *click-through rate* »), c'est-à-dire le rapport entre le nombre de clics sur une vidéo et le nombre de fois où la vignette de la vidéo est apparue sur l'écran des utilisateurs, constitue une mesure de son attractivité.

[1] Voir par exemple : Le Figaro. *YouTube permet de contrôler (un peu) ses algorithmes de recommandations.* 27/06/2019.
[2] https://www.youtube.com/creators/
[3] Voir par exemple : How-To Geek. *How does the YouTube Algorithm Work?.* 10/09/2018.

Critères de recommandation à l'échelle d'une vidéo

Source : YouTube avec ajouts CSA.

À l'échelle d'une session de lecture de plusieurs vidéos, des critères sont utilisés afin de caractériser des parcours de consommation similaires entre utilisateurs et de proposer les vidéos qu'un utilisateur est susceptible de souhaiter regarder. Un utilisateur A, ayant un comportement proche de celui d'un utilisateur B, a de fortes chances de se voir recommander une vidéo qui a été volontairement regardée par l'utilisateur B. Parmi les critères permettant d'évaluer cette proximité comportementale entre utilisateurs figurent[1] :

- le nombre de visionnages d'une même vidéo par un utilisateur,
- l'intervalle de temps entre les lancements d'une même vidéo,
- le temps de lecture de chaque vidéo lancée[2],
- la durée totale de la session de visionnage,
- le temps d'enchaînement entre les vidéos au cours de la session[3],

[1] Liste apparue dans plusieurs sources, dont : tubefilter. *Reverse Engineering the YouTube Algorithm: Part I*. 23/06/2016.
[2] Notamment en proportion de la durée totale de la vidéo.
[3] Qui varie de 0 (en cas de lecture automatique) au temps pris par l'utilisateur pour choisir la vidéo suivante.

- la durée de connexion sur la plateforme,
- la période de la journée à laquelle la session commence et s'arrête.

Ces critères peuvent également être utilisés pour comparer la session d'un utilisateur à ses sessions précédentes et déterminer le comportement de lecture de cet utilisateur à un instant t afin de lui proposer des contenus adaptés.

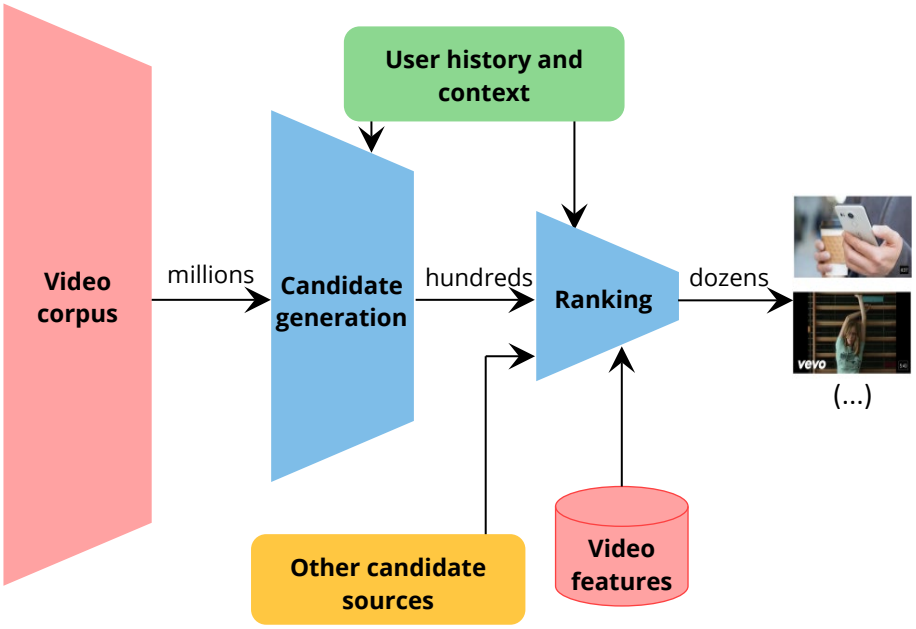
L'ensemble de ces critères a été schématisé par une équipe de recherche de Google lors d'une publication sur l'utilisation du « *machine learning* » dans le but d'améliorer le système de recommandations de YouTube et de mieux prédire les vidéos qui seront regardées le plus longtemps et par quel utilisateur[1].

Pour être recommandées, les vidéos doivent passer par deux étapes de filtrage. La première est appelée « *candidate generation* » et est définie comme l'historique de l'utilisateur sur YouTube (vidéos regardées et comportements de lecture). La seconde est celle du classement (« *ranking* »). Elle consiste à croiser les données de l'utilisateur avec les données des autres utilisateurs (« filtrage collaboratif ») et les caractéristiques des vidéos. L'algorithme assigne ensuite un score à chaque vidéo selon une fonction d'objectifs définie par YouTube. Les vidéos recommandées sont celles qui ont obtenu le meilleur score.

Les chercheurs indiquent que des centaines de caractéristiques sont prises en compte dans le modèle de classement des vidéos, parmi lesquelles le « *click-through rate* » ou encore leur date de publication. Il est aussi expliqué qu'une importance particulière est accordée aux critères décrivant le comportement de l'utilisateur avec une vidéo ou des vidéos similaires.

[1] Paul Covington, Jay Adams, Emre Sargin. *Deep neural networks for YouTube Recommendations*. 09/2016.

Architecture du système de recommandation de YouTube



Source : Paul Covington, Jay Adams, Emre Sargin. Deep neural networks for YouTube Recommendations. 09/2016.

Une enquête du *Pew Research Center*[1] a fait émerger deux autres critères vraisemblablement pris en compte dans le classement des vidéos sur la plateforme YouTube. L'étude a montré qu'au fil des recommandations, l'algorithme de YouTube recommande des vidéos de plus en plus longues et avec de plus en plus de vues. Ceci semble cohérent avec l'objectif, pour la plateforme, de maximiser le temps de visionnage de ses utilisateurs[2]. Même si elle n'est pas regardée dans son intégralité, une vidéo plus longue sera probablement regardée plus longtemps (à proportion de temps de lecture égale). De même, une vidéo plus populaire sera plus susceptible de plaire aux utilisateurs et donc d'être regardée plus longtemps.

Cette étude du *Pew Research Center* repose sur un nombre significatif de chaînes YouTube populaires sur la plateforme et éditées en anglais (30 481). À partir d'une des cinq vidéos les plus vues de chaque chaîne, l'étude a consisté à choisir aléatoirement une des cinq premières vidéos recommandées et ainsi de suite jusqu'à totaliser 5 vidéos (y compris la vidéo de départ).

[1] Pew Research Center. *Many Turn to YouTube for Children's Content, News, How-To Lessons*. 07/11/2018.
[2] Voir par exemple : MIT Technology Review. *YouTube is experimenting with ways to make its algorithm even more addictive*. 27/09/2019.

L'étude présentée ici par le Conseil suit une logique similaire, présentée dans la partie suivante. Toutefois, elle s'emploie à l'analyse des opinions véhiculées par les vidéos[1] et porte sur un volume de données moins important.

[1] Le journal de presse en ligne *Buzzfeed News* a également mené une enquête similaire, toutefois avec un volume restreint de vidéos et sur un seul sujet. À partir de quelques dizaines de requêtes sur le thème de la vaccination, les journalistes se sont attachés à identifier les vidéos anti-vaccination et leur ordre d'apparition. Source : Buzzfeed News. *YouTube Continues To Promote Anti-Vax Videos As Facebook Prepares To Fight Medical Misinformation*. 20/02/2019.

Principes de l'étude

Les données utilisées pour l'étude ont été recueillies selon un *modus operandi* conçu pour collecter et essayer de comprendre les recommandations proposées par l'algorithme de YouTube pour une vidéo de départ donnée. À partir de vidéos portant sur différents thèmes choisis par l'équipe en charge du projet pour la diversité de points de vue qu'ils offrent, les URLs et les métadonnées (titre des vidéos, chaîne, date de publication, etc.) des dix vidéos successivement recommandées par la lecture automatique de YouTube ont été récoltées afin de caractériser ces recommandations.

Choix de YouTube

La plateforme de partage de vidéos YouTube a été choisie pour plusieurs raisons.

Elle occupe une place majeure dans la consommation de contenus vidéo sur Internet : elle a rassemblé plus de 23 millions de vidéonautes uniques en France en juin 2018[1], représentant 82 % de la population ayant consulté une vidéo. À titre comparatif, Facebook, la deuxième plateforme la plus visitée[2], totalise un nombre d'utilisateurs uniques de 11 millions sur la même période[3]. En outre, toujours en juin 2018, près d'un milliard de vidéos ont été visionnées sur YouTube pour un total de plus de 44 millions d'heures[4] (soit deux fois plus que Facebook)[5].

Ensuite, YouTube propose une grande diversité dans les thématiques abordées, les producteurs de contenus (chaînes de télévision, professionnels d'Internet, vidéastes amateurs, etc.), les formats utilisés et la durée des contenus. La diversité des thèmes est un critère important dans la mesure où l'étude vise à analyser les vidéos recommandées à partir d'une sélection de vidéos sur des sujets divers (cf. *infra*).

De plus, YouTube rend public un grand nombre de données permettant d'appréhender la typologie des vidéos : nombre de vues des vidéos et nombre d'abonnés aux chaînes (popularité), date de publication (ancienneté), nombre de commentaires (intensité des discussions suscitées par la vidéo), nombre de « j'aime » et de « je n'aime pas » (engagement des vidéonautes,

[1] Le mois de juin 2018 est présenté ici car l'expérience a eu lieu durant ce mois.
[2] *Brand Players* : Agrégation de l'audience du ou des *player(s)* vidéo d'un même diffuseur quel que soit le site sur lequel il est diffusé (définition Médiamétrie//NetRatings).
[3] La deuxième position de ce classement est habituellement occupée par Dailymotion (avec plus de 11 millions de vidéonautes uniques par mois). Toutefois, pour des raisons techniques, le nombre de vidéonautes uniques de cette plateforme n'a pas pu être compté sur le mois de juin 2018.
[4] *Brand Support* : Agrégation de l'audience des pages d'un site sur lesquelles des vidéos ont été vues, quel que soit le *player* sur lequel est jouée la vidéo (définition Médiamétrie//NetRatings).
[5] Médiamétrie//NetRatings. *L'Audience Vidéo Ordinateur en France en juin 2018*. 11/01/2019.

satisfaction et approbation) ou encore les *tags*[1] permettant de thématiser chaque vidéo. La grande majorité de ces données a été utilisée pour l'analyse des vidéos recommandées.

Enfin, YouTube dispose d'une fonctionnalité de lecture automatique, qui s'appuie sur un algorithme de recommandation, enchaînant les vidéos sans que le vidéonaute n'ait à intervenir. Cette fonctionnalité, qui permet d'imiter un comportement passif pouvant se rapprocher dans certains cas de celui du téléspectateur en télévision, a été retenue pour l'enchaînement des recommandations étudiées[2]. D'autres similitudes entre YouTube et la télévision peuvent également être soulignées, notamment la présence de vidéos s'apparentant à des programmes de flux et d'autres à des programmes de stock[3] (de par leurs statistiques de consommation) et l'éditorialisation poussée de certaines chaînes (publication à des heures précises pour créer de l'attente)[4].

La lecture automatique sur YouTube

Source : YouTube avec ajouts CSA.

Par ailleurs, il est important de relever que la démarche du Conseil consistant à récolter les URL de vidéos recommandées à différentes personnes par un algorithme, ne vient en rien perturber le comptage statistique du nombre de vues des vidéos de la plateforme. Même à supposer que la

[1] Les *tags* des vidéos ne sont pas visibles sur la page YouTube des vidéos mais sont consultables grâce à l'API de Google.
[2] À noter que dans les questionnaires sur les habitudes et préférences de consommation des participants à l'étude, 65 % d'entre eux utilisant YouTube ont indiqué « *suivre la lecture automatique de temps en temps* », 24 % « *ne la suivent jamais* », 5 % « *la suivent la plupart du temps* » et 5 % « *systématiquement* ».
[3] Les programmes de stock sont diffusables plusieurs fois car après leur première diffusion, ils conservent un intérêt pour le téléspectateur (exemple : cinéma). À l'inverse, les programmes de flux ne présentent un intérêt qu'éphémère (exemple : journal télévisé).
[4] CSA. *Vidéos en ligne ou télévision chez les jeunes publics : étude économétrique*. PP19-21. 01/2018.

totalité des participants (42) se soient vus recommandés les mêmes vidéos[1], leur nombre restreint n'augmenterait qu'à la marge le nombre de vues total de chaque vidéo, le nombre de vues moyen des vidéos recommandées lors de l'étude étant de plus d'un million[2]. En outre, YouTube utilise des techniques algorithmiques lui permettant d'actualiser à la baisse le nombre de vues s'il détecte des vues manifestement fictives[3].

Le modèle économique de YouTube, reposant sur la popularité des vidéos notamment mesurée en nombre de vues qui conditionne la vente d'espaces publicitaires aux annonceurs et le partage de revenus avec les créateurs, ne se trouve donc pas affecté par cette étude.

Échantillon de population

Afin de réunir un nombre de participants suffisamment grand pour disposer d'un large échantillon de vidéos recommandées, un appel à participation volontaire a été lancé auprès des agents du Conseil au début du mois de juin 2018.

La seule condition pour prendre part à l'étude était de disposer d'un compte Google[4]. En effet, si un internaute est connecté à son compte Google lorsqu'il regarde des vidéos sur YouTube, son adresse de messagerie électronique Google est automatiquement utilisée pour créer un identifiant YouTube. Cet identifiant est par la suite réutilisé à chaque passage sur la plateforme, sauf si l'utilisateur fait la démarche de se déconnecter de son compte Google. Ainsi, l'historique de consommation sur la plateforme est disponible sur n'importe quel écran dès lors que l'utilisateur est connecté à son compte Google. Cet élément est important au vu de l'objectif de l'étude d'identifier notamment l'influence des profils de consommation sur les recommandations de l'algorithme.

Après une présentation des principes et objectifs de l'étude et la mise en place d'un protocole garantissant la protection de leurs données personnelles conformément à la réglementation en vigueur (cf. *infra*), 38 agents ont décidé de participer à l'étude.

Bien qu'il ait pu être démontré que les données sociodémographiques (sexe, âge, catégorie professionnelle, etc.) pouvaient jouer un rôle important dans les comportements d'accès à l'information (cf. *supra*), aucune information de ce type n'a été recueillie sur l'échantillon de population, afin que les participants aient la garantie de ne pas pouvoir être identifiés, conformément aux dispositions du Règlement général sur la protection des données (RGPD).

Aux agents du Conseil ont été ajoutés quatre profils fictifs créés spécialement pour l'étude, deux centrés sur la thématique « football », les deux autres sur les thématiques « vie pratique », « mode » et « maquillage ». Les profils fictifs, pour lesquels les préférences et habitudes de

[1] Dans les faits, certaines vidéos ont été recommandées au maximum 57 fois.
[2] Moyenne réalisée sur un échantillon sans doublon.
[3] Site Internet de YouTube. *Méthode de comptabilisation des visionnages des vidéos*. 30/01/2018. https://support.google.com/youtube/answer/2991785?hl=fr.
[4] YouTube appartient à Google depuis octobre 2006.

consommation sont particulièrement marquées, permettent, par comparaison avec les profils réels, d'apprécier l'adhérence des vidéos recommandées aux intérêts de l'utilisateur.

En pratique, des profils fictifs Google (et donc YouTube) ont été créés et ont été nourris d'un nombre très important de vidéos portant sur la thématique choisie au cours de plusieurs sessions de visionnage. Les historiques de navigation de l'ordinateur utilisé avaient préalablement été supprimés. Ces comptes ont par la suite suivi le même principe de lancement des vidéos de points de départ que pour n'importe quel participant réel.

Choix des thèmes et vidéos de départ

Le principe de l'expérience consiste à lancer, à partir du compte YouTube de chaque participant, des vidéos sur divers sujets. Ces vidéos de départ sont les mêmes pour tous les participants. À partir de chaque vidéo de départ, les 10 vidéos successivement recommandées par la lecture automatique sont recueillies et par la suite analysées selon les prismes de la diversité des points de vue exprimés et de la proximité avec les centres d'intérêt des participants.

Les vidéos de départ portent sur 23 thèmes présélectionnés, recensés dans le tableau ci-dessous. Ces thèmes portent sur des sujets sociétaux qui suscitent de forts clivages d'opinions. Ces différences de point de vue sont désignées par « opinion A » et « opinion B » et peuvent représenter diverses formes de clivage, précisées pour chaque thème dans le tableau.

En outre, il est important de souligner que la démarche du Conseil ne s'inscrit pas dans le sillon des initiatives de vérification de fait (« *fact-checking* ») et de lutte contre la manipulation de l'information. Les clivages d'opinions sont abordés au sens strict, sans parti-pris sur les opinions exprimées, et ce même si l'intérêt public voudrait parfois qu'une opinion soit plus mise en avant qu'une autre. Les points de vue exprimés ne sont soumis à aucun jugement et il n'est pas question ici de déterminer si une opinion portée dans les vidéos est vraie ou fausse. La qualification des opinions exprimées dans les vidéos est purement descriptive.

Pour chacun des thèmes choisis, entre 3 et 5 vidéos ont été sélectionnées, portant à 101 le nombre de vidéos de départ (liste complète en annexe). Pour chaque thème, les vidéos présentent une diversité de points de vue selon la nomenclature ci-dessous et dans des proportions équilibrées dès lors que la disponibilité des vidéos le permet. Des vidéos neutres, qui concourent au débat en présentant les arguments des différentes parties, ont été ajoutées lorsque cela était possible.

Liste des thèmes sélectionnés et points de vue associés

Thèmes	Opinion A	Opinion B
« Le grand remplacement »	Théorie vraie (2)	Théorie fausse (2)
Bitcoin et cryptomonnaies	Opportunité (2)	Danger (2)
Corrida	Pour (2)	Contre (2)
Drague dans la rue	Pour (2)	Contre (2)
Droits sexuels et reproductifs	Pour (2)	Contre (1)
Éducation sexuelle à l'école	Pour (2)	Contre (2)
Épilation	Pour (2)	Contre (2)
Euro	Bénéfique (2)	Néfaste (1)
Homosexualité	*Gay-friendly*[1] (2)	Homophobie (2)
Intelligence artificielle	Opportunité (2)	Danger (1)
Laïcité	Libérale (2)	Stricte (2)
Machisme-Féminisme	Féminisme (2)	Machisme (2)
Missions sur la lune	Ont eu lieu (2)	N'ont pas eu lieu (2)
Mort de Michael Jackson	A eu lieu (2)	N'a pas eu lieu (2)
Réchauffement climatique	Existe (2)	N'existe pas (2)
Réforme de la SNCF	Pour (2)	Contre (2)
Sectes	Choix de vie (1)	Danger (2)
SIDA	Origine naturelle (2)	Virus créé artificiellement (2)
Téléphone à l'école	Pour (1)	Contre (2)
Territoires palestiniens occupés	Un État (1)	Deux États (2)
Vaccins	Pour (2)	Contre (2)
Véganisme	Pour (2)	Contre (carnisme) (2)
Vie extraterrestre	Existe (2)	N'existe pas (2)

Note : le chiffre entre parenthèse indique le nombre de vidéos de départ portant l'opinion associée.

Le choix des vidéos de départ devait satisfaire également à d'autres contraintes : être en langue française, avec un nombre de vues minimum pour les vidéos de départ fixé à un seuil volontairement bas de 200 vues[2] ; il a en effet été constaté au cours de tests liminaires que YouTube recommande parfois des vidéos avec peu de vues.

[1] *Gay-friendly* (que l'on peut traduire par « amical envers les homosexuels ») est un anglicisme se référant à des lieux, des politiques, des personnes ou des institutions qui sont ouverts et accueillants envers les membres de la communauté LGBT.
[2] À titre de comparaison, le nombre moyen de vues de l'échantillon de vidéos de l'étude est de 932 388. Seules 0,4 % des vidéos de l'échantillon ont moins de 200 vues. Très peu de vidéos ont un nombre de vues en dessous de ce niveau, ce qui en fait une tranche basse pertinente.

Ensuite, les vidéos de départ ont été lancées dans un ordre aléatoire mais identique pour chaque participant. Une fois que les titres des 10 vidéos successivement recommandées par la lecture automatique après le lancement d'une vidéo de départ ont été connus, la vidéo de départ suivante a été lancée dans l'ordre prédéfini. Pour prendre en compte le critère de « *watch time* » de YouTube, le temps de lecture de chaque vidéo a été fixé de manière aléatoire entre 10 et 45 secondes. Du fait du nombre de vidéos de départ et du nombre de participants, la collecte des URL des vidéos, réalisée fin juin 2018, a duré 10 jours.

Questionnaire sur les habitudes et préférences de consommation sur YouTube

Afin de déterminer si les vidéos recommandées aux participants par l'algorithme de recommandation de la lecture automatique avaient un lien avec leurs centres d'intérêt, les participants ont rempli un questionnaire sur leurs habitudes et préférences de consommation sur YouTube[1].

Les quelques statistiques d'habitudes et de préférences présentées ci-dessous se basent sur un échantillon de population limité et sont de ce fait à prendre avec précaution. En revanche, le nombre de vidéos recommandées (39 000) constitue un échantillon suffisamment important pour une analyse robuste des résultats présentés en troisième partie de cette étude.

Les participants ont d'abord été interrogés sur le type d'équipement utilisé pour regarder des vidéos sur YouTube. L'ordinateur personnel arrive en tête (84 % des répondants[2]), devant le *smartphone* (81 %) et l'ordinateur professionnel (68 %). Plus d'un tiers des répondants ont déclaré utiliser ces trois équipements pour regarder des vidéos sur YouTube.

L'utilisation de l'ordinateur personnel coïncide avec les horaires d'utilisation de YouTube : la soirée (de 20 heures à 23 heures) est le moment privilégié pour regarder des contenus (86 % des répondants) lorsque les participants sont chez eux et qu'ils sont plus susceptibles d'avoir recours à leur ordinateur personnel. Par ailleurs, près de 54 % des participants ont déclaré consommer des vidéos sur YouTube de 10 heures à 17 heures sur leur ordinateur professionnel.

Un des objectifs de l'étude étant de déterminer si l'historique de consommation sur YouTube a une influence importante sur les recommandations, les participants ont été invités à choisir parmi plusieurs grandes thématiques[3] les types de contenus qu'ils consultent le plus souvent sur YouTube. Les vidéos musicales sont les plus visionnées (78 % des répondants) ; les contenus humoristiques arrivent en deuxième position (70 %), devant les contenus d'information (65 %).

[1] Ce questionnaire est disponible en annexe.
[2] Hors profils fictifs et répondant n'utilisant pas YouTube.
[3] Information, politique, musique, humour, sport.

Enfin, 70 % des participants indiquent suivre « *de temps en temps* » les recommandations faites par l'algorithme. De même, la lecture automatique est utilisée « *de temps en temps* » par 65 % des répondants[1]. Enfin, les participants ont pu donner leur avis sur les recommandations qui leur sont faites : 86 % d'entre eux estiment que l'algorithme leur propose des contenus « *qui se ressemblent* ». Parmi eux, 47 % y sont indifférents, 31 % disent être gênés et 22 % apprécient les recommandations.

Conformité avec le Règlement général sur la protection des données (RGPD)

Les participants ont utilisé leurs identifiant et mot de passe Google pour se connecter à leur compte avant le lancement des vidéos de départ. Ces informations n'ont pas été sauvegardées ni fait l'objet d'une collecte ou d'un traitement par le Conseil.

En revanche, deux autres types de données à caractère personnel ont été traités :

- les informations fournies dans le questionnaire sur les préférences et habitudes de consommation sur YouTube ;
- les vidéos suggérées par la lecture automatique de YouTube.

Ces données constituent la matière principale de l'étude et ont été soumises à un traitement statistique. Aucune donnée sensible (au sens du RGPD) n'a été traitée.

À partir de la rédaction d'une analyse d'impact relative à la protection des données qui a permis d'évaluer l'origine, la nature, la particularité et la gravité des risques potentiels de l'expérience pour les droits des personnes, deux types de mesures ont été prises :

- le recueil du consentement libre, spécifique, éclairé et univoque des participants à l'expérience en les informant de leur droits et du caractère facultatif de leur participation[2] ; et,
- la mise en place de mesures de sécurité organisationnelles et techniques pour garantir aux participants un degré de confidentialité maximal des réponses apportées ou générées dans le cadre expérimental (minimisation et pseudonymisation des données collectées).

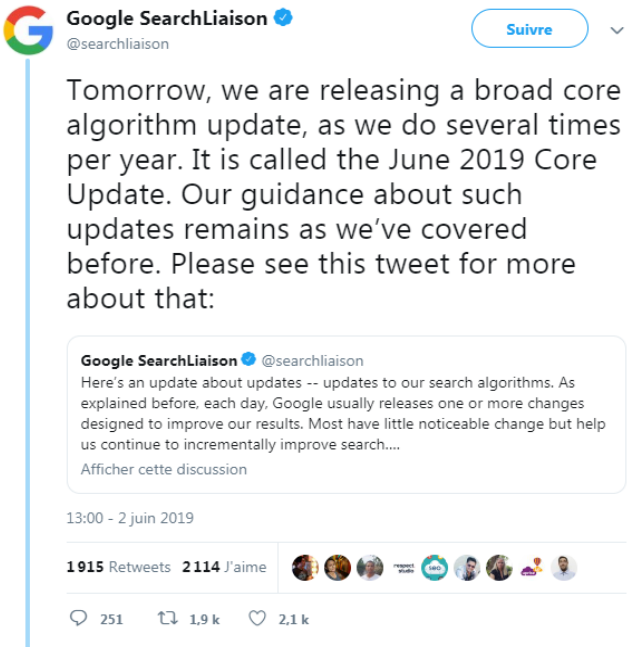
[1] En complément, 15 % des répondants à l'enquête du *Pew Research Center* (4 594, États-Unis) affirment suivre les recommandations régulièrement, 66 % les suivre de temps en temps et 19 % ne jamais les suivre. Source : Pew Research Center. *Many Turn to YouTube for Children's Content, News, How-To Lessons*. 07/11/2018.
[2] Art. 6.1.a. Règlement (UE) 2016/679.

Éléments d'attention sur le mode opératoire de l'étude

En dehors des limites imposées par le RGPD énoncées plus haut, plusieurs points d'attention concernant cette étude doivent être relevés.

Tout d'abord, l'analyse porte sur un outil difficilement cognoscible. L'analyse a été menée à un moment donné de développement de l'algorithme et ne préjuge pas des modifications qui ont pu y être apportés ultérieurement. En effet, les plateformes mettent régulièrement à jour leurs algorithmes, comme en atteste la communication récente de Google dans l'exemple ci-dessous. YouTube a précisé au Conseil lors de la phase d'échange que, dans une démarche constante d'amélioration de son dispositif de recommandations, son algorithme avait notamment fait l'objet d'une centaine de modifications en 2018.

Annonce par Google le 2 juin 2019 d'une mise à jour de son algorithme



Source : compte Twitter certifié de GoogleSearchLiaison.
Traduction libre : « Demain, nous lancerons une mise à jour importante de notre algorithme, comme nous le faisons plusieurs fois par an. Cette mise à jour s'appelle « June 2019 Core Update *». Nos conseils concernant ces mises à jour restent les mêmes que ceux décrits précédemment. (...) ».*

Par ailleurs, les effets du référencement payant des chaînes et de leurs vidéos, qui entre également dans le système de pondération des critères de recommandation, n'ont pas pu être isolés, les accords de référencement étant couverts par le secret des affaires.

De même, Google et YouTube appartenant au même groupe et les données de consommation pouvant être mutualisées entre services, il ne peut être exclu que les recherches sur Google aient une influence sur les recommandations de l'algorithme de YouTube et inversement[1]. Lors de la phrase d'échange avec les représentants de YouTube, ces derniers ont indiqué au Conseil que tel n'était pas le cas.

De plus, l'hypothèse peut être émise que l'heure à laquelle la collecte des URLs des vidéos a été effectuée a pu avoir un impact sur les recommandations, sans qu'il soit possible d'évaluer cet éventuel impact. Par exemple, certains contenus pourraient être plus intensément recommandés en soirée, en fonction de leur type ou encore de leur durée.

En outre, YouTube a tenu à rappeler l'importance de la consommation des autres utilisateurs dans les critères pris en compte par l'algorithme. Ainsi, l'algorithme se basant notamment sur ce que les autres utilisateurs ont regardé, le fait que les participants aient eu à regarder les mêmes vidéos de départ a pu conduire l'algorithme à leur faire des recommandations plus similaires.

Le Conseil est également conscient des questions soulevées par la représentativité de l'échantillon de participants. Pour des raisons pratiques, cet échantillon n'a pu être ouvert à des personnes extérieures au Conseil. L'automatisation du processus de collecte des données nécessitant une sédentarisation des outils informatiques, il était en effet plus commode que les volontaires puissent y accéder sans avoir à se déplacer. Cela a en outre facilité le contrôle du respect des règles relatives à la protection des données personnelles.

Enfin, le choix d'analyser la seule fonctionnalité de lecture automatique concentre l'analyse sur une des dimensions du dispositif de recommandation de YouTube. Cet outil a été choisi car il imite un comportement passif de la part du vidéonaute, assimilable à celui du téléspectateur qui regarde les programmes successifs d'une même chaîne de télévision. Toutefois, sur YouTube comme devant la télévision, l'utilisateur peut également adopter un comportement proactif, respectivement en recherchant ou en sélectionnant dans les recommandations la vidéo suivante et en changeant de chaîne. La part des comportements passif et actif dans l'ensemble du temps de visionnage (sur YouTube et devant la télévision) n'étant ni connu ni modélisable, un arbitrage a dû être fait.

[1] Il semble que ce soit le cas pour *Google News* par exemple, le centre d'aide du portail d'actualités indiquant que « *les algorithmes personnalisent votre interface utilisateur en fonction de vos paramètres Google Actualités, comme vos centres d'intérêt et vos sources, ainsi qu'en fonction de votre activité passée sur Google, par exemple dans la recherche Google et sur YouTube.* »
Source : https://support.google.com/googlenews/answer/9005749?hl=fr.

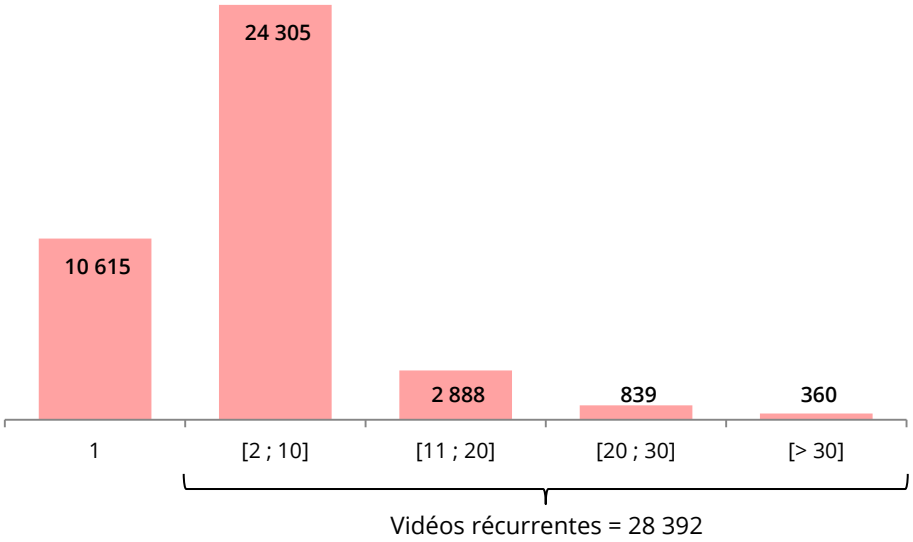
Résultats

Les recommandations récoltées ont été analysées sur la base de plusieurs critères et indicateurs. Ces éléments et les conclusions qui en découlent sont présentés dans cette partie.

Typologie des vidéos recommandées

Au total, 39 007 vidéos ont été recommandées aux participants. Certains participants se sont vus recommander les mêmes vidéos : 6 086 vidéos distinctes ont été recommandées à plusieurs reprises. Certaines vidéos ont même été vues plusieurs fois par un même participant (nombre d'occurrences supérieur au nombre de participants, soit 42). Le nombre total de vidéos distinctes (vidéos vues une seule fois et première occurrence des vidéos vues plusieurs fois) s'élève à 16 701, soit 43 % de l'ensemble des vidéos[1].

Nombre de vidéos recommandées et nombre d'occurrences des recommandations



Base : total des vidéos recommandées (39 007).
Note : une vidéo récurrente recommandée 9 fois sera comptabilisée autant de fois.

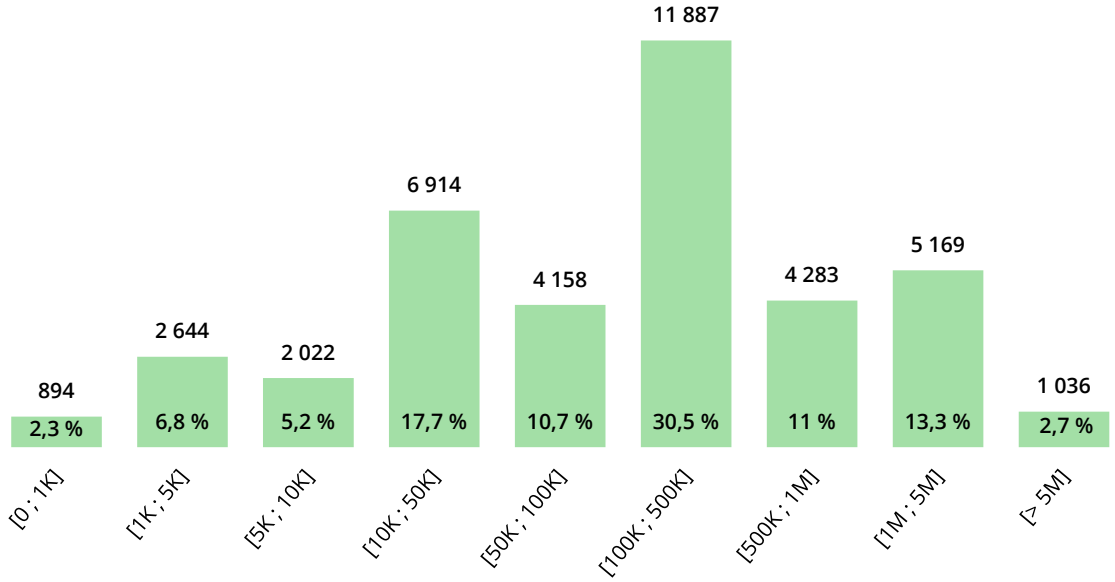
Popularité des vidéos recommandées

Les vidéos recommandées aux participants ne sont pas les plus vues sur la plateforme. Le nombre moyen de vues des vidéos est de 932 388. Toutefois, la moitié des vidéos recommandées concentre chacune moins de 150 000 vues. Plus de 30 % des vidéos rassemblent

[1] Sauf mention contraire, les résultats sont calculés sur la base de l'ensemble des vidéos afin de prendre en compte toutes les vidéos ayant été recommandées à tous les participants.

entre 100 000 et 500 000 vues et 18 % des vidéos comptent entre 10 000 et 50 000 vues. Il aurait été intéressant de comparer cette répartition du nombre de vues des vidéos de l'étude avec celle de l'intégralité des vidéos présentes sur YouTube. Le Conseil ne dispose toutefois pas de cette information.

Répartition des vidéos recommandées par la lecture automatique par nombre de vues



Base : total des vidéos recommandées (39 007).
Note : métadonnées collectées le 04 juillet 2018.

La répartition des vidéos recommandées par nombre de vues pour chacun des thèmes est, dans la majorité des cas, similaire à celle de l'ensemble des thèmes et des vidéos analysés[1]. Seuls 5 thèmes se distinguent : les thèmes « Téléphone à l'école » et « Euro » présentent un nombre plus important de vidéos avec peu de vues ; *a contrario*, les thèmes « Drague dans la rue », « Machisme/féminisme » et « Mort de Michael Jackson » présentent un nombre plus important de vidéos avec beaucoup de vues.

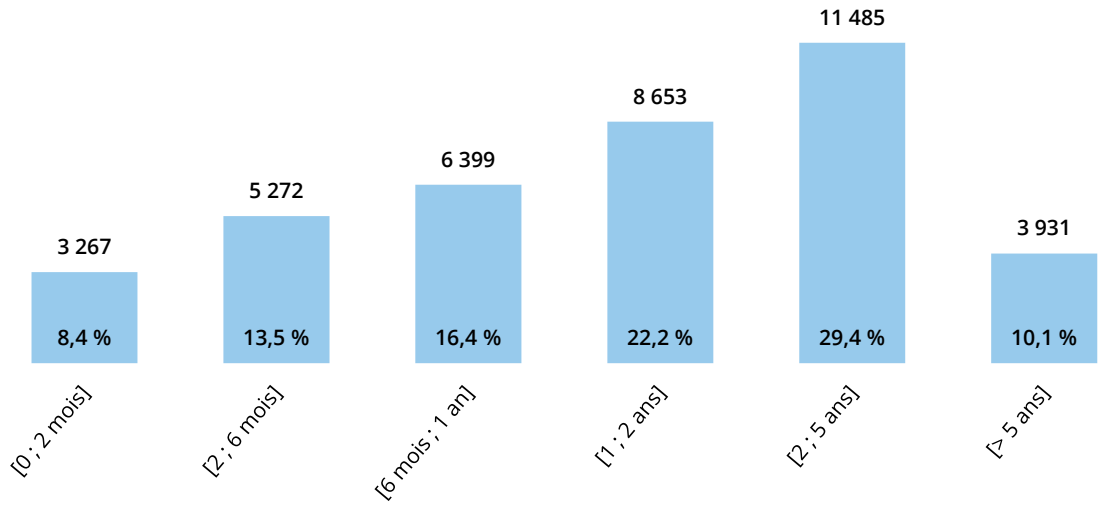
Les disparités observées selon les thèmes de départ pourraient s'expliquer par la popularité du thème lui-même. Cependant, pour chaque vidéo de départ, le nombre de vidéos portant sur ce même thème est assez restreint (voir page 40). Ce faisant, la popularité du thème ne peut être retenue comme explication principale à la courbe de répartition de l'ensemble des vidéos recommandées. En revanche, l'algorithme propose souvent des vidéos de la même chaîne (voir page 36), ce qui pourrait, selon sa notoriété et le nombre de vues des vidéos qu'elle agrège, modifier la courbe de répartition du nombre de vues.

[1] De même, la courbe de répartition du nombre de vues des vidéos uniques est similaire à celle pour l'ensemble des vidéos.

Date de publication des vidéos recommandées

Les vidéos recommandées ne sont pas non plus les vidéos les plus récentes. Leur ancienneté moyenne est d'un peu plus de 2 ans, et la médiane se situe à 1 an et 4 mois. Près de 30 % des vidéos ont été publiées entre 2 et 5 ans avant la collecte des métadonnées (qui est intervenue postérieurement à la collecte des vidéos).

Répartition des vidéos recommandées par la lecture automatique selon leur ancienneté
(Temps écoulé depuis leur date de publication)



Base : total des vidéos recommandées (39 007).
Note : date de référence fixée au 4 juillet 2018.

La répartition des vidéos recommandées selon leur ancienneté diffère sensiblement selon les thèmes. Par exemple, les vidéos sur l'euro et sur la corrida sont en moyenne plus anciennes (respectivement 32 % de vidéos de plus de 5 ans et 60 % de vidéos de plus de 2 ans). Au contraire, les vidéos sur les cryptomonnaies et l'intelligence artificielle sont plus récentes (respectivement 62 % de vidéos de moins d'un an et 75 % de vidéos de moins de deux ans). Ce phénomène peut être lié à l'actualité et à la présence de ces thèmes dans le débat public. Il s'agit cependant d'un effet indirect : l'algorithme peut proposer des vidéos récentes mais pas nécessairement en lien avec le thème de départ, et la récence d'un thème semble appeler des recommandations de vidéos récentes sans qu'elles portent sur le même sujet.

Recoupement entre le nombre de vues et la date de publication

Plus de 10 % des vidéos recommandées et qui ont été publiées moins de deux mois avant l'étude comptent plus d'un million de vues. Ce taux ne monte qu'à 16 % lorsqu'est pris en compte l'ensemble des vidéos (quelle que soit leur date de publication), ce alors même que les vidéos plus anciennes ont eu plus de temps pour accumuler des vues. Après une période d'augmentation intense du nombre de vues juste après la publication des vidéos recommandées, la croissance du nombre de vues ralentit ainsi au fil du temps.

Parmi les vidéos de moins de deux mois les plus recommandées, il semble que l'algorithme ne recommande pas en priorité celles ayant fait l'objet d'un phénomène de *buzz* (un nombre de vues important en très peu de temps) au moment de la collecte des données.

Liste des vidéos de moins de deux mois les plus recommandées
(Nombre d'occurrences supérieur à 20 en ordre décroissant)

Titre[1]	Vues	Occurrences
Interdiction des portables au collège	757	35
Le portable déjà interdit au collège de Truchtersheim	1 215	34
On a visité le repaire d'un mineur de cryptomonnaies	297 997	33
Comment muscler et libérer votre cerveau ? - Idriss ABERKANE	91 184	31
Draguer une fille devant son copain -Part 2 (IbraTV)	3 402 187	28
Banlieues : la méthode Macron	133	26
Téléphones portables à l'école - On nous demande d'interdire une (...)	1 382	22
J'ai reçu ma machine de bitcoin ! feat. Hasheur	785 414	21
À Besançon, un collège a déjà interdit les portables dans son enceinte	698	20

Base : total des vidéos recommandées (39 007).
Lecture : la vidéo « Interdiction des portables au collège » a été vue 757 fois sur YouTube et recommandée 35 fois aux participants.

En outre, les vidéos totalisant un nombre de vues plus grand que le maximum de la liste ci-dessus sont pour près de la moitié d'origine internationale. À noter également, dans cette catégorie, une proportion importante de vidéos sur la Coupe du monde de football qui s'est tenue de mi-juin à mi- juillet 2018, au moment de la collecte des données. Elles proviennent de la même chaîne (FIFATV) et ont été recommandées à 3 participants, dont 2 ayant indiqué regarder respectivement des extraits de compétitions sportives ou du football[2]. Elles n'ont toutefois pas été recommandées aux profils fictifs orientés « football ».

Enfin, ayant fait le *buzz* à l'époque[3], la vidéo « *Childish Gambino - This Is America (Official Video)* » se distingue avec plus de 320 millions de vues (à l'échelle mondiale), deux mois après sa publication. Ce vidéoclip d'un musicien américain n'est toutefois recommandé qu'à une seule reprise dans la totalité des vidéos, à un participant ayant indiqué regarder régulièrement sur YouTube des vidéoclips de RnB et Pop provenant des États-Unis. La deuxième vidéo de moins de deux mois la

[1] Ici comme pour la suite du document, les titres des vidéos seront laissés dans la forme rédactionnelle adoptée par leurs auteurs, sans modification.
[2] Le troisième participant qui s'est vu recommandé ces vidéos n'a pas renseigné ses thématiques les plus consultées sur la plateforme.
[3] Voir par exemple : Europe 1. *La vidéo choc "This is America" de Childish Gambino fait le buzz sur Internet*. 08/05/2018.

plus vue totalise « seulement » 60 millions de vues (vidéoclip de RnB étatsuniens, recommandé à la même personne).

Vidéos les plus recommandées

Les vidéos les plus recommandées (nombre d'occurrences des recommandations supérieur à 40[1]) portent sur les thèmes sélectionnés pour les vidéos de départ. Cela signifie que sur un sujet donné, l'algorithme peut proposer des vidéos identiques à différents vidéonautes, qui ont donc accès à la même information. Elles apparaissent le plus souvent tout de suite après la vidéo de départ (1ère itération), plus rarement en deuxième position.

Liste des vidéos les plus recommandées par l'algorithme
(Nombre d'occurrences des recommandations supérieur à 40 en ordre décroissant)

Titre	Occ.	Ité.
Téléphones portables et très jeunes enfants - hi-tech	57	1 (26)
« Grand remplacement » : complot, fantasme ou réalité ?	57	1 (57)
L'homme a t-il réellement marché sur la Lune ?	56	1 (46)
Faut-il interdire l'usage du téléphone portable dans les établissements scolaires ?	53	1 (32)
Complément d'enquête. Les millionnaires du bitcoin - 12 octobre 2017 (France 2)	50	2 (26)
Comment sécuriser vos Bitcoins ?	45	1 (37)
SNCF : D'où vient (vraiment) la dette ?	44	1 (34)
Galères de train : qui veut tuer le rail français ?	43	2 (33)
Mickael Jackson en vie (retour le 07/07/2016)	42	2 (35)
Peut-on donner la parole à un théoricien du « Grand remplacement » sur la radio publique ?	42	1 (25)
Martin Weill rencontre le rédacteur en chef de Valeurs Actuelles - Quotidien du 28 Avril	41	1 (30)
Le 3 minutes - Laïcité à l'école - #flashtalk	40	1 (34)
Marion Sigaut - Le programme LGBT de Macron	40	2 (18)
Mort de Michael Jackson, les rumeurs les plus folles	40	2 (10)

Base : total des vidéos recommandées (39 007).
Légende : Occ. = occurrences, Ité. = itération.
Lecture : la vidéo « Téléphones portables et très jeunes enfants - hi-tech » a été recommandée à 57 reprises, dont 26 fois juste après la vidéo de départ (1ère itération).
Note : la vidéo "« Grand remplacement » : complot, fantasme ou réalité ?" est également une vidéo de départ (non comptabilisée).

[1] Un nombre d'occurrences des recommandations supérieur à 40 équivaut à 0,10 % de la totalité des vidéos et a été choisie pour ne pas tomber en dessous d'un zéro après la virgule.

Les vidéos les plus recommandées comptabilisent un nombre de vues relativement limité. Seules deux d'entre elles dépassent les 400 000 vues, certainement du fait de la popularité de leur sujet : Michael Jackson. Une des vidéos a même été vue moins de 1 000 fois (sur le thème du téléphone à l'école). De la même manière, les vidéos ont reçu plus de « j'aime » que de « je n'aime pas », comme c'est souvent le cas sur la plateforme.

Les « j'aime » et « je n'aime pas » semblent ne pas avoir d'effet sur la sélection des vidéos que l'algorithme recommande le plus. Ceci peut être la conséquence d'une importance moindre accordée par l'algorithme à l'engagement (« j'aime » / « je n'aime pas ») aux premiers stades des recommandations, qui s'appuient *a priori* d'abord sur le thème de la vidéo de départ et les *tags* associés (voir page 56). Il convient toutefois de s'interroger sur le poids accordé aux « j'aime » et « je n'aime pas » dans les recommandations suivantes.

Nombre de vues, « j'aime » (👍) et « je n'aime pas » (👎) des vidéos les plus recommandées

(Ordre similaire au tableau précédent)

Titre	Nb de vues	👍	👎	Taux de 👎 sur réactions
Téléphones portables et très jeunes enfants - hi-tech	**4 486**	25	2	7,4 %
« Grand remplacement » : complot, fantasme ou réalité ?	**22 589**	583	40	6,4 %
L'homme a t-il réellement marché sur la Lune ?	**124 202**	485	56	10,4 %
Faut-il interdire l'usage du téléphone portable dans les (...)	**741**	5	0	0 %
Complément d'enquête. Les millionnaires du bitcoin (...)	**203 965**	1 209	113	8,5 %
Comment sécuriser vos Bitcoins ?	**206 314**	4 286	123	2,8 %
SNCF : D'où vient (vraiment) la dette ?	**68 542**	4 728	89	1,8 %
Galères de train : qui veut tuer le rail français ?	**22 539**	1 428	19	1,3 %
Mickael Jackson en vie (retour le 07/07/2016)	**577 937**	2 582	508	16,4 %
Peut-on donner la parole à un théoricien du « grand (...)	**282 668**	2 162	1 576	42,2 %
Martin Weill rencontre le rédacteur en chef de Valeurs (...)	**294 255**	2 154	428	16,6 %
Le 3 minutes - Laïcité à l'école - #flashtalk	**1 581**	4	1	20 %
Marion Sigaut - Le programme LGBT de Macron	**71 506**	867	739	46 %
Mort de Michael Jackson, les rumeurs les plus folles	**419 847**	5 532	397	6,7 %

Lecture : la vidéo « L'homme a t-il réellement marché sur la Lune ? » totalise 124 202 vues, 485 « j'aime » et 56 « je n'aime pas », ces derniers représentent 6,4 % de l'ensemble des réactions (« j'aime » + « je n'aime pas »).

Le nombre d'abonnés aux chaînes ayant publié les vidéos les plus recommandées va de 200 à plus de 700 000 et ne permet pas de dégager des caractéristiques communes. Le nombre d'abonnés à une chaîne n'est pas toujours déterminant pour le nombre de vues des vidéos publiées par cette chaîne : la vidéo totalisant le plus de vues (voir tableau ci-dessus) provient

d'une chaîne avec moins de 2 000 abonnés. De même, la chaîne avec le moins d'abonnés totalise plus de 120 000 vues sur sa vidéo.

Nombre d'abonnés des chaînes ayant publié les vidéos les plus recommandées
(Ordre similaire aux tableaux précédents)

Titre	Chaîne	Nb d'abonnés
Téléphones portables et très jeunes enfants - hi-tech	euronews (en français)	345 967
« Grand remplacement » : complot, fantasme ou réalité ?	Radio Sputnik	12 365
L'homme a t-il réellement marché sur la Lune ?	MrKershouw	199
Faut-il interdire l'usage du téléphone portable dans les (...)	Kaina.tv	794
Complément d'enquête. Les millionnaires du bitcoin (...)	Complément d'enquête	33 461
Comment sécuriser vos Bitcoins ?	STEVEN	701 930
SNCF : D'où vient (vraiment) la dette ?	Le Fil d'Actu - Officiel	73 808
Galères de train : qui veut tuer le rail français ?	Le Fil d'Actu - Officiel	73 808
Mickael Jackson en vie (retour le 07/07/2016)	Stv event saji tv	1 728
Peut-on donner la parole à un théoricien du « grand (...)	Quotidien	324 872
Martin Weill rencontre le rédacteur en chef de Valeurs (...)	Quotidien	324 872
Le 3 minutes - Laïcité à l'école - #flashtalk	#flashtalk	2 551
Marion Sigaut - Le programme LGBT de Macron	ER Lille	2 494
Mort de Michael Jackson, les rumeurs les plus folles	Investigations et Enquêtes	332 403

Base : total des vidéos recommandées (39 007).

En conclusion de cette analyse relatives aux vidéos récurrentes, l'algorithme semble d'abord se focaliser sur le thème de départ dans les premières itérations, indépendamment du nombre de vues, du nombre de réactions (« j'aime » / « je n'aime pas ») ou du nombre d'abonnés aux chaînes qui les publient.

Chaînes les plus recommandées

La chaîne la plus recommandée est celle de l'émission de France 2 « On n'est pas couché », qui représente 2 % de l'ensemble des vidéos recommandées (soit 798 vidéos). D'une manière générale, les éditeurs de télévision comptent pour près de la moitié des chaînes les plus recommandées listées ci-dessous (8 sur 17). En outre, toutes les chaînes présentes dans cette liste portent le badge de validation YouTube[1]. Seules deux de ces chaînes comptent moins de

[1] Lorsque l'un des marqueurs de validation (ou) s'affiche à côté du nom d'une chaîne YouTube, cela signifie que la chaîne appartient à un créateur reconnu ou qu'il s'agit de la chaîne officielle d'une marque,

100 000 abonnés (le média Brut et le dessin animé jeunesse « Il était une fois »). La chaîne TEDx Talks comptabilise le plus grand nombre d'abonnés, du fait très vraisemblablement de sa dimension internationale.

Liste des chaînes les plus recommandées par l'algorithme
(Nombre d'occurrences des recommandations supérieur à 200 en ordre décroissant)

Chaînes	Abonnés	Occurrences	Enchaînements successifs
On n'est pas couché	359 612	798	505
C'est mon choix - La chaîne officielle	826 162	635	581
Europe 1	195 345	513	345
Brut	91 267	498	374
Quotidien	324 872	419	306
C à vous	156 595	366	136
C'est pas sorcier	611 123	343	296
Lama Faché	3 694 933	321	173
L'Effet Papillon	247 691	314	256
TEDx Talks	13 209 307	261	220
Ça commence aujourd'hui	101 873	259	190
iletaitunefoisVEVO	43 740	238	205
madmoiZelle	422 306	236	201
1 jour, 1 question	110 844	229	188
SYMPA	1 471 831	225	117
France Inter	223 518	210	149
Le Monde	224 203	205	124

Base : total des vidéos recommandées (39 007).
Lecture : parmi les 798 vidéos recommandées provenant de la chaîne « On n'est pas couché », 505 ont été proposées toute de suite après une vidéo provenant de cette même chaîne.

Ces nombres élevés d'occurrences renvoient à la « puissance des chaînes » sur YouTube. Le parcours qui mène aux chaînes puissantes est toujours le même. Sur un sujet donné, l'algorithme enchaîne les vidéos de différentes chaînes jusqu'à aboutir à une chaîne « puissante » pour y rester pendant quelques vidéos, voire ne plus la quitter, quel que soit le sujet traité. La puissance de la chaîne se manifeste lorsque l'algorithme recommande une chaîne pour la première fois et lorsqu'ensuite il continue de la recommander. Elle prend alors le dessus sur le sujet de départ.

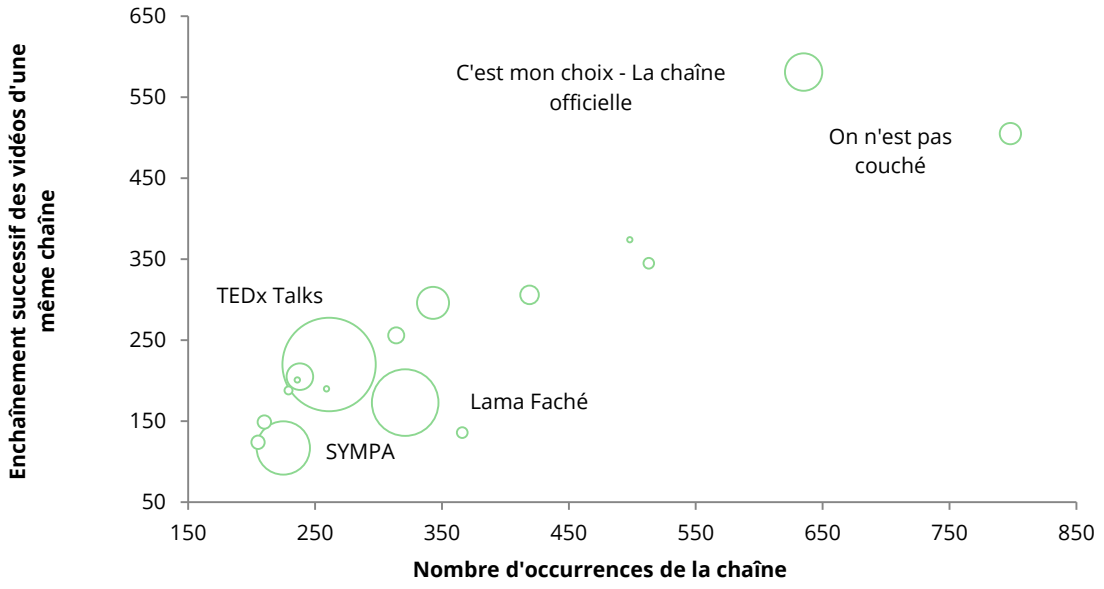
d'une entreprise ou d'une organisation.
Source : https://support.google.com/youtube/answer/3046484?hl=fr.

Sur l'ensemble des vidéos recommandées, 41 % des vidéos lues proviennent de la même chaîne que celle de la vidéo lue à l'itération directement précédente. Cette moyenne est tirée vers le haut par les vidéos les plus recommandées, pour lesquelles ce chiffre atteint 72 %. En moyenne, les participants arrivent sur une chaîne « puissante » à partir de la sixième itération puis y restent jusqu'à la dixième. Ils voient donc au total cinq vidéos successives provenant de la même chaîne. Bien que les thèmes abordés par une même chaîne varient, l'approche éditoriale (orientation des idées, traitement des sujets, images utilisés, etc.) est généralement constante. Un phénomène de redondance, consistant à se voir recommander des vidéos qui suivent les mêmes standards éditoriaux pendant une longue période de temps, pourrait donc apparaître ici. Ce phénomène, par certains aspects, pourrait se rapprocher de celui dit de chambre d'écho.

La puissance d'une chaîne peut provenir de plusieurs facteurs. Son animation (fréquence de publication, homogénéité dans le format des titres, etc.) joue potentiellement un rôle important. Le pourcentage d'utilisateurs de YouTube qui, après avoir regardé une vidéo, cliquent volontairement sur une vidéo recommandée provenant de la même chaîne, signe de satisfaction de l'utilisateur, peut également avoir un impact.

Le graphique ci-dessous rapproche le nombre d'abonnés aux chaînes (taille des points) et le nombre de vidéos qui se succèdent provenant de chaque chaîne. Il montre que le nombre d'abonnés à une chaîne n'est pas forcément déterminant dans la « puissance » de cette chaîne. Si le nombre d'abonnés à une chaîne était l'unique critère de mise en avant de celle-ci, les chaînes « TEDx Talk », « Lama Faché » et « SYMPA » auraient occupé l'espace supérieur droit du graphique, et non « C'est mon choix – La chaîne officielle » et « On n'est pas couché ».

Répartition des chaînes les plus recommandées en fonction de leur nombre d'occurrences et des enchaînements successifs de leurs vidéos



Note : la taille des points représente le nombre d'abonnés à la chaîne.

L'algorithme dévie d'une logique de recommandation par thème vers une logique par chaîne. Il arrive parfois qu'il suive à nouveau une logique par thème ou même combine un même sujet et une même chaîne. Dans tous les cas, une clé de jointure apparaît entre chaque vidéo : au-delà de la chaîne, l'algorithme semble recommander la vidéo suivante en sélectionnant des mots-clés présents dans les *tags* ou dans le titre de la vidéo actuellement lue (cf. *infra*).

Exemple de chemin conduisant à une des chaînes les plus recommandées et y maintenant le participant

Ité.	Titre	Chaîne
0	Israël–Palestine : comprendre le conflit par les cartes	Le Figaro
1	Israël / Palestine – Le conflit expliqué sur carte – Partie 2	Histoire Géo
2	Israël / Palestine – L'Histoire du peuple juif – Partie 1	Histoire Géo
3	Pourquoi Israéliens et Palestiniens se font-ils la guerre ? Le Professeur Gamberge	Les Z'animés
4	C'est où la bande de Gaza ? – 1 jour, 1 question	1 jour, 1 question
5	C'est où Jérusalem ? – 1 jour, 1 question	1 jour, 1 question
6	À quoi ça sert la reine d'Angleterre ? – 1 jour, 1 question	1 jour, 1 question
7	À quoi ça sert la tour Eiffel ? – 1 jour, 1 question	1 jour, 1 question
8	Qui a inventé le foot ? – 1 jour, 1 question	1 jour, 1 question
9	Qui était le premier homme sur terre ? – 1 jour, 1 question	1 jour, 1 question
10	Pourquoi les dinosaures ont disparu ? – 1 jour, 1 question	1 jour, 1 question

La puissance des chaînes dépend également du sujet abordé. Par exemple, les chaînes traitant exclusivement de véganisme seront recommandées de manière plus fréquente sur ce thème que sur un autre. C'est certainement pour cette raison que les chaînes listées dans le tableau page 36 sont les plus recommandées, du fait de la diversité des sujets qu'elles traitent.

Place des éditeurs « traditionnels »

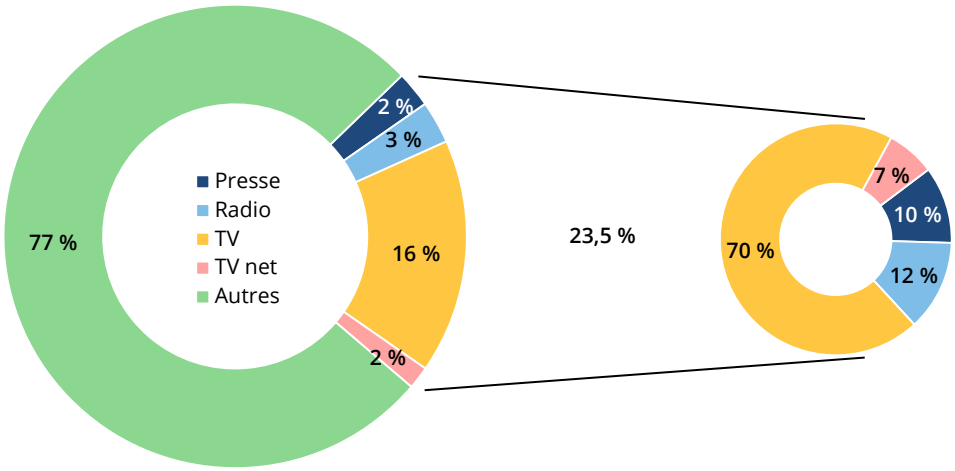
Plus de 23 % des vidéos recommandées sont issues d'éditeurs traditionnels de presse écrite, de radio et de télévision. Ce chiffre inclut les programmes hors antenne des groupes audiovisuels et ceux de leurs réseaux multi-chaînes1 (MCN) (catégorie intitulée « TV net » dans le graphique). Il s'agit par exemple de la chaîne Culturebox éditée par France Télévisions, de Golden Moustache du groupe M6 ou de Studio Bagel pour le groupe Canal Plus.

Parmi les éditeurs traditionnels, la catégorie des éditeurs de télévision est la plus représentée dans la mesure où ils proposent la même nature de contenus que les plateformes de partage de vidéos. Ils comptent pour 16 % de l'ensemble des vidéos recommandées (18 % en comptant les vidéos de leurs MCN). La radio arrive en deuxième position, loin derrière la télévision, avec seulement

[1] Un réseau multi-chaînes, ou MCN pour *multi-channel network* en anglais, désigne une société qui accompagne des créateurs de contenus sur les plateformes de partage de vidéos et les réseaux sociaux.

3 % des vidéos recommandées. La presse écrite imprimée représente 2,5 % des vidéos recommandées.

Répartition des vidéos recommandées par type d'éditeurs



Base : total des vidéos recommandées (39 007).

Ces résultats, en particulier pour la télévision, sont le fruit des stratégies numériques de nombreux groupes audiovisuels qui ont recours à la fois à leurs propres services en ligne (sites Internet et applications) et aux plateformes tierces telles que YouTube pour diffuser leurs contenus. Certains éditeurs choisissent de créer une chaîne YouTube portant le nom de leur groupe, de leur(s) chaîne(s) ou de leur(s) programme(s) phare(s). Toutes ces chaînes YouTube éponymes sont certifiées par la plateforme.

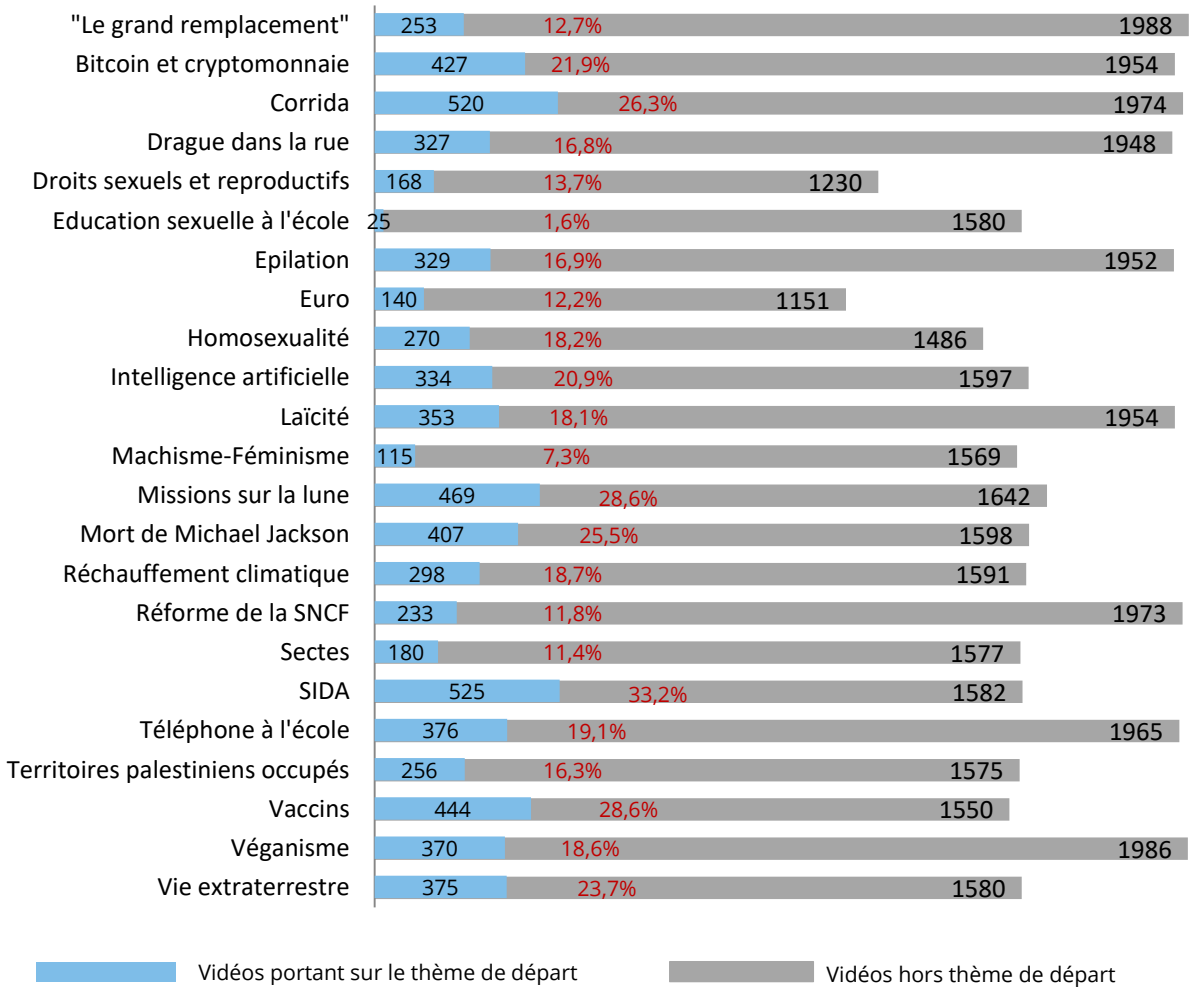
Offre de vidéos portant sur le thème de départ

Afin de mesurer la profondeur de l'offre associée aux 23 thèmes sélectionnés, un travail de qualification des vidéos a été effectué « à la main ». À partir de chaque vidéo de départ portant sur un thème donné, les vidéos recommandées ont été qualifiées selon la nomenclature associée à ce thème (voir page 24). Les nombres de vidéos portant l'opinion A, l'opinion B et une approche neutre ont ensuite été additionnés pour obtenir le nombre total de vidéos apportant une contribution au débat identifié sur le thème de départ.

Profondeur de traitement des thèmes sélectionnés

Plus de 18 % de l'ensemble des vidéos recommandées par la lecture automatique concourent au débat associé au thème de départ. Même si la médiane est similaire à la moyenne, ce chiffre diffère dans des proportions importantes selon les thèmes (voir graphique ci-après). Le sujet traité le plus en profondeur est celui du SIDA : 33 % des vidéos recommandées à partir de vidéos de départ traitant du SIDA portent sur ce thème. À l'opposé, le thème de l'éducation sexuelle à l'école n'est que très peu discuté : moins de 2 % des vidéos recommandées le traite.

Profondeur de traitement des thèmes sélectionnés



Lecture : à partir des vidéos de départ portant sur le thème de la corrida, 1 974 vidéos ont été recommandées, dont 26,3 % portent effectivement sur ce thème.

Le nombre de vidéos recommandées sur chaque thème peut être en partie lié au nombre total de vidéos portant sur ce thème disponibles sur YouTube. Certains sujets se prêtent plus aisément à un traitement vidéo. L'actualité passée peut aussi avoir favorisé la production d'un volume considérable de vidéos sur un sujet et constituer ainsi un stock important de contenus mobilisable lorsqu'un utilisateur s'y intéresse à nouveau (c'est le cas par exemple pour le thème de l'euro, dont la proportion de vidéos associées de plus de 5 ans est importante).

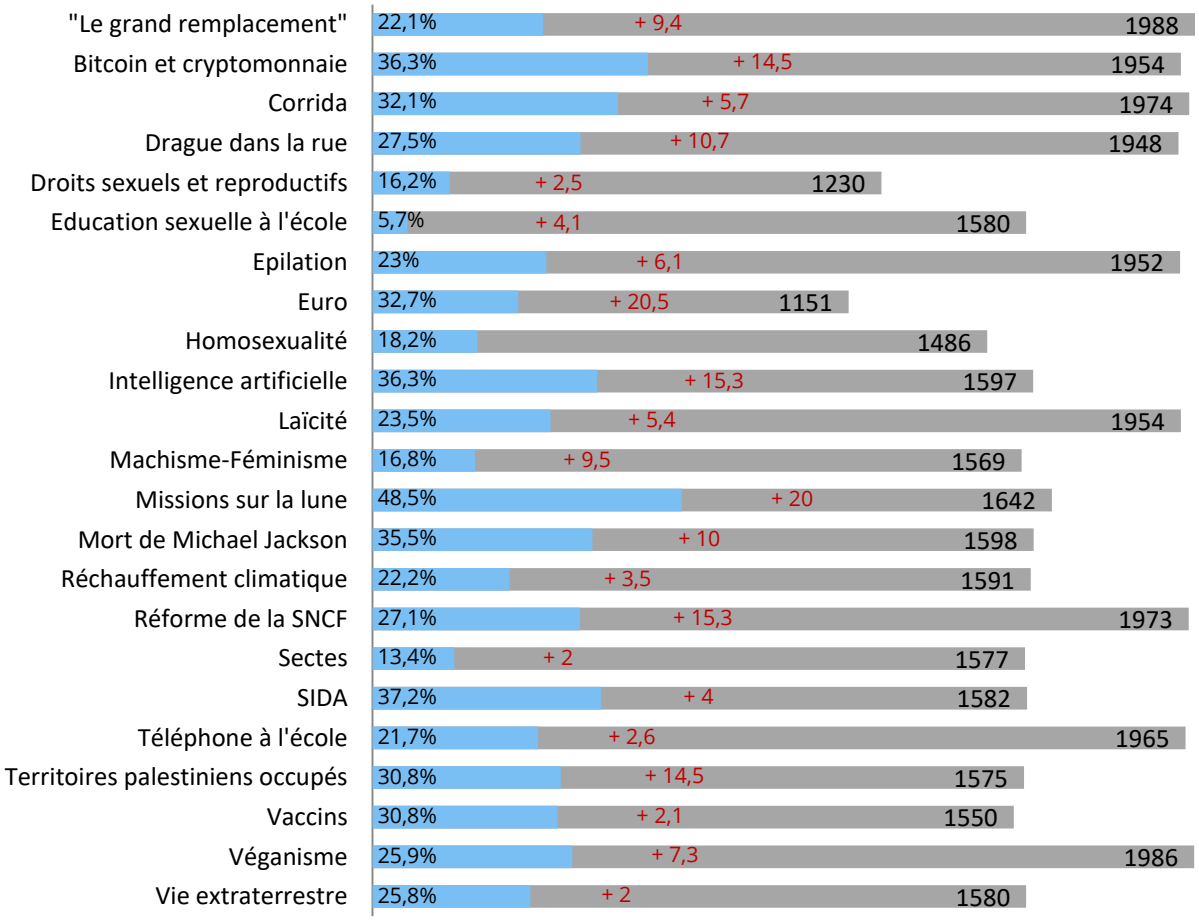
En outre, le dynamisme de la communauté autour des différents thèmes joue un rôle essentiel[1]. Certains sujets sont débattus par des communautés plus dynamiques et plus communicantes que d'autres, plus enclines à utiliser massivement les outils numériques à leur disposition afin de

[1] L'importance de la communauté dans la représentation des opinions sur YouTube a récemment été mise en évidence par un article de recherche de l'université de Penn State. Voir : Kevin Munger & Joseph Phillips. Penn State Political Science. *A supply and demand framework for YouTube politics*. 01/10/2019.

toucher un public plus large. Certaines communautés, moins représentées par les médias traditionnels, utilisent le média Internet comme caisse de résonnance dans la population. Ceci explique que les sujets poussés par une communauté active sont davantage représentés dans les vidéos recommandées. Cette dynamique communautaire peut s'exprimer par la fréquence de mise en ligne de vidéos, la densité des commentaires ou par l'intensité des réactions du type « j'aime » / « je n'aime pas », etc.

Pour avoir une vision plus globale de ce que l'algorithme peut proposer autour d'un thème, une quatrième qualification a été ajoutée. Elle prend en compte les vidéos portant sur le sujet mais ne relevant pas du débat caractérisé initialement. De plus, elle élargit le thème de départ à des sujets connexes (par exemple : la *blockchain* pour les vidéos sur les cryptomonnaies). Dans cette configuration élargie, 27 % des vidéos recommandées traitent de manière plus ou moins pointue le thème de départ (la médiane est au même niveau).

Profondeur de traitement des thèmes élargis



Même si la part de vidéos recommandées portant sur le thème de départ paraît faible, leur nombre par thème est plutôt élevé (449 vidéos en moyenne). Ceci est notamment dû au fait que ces statistiques comptent les récurrences (les vidéos ayant été recommandées à plusieurs participants). Parmi les vidéos les plus recommandées sur chaque thème (nombre d'occurrences

supérieur à 10), près de 60 % portent sur le thème de départ dans son univers élargi. Ce chiffre monte à près de 93 % sur le thème de l'euro et tombe à 11 % sur le thème de l'éducation sexuelle à l'école.

Vidéos uniques et vidéos récurrentes sur les thèmes sélectionnés

Le nombre de vidéos uniques portant sur les thèmes sélectionnés s'élève en moyenne à 129 vidéos par thème, soit 29 % des vidéos thématiques. Il semble qu'au fil des recommandations et des sessions de visionnage sur un thème, l'algorithme atteigne un plafond au-delà duquel il cesse de recommander des vidéos sur ce thème.

Nombre de vidéos uniques et de vidéos récurrentes pour chaque thème
(Ordre alphabétique)

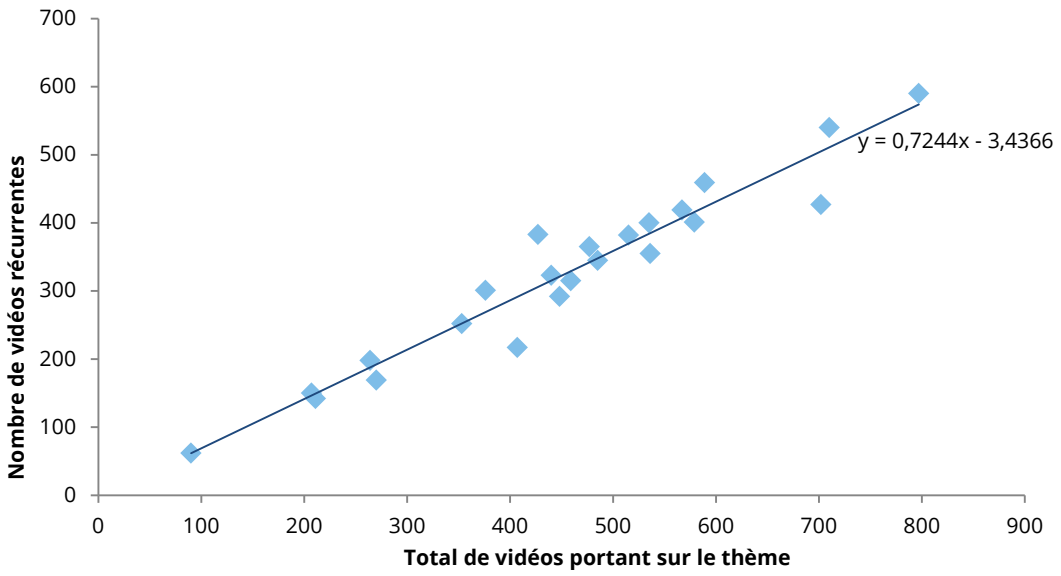
Thèmes	Vidéos uniques		Vidéos récurrentes	Total
« Le grand remplacement »	117	26,6 %	323	440
Bitcoin et cryptomonnaies	170	23,9 %	540	710
Corrida	275	39,2 %	427	702
Drague dans la rue	181	33,8 %	355	536
Droits sexuels et reproductifs	57	27,5 %	150	207
Éducation sexuelle à l'école	28	31,1 %	62	90
Épilation	156	34,8 %	292	448
Euro	75	19,9 %	301	376
Homosexualité	101	37,4 %	169	270
Intelligence artificielle	178	30,7 %	401	579
Laïcité	144	31,4 %	315	459
Machisme-Féminisme	66	25,0 %	198	264
Missions sur la lune	207	26,0 %	590	797
Mort de Michael Jackson	148	26,1 %	419	567
Réchauffement climatique	101	28,6 %	252	353
Réforme de la SNCF	135	25,2 %	400	535
Sectes	69	32,7 %	142	211
SIDA	130	22,1 %	459	589
Téléphone à l'école	44	10,3 %	383	427
Territoires palestiniens occupés	140	28,9 %	345	485
Vaccins	112	23,5 %	365	477
Véganisme	133	25,8 %	382	515
Vie extraterrestre	190	46,7 %	217	407

Ce tableau est représenté par les deux graphiques ci-après. Le premier montre que, pour un thème donné, le nombre de vidéos récurrentes augmente avec le nombre total de vidéos portant

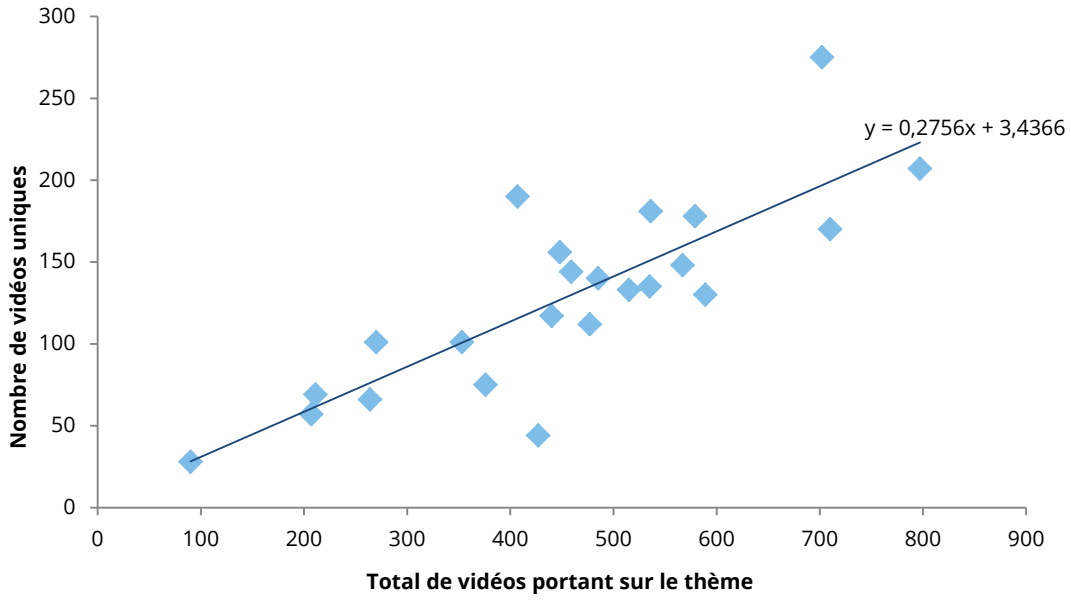
sur ce thème. Le second graphique montre que le nombre de vidéos uniques sur un thème donné progresse plus lentement que le nombre de vidéos récurrentes. Autrement dit, plus le nombre de vidéos sur un thème donné augmente, plus la part de vidéos recommandées plusieurs fois est importante, et plus la part de vidéos recommandées une seule fois baisse.

Sur un thème, l'algorithme tend à proposer aux différents utilisateurs un même corpus restreint de vidéos qui peut venir nourrir un phénomène de redondance dès lors que les utilisateurs sont davantage soumis aux mêmes vidéos qu'à des vidéos diversifiées. Les participants à l'expérience ont ainsi pour la plupart vu des mêmes vidéos. Ce phénomène, assimilable pour partie à celui dit de chambre d'écho, apparaîtrait dans les premières recommandations qui ont pour caractéristique de rester sur le thème de départ. Ce phénomène à ici pour spécificité de se manifester à l'échelle du groupe de participants plus qu'à l'échelle de chaque participant.

Répartition des thèmes en fonction du nombre de vidéos récurrentes



Répartition des thèmes en fonction du nombre de vidéos uniques



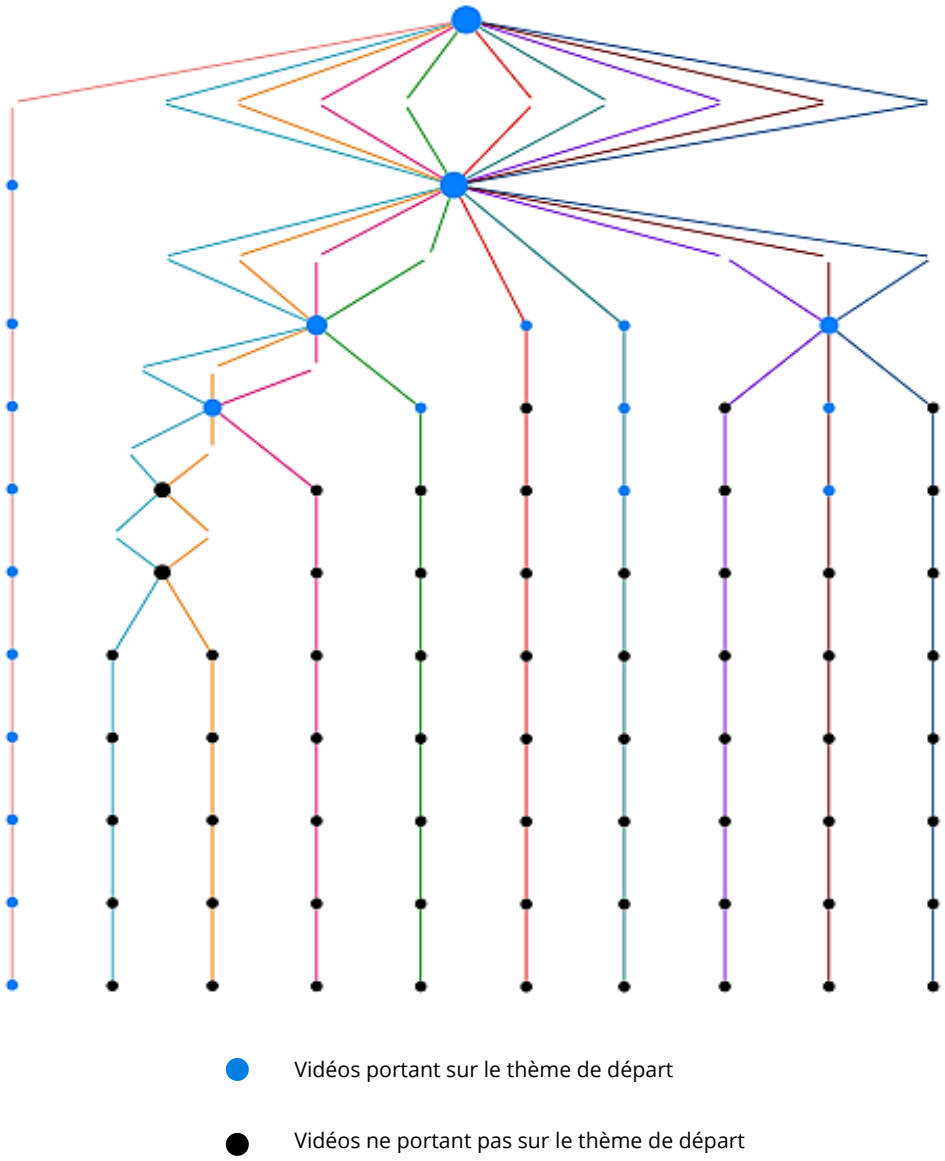
Le rang de l'itération à partir duquel l'algorithme dévie totalement du thème de départ a été calculé pour mesurer le volume d'information rendu accessible aux participants sur ce thème. Le thème de départ (élargi) est abandonné assez vite, en moyenne juste après la troisième recommandation. C'est le thème de l'éducation sexuelle à l'école qui montre la dérive la plus rapide par l'algorithme (en moyenne au bout de 1,5 itération) ; à l'opposé l'algorithme reste plus longtemps sur le thème des missions lunaires (déviation en moyenne à la cinquième itération).

Dans l'arbre de recommandation ci-dessous est représenté le chemin unique de 10 participants, échantillonnés pour l'exemple. Les points représentent les 10 vidéos recommandées par la lecture automatique après le lancement de la vidéo de départ « *BITCOIN : Comment ça marche et où en acheter ?* 🚀 ». La taille des points varie en fonction du nombre de participants ayant regardé la vidéo. Ces points sont bleus lorsque la vidéo porte sur le thème de départ. Pour certains participants, comme celui en violet ou en bleu foncé (le plus à droite), l'algorithme a dévié à la deuxième itération. À l'inverse, le participant le plus à gauche n'a vu que des vidéos sur le thème des cryptomonnaies[1].

[1] Ce participant n'a pas indiqué être intéressé par le sujet des cryptomonnaies, aussi bien dans la question sur les thématiques préférées sur YouTube que dans celle sur les précédentes recherches sur la plateforme.

Exemple n° 1 d'arbre de recommandations

BITCOIN : Comment ça marche et où en acheter ?



Source : CSA.

Des croisements entre certains thèmes de la liste définie initialement sont apparus à plusieurs reprises dans les recommandations. Ainsi, par exemple, sur le thème des missions lunaires, 21 % des vidéos portent sur l'existence d'une vie extraterrestre, sujet assez proche. D'autres rapprochements sont plus difficiles à expliquer. Une hypothèse pourrait être liée à l'historique de consommation des participants au cours de la collecte des données : au fur et à mesure du lancement des vidéos de départ, l'algorithme retient que des vidéos sur tel ou tel sujet ont été vues, et en recommande donc d'autres par la suite.

Autres thèmes recommandés

(Thème le plus recommandé parmi les autres thèmes identifiés)

Thèmes	%	Autre thème récurrent	%
Droits sexuels et reproductifs	55,7 %	Homosexualité	24,2 %
Éducation sexuelle à l'école	55,6 %	SIDA	17,3 %
Missions sur la lune	78,8 %	Vie extraterrestre	20,6 %
Sectes	85,1 %	Véganisme	8,5 %
Vaccins	91 %	SIDA	6,5 %

Lecture : à partir d'une vidéo sur le thème des droits sexuels et reproductifs, près de 56 % des vidéos recommandées traitant d'un des 23 thèmes sélectionnés pour l'étude portent bien sur ce thème de départ des droits sexuels et reproductifs ; parmi les 44 % restant, 24 % portent sur le thème de l'homosexualité.

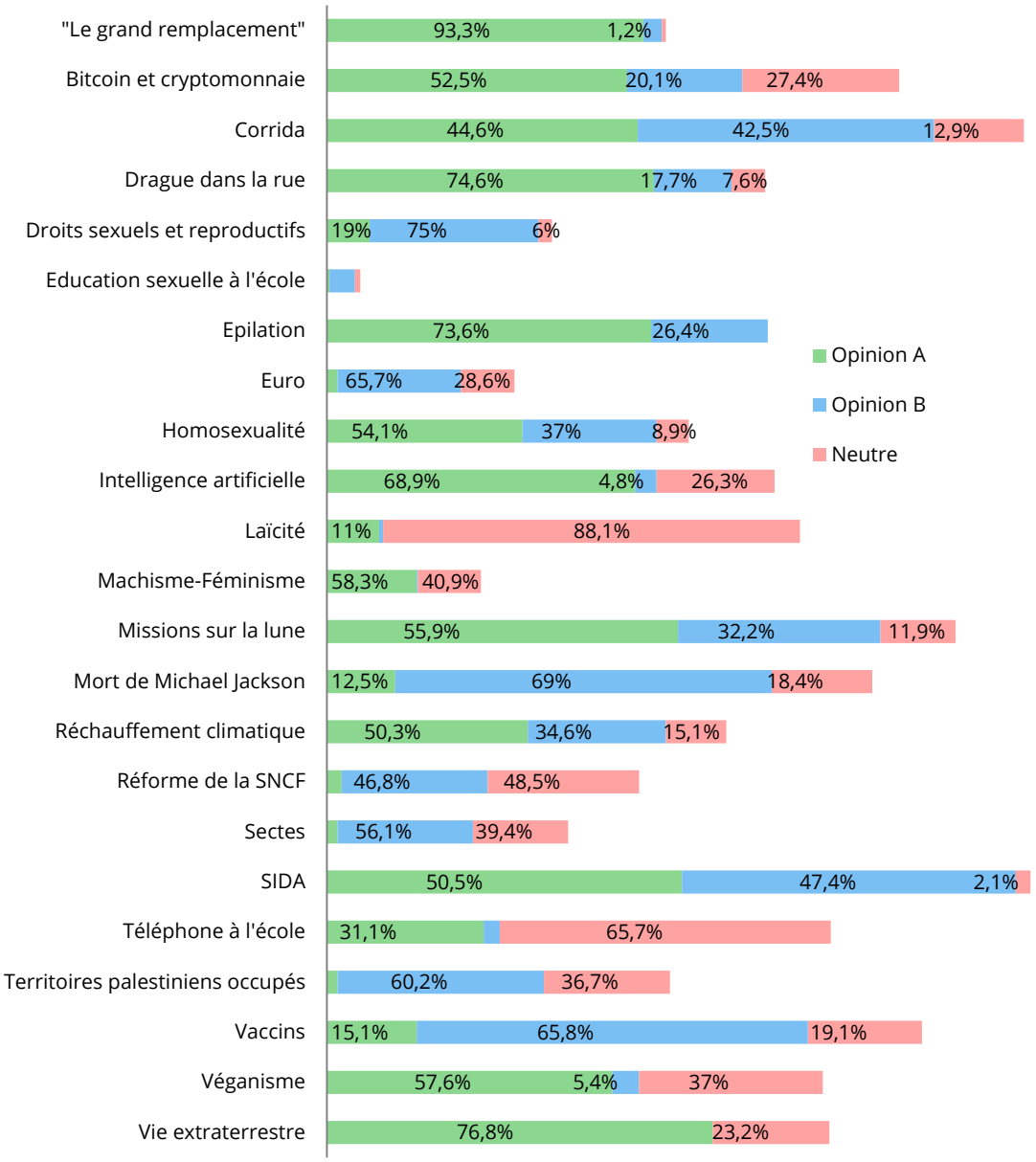
Diversité des opinions exprimées dans les vidéos recommandées

Il est important de noter à titre préliminaire que la qualification des opinions portées dans les vidéos recommandées peut présenter une part de subjectivité, que l'équipe en charge du projet a toutefois tenté de réduire au maximum au cours de discussions et de débats jusqu'à l'obtention d'un consensus[1].

Les résultats du tableau ci-dessous montrent à titre d'exemple que sur le thème de la laïcité, 88 % des vidéos recommandées concourent au débat sur ce sujet de manière neutre, en expliquant les tenants et aboutissants de ce principe constitutionnel. Du côté des opinions exprimées, la définition libérale de la laïcité est prônée dans 11 % des vidéos recommandées qui traitent ce sujet. Très peu de vidéos sont en faveur de la définition stricte de la laïcité (moins de 1 %). En outre, la représentation des opinions dans les vidéos recommandées peut varier en fonction de l'opinion portée dans la vidéo de départ.

[1] Après une première observation des vidéos recommandées, de leur type et de leur sujet, l'équipe projet s'est réunie afin d'arrêter une qualification pour chaque catégorie de vidéo.

Diversité des opinions dans les vidéos recommandées portant sur le thème de départ
(En nombre de vidéos et en pourcentage par thème, ordre alphabétique des thèmes)

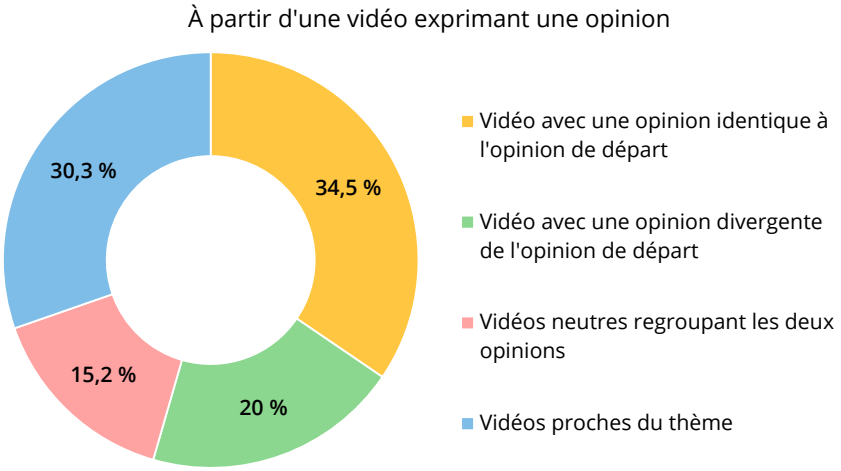


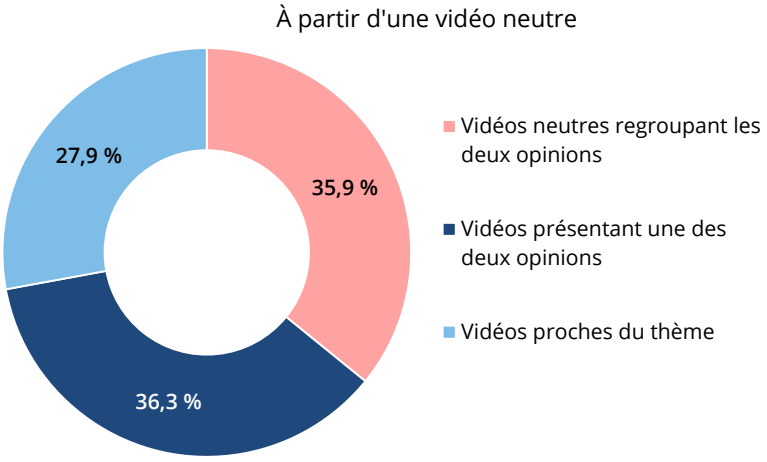
Note : compte tenu du nombre de vidéos très bas sur le thème de l'éducation sexuelle à l'école, ces résultats sont à interpréter avec précaution.

Comme cela a été évoqué plus haut, la communauté autour d'un sujet joue un rôle essentiel dans la place occupée par ce sujet sur la plateforme. Les opinions en faveur de certains courants de pensée sont portées par des communautés dynamiques et communicantes. Ces communautés (végane, théorie du grand remplacement, anti-vaccination, etc.), qui trouvent peut-être un écho moindre dans les médias traditionnels, utilisent YouTube comme média pour partager et diffuser leurs idées à un large public.

Certaines opinions sont donc surreprésentées dans les vidéos recommandées. Il est alors probable qu'à partir de n'importe quelle opinion sur un thème donné, l'algorithme emmène l'utilisateur vers des vidéos portant sur des opinions plus largement exposées sur la plateforme. Dans les faits, plus d'un tiers des vidéos recommandées restent sur l'opinion exprimée dans la vidéo de départ et 20 % des vidéos recommandées proposent une opinion divergente. À partir d'une vidéo neutre, la part des vidéos également neutres et celle des vidéos exprimant une des deux opinions sont identiques. Les vidéos recommandées qui se rapprochent du thème, mais ne le traitent pas explicitement (dites « proches »), représentent plus de 30 % des vidéos recommandées. La part importante de ces vidéos est le fait des mots-clés qui lient les vidéos entre elles selon leur type ou leur sujet ; ces mots-clés peuvent être précis mais aussi plus généraux.

Opinions dans les vidéos recommandées par rapport à l'opinion de la vidéo de départ





En outre, 60 % de l'ensemble des vidéos exprimant une opinion (opinions A et B, tous thèmes de départ confondus) sont recommandées dans les itérations suivant une vidéo de départ portant sur cette même opinion. Ainsi, l'opinion initiale reste dans l'ensemble majoritaire dans les vidéos

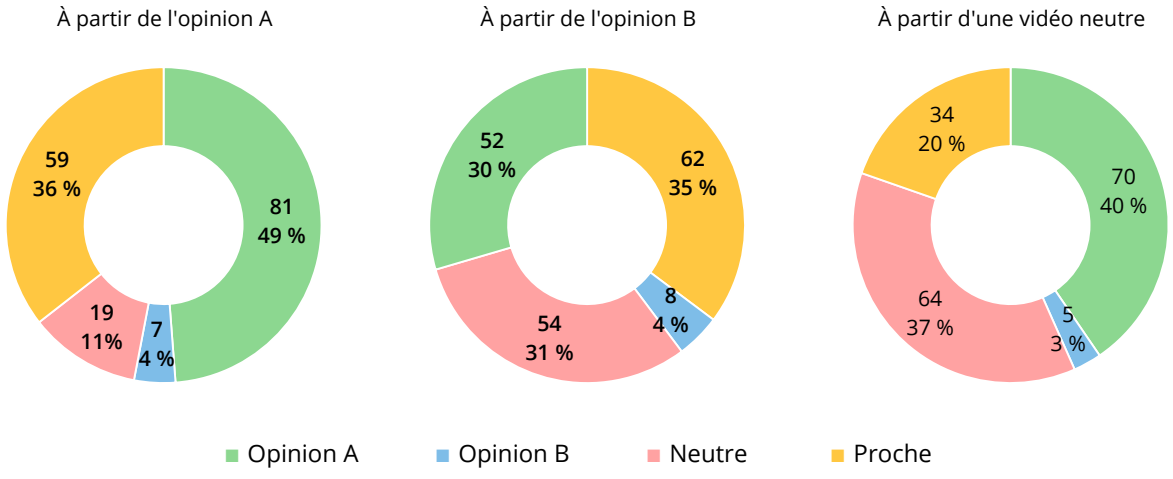
recommandées, au détriment de la diversité des opinions proposées. Cette situation pourrait déboucher sur un phénomène de chambre d'écho.

Par ailleurs, la répartition des opinions diffère selon les thèmes en fonction du nombre de vidéos disponibles par opinion. Ainsi, sur le thème du véganisme pour lequel une des deux opinions est surreprésentée, l'algorithme propose systématiquement un nombre plus important de vidéos en faveur du véganisme (opinion A). À noter également que dans ce thème, les vidéos neutres sont proposées dans des proportions bien plus importantes que la moyenne à partir des vidéos contre le véganisme (18 %).

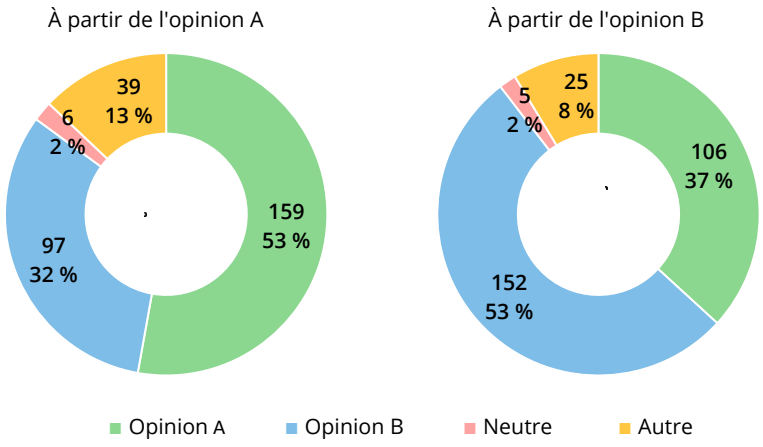
Un autre exemple, celui du SIDA, témoigne d'une certaine parité entre le nombre de vidéos dénonçant l'origine artificielle du virus (opinion B) et le nombre de vidéos présentant l'origine du virus comme naturelle (opinion A). Tout comme pour l'ensemble des thèmes, 53 % de ces vidéos sont recommandées suite à une vidéo de départ véhiculant la même opinion. Ce phénomène peut provenir des *tags* qui peuvent porter la marque de l'opinion de la vidéo et ainsi lier prioritairement les vidéos exprimant le même point de vue.

Opinions dans les vidéos recommandées
en fonction de l'opinion de la vidéo de départ

Exemple du véganisme



Exemple du SIDA



Note : il n'y a pas de vidéo de départ « neutre » sur le thème du SIDA.

Enfin, 44 % des recommandations expriment la même opinion que la vidéo directement précédente (vidéo de départ ou non). Dans le détail, ce chiffre est de plus de 52 % pour les vidéos portant une opinion et de 26 % pour les vidéos neutres. Pour les vidéos exprimant une opinion, les participants avaient donc une chance sur deux de rester sur cette même opinion.

Cependant, la faible proportion de vidéos neutres qui se suivent (par rapport à celle des vidéos portant une opinion) pourrait suggérer que ces vidéos prises individuellement sont intercalées entre des vidéos exprimant d'autres opinions, permettant de balancer les débats. Par ailleurs, les recommandations de l'algorithme ne portent finalement que très peu le thème de départ (déviation en moyenne au bout de la troisième itération).

La suite de l'étude vise à apporter des éléments sur les parcours de consommation à partir du moment où les recommandations dévient du thème de départ.

Influence des profils personnels sur les vidéos recommandées

Individualisation

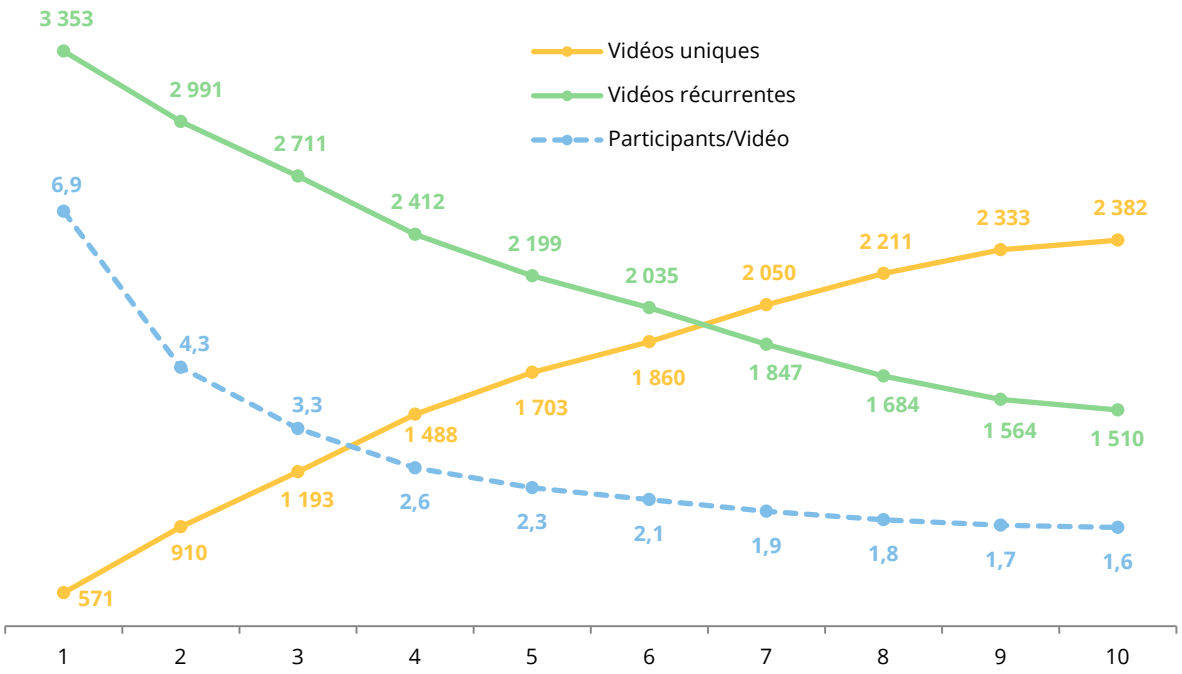
Comme indiqué précédemment, 16 701 vidéos uniques ont été comptabilisées sur un total de 39 007 vidéos vues. Le nombre important de répétitions indique que les participants se voient recommander les mêmes vidéos, indépendamment de leur profil. En moyenne, chaque vidéo est recommandée à 2 ou 3 participants (2,3 en moyenne). Cependant, certaines vidéos ont été recommandées à de nombreuses reprises, jusqu'à 57 fois, ce qui signifie que certains participants ont vu plusieurs fois la même vidéo dans la mesure où l'expérience rassemble moins de 57 participants.

Au fil des itérations, le nombre de vidéos uniques augmente progressivement ; inversement, le nombre de vidéos récurrentes diminue. En moyenne, juste après la vidéo de départ (première

itération), près de 7 participants regardent une même vidéo. À l'opposé, en bout de chemin (itération n° 10), entre 1 et 2 participant(s) se voient proposer une même vidéo. Ces éléments laissent entrevoir une certaine individualisation des recommandations : chaque participant regarde une vidéo distincte de celles recommandées aux autres participants. Les indications données par YouTube précisent que l'algorithme étant construit de sorte à recommander des vidéos à chaque utilisateur pris individuellement, les recommandations s'individualisent de manière statistique.

Une asymptote apparaît autour de la valeur 1,6 alors que le nombre de vidéos uniques semble continuer de progresser, suggérant que les recommandations ne seront jamais totalement individuelles.

Vidéos récurrentes, vidéos uniques et nombre de participant(s) par vidéo à chaque itération

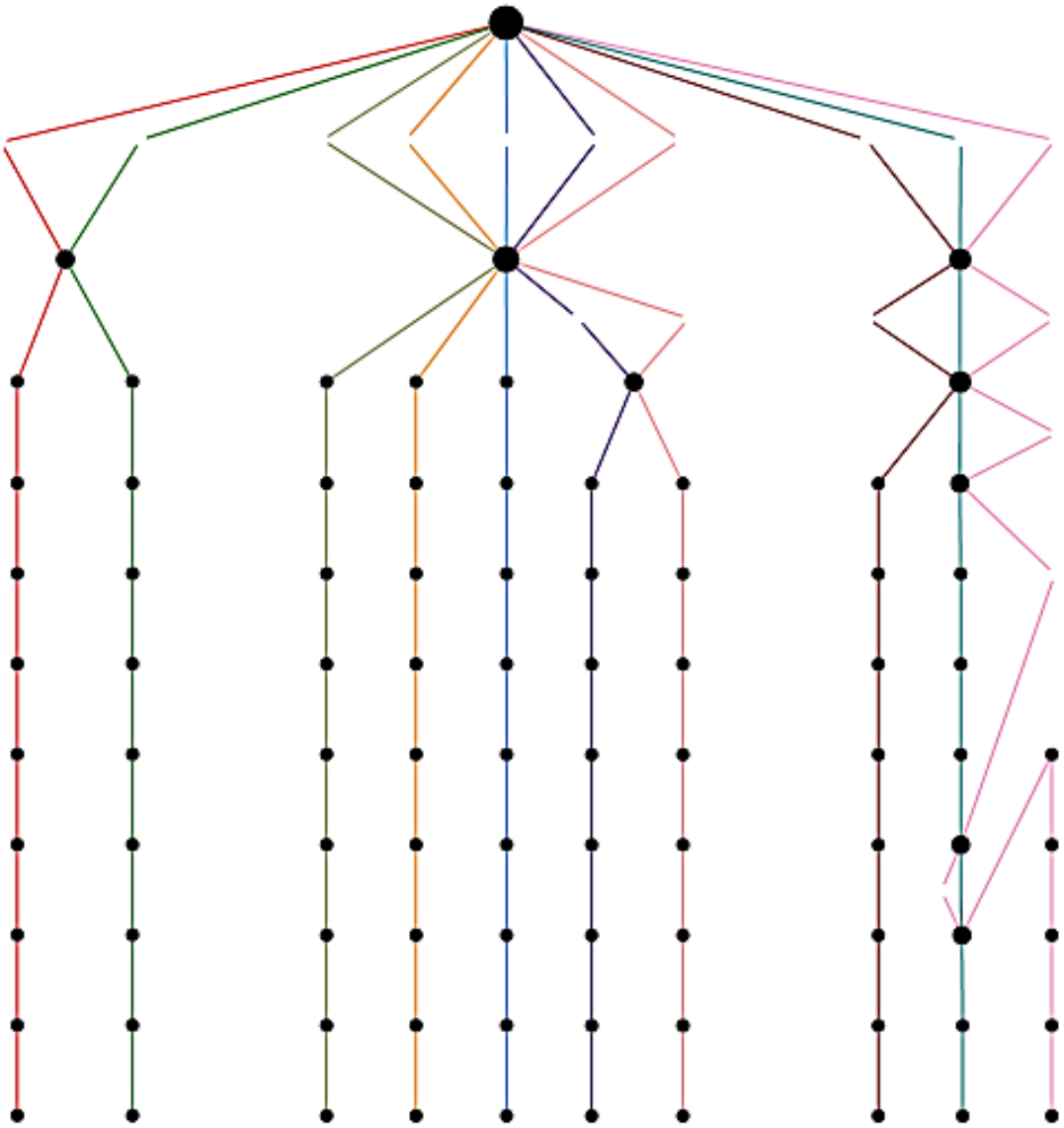


Base : ensemble des vidéos recommandées (39 007).
Note : une vidéo récurrente est comptabilisée une fois en tant que vidéo unique et (n-1) fois en tant que vidéos récurrentes. À titre d'exemple, une vidéo recommandée 10 fois sera comptabilisée comme une vidéo unique et 9 fois en tant que vidéo récurrente.

L'individualisation est particulièrement visible dans l'exemple ci-dessous. L'arbre représente le parcours de consommation de 10 participants sélectionnés au hasard, à partir de la vidéo de départ « *" 2 degrés avant la fin du monde " - #DATAGUEULE* » sur le thème du réchauffement climatique. Les points noirs représentent les vidéos recommandées ; leur taille varie en fonction du nombre de participants les ayant regardées. Il y a autant de vidéos à la racine de l'arbre que de participants sélectionnés pour l'exemple. À l'inverse, à la première itération, seules 3 vidéos ont été recommandées pour 10 participants.

Exemple n° 2 d'arbre de recommandations

« 2 degrés avant la fin du monde » - #DATAGUEULE



Source : CSA.

L'individualisation ne prend pas en compte la correspondance entre le profil du participant et les vidéos qui lui sont recommandées, contrairement à la personnalisation[1].

Personnalisation des recommandations selon le profil des utilisateurs

Pour analyser la correspondance entre les recommandations de l'algorithme et le profil des utilisateurs établi à partir de l'expression de leurs centres d'intérêt sur YouTube, un échantillon de six participants a été sélectionné (représentant 16 % des participants), avec, parmi ceux-ci, un profil fictif (vie pratique, mode et maquillage), deux profils avec une ou deux thématiques très précise(s) (exemple : « *chroniques humoristiques de France Inter* ») et enfin trois profils avec plusieurs thématiques différentes, diversifiées et plus englobantes (exemple : « *jeux vidéo* », « *sciences* » et « *piano* »).

Ensuite, une recherche approfondie par mots-clés a été réalisée en fonction des grandes thématiques[2] cochées par les participants (septième question du questionnaire sur les habitudes et préférences de consommation sur YouTube), ainsi que de leurs centres d'intérêt sur la plateforme (question n° 8). La thématique « Information », cochée par 4 des 6 participants est prise dans une acception large comprenant les médias d'actualités et d'informations ainsi que tout contenu apportant de la connaissance (documentaires, reportages, enquêtes, etc.)[3].

En moyenne, seulement 11 % des vidéos recommandées aux participants correspondent à leurs habitudes et préférences déclarées de consommation sur YouTube. Comme pour l'ensemble des vidéos, les éditeurs traditionnels (presse, télévision, radio) sont surreprésentés. Ceci explique également la proportion importante de vidéos pour la thématique « Information ». Le taux de correspondance le plus bas est celui du profil fictif (ce qui, après vérification, vaut aussi pour les autres profils fictifs).

Nombre de vidéos recommandées se rapprochant du profil des participants

Participant 1 (diversifié)				**/ total des vidéos recommandées :**	**12 %**
Thèmes d'intérêt pour le participant	Information	Musique	Humour	France TV	Tutoriels informatiques
Vidéos recommandées sur le thème	72	0	24	22	0

[1] L'individualisation est une condition nécessaire à la personnalisation.
[2] Information, politique, musique, humour et sport.
[3] Dans l'enquête du *Pew Research Center* citée plus haut, 1 répondant sur 5 indique avoir recours à YouTube afin de mieux comprendre ce qui se passe autour d'eux dans le monde. Source : Pew Research Center. *Many Turn to YouTube for Children's Content, News, How-To Lessons*. 07/11/2018.

Participant 2 (diversifié)					/ total des vidéos recommandées :	15 %
Thèmes d'intérêt pour le participant	Information	Musique	Humour	Jeux vidéo	Sciences	Piano
Vidéos recommandées sur le thème	113	0	17	0	15	1

Participant 3 (spécialisé)					/ total des vidéos recommandées :	17,8 %
Thèmes d'intérêt pour le participant	Politique	Musique	Humour	Humour de France Inter	Clips RnB pop us	Clips Rap français
Vidéos recommandées sur le thème	59	1	57	54	3	0

Participant 4 (fictif)			/ total des vidéos recommandées :	2,9 %
Thèmes d'intérêt pour le participant	Vie pratique	Mode	Maquillage	
Vidéos recommandées sur le thème	17	8	4	

Participant 5 (diversifié)					/ total des vidéos recommandées :	24,6 %
Thèmes d'intérêt pour le participant	Information	Musique	Humour	Interviews	Documentaires	Multi-critères
Vidéos recommandées sur le thème	113	0	35	22	24	43

Participant 6 (spécialisé)				/ total des vidéos recommandées :	16,4 %
Thèmes d'intérêt pour le participant	Information	Politique	Musique	Blockchain	Economie
Vidéos recommandées sur le thème	125	19	3	14	0

29 participants ont déclaré écouter de la musique et regarder des vidéoclips sur YouTube. Un tiers des participants ont indiqué que ce type de contenus constituait leur consommation principale sur la plateforme. Pourtant, parmi eux, seulement 7 se sont vus proposer ce type de contenus par l'algorithme, pour un nombre de vidéos musicales de 17. Au total, seulement 19 vidéoclips ou concerts ont été recensés dans l'ensemble des vidéos pour l'ensemble des participants. Il est possible que des vidéos musicales ne soient proposées par l'algorithme que

lorsque la vidéo de départ est également musicale. L'algorithme détecterait alors que l'utilisateur est venu sur la plateforme explicitement pour écouter de la musique ou regarder des vidéoclips, et poursuivrait avec ce type de contenu très marqué[1].

Sur les thèmes sélectionnés

Dans le questionnaire sur les habitudes et préférences de consommation, les participants étaient interrogés sur leurs recherches passées sur YouTube sur les thèmes sélectionnés pour l'étude dans l'objectif de comprendre s'ils se voyaient recommander plus de vidéos sur ces sujets. Au global, la profondeur de l'offre associée à des sujets est très légèrement supérieure lorsque les individus ont déjà effectué des recherches sur ces mêmes sujets (28 % contre 26 % pour l'ensemble des participants).

Néanmoins, cette faible différence de taux peut masquer des réalités très différentes selon les thèmes qui empêchent de tirer des conclusions de ces résultats. Il ne peut être exclu que, bien que des recherches sur les sujets aient été faites par les participants, leur antériorité lointaine ou leur très faible récurrence aient rendu caduque le critère de consommations passées, de sorte que YouTube ne reconnaisse pas ces participants comme ayant un attrait particulier pour ces sujets.

De façon paradoxale, un des participants qui avait indiqué consulter régulièrement des vidéos documentaires sur l'astronomie s'est vu proposer moins de vidéos sur le thème des missions lunaires que la moyenne (35 % contre 49 %).

La personnalisation plutôt limitée des recommandations peut s'expliquer de deux manières principales. Premièrement, il est possible que les vidéos recommandées par la lecture automatique et celles recommandées dans le bandeau à droite de la page YouTube soient sélectionnées par deux algorithmes différents. Contrairement aux vidéos de la lecture automatique, certaines vidéos portent en effet la mention « recommandée pour vous » : il est probable que l'algorithme personnalise ces vidéos-là.

[1] Cette hypothèse ne peut être confirmée car aucune des vidéos de départ n'est une vidéo musicale.

Vidéo « recommandée pour vous » par YouTube

Source : YouTube avec ajouts CSA.

Deuxièmement, il est possible que, du fait du visionnage systématique des vidéos recommandées par la lecture automatique, l'algorithme ne reconnaisse aucun schéma de consommation passé, de l'utilisateur lui-même comme des autres utilisateurs, et n'ait pas suffisamment d'information pour personnaliser les recommandations.

Ingénierie inversée (*reverse engineering*)

L'analyse menée par le Conseil permet de dégager quelques grands principes de fonctionnement de l'algorithme de recommandation de YouTube. Ce dernier semble suivre une logique en deux temps.

① Sur le thème de départ

L'algorithme de recommandation propose des contenus portant sur le thème sélectionné. Pour ce faire, il se base sur les vidéos regardées par les autres utilisateurs qui ont visionné la même vidéo de départ. Les vidéos recommandées aux participants sont donc probablement celles qui, après la vidéo de départ, ont reçu le plus de clics de la part de ces utilisateurs ou pour lesquelles l'algorithme a constaté un temps de lecture significatif. L'algorithme croise cette information avec les mots-clés décrivant les vidéos, grâce à leur titre et à leurs *tags*. Ainsi, il est en mesure de recommander les vidéos les plus pertinentes en lien avec le sujet de départ.

Sur un sujet donné, l'algorithme recommande des vidéos, sans tenir compte des opinions exprimées, qui ne sont pas toujours présentes dans le titre ou les *tags* et qui ne sont en outre identifiables qu'en les regardant. La surreprésentation de certains courants de pensée dans les vidéos recommandées tient donc au dynamisme des communautés qui portent ces opinions et à

leur capacité à s'approprier les outils numériques de communication. Ce mécanisme de sélection crée donc un biais dans la diversité des opinions représentées.

La disponibilité des vidéos et la surreprésentation de certaines opinions peuvent faire apparaître un potentiel risque de chambre d'écho. De par leur construction, les algorithmes de recommandation sont structurés pour proposer successivement des contenus similaires lorsqu'ils sont susceptibles de plaire à l'utilisateur. Dès lors, en fonction de la répartition des opinions dans le total de vidéos sur un sujet, l'utilisateur peut se trouver confronté à une forme de récurrence des idées et des orientations. C'est d'autant plus vrai dans le cas de vidéos renvoyant à des sujets portés par des communautés actives.

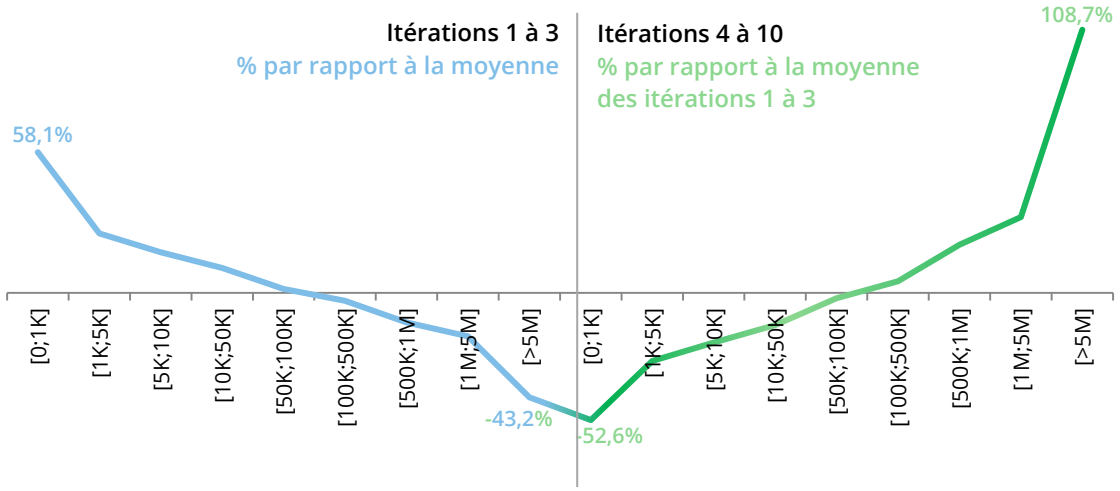
Les premières recommandations sur le thème de départ sont choisies par l'algorithme quasi-indépendamment du profil des utilisateurs et des caractéristiques des vidéos (chaîne, vues, dates, réactions, etc.). Toutefois, après la troisième vidéo recommandée sur le thème de départ, la lecture automatique s'en éloigne et commence à s'appuyer sur ces différents critères pour proposer des recommandations. Plus cette seconde étape s'opère tôt dans les itérations et moins l'algorithme permet de s'informer de façon pluraliste sur un sujet donné.

② *Sur le profil des utilisateurs*

Dans un second temps, l'algorithme dévie du thème de départ et individualise les recommandations pour proposer des vidéos censées plaire aux utilisateurs et maximiser le temps qu'ils passent sur la plateforme. Pour ce faire, l'algorithme recommande des vidéos de plus en plus populaires (nombre de vues) et de plus en plus récentes. L'identification de ce phénomène découle d'une observation menée avant et après la troisième itération, qui constitue un « pivot » : c'est à ce moment-là que l'algorithme dévie du thème de départ et individualise les recommandations.

Concernant la progression du nombre de vues au fil des itérations, 48 % des vidéos de moins de 1 000 vues ont été recommandées au cours des trois premières itérations, alors que ce taux n'est que de 21 % pour les vidéos de plus d'un million de vues. L'observation du nombre de vues des recommandations au-delà de la troisième itération par rapport à la moyenne des trois premières itérations vient confirmer ces constats : près de 80 % des vidéos de plus d'un million vues ont été recommandées après cette troisième itération, ce qui représente une augmentation par rapport à la moyenne des trois premières itérations de 70 %. Cette augmentation monte jusqu'à près de 109 % (plus du double) pour les vidéos de plus de cinq millions de vues.

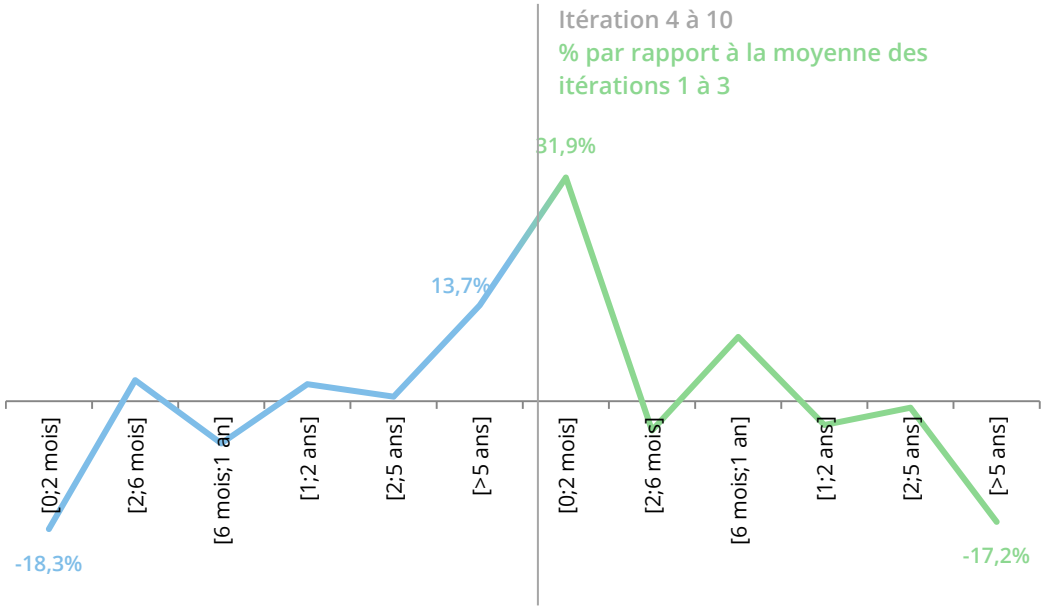
Pourcentages d'évolution du nombre de vues des vidéos avant et après la troisième itération



Base : ensemble des vidéos recommandées (39 007).
Lecture : par rapport à l'ensemble des itérations (1 à 10), il y a 58 % de vidéos de moins de 1000 vues en plus au cours des 3 premières itérations. A contrario, *le nombre de vidéos comptant plus de 5 millions de vues a plus que doublé (pourcentage supérieur à 100 %) entre les itérations 1 à 3 et les itérations 4 à 10.*

Un constat similaire peut être fait pour la date de publication des vidéos recommandées, toutefois moins marqué que pour le nombre de vues : 75 % des vidéos de moins de deux mois ont été recommandées après la troisième itération alors qu'elles sont 66 % de plus de cinq ans. Plus particulièrement, après la troisième itération, il y a 32 % en plus de vidéos de moins de deux mois qu'au cours des trois premières itérations.

Pourcentages d'évolution des dates de publication des vidéos recommandées

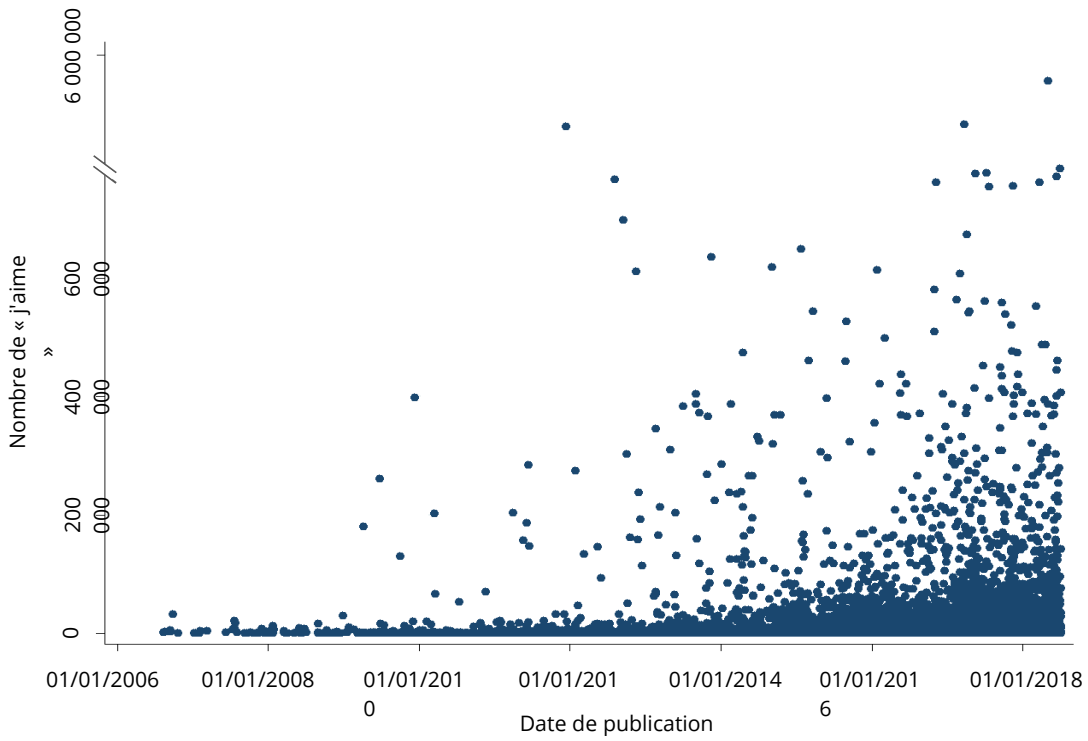


Base : ensemble des vidéos recommandées (39 007).
Lecture : par rapport à l'ensemble des itérations (1 à 10), il y a 18 % de vidéos de moins de 2 mois en moins au cours des 3 premières itérations. A *contrario, le nombre de vidéos de moins de 2 mois a augmenté de près de 32 % entre les itérations 1 à 3 et les itérations 4 à 10.*

Par ailleurs, l'observation croisée de la date de publication et du nombre de « j'aime » des vidéos montre que ces deux variables ont tendance à croître ensemble. Dès lors, tout comme la date de publication, il est vraisemblable que le nombre de « j'aime »[1] soit lui aussi un des nombreux critères pris en compte par l'algorithme pour déterminer ses recommandations.

[1] Le nombre de « je n'aime pas » pourrait avoir un effet inversement similaire. Toutefois, le nombre de « je n'aime pas » est généralement bien plus petit que le nombre de « j'aime », ce dernier étant beaucoup plus usuel, notamment sur les réseaux sociaux, et donc plus facilement utilisé.

Répartition des vidéos en fonction de leur date de publication et du nombre de « j'aime »



Base : ensemble des vidéos recommandées (39 007).

Enfin, il a pu être établi dans une enquête similaire que la durée des vidéos augmente au fil de la navigation sur YouTube[1]. L'objectif sous-jacent serait de garder les utilisateurs sur la plateforme : une vidéo plus longue sera statistiquement regardée plus longtemps. Toutefois, l'échantillon sur lequel s'appuie l'étude du Conseil[2] ne permet pas de venir confirmer de manière évidente ce phénomène. Les résultats montrent une très légère mais constante diminution de la durée moyenne des vidéos recommandées au fil des itérations (près de 19 minutes à la première itération contre un peu plus de 16 minutes à la dixième). De même, la proportion de vidéos de plus de 30 minutes diminue de 18 % entre la première et la dernière itération.

Les vidéos s'individualisent sans pour autant correspondre aux préférences et habitudes des utilisateurs. Que ce soit pour les thématiques les plus consultées par les participants sur YouTube ou pour les recherches déjà effectuées par les participants sur les thèmes de départ, l'algorithme ne semble pas aller au-delà de l'individualisation et ne personnalise pas les recommandations. Cette conclusion illustre l'importance du comportement et des parcours de consommation des utilisateurs sur la plateforme. Il serait intéressant de comparer ces résultats

[1] Pew Research Center. *Many Turn to YouTube for Children's Content, News, How-To Lessons*. 07/11/2018.
[2] À noter que la durée des vidéos n'a pas pu être recueillie pour 6,2 % de l'ensemble de l'échantillon de vidéos, ce qui n'est pas de nature à affecter les enseignements tirés des résultats.

pour la lecture automatique avec le degré de personnalisation des vidéos labellisées « recommandée pour vous ».

Pour proposer du contenu supposé pertinent aux utilisateurs, et en dehors de critères de popularité et de récence des vidéos, l'algorithme s'appuie sur les mots-clés présents dans le nom des chaînes, les titres et les *tags* afin de lier les vidéos entre elles par un dénominateur commun. Il a pu être montré que la puissance des chaînes, c'est-à-dire la probabilité pour qu'un utilisateur qui vient de regarder une vidéo d'une chaîne donnée clique ensuite sur une vidéo de la même chaîne, est un facteur important de recommandation. Au-delà, une fois que l'ensemble des vidéos disponibles sur le thème de départ a été consommé, l'algorithme dévie de ce thème en le reliant à une chaîne ou à un sujet plus global.

Conclusion

Ces travaux visent à alimenter le débat public relatif à la responsabilité des plateformes de partage de contenus dans nos sociétés démocratiques. Ils fournissent un éclairage sur les effets de l'algorithme de recommandation utilisé par YouTube pour sa fonctionnalité de lecture automatique sur l'accès à l'information, dans les limites imposées par l'exercice : l'observation *a posteriori* des recommandations de la plateforme ne permet pas d'isoler les poids et les effets de chacun des critères pris séparément.

Même si ces observations ne constituent pas des conclusions pouvant nécessairement être généralisées à l'ensemble des algorithmes, elles confirment le rôle central joué par les plateformes dans l'accès à l'information dès lors qu'elles disposent de fonctions d'ordonnancement et de recommandation de contenus. En particulier, la mise en lumière de phénomènes qui par certains aspects pourraient être assimilés à ceux dits de chambre d'écho, renforcés par l'individualisation des recommandations, peut interroger sur la capacité des algorithmes à proposer aux utilisateurs des contenus exprimant une diversité de points de vue.

En cohérence avec les initiatives en cours reposant pour la plupart sur des outils de corégulation, cette étude souligne la nécessité de renforcer, d'une part, l'éducation aux médias et, d'autre part, les relations entre les pouvoirs publics et les plateformes, afin d'améliorer la transparence des algorithmes et d'assurer aux utilisateurs une information claire sur les recommandations qui leur sont faites.

Bibliographie

Institutions

Conseil national du numérique. Avis n° 2014-2. *Neutralité des plateformes. Réunir les conditions d'un environnement numérique ouvert et soutenable.* 05/2014.

Commission nationale informatique & libertés. *Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle*. 12/2017.

CSA Lab. *Le rôle des données et des algorithmes dans l'accès aux contenus*. 01/2017.

Conseil supérieur de l'audiovisuel. *Vidéos en ligne ou télévision chez les jeunes publics : étude économétrique*. 01/2018.

Conseil supérieur de l'audiovisuel. *Plateformes et accès aux contenus audiovisuels. Quels enjeux concurrentiels et de régulation ?* 09/2016.

European Commission. *A multi-dimensional approach to disinformation*. 03/2018.

Instituts de sondage

Ifop pour Fondation Jean Jaurès et Conspiracy Watch. *Enquête sur le complotisme*. 01/2019.

Ifop pour CNIL. *Notoriété et attentes vis-à-vis des algorithmes*. 01/2017.

Kantar pour La Croix. *Baromètre de la confiance dans les médias*. 01/2019.

Médiamétrie pour DGMIC. *Les jeunes et l'information. Synthèse*. 07/2018.

Médiamétrie. *La place croissante des réseaux sociaux dans l'accès à l'information*. 29/11/2016.

Centres de recherche

Paul Covington, Jay Adams, Emre Sargin. Google. *Deep neural networks for YouTube Recommendations*. RecSys'16. 09/2016.

Pew Research Center. *Many Turn to YouTube for Children's Content, News, How-To Lessons*. 07/11/2018.

Reuters Institute for the Study of Journalism. University of Oxford. *Digital news report 2018*. 06/2018.

Presse

The Guardian. *Fiction is outperforming reality*. 02/02/2018.

The New York Times. *YouTube, the great radicalizer*. 10/03/2018.

Tubefilter. *Reverse engineering the YouTube algorithm*. *Part I*. 06/2016.

Tubefilter. *Reverse engineering the YouTube algorithm*. *Part II*. 02/2017.

Annexes

Liste des vidéos de départ sélectionnées

La liste des vidéos de départ sélectionnées répond à un objectif de diversité des points de vue, avec, si la disponibilité des vidéos le permettait, autant de vidéos portant l'opinion A que de vidéos portant l'opinion B (voir nomenclature des points de vue au troisième point de la partie 2 de l'étude). Des vidéos neutres, qui concourent au débat en présentant les arguments des différentes parties, ont été ajoutées, lorsque cela était possible.

Certains titres de vidéos pouvant être trompeurs, les vidéos ont toutes été visionnées afin de déterminer le point de vue qu'elles présentent. Certaines vidéos peuvent ne plus être disponibles aujourd'hui sur la plateforme.

Les thèmes portant la mention « W » sont plutôt discutés sur Internet, le « M » réfère un sujet principalement abordé dans les médias traditionnels, le « m » renvoie à un traitement médiatique plus restreint (notamment en fonction de l'actualité) et à une présence sur le web.

« Le grand remplacement »	**M**
« Le Grand remplacement stupéfie le monde entier! » - Renaud Camus	Opinion A
Le Grand remplacement des blancs Européens est voulu et facilité	Opinion A
« Grand remplacement » : complot, fantasme ou réalité ?	Neutre
Les détricoteuses : Le grand méchant remplacement	Opinion B
Théorie du grand remplacement - Désintox - ARTE	Opinion B

Bitcoin et cryptomonnaies	**m**
BITCOIN : Comment ça marche et où en acheter ?	Opinion A
Questions Réponses Bitcoin Ethereum & Blockchain	Opinion A
La bitcoin mania : révolution ou bulle prête à éclater ?	Neutre
10 Choses qui prouvent que le bitcoin est voué à l'échec | Lama Faché	Opinion B
L'arnaque des cryptomonnaies et scandales à venir | Bitcoin - Ethereum	Opinion B

Conflit Israélo-Palestinien	**M**
Tel-Aviv-sur-Seine : entretien avec un citoyen pro-israélien	Opinion A
Israël-Palestine : comprendre le conflit par les cartes	Neutre
Palestine: un député français ose dire la vérité sur Israël	Opinion B
Gaza : les pressions sur Israël continuent	Opinion B

Corrida	**m**
250 réponses sur la Tauromachie (Arles)	Opinion A
Découvre la tauromachie en t'amusant !	Opinion A
La corrida mise à mort ? Avec Jean-Pierre Garrigues (CRAC Europe)	Neutre
Journée de mobilisation contre la corrida	Opinion B
Reportage sur le combat contre la corrida dans l'écho des lois	Opinion B

Drague dans la rue	**m**
Comment aborder une fille dans la rue (caméras cachées)	Opinion A
Ma réaction quand une fille n'est pas intéressée et vous dit NON	Opinion A
#53: La Limite entre « Séduire » et « Harceler » (Exp. Sociale)	Neutre
Harcèlement de rue | expérience sociale (IbraTV)	Opinion B
Harcèlement de rue - Cam Clash	Opinion B

Droits sexuels et reproductifs	**m**
Droits sexuels et reproductifs	Opinion A
Droits sexuels et reproductifs – animation	Opinion A
La déclaration internationale des droits sexuels (Marion Sigaut)	Opinion B

Éducation sexuelle à l'école	**m**
Éducation à la vie sexuelle et affective en école primaire	Opinion A
Être jeune : l'éducation complète à la sexualité	Opinion A
Geoffroy Lejeune : « Il est urgent de supprimer l'éducation sexuelle à l'école » (LCI,20/11/17,17h49)	Opinion B
M. BERGER - Les dangers de l'éducation à la sexualité pour les enfants et les adolescents	Opinion B

Épilation	**W**
Poils, Hygiène et Épilation !!!	Opinion A
Je n'ai plus de poils... ☹ Mon Secret !	Opinion A
Micro trottoir : Êtes-vous pour ou contre l'épilation des poils ?	Neutre
Pourquoi j'ai arrêté de m'épiler	Opinion B
Voici pourquoi il ne faut plus s'épiler !	Opinion B

Euro	**M**
Avec l'euro tout a augmenté sauf mon pouvoir d'achat ! Vraiment ? #DecodeursUE	Opinion A
France 24 JT de l'éco - L'Euro un accélérateur de la hausse des prix	Opinion A
La catastrophe de l'Euro - Avec Jean-Jacques Rosa (21 novembre 2014) - 1ère partie	Opinion B

Homosexualité	**m**
L'homosexualité n'est pas contre-nature mais l'homophobie oui	Opinion A
Pourquoi les rapports homosexuels ne sont pas « contre-nature »	Opinion A
Comment vaincre l'homosexualité ? - David Storm	Opinion B
De la réalité cruelle de la pratique de l'homosexualité !	Opinion B

Intelligence artificielle	**m**
La réalité de l'intelligence artificielle, c'est ça	Opinion A
Laurent Alexandre : Intelligence artificielle [EN DIRECT]	Opinion A
Intelligence artificielle : CESPEDES VS Dr ALEXANDRE	Neutre
[Documentaire] Lintelligence artificielle: Le règne des robots	Opinion B

Laïcité	**M**
#generationlaicité - Le port de signes distinctifs à l'école	Opinion A
Quand elle impose de manger du porc, la laïcité, c'est la boucherie	Opinion A
La laïcité un concept à géométrie variable ?	Neutre
« Radicalisation de la laïcité »: « On ne gagne jamais à stigmatiser quiconque », estime Griveaux	Opinion B
Marine Le Pen demande un sacrifice aux Juifs de France (interdiction du port de la kippa)	Opinion B

Machisme/féminisme	**M**
#TesFéministeMais... tu suces ? (Marion Seclin)	Opinion A
[France] Sexisme au travail : on s'laisse pas faire !	Opinion A
« Les femmes aiment les machos mais elles ne se l'admettent pas » - GIUSEPPE LFM SHOW	Opinion B
C'est mon choix : Je suis macho, et alors	Opinion B

Missions sur la lune	**W**
Top 5 des preuves irréfutables qu'on n'a jamais marché sur la Lune	Opinion A
Le dernier homme sur la lune - APOLLO 17	Opinion A
★ Missions lunaires : Une Chose étrange s'est passée avec Apollo 17...	Opinion B
Conquête de la lune : Supercherie dévoilée par les principaux acteurs de l'époque	Opinion B

Mort de Michael Jackson	**W**
20h France 2 du 26 juin 2009 - Mort de Michael Jackson - Archive vidéo INA	Opinion A
Michael Jackson - L'histoire d'une tragédie	Opinion A
Michael Jackson n'est pas mort !	Opinion B
7 Choses étranges sur la mort de Michael Jackson	Opinion B

Réchauffement climatique	**M**
« 2 degrés avant la fin du monde » - #DATAGUEULE	Opinion A
J'ai vu les preuves du réchauffement climatique	Opinion A
Réchauffement Climatique - On vous ment	Opinion B
Réchauffement Climatique c'est faux la Banquise découpée au Laser	Opinion B

Réforme de la SNCF	**M**
« La SNCF est une anomalie en Europe » - La réforme de la SNCF vue par les médias étrangers	Opinion A
VIDEOS. Grève SNCF : soutien, colère, résignation... qu'en pensent les usagers ?	Opinion A
Le TFUE ou la privatisation de la SNCF	Neutre
Ce que les médias ne disent pas sur la SNCF	Opinion B
Le scandale de la privatisation de la SNCF - Vincent Lapierre	Opinion B

Sectes	**m**
C'est Mon Choix : Je suis devenu adepte d'une secte	Opinion A
C'est quoi une secte ? - le Professeur Gamberge	Neutre
Vivre dans une secte et s'en sortir, Amoreena Winkler témoigne - Mille et une vies	Opinion B
Les sectes et leurs dangers	Opinion B

SIDA	**m**
La vérité sur Le SIDA !	Opinion A
Grandes tueuses : Sida	Opinion A
Le SIDA, un complot médical de masse	Opinion B
LE SIDA - Le plus gros mensonge du 21eme siecle-.wmv	Opinion B

Téléphone à l'école	**m**
« Interdire le téléphone portable dans les écoles ? Une mauvaise idée »	Opinion A
JMI L'émission 16 - Pour ou contre le droit au téléphone au collège ?	Neutre
Que pensent les enfants du téléphone portable ? - La Maison des Maternelles	Neutre
Les téléphones portables bientôt interdits à l'école !!	Opinion B
Parents : Pour ou contre le téléphone portable à l'école ??	Opinion B

Vaccins	**m**
La vérité sur les vaccins - Dossier #11 - L'Esprit Sorcier	Opinion A
9 Bullshits sur les vaccins [MYTHOBUSTER]	Opinion A
L'arnaque de la vaccination	Opinion B
Vaccins: mort, autisme ou abrutissement assuré?...	Opinion B

Véganisme	**m**
Les Carnistes ont-ils un Cerveau ? Pas sûr ...	Opinion A
Pourquoi devenir végétarien ? - L214	Opinion A
Faut-il arrêter de manger de la viande ? Débat	Neutre
Le végétarisme est-il vraiment plus éthique que le carnisme ?, Frédéric Bisson	Opinion B
Manger de la viande n'est pas un crime, et s'en abstenir n'est pas une vertu	Opinion B

Vie extraterrestre	**W**
Documentaire // L'Armée américaine cache l'existence d'une vie extraterrestre // 2018	Opinion A
6 témoignages d'astronautes sur les extraterrestres	Opinion A
C'est pas sorcier - Vie extraterrestre	Opinion B
7 minutes pour comprendre : La vie extraterrestre (Partie 1)	Opinion B

Questionnaire sur les habitudes et préférences de consommation sur YouTube

FICHE PROFIL

Les volontaires doivent posséder un compte Google (Gmail, YouTube…).

Identifiant : _____

1. À quelle fréquence regardez-vous la télévision ?

- ❑ Quotidiennement (entre 25 et 31 fois par mois)
- ❑ Plusieurs fois par semaine (entre 15 et 25 fois par mois)
- ❑ Occasionnellement (entre 5 et 15 fois par mois)
- ❑ Rarement (moins de 5 fois par mois)

2. Utilisez-vous YouTube ?

- ❑ Oui (veuillez répondre aux questions suivantes)
- ❑ Non (répondez seulement aux questions 8, 9 et 10)

3. Sur YouTube, avez-vous modifié d'une quelconque manière vos paramètres de recommandations ?

- ❑ Oui
- ❑ Non

4. Vous utilisez YouTube sur votre (plusieurs réponses possibles) :

- ❑ Ordinateur professionnel (fixe ou portable)
- ❑ Ordinateur personnel (fixe ou portable)
- ❑ Smartphone
- ❑ Tablette
- ❑ Autres, précisez : _____

5. À quelle fréquence utilisez-vous YouTube ?

- ❑ Quotidiennement (entre 25 et 31 fois par mois)
- ❑ Plusieurs fois par semaine (entre 15 et 25 fois par mois)
- ❑ Occasionnellement (entre 5 et 15 fois par mois)
- ❑ Rarement (moins de 5 fois par mois)

6. À quel(s) moment(s) de la journée utilisez-vous le plus YouTube (plusieurs réponses possibles) ?

- ❑ Entre 8 heures et 10 heures
- ❑ Entre 10 heures et 13 heures
- ❑ Entre 13 heures et 17 heures
- ❑ Entre 17 heures et 20 heures
- ❑ Entre 20 heures et 23 heures

7. Sur YouTube, vous regardez plutôt des vidéos à caractère (plusieurs réponses possibles) :

- ❑ Informatif
- ❑ Politique
- ❑ Musical
- ❑ Humoristique
- ❑ Sportif
- ❑ Autres, précisez : _____

8. Sur YouTube, quelles sont les trois thématiques de vidéo (thème comme le football, chaîne comme Golden Moustache, sujet comme la neutralité du net, personnalité comme Beyonce, etc.) que vous consultez le plus :

① _____
② _____
③ _____

9. Dans la liste suivante, à quel(s) thème(s) portez-vous un intérêt particulier (zéro ou plusieurs réponse(s) possible(s)) ?

- ❑ Conquête spatiale
- ❑ Développement durable
- ❑ Diplomatie internationale
- ❑ Europe
- ❑ Laïcité
- ❑ Protection des minorités
- ❑ Égalité des sexes
- ❑ Éducation sexuelle
- ❑ Santé
- ❑ Protection des animaux
- ❑ Alimentations alternatives
- ❑ Intelligence artificielle
- ❑ Blockchain

10. Dans la liste suivante, pour quel(s) thème(s) avez-vous déjà effectué une recherche Google ou YouTube (zéro ou plusieurs réponse(s) possible(s)) ?

- ❑ Conquête spatiale
- ❑ Développement durable
- ❑ Diplomatie internationale
- ❑ Europe
- ❑ Laïcité
- ❑ Protection des minorités
- ❑ Égalité des sexes
- ❑ Éducation sexuelle
- ❑ Santé
- ❑ Protection des animaux
- ❑ Alimentations alternatives
- ❑ Intelligence artificielle
- ❑ Blockchain

11. Sur YouTube, vous regardez plutôt des vidéos publiées par un (plusieurs réponses possibles) :

- ❑ « Amateur » (vidéo réalisée avec de petits moyens)
- ❑ « Amateur professionnalisé » (exemple : Enjoy Phoenix)
- ❑ « Professionnel de l'Internet » (exemple : Brut)
- ❑ « Professionnel de la télévision » (exemple : The Voice, L'Effet Papillon, etc.)

12. Si vous avez coché la case « Professionnel de la télévision », à quelle fréquence consultez-vous ce type de contenu sur YouTube ?

- ❑ Systématiquement
- ❑ La plupart du temps
- ❑ De temps en temps

13. Sur YouTube, vous arrive-t-il de laisser la lecture automatique enchaîner les vidéos ?

- ❑ Systématiquement
- ❑ La plupart du temps
- ❑ De temps en temps
- ❑ Jamais

14. Sur YouTube, avez-vous l'habitude de regarder des vidéos sélectionnées à partir des recommandations ?

- ❑ Systématiquement
- ❑ La plupart du temps
- ❑ De temps en temps
- ❑ Jamais

15. Sur YouTube, avez-vous l'impression que l'algorithme vous recommande des vidéos qui se « ressemblent » ?

- ❑ Oui, et cela me plaît
- ❑ Oui, et cela me dérange
- ❑ Oui, et cela m'est indifférent
- ❑ Non