



# L'Essentiel

Les études de l'Hadopi en 10 minutes chrono - #20 - Avril 2020

## L'ÉVALUATION DES TECHNOLOGIES DE RECONNAISSANCE DE CONTENUS

le cas des systèmes d'empreintes appliqués aux contenus audiovisuels

L'article 17 de la directive européenne sur le droit d'auteur dans le marché numérique devrait donner une portée renforcée aux outils de reconnaissance de contenus déjà mis en place par les plateformes, en les transformant en dispositifs prévus et encadrés par le droit.

Dans ce contexte, le Conseil supérieur de la propriété littéraire et artistique (CSPLA), le Centre du cinéma et de l'image animée (CNC) et l'Hadopi se sont associés pour mener une mission conjointe, visant notamment à réaliser un état de l'art des technologies existantes en la matière et à évaluer leur efficacité.

Il est apparu que l'évaluation de ces technologies de reconnaissance de contenus pouvait s'effectuer selon trois types de critères : leur robustesse, c'est-à-dire leur capacité à reconnaître de manière effective et automatique les contenus protégés, y compris lorsque ces derniers subissent des altérations ; leur praticité, ce qui correspond aux modalités selon lesquelles les ayants droit peuvent les utiliser ; leur finesse, ce qui consiste à prendre en compte leurs éventuels effets de bord sur les usages légitimes des consommateurs.

Une première évaluation, relative aux systèmes de reconnaissance par empreintes appliqués aux contenus audiovisuels, a pu être mise en œuvre à titre expérimental.

### Les outils de reconnaissance automatique de contenus : l'importance des systèmes d'empreintes

Les systèmes de reconnaissance de contenus peuvent s'appliquer à différents types de contenus : musique, vidéo, image, texte. Des technologies d'identification de vidéos sont aujourd'hui déployées sur les principales plateformes de partage de contenus telles que Facebook (avec son outil Rights Manager), YouTube (Content ID) ou Dailymotion (qui utilise INA-Signature).

La technique de reconnaissance de contenus par empreintes numériques, ou "fingerprinting", est fondée sur la comparaison de représentations numériques simplifiées de contenus (images fixes, vidéos ou pistes sonores).

Plutôt que d'évaluer directement la comparaison entre deux documents, ces outils observent leurs éventuelles similarités en comparant leurs empreintes. La similarité des documents s'établit donc au travers de la création d'une empreinte pour chaque document

et de la métrique permettant de comparer ces empreintes.

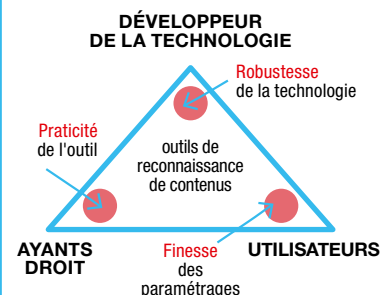
Les empreintes (qui ne contiennent que les éléments caractéristiques d'un contenu, selon des critères choisis) sont plus légères et plus faciles à manipuler que les documents eux-mêmes. Elles sont également plus robustes, car résistantes à de faibles modifications ou altérations des contenus.

Un système de reconnaissance de contenus se compose en général d'une base de référence où sont stockées les empreintes de tous les documents à identifier. Une fois cette base créée, le système est utilisé comme un moteur de recherche : lorsque l'on veut vérifier si un document inconnu est présent ou non dans la base de référence, on extrait d'abord son empreinte avant de la comparer avec les empreintes pré-calculées et stockées dans la base. Cette comparaison permet ensuite de dire si l'on a trouvé des documents similaires.

#### FOCUS

#### LES ENJEUX D'ÉVALUATION DES OUTILS DE RECONNAISSANCE DE CONTENUS : UNE MÉTHODOLOGIE REPOSANT SUR TROIS CRITÈRES

L'évaluation des technologies de reconnaissance de contenus repose sur trois types de critères. Les capacités et la robustesse de la technologie, mesurées par les tests réalisés dans le cadre de la mission, sont certes essentielles, mais elles n'en sont qu'un aspect. Pour une évaluation complète, il faut aussi prendre en compte les fonctionnalités offertes aux ayants droit et la commodité de leur mise en œuvre (nombre d'offres, simplicité d'utilisation, etc.). Enfin, la finesse dont les ayants droit font preuve dans l'usage des outils pour tenir compte des exceptions a été appréhendée en interrogeant directement les internautes sur leurs usages.



#### MÉTHODOLOGIE

Les tests techniques ont été réalisés par l'Hadopi avec l'aimable autorisation de Gaumont et TF1, et la coopération de YouTube, Facebook, INA et Audible Magic. Étude quantitative en ligne confiée à OpinionWay, en deux temps : interrogation d'un échantillon de 3 040 internautes âgés de 15 ans et plus, représentatif de la population française (terrain du 19 août au 4 septembre 2019), puis interrogation d'un échantillon de 1 445 internautes ayant partagé des contenus audio et vidéo, incluant une base de 285 internautes ayant été effectivement confrontés à des mesures de blocage (terrain du 21 octobre au 15 novembre 2019).

**FIGURE 1 : EXEMPLE FICTIF DE LA MISE EN CORRESPONDANCE PARTIELLE DES EMPREINTES DE DEUX IMAGES**



La photo originale (à gauche) a été transformée pour obtenir la photo de droite :

- zoom
- recadrage
- changement de couleur
- changement du format de fichier

Leur format et leur apparence ne sont donc pas identiques en tout point.

Or, la comparaison des empreintes numériques des deux images, obtenues à partir de points saillants des photos (points jaunes), permet d'établir des ressemblances suffisamment caractéristiques pour déterminer qu'il s'agit bien du même contenu : le système de reconnaissance identifie bien qu'il s'agit de la même image.

## I Une robustesse avérée des systèmes de reconnaissance par empreintes

Les tests techniques réalisés sur trois plateformes démontrent globalement des capacités élevées de ces technologies, tant pour les contenus audio que vidéo.

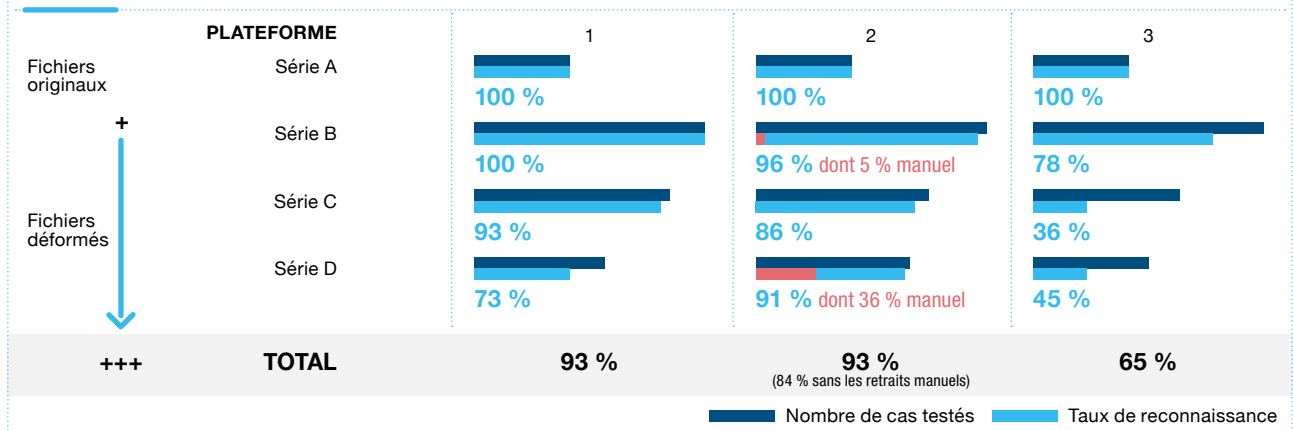
Des œuvres ont été mises en ligne sur chacune des trois plateformes, comme pourrait le faire un de leurs utilisateurs, afin de mesurer la capacité de reconnaissance des outils mis en place par ces plateformes.

Quatre séries de tests ont été réalisées pour évaluer la robustesse des technologies : d'abord des tests consistant en la mise en ligne des fichiers originaux (série A), puis des tests consistant en la mise en ligne de fichiers ayant subi des déformations, de légères à sévères (séries B à D), avec des altérations allant de la simple modification de la vitesse de la vidéo jusqu'à la combinaison d'effets multiples, tels que

rotation de l'image, zoom, filtres, etc. Au total, plus de 150 extraits représentant près de 20 heures de vidéo ont été testés.

On observe que les outils mis en place par les plateformes sont capables de reconnaître un grand nombre d'œuvres, y compris lorsque les altérations qu'elles ont subies rendent en réalité impossible leur visionnage.

**FIGURE 2 : RÉSULTATS DES TESTS RÉALISÉS**



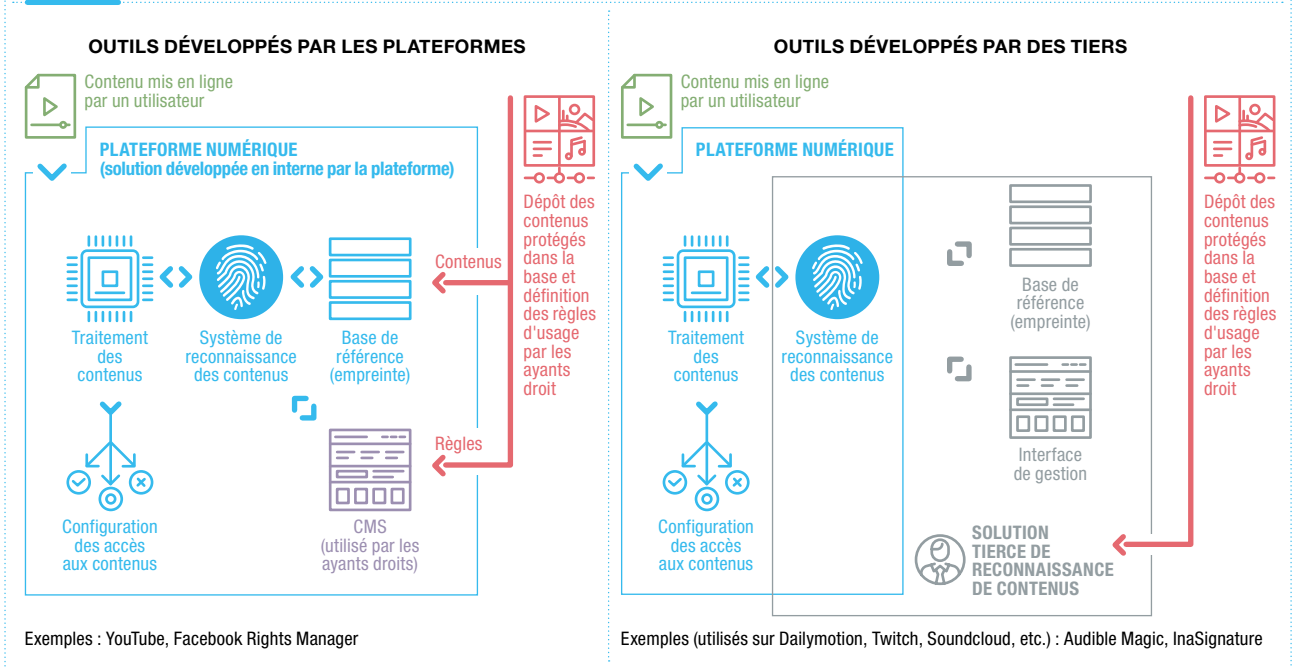
## I Des modèles d'organisation fermés

Il n'existe pas de solution unique ou dominante, certaines plateformes ayant choisi de développer en interne une solution de reconnaissance de contenus sur mesure, tandis que, en parallèle, plusieurs prestataires proposent des services tiers équivalents.

À ce stade, ces prestataires semblent tous avoir adopté leurs propres formats et leurs propres standards, et cette hétérogénéité risque, avec le développement de nouveaux outils de détection, d'alourdir la tâche des ayants droit qui trouveraient avantage à une certaine uniformisation des pratiques.

Pour autant, la mise en place de systèmes ouverts ou cogérés posent des difficultés, notamment en matière de sécurité.

**FIGURE 3 : FONCTIONNEMENT SCHÉMATIQUE DES OUTILS DE RECONNAISSANCE DE CONTENUS**



## Une compréhension globalement plutôt bonne par les internautes des règles relatives au respect du droit d'auteur sur les plateformes

### UNE BONNE CONNAISSANCE DES RÈGLES DE BLOCAGE

Les deux tiers des internautes semblent maîtriser le principe d'autorisation nécessaire pour partager un contenu. Interrogés sur une série d'assertions dont ils doivent mesurer le caractère "vrai" ou "faux", les internautes semblent avoir un certain niveau de connaissance et de compréhension des principes applicables en matière de respect du droit d'auteur.

**Ainsi, 87 % des internautes savent que les plateformes peuvent retirer un contenu et 75 % savent qu'elles peuvent également empêcher la mise en ligne de contenus culturels.**

**FIGURE 4 : OPINION DES INTERNAUTES SUR LA MISE EN LIGNE DE CONTENUS - BASE : INTERNAUTES ÂGÉS DE 15 ANS ET PLUS**

|  | ❌    | ✅    |
|--|------|------|
| Mettre en ligne des extraits de musique ou de film d'un autre auteur ne nécessite pas une autorisation                     | 31 % | 69 % |
| Mettre en ligne des partitions ne nécessite pas une autorisation   | 32 % | 68 % |
| Poster une vidéo en direct de son téléviseur pendant la retransmission d'une compétition ne nécessite pas une autorisation | 35 % | 65 % |

Rouge : réponse fausse | Vert : réponse correcte

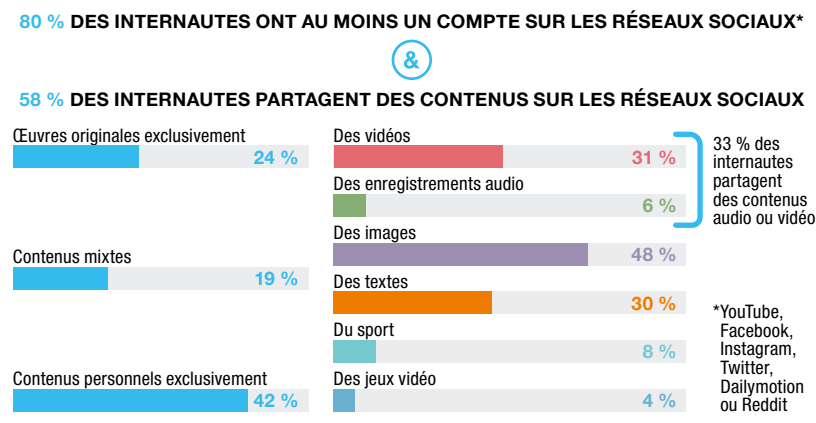
## Le partage de contenus, une pratique importante des internautes

### PLUS DE LA MOITIÉ DES INTERNAUTES PARTAGE DES CONTENUS SUR LES PLATEFORMES

De manière globale, 58 % des internautes partagent des contenus sur au moins l'une des six plateformes et réseaux considérés, quel que soit le type de contenus. On distingue trois types de contenus postés :

- les contenus exclusivement personnels de l'utilisateur ;
- les contenus émanant exclusivement d'autres personnes que l'utilisateur, qualifiés d'œuvres originales ;
- les contenus mixtes, mêlant des contenus personnels et des contenus émanant de tiers, comme un film de vacances qu'accompagne la musique d'un artiste.

**FIGURE 5 : PARTAGE DE CONTENUS - BASE : INTERNAUTES ÂGÉS DE 15 ANS ET PLUS**



# Des effets de bord limités sur les usages des internautes

## UNE MINORITÉ D'INTERNAUTES VOIT SES CONTENUS BLOQUÉS

Parmi les internautes qui partagent des contenus (tous types de contenus confondus), 11 % ont déjà reçu un message de blocage de leurs contenus pour des raisons liées au respect du droit d'auteur, soit 6 % des internautes. Les messages reçus faisaient suite, dans 51 % des cas, à la mise en ligne d'œuvres originales (c'est-à-dire des

contenus émanant exclusivement d'autres personnes), dans 29 % des cas à la mise en ligne de contenus mixtes.

## DES CONTESTATIONS BIEN GÉRÉES

Moins de la moitié (43 %) des partageurs bloqués ont contesté le blocage de leurs contenus et 27 % des partageurs ont contesté plusieurs blocages. Au final, 3 % des internautes ont contesté une mesure de blocage.

FIGURE 6 : SCHÉMA RÉCAPITULATIF – BLOCAGE ET CONTESTATION DE BLOCAGE LIÉ À DES QUESTIONS LIÉES AU DROIT D'AUTEUR



La grande majorité des partageurs bloqués (89 %) déclare avoir compris la raison de la réception du message les informant de la mesure de blocage. Cependant, un tiers (34 %) des bloqués estime que les blocages de ces contenus, en général, n'étaient pas justifiés, au motif principal qu'il s'agissait d'un extrait (ce que déclare la moitié de ces 34 % d'individus bloqués). Enfin, si l'on se focalise sur les seuls

internautes ayant partagé un contenu audio ou vidéo, on note que le blocage d'un tel contenu a des conséquences sur leurs comportements. À la suite du blocage de leurs contenus, la moitié des partageurs se sont renseignés sur les règles en matière de partage de contenus (47 %). Une autre moitié (55 %) des internautes bloqués a déclaré faire plus attention à la citation des sources (crédits, mentions, etc.) lors de la

FIGURE 7 : ACTIONS DÉCLARÉES SUITE AU BLOCAGE DU DERNIER CONTENU AUDIO OU VIDÉO PARTAGÉ – BASE : INTERNAUTES 15 ANS ET PLUS AYANT VU LEUR DERNIER CONTENU AUDIO OU VIDÉO PARTAGÉ BLOQUÉ

Arrêt du partage du type de contenus pour lequel ils ont reçu un message de blocage

53 %

Utilisation d'autres plateformes pour partager des contenus

44 %

Prise de mesures pour contourner les blocages

34 %

Davantage d'attentions portées à la citation des sources lors du partage de contenus (crédits, mentions, sources, etc.)

55 %

Prise de renseignements sur les règles en matière de partage de contenus

47 %

mise en partage de contenus, une action en réponse qui met en évidence un problème de compréhension des règles d'autorisation du partage de contenus protégés par le droit d'auteur : la sensibilisation et l'éducation des internautes au droit d'auteur apparaissent comme une nécessité.

## ENSEIGNEMENTS CLÉS

- Les systèmes de reconnaissance de contenus sont globalement efficaces, y compris lorsque les œuvres mises en ligne sont fortement déformées.
- La multiplicité des systèmes existants, qu'ils soient développés en interne par les plateformes ou proposés par des prestataires, alourdit la tâche des ayants droit et ne permet pas d'uniformiser les pratiques.
- La finesse des systèmes est satisfaisante : si 6 % des internautes ont déjà reçu des messages de blocage de leurs contenus par les plateformes pour des raisons liées au respect du droit d'auteur, seulement 3 % des internautes ont contesté une mesure de blocage.
- 89 % des internautes ayant déjà vu un de leurs contenus mis en ligne faire l'objet d'une mesure de blocage ont compris la raison du blocage.

## RETROUVEZ LES RAPPORTS COMPLETS DES ÉTUDES SUR LE SITE DE L'HADOPI

[www.hadopi.fr](http://www.hadopi.fr)

### Derniers numéros parus :

- L'Essentiel #19 : L'offre OTT de contenus sportifs : un marché émergent en cours de structuration
- L'Essentiel #18 : Baromètre de la consommation de biens culturels dématérialisés 2019 : une consommation légale payante en hausse et des pratiques illicites en recul
- L'Essentiel #17 : 16 % des internautes utilisent régulièrement les réseaux sociaux pour accéder à des contenus audiovisuels et sportifs de manière illicite

Contactez-nous pour recevoir par courriel les avis de parution (10 numéros par an environ) : [presse@hadopi.fr](mailto:presse@hadopi.fr)

Directeur de la publication : Denis Rapone

Contributeurs à ce numéro : Didier Wang, Raphaël Berger, Marion Ozanne et Carla Menaldi sur la base du rapport sur "les outils de reconnaissance des contenus protégés par les plateformes de partage en ligne : état de l'art et propositions", réalisé par le CSPLA, l'Hadopi et le CNC

Chargé de publication : Martin de Beaune

Réalisation : [agencezebra.com](http://agencezebra.com)

ISSN 2646-831X

Hadopi

HAUTE AUTORITÉ POUR LA DIFFUSION DES ŒUVRES ET LA PROTECTION DES DROITS SUR INTERNET

4, rue du Texel - 75014 PARIS - [www.hadopi.fr](http://www.hadopi.fr)

Contact : [presse@hadopi.fr](mailto:presse@hadopi.fr) @InsidOpi