



Intelligence artificielle

Comment  
procéder  
à l'*opt-out*  
pour vos œuvres ?

---

Guide  
à l'attention  
des membres  
de l'@dagp

# Avant-propos

Face à l'essor ultra rapide des technologies liées à l'intelligence artificielle (IA), et en particulier celles relevant des IA dites génératives produisant des textes et des images, l'ADAGP, société d'auteur dans les arts visuels, vigilante et impliquée face à ces évolutions technologiques, a souhaité mettre à disposition de ses adhérents ce guide pratique.

*Comment procéder à l'opt-out pour vos œuvres ?* a pour but d'informer les artistes sur les possibilités techniques à leur disposition afin de protéger au mieux leurs œuvres en s'opposant à la fouille de données des IA sur internet.

*Mise à jour : février 2024*

# Sommaire

<b>Introduction</b>	<b>4</b>
<b>1. Je montre mes images sur mon site</b>	<b>6</b>
<b>Je change la configuration générale de mon site</b>	<b>7</b>
<i>Le fichier robots.txt</i>	7
<i>Le fichier ai.txt</i>	10
<i>Le fichier tdmrep.json</i>	12
<i>Réponse http dans l'entête (TDMRep)</i>	14
<i>Blocage des robots d'exploration à l'aide du fichier .htaccess</i>	15
<i>Blocage des robots à l'aide d'outils spécifiques de gestion du trafic sur le serveur web</i>	16
<b>Je change la configuration au niveau des pages de mon site</b>	<b>17</b>
<i>Protocole TDMRep</i>	18
<i>Balises nocache et noarchive</i>	18
<i>Balises noai et noimageai</i>	19
<b>J'exprime mon refus dans les conditions générales du service</b>	<b>20</b>
<b>J'interviens au niveau des images</b>	<b>21</b>
<b>2. Je partage mes images</b>	<b>22</b>
<b>J'interviens sur les métadonnées</b>	<b>23</b>
<b>J'utilise des outils visant à perturber le fonctionnement des IA</b>	<b>25</b>
<i>Je veux protéger mon style : Glaze</i>	25
<i>Je veux empoisonner les IA : NightShade</i>	26
<b>Autres ressources</b>	<b>27</b>

# Introduction

Les intelligences artificielles (IA) sont aujourd'hui capables de générer instantanément des images, sur la base d'une simple phrase décrivant le résultat attendu (*prompt*).

Pour pouvoir générer ces ersatz de création, les IA n'inventent rien : il leur faut plagier la création humaine, en s'inspirant de centaines de millions d'images et métadonnées aspirées automatiquement sur internet.

En 2019, le législateur européen, voulant favoriser le développement des IA européennes, a créé une nouvelle exception au droit d'auteur pour autoriser des « fouilles de données » : par principe, l'utilisation par les IA des images trouvées sur internet est autorisée, sans contrepartie financière, à moins que l'auteur ait exprimé son opposition (*opt-out*).

Ce mécanisme a cependant été imaginé sans que le législateur ait à l'esprit le cas des IA génératives, peu développées à l'époque mais qui viennent aujourd'hui menacer l'activité des auteurs professionnels. Cinq ans seulement après son adoption, le texte européen est déjà dépassé, juridiquement, techniquement et économiquement.

L'ADAGP milite depuis plusieurs mois pour que le cadre juridique soit révisé – ou à tout le moins aménagé – de manière à assurer le respect des droits des auteurs.



**Pour l'heure, et bien que cela ne suffise malheureusement pas à résoudre toutes les difficultés attachées au développement de ces IA génératives, l'ADAGP a publié une déclaration générale d'*opt-out* pour l'ensemble des auteurs dont elle gère les droits de reproduction.**

**En complément, elle recommande à ses membres, notamment ceux dont elle ne gère que les droits collectifs, de prendre des mesures pour reprendre le contrôle de leurs œuvres en signifiant leur opposition à la fouille de données.**

L'article L. 122-5-3 du code de la propriété intellectuelle, qui transpose en droit français l'article 4 de la directive de 2019 sur le droit d'auteur dans le marché unique numérique, prévoit que l'opposition doit être faite « de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne ».

**Il faut donc exprimer l'*opt-out* au moyen de procédés compréhensibles par les robots d'exploration des IA (voir l'encadré ci-dessous).**

Cela peut se faire de façon globale, par un paramétrage du serveur hébergeant le site web, par une intervention sur les pages du site ou par une intervention au niveau des images mises à disposition. Ce guide a pour vocation de décrire les différentes solutions disponibles.



### **Qu'est-ce qu'un robot d'exploration ?**

Les robots d'exploration (ou *crawler*) sont des programmes informatiques qui permettent d'explorer et analyser automatiquement des sites web en suivant des liens d'une page web à une autre.

Les robots exécutent les fonctions définies dans le code de leur programme de façon autonome : ils analysent les informations contenues dans les pages des sites web et vont explorer d'autres pages web à partir des liens qu'ils ont pu trouver dans les pages précédentes.

Il convient de distinguer deux types de robots d'exploration :

- D'une part **les robots des moteurs de recherche** (Google, Bing, etc.), qui vont collecter les contenus des sites (textes, images, etc.) et les métadonnées associées (titre, mots-clés, description, etc.) afin de pouvoir les inclure dans les résultats du moteur de recherche. Ce sont eux qui permettent le référencement des pages du site dans le moteur de recherche.
- D'autre part, les **robots des moteurs d'intelligence artificielle**, qui collectent les contenus publiés sur les pages des sites internet afin d'alimenter la base d'apprentissage du moteur d'intelligence artificielle.

Il arrive que les mêmes sociétés exploitent à la fois des robots destinés à alimenter les IA et des robots destinés à référencer les sites internet. Il faut donc être prudent dans l'utilisation des outils d'*opt-out* afin qu'un site ne se retrouve pas "déréféréncé" des moteurs de recherche alors que l'auteur voulait uniquement s'opposer à l'utilisation des œuvres pour entraîner les IA.

# 1.

## Je montre mes images sur mon site

Le site web peut être configuré de manière à faire obstacle à l'utilisation des contenus publiés à des fins d'entraînement des IA. Cela peut se faire au moyen d'un fichier de directives à destination des robots ou en déployant des mesures techniques permettant de leur bloquer l'accès au site.

Un fichier de directives pour les robots est un fichier au format texte contenant des consignes à destination des robots d'exploration leur indiquant ce qu'ils ont le droit de faire ou non sur un site. Ces fichiers sont enregistrés sur le serveur web, comme tous les autres fichiers du site. Ils sont pour la plupart placés à la racine du site.

Lorsqu'un robot d'exploration commence à moissonner un site, il va (normalement) commencer par lire le contenu du fichier de directives pour savoir ce qu'il est autorisé ou non à faire.

Les mesures techniques de blocage visent, elles, à empêcher purement et simplement l'accès aux robots d'exploration.

# Je change la configuration générale de mon site

## Le fichier robots.txt

Le fichier robots.txt est un standard reconnu par la plupart des robots d'exploration utilisés par les moteurs de recherche.

Il permet de préciser aux robots quels répertoires peuvent être indexés ou non.

Il peut également contenir des liens vers des pages du site que l'on souhaite faire référencer par les moteurs de recherche.

Certains robots d'exploration étant dédiés au moissonnage de contenus pour alimenter les IA (c'est le cas, par exemple, du robot de Google dédié à ses IA : Google Extended), il est possible d'utiliser le fichier robots.txt pour leur interdire d'aspirer les contenus du site. Il convient toutefois d'être vigilant et de ne pas interdire par erreur l'accès à un robot d'exploration dédié à un service de référencement car les conséquences seront lourdes : le site disparaîtra des résultats du moteur de recherche concerné !



## Mise en place de l'*opt-out* avec le fichier « robots.txt »

1) S'il n'existe pas déjà sur le serveur web, créer le fichier texte robots.txt sur l'ordinateur à l'aide d'un éditeur texte (par exemple : bloc note), ou récupérer le fichier existant sur le serveur web (*www.MonNomDeDomaine.fr/robots.txt*) à l'aide de votre logiciel FTP habituel afin de pouvoir le modifier dans l'éditeur de texte.

2) Ajouter des instructions à ce fichier afin de limiter les contenus auxquels le robot est autorisé à accéder, par exemple :

*Pour interdire l'exploration de l'ensemble du site au robot NomDuRobot :*

```
User-agent : NomDuRobot  
Disallow : /
```

*Pour interdire l'exploration de « répertoire-1 » et du sous-répertoire « sous-repertoire-D » :*

```
User-agent : NomDuRobot  
Disallow : /repertoire-1  
Disallow : /repertoire-2/sous-repertoire-D
```

*Pour interdire l'exploration d'une page (URL spécifique) :*

```
User-agent : NomDuRobot  
Disallow : /page.html
```

Il existe un large éventail d'instructions que vous pouvez utiliser pour signaler aux robots logiciel que vous vous opposez à la fouille de données. Les professionnels du web peuvent également vous y aider.

Dans les exemples précédents, les commandes s'adressent à un robot spécifique (NomDuRobot). Il faut donc répéter les commandes pour chacun des robots auxquels on veut interdire l'accès.

NB : il est également possible de donner des commandes pour l'ensemble des robots en indiquant User-Agent: \* mais dans ce cas, vous interdisez également l'accès aux robots d'exploration des moteurs de recherche ce qui nuira au référencement de votre site.



**L'ADAGP recommande d'ajouter les commandes suivantes à votre fichier robots.txt pour indiquer votre opposition à la fouille des données par les principaux robots de moteurs d'IA connus (les lignes commençant par le symbole # sont des commentaires) :**

```
#Common Crawl dataset : utilisé par GPT et d'autres moteurs IA  
User-agent: CCBot  
Disallow: /  
  
#img2dataset: utilisé par certains robots pour moissonner les images  
User-agent: img2dataset  
Disallow: /  
  
#GPTBot : le robot de OpenAI (ChatGPT)  
User-agent: GPTBot  
Disallow: /  
  
#robots des moteurs d'IA générative de Google (Bard, Vertex AI)  
User-agent: Google-Extended  
Disallow: /
```

```
#anthropic-ai: utilisé pour certains moteurs IA tels que Claude
User-agent: anthropic-ai
Disallow: /

#Bytespider: utilisé par ByteDance (TikTok) pour ses outils d'IA tels que Doubao
User-agent: Bytespider
Disallow: /

#FacebookBot: utilisé par Meta pour entraîner ses moteurs de reconnaissance de parole
User-agent: FacebookBot
Disallow: /

#Omgili: utilisé par Webz.io pour constituer une base de données vendue notamment
#à des moteurs IA
User-agent: omgili
Disallow: /
```

La difficulté réside dans le fait que cette liste n'est pas exhaustive, le fichier doit être mis à jour régulièrement pour prendre en compte l'arrivée de nouveaux robots. Le site <https://darkvisitors.com/> propose une liste de user-agent connus pour alimenter des moteurs d'IA, et envoie des notifications par courriel aux utilisateurs inscrits pour signaler l'arrivée d'un nouveau robot dans la liste.

**3) Une fois les modifications enregistrées, télécharger le fichier robots.txt** sur le serveur à la racine du site web à l'aide du client FTP habituel.

**4) Pour vérifier que votre fichier robots.txt modifié a bien été enregistré sur votre site, vous pouvez l'ouvrir dans votre navigateur.** Par exemple si votre site a pour domaine *www.MonNomDeDomaine.fr*, l'URL à saisir dans le navigateur est : *www.MonNomDeDomaine.fr/robots.txt*

Il convient de noter que tous les robots d'exploration ne suivent pas les directives du fichier robots.txt donc il ne constitue pas une garantie que le site sera épargné de toute fouille de données par des moteurs d'IA et n'est pas rétroactif pour les images qui auraient été identifiées par un robot préalablement à sa mise en place.



**Pour en savoir plus sur le fichier robots.txt**

<https://developers.google.com/search/docs/crawling-indexing/robots/create-robots-txt?hl=fr>

# Le fichier ai.txt

Le fichier ai.txt est une initiative portée par la société Spawning.ai, qui cherche à rendre les moteurs d'IA plus responsables.

Tout comme le fichier robots.txt, il permet de préciser quelles sont les données dont vous autorisez ou non l'utilisation par les moteurs d'intelligence artificielle.

Les principales différences avec le fichier robots.txt sont les suivantes :

- **Le fichier ai.txt s'adresse uniquement aux moteurs d'IA, il ne concerne pas les moteurs de recherche** (indexation, référencement des sites web)
- **Les directives qui sont enregistrées dans le fichier ai.txt ne sont pas lues par le robot au moment où il explore les pages du site, mais par le moteur d'IA (via spawning.ai) au moment où il télécharge les contenus du site** (préalablement référencés par un robot d'exploration dans une base de données) pour alimenter sa base d'apprentissage.

En théorie, cela signifie que même si l'adresse d'une image de votre site a déjà été référencée par un robot d'exploration dans une base de données comme LAION 5B (il s'agit d'une très grande base de données contenant les liens vers plusieurs millions d'images, couramment utilisée par les moteurs d'IA pour alimenter leur base d'apprentissage), au moment où le moteur d'IA va télécharger l'image dont il a trouvé l'adresse dans la base LAION 5B, il va interroger la base de spawning.ai, qui pourra lui répondre en fonction des informations trouvées dans le fichier ai.txt s'il est autorisé à récupérer cette image, alors que le fichier robots.txt serait inefficace une fois l'image déjà référencée après le passage du robot d'exploration.

Mais ce mécanisme ne fonctionne donc qu'avec les moteurs d'IA respectueux qui prennent la peine d'interroger spawning.ai (à ce jour uniquement stability.ai et Hugging Face).



## Mise en place de l'*opt-out* avec le fichier « ai.txt »

1) Créer le fichier ai.txt

Le lien suivant vous donne accès un générateur (en anglais) qui crée un fichier ai.txt selon les paramètres que vous précisez : <https://site.spawning.ai/spawning-ai-txt#ai-text-generator>

2) Une fois le fichier créé, télécharger le fichier ai.txt sur le serveur à la racine du site web à l'aide du client FTP habituel.



**Pour en savoir plus sur le fichier ai.txt**

<https://site.spawning.ai/spawning-ai-txt>

# Le fichier tdmrep.json

Le fichier tdmrep.json est un des modes d'expression de l'*opt-out* proposé dans le projet de protocole TDMRep porté par le W3C (*World Wide Web Consortium*), l'organisme qui définit les normes de l'Internet.

Le protocole TDMRep (*Text and Data Mining Reservation Protocol*) est le protocole permettant d'exprimer ses réserves à la fouille de données que le W3C a élaboré en réponse à l'obligation par la directive européenne de permettre l'*opt-out* dans un format lisible par les machines.

Le protocole TDMRep prévoit que l'*opt-out* peut être effectué à différents niveaux : globalement au niveau du site, dans le fichier tdmrep.json, ou au niveau du serveur web (voir ci-après TDM Header Field in http Responses), ou encore au niveau de chaque page du site (voir ci-après TDM Metadata in HTML).

À l'instar du fichier robots.txt, le fichier tdmrep.json permet de donner des consignes aux robots d'exploration sur les fichiers auxquels ils ont accès ou non, mais il ne concerne que les robots d'IA et s'attache exclusivement à traiter la question de l'*opt-out*.

Ce protocole, toujours en cours d'élaboration, n'a pas encore été adopté à ce jour, et n'a donc aucun effet sur les robots d'exploration. Après son adoption, il pourrait cependant devenir un standard étant donné l'importance du W3C sur Internet.



## Mise en place de l'*opt-out* avec le fichier « tdmrep.json »

- 1) Créer le fichier tdmrep.json sur l'ordinateur à l'aide d'un éditeur texte (par exemple : bloc note).
- 2) Ajouter des instructions à ce fichier afin de préciser la portée de l'*opt-out* :

Pour préciser que l'opt-out porte sur tous les fichiers du serveur :

```
[
  {
    «location»: «/»,
    «tdm-reservation»: 1
  }
]
```

L'exemple suivant indique un opt-out sur l'ensemble du sous-répertoire `repertoire-b/html`, avec la présence d'un fichier précisant la politique d'opt-out détaillée dans un fichier à part, et un opt-out sur tous les fichiers images de type `jpg` dans le sous répertoire `repertoire-b/image`

```
[
  {
    «location»: «/repertoire-b/html/»,
    «tdm-reservation»: 1,
    «tdm-policy»:»https://www.monnomdedomaine/policies/policy.json«
  },
  {
    «location»: «/repertoire-b/images/*.jpg»,
    «tdm-reservation»: 0
  }
]
```

Il existe un large éventail d'instructions et d'expressions régulières permettant d'indiquer de façon générique la portée de l'opt-out par rapport à un ensemble de règles (par exemple concerne les fichiers dont le nom contient "IMA", etc.).

La syntaxe des instructions est décrite en détail sur le site du W3C.

**3) Une fois les modifications enregistrées, télécharger le fichier `tdmrep.json` sur le serveur web** dans le dossier à l'aide du client FTP habituel.



**Pour en savoir plus sur le protocole TDMRep**

<https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20220216/>

# Réponse http dans l'entête (TDMRep)

Le protocole TDMRep prévoit également que la notification de l'*opt-out* puisse être envoyée directement par le serveur web dans les données d'entête http qu'envoie le serveur en réponse à une demande (par exemple quand un navigateur demande à charger une page du site).

Il s'agit de données techniques, comme la date-heure, le type de fichier, placées au début des données envoyées par le serveur web à chaque envoi.

NB : ce protocole est déjà pris en compte par Spawning.ai, donc les moteurs d'IA qui interrogent spawning.ai avant de récupérer un fichier sauront qu'ils ne peuvent pas le faire sur les site ayant mis en place ce protocole.



## Mise en place de l'*opt-out* TDMRep dans les données d'entête

NB : il faut pour cela avoir accès à la configuration du serveur web qui héberge le site.

Il suffit alors de configurer le serveur web pour qu'il ajoute systématiquement l'instruction *tdm-reservation* : 1 dans les données d'entête qu'il envoie à chaque fois qu'il répond à une demande, de telle sorte que les données d'entête ressemblent à ceci :

```
HTTP/1.1 200 OK
Date: Wed, 14 Jul 2021 12:07:48 GMT
Content-type: image/jpg
tdm-reservation: 1
```



**Pour en savoir plus sur le protocole TDMRep dans les réponses http**

<https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20220216/#sec-tdm-header>

# Blocage des robots d'exploration à l'aide du fichier .htaccess

Les fichiers .htaccess sont des fichiers qui modifient la configuration du serveur web (pour les serveurs Apache) au niveau de chaque répertoire où ils sont placés.

Le fichier .htaccess contient des directives de configuration pour le serveur web qui s'appliquent au répertoire où il est placé ainsi qu'à tous ses sous-répertoires. De cette manière, il est possible d'essayer de bloquer l'accès aux robots d'exploration aux fichiers d'un répertoire et de ses sous-répertoires.



## Mise en place du blocage des robots avec le fichier « .htaccess »

1) S'il n'existe pas déjà sur le serveur, **créer le fichier .htaccess sur l'ordinateur à l'aide d'un éditeur texte** (par exemple : bloc note) ou récupérer le fichier existant sur le serveur à l'aide de votre logiciel FTP habituel afin de pouvoir le modifier dans l'éditeur de texte.

2) **Ajouter des instructions suivantes dans le fichier :**

```
RewriteEngine On
RewriteCond %{HTTP_USER_AGENT} (anthropic-ai|Bytespider|CCBot|ChatGPT-User|FacebookBot|GPTBot|Omgilibot)
[NC]
RewriteRule ^ - [F]
```

*Suivant l'exemple ci-dessus, si le serveur détecte que la machine qui demande à lire un fichier est un des robots listés (anthropic-ai ou Bytespider ou CCBot ou ChatGPT-User ou FacebookBot ou GPTBot Omgilibot), il envoie un code de statut 403 Forbidden (interdit) à la place du fichier demandé.*

Il existe un large éventail d'instructions et d'expressions régulières permettant par exemple de remplacer tous les fichiers images par un autre fichier (votre logo par exemple).

3) **Une fois les modifications enregistrées, télécharger le fichier .htaccess sur le serveur web**, à la racine du site ou dans le dossier dans lequel sont les fichiers qu'on ne veut pas que le robot puisse lire à l'aide du client FTP habituel.

# Blocage des robots à l'aide d'outils spécifiques de gestion du trafic sur le serveur web

D'autres outils réservés aux informaticiens - car devant être installés sur la machine où se trouve le serveur web - permettent également de tenter de bloquer les robots, soit en s'appuyant sur des listes prédéfinies de noms d'agent (ceux que l'on renseigne dans les fichiers robots.txt par exemple), ou en bloquant toutes les requêtes provenant d'adresses IP connues pour être celles de robots (mais ces listes sont difficiles à maintenir car les robots utilisent de nombreuses adresses IP qui changent souvent).

On peut par exemple citer **le projet Nginx ultimate bad bot blocker** dont l'objectif, plus large que l'*opt-out* est d'empêcher l'accès à toute une série de robots qui parasitent les sites.

D'autres outils de sécurité, qui peuvent se trouver au niveau des *firewall* essaient de détecter de façon dynamique les robots en analysant le trafic (en repérant les adresses IP qui envoient beaucoup de requêtes rapprochées) pour les bloquer.

C'est le cas de **l'outil Kuduru** développé par la société Spawning.ai (à l'initiative de ai.txt). Le principe de Kuduru est que tous les logiciels Kuduru installés sur des serveurs web partagent en temps réel leurs informations sur les adresses IP qui font beaucoup de requêtes afin de détecter plus efficacement les adresses IP utilisées par les robots et de bloquer les adresses IP des robots ainsi identifiées.



**Pour en savoir plus sur Kuduru (en anglais)**

<https://kudurru.ai/>

# Je change la configuration au niveau des pages de mon site

De la même façon qu'on peut donner des consignes générales au niveau du site pour notifier aux robots d'exploration son opposition à la fouille de données, il est possible d'indiquer au niveau de chaque page son refus de voir les données utilisées par un moteur d'intelligence artificielle en y intégrant des balises à l'intérieur du code HTML des pages du site. Ces informations, qui ne seront pas visibles par un humain dans la page web affichée, seront lues par la machine quand elle récupérera la page.

Là encore, il n'existe pas encore de standard universellement adopté, donc pas de garantie que ces consignes soient respectées.

Il existe néanmoins plusieurs possibilités techniques (qui peuvent être cumulées pour mettre toutes les chances de son côté) pour signifier votre refus.

## Protocole TDMRep

Le protocole TDM reservation prévoit également que l'on puisse signifier son opposition à la récupération des données par les moteurs d'IA au niveau de l'entête des pages HTML.

La balise tdm-reservation indique son opposition à l'utilisation des contenus de la page elle-même pour l'entraînement des IA, et la balise noimageai indique son opposition à l'utilisation des fichiers images affichés dans la page pour l'entraînement des IA.

Il convient d'ajouter dans la section entête de la page (entre les balises <HEAD> </HEAD>) le code HTML suivant :

```
<meta name="tdm-reservation" content="1"
```



### Pour en savoir plus

<https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20220216/#sec-tdm-html-meta>

## Balises nocache et noarchive

Ces balises sont habituellement utilisées pour signifier aux robots d'exploration des moteurs de recherche qu'ils ne doivent pas garder les données recueillies en cache pour la première, et qu'ils ne doivent pas récupérer du tout de données pour la seconde.

En s'opposant au fait que le robot puisse télécharger l'ensemble des données de la page pour les garder en archive, on lui interdit également de les récupérer pour alimenter une base d'IA. Ces balises sont bien prises en compte par OpenAI.

Il convient d'ajouter dans la section entête de la page (entre les balises <HEAD> </HEAD>) le code HTML suivant :

```
<meta name="robots" content="noarchive, nocache">
```

# Balises noai et noimageai

Ces balises ont été introduites par DeviantArt pour signifier spécifiquement aux robots d'IA qu'on ne souhaite pas que le contenu de la page soit utilisé pour entraîner des intelligences artificielles avec un formalisme identique aux balises nocache et noarchive.

La balise noai indique son opposition à l'utilisation des contenus de la page elle-même pour l'entraînement des IA, et la balise noimageai indique son opposition à l'utilisation des fichiers images affichés dans la page pour l'entraînement des IA.

Il convient d'ajouter dans la section entête de la page (entre les balises <HEAD> </HEAD>) le code HTML suivant :

```
<meta name="robots" content="noai, noimageai">
```

La plupart des solutions de gestion de contenus des sites web proposent un paramétrage ou des *plug-in* permettant d'insérer automatiquement ces balises sur toutes les pages, des tutoriels sont disponibles en ligne pour les différents gestionnaires de contenus.

Il est également possible pour ceux qui ont accès au paramétrage du serveur web de configurer celui-ci pour qu'il intègre automatiquement ces balises dans les réponses http envoyées.



## Pour en savoir plus sur les balises meta robots

<https://robots-txt.com/meta-robots>

<https://robots-txt.com/x-robots-tag>

<https://developers.google.com/search/docs/crawling-indexing/robots-meta-tag?hl=fr>

# J'exprime mon refus dans les conditions générales du service

L'article R. 122-8 du code de la propriété intellectuelle prévoit que l'*opt-out* peut être exprimé « au moyen de procédés lisibles par machine, y compris des métadonnées, et par le recours à des conditions générales d'utilisation d'un site internet ou d'un service ».

Il est donc en principe possible de formuler l'*opt-out* au sein des conditions générales d'utilisation du site. Afin d'accroître les chances que les robots d'exploration en tiennent compte, il est possible de compléter le texte d'une balise conforme au protocole TDMRep (*voir page 18*).

Voici un exemple de texte qui pourrait être placé dans les conditions générales d'utilisation du site :

*« Les œuvres publiées sur ce site web sont protégés par le droit d'auteur et ne peuvent en aucun cas être réutilisées sans l'autorisation expresse de leur auteur.*

*Les images présentes sur ce site, ainsi que les données qui s'y rapportent, tels que le titre ou le nom de l'auteur, ne peuvent en aucune façon faire l'objet, en tout ou partie, de reproductions, sous quelque forme que ce soit, réalisées en vue de fouilles de textes et de données, à l'exception de celles menées à des fins exclusives de recherche scientifique par un organisme visé à l'article L. 122-5-3, II du code de la propriété intellectuelle. Cette interdiction, exprimée dans le cadre du dispositif prévu à l'article R. 122-8 du code de la propriété intellectuelle, s'applique en particulier à l'utilisation des œuvres pour alimenter ou entraîner des dispositifs d'intelligence artificielle conçus ou adaptés pour générer des créations, telles que des images ou des contenus audiovisuels, destinées à une diffusion publique.*

*Cette interdiction est également exprimée, conformément au protocole TDMRep accessible à l'adresse [www.w3.org/2022/tdmrep](http://www.w3.org/2022/tdmrep), de la manière suivante : <TDM-RESERVATION : 1>.*

*Pour toute demande de reproduction ou de communication au public d'une œuvre, contacter l'ADAGP ([www.adagp.fr](http://www.adagp.fr)). »*

Le texte en français peut être doublé d'une version en anglais :

*« The works published on this website are protected by copyright and may not, under any circumstances, be reused without the express permission of their author.*

*The images on this site, along with associated data such as title or author's name, may not in any way be subject, in whole or in part, to reproductions in any form, carried out for the purpose of text and data mining, except for those conducted exclusively for scientific research purposes by an entity referred to in Article L. 122-5-3, II of the Intellectual Property Code. This prohibition, as expressed within the framework provided in Article R. 122-8 of the Intellectual Property Code, particularly applies to the use of works to feed or train artificial intelligence systems designed or adapted to generate creations, such as images or audiovisual content, intended for public dissemination.*

*This prohibition is also expressed, in accordance with the TDMRep protocol accessible at [www.w3.org/2022/tdmrep](http://www.w3.org/2022/tdmrep), as follows: <TDM-RESERVATION: 1>.*

*For any request for reproduction or public communication of a work, please contact the ADAGP ([www.adagp.fr](http://www.adagp.fr)). »*

## J'interviens au niveau des images

Il est également possible d'**intervenir directement au niveau des images**, ce qui permet aussi d'agir les rediffusions sur d'autres sites que celui de l'auteur.

*Les techniques pouvant être mises en œuvre sont décrites à la section suivante.*

## 2. Je partage mes images

Bien souvent, les images des œuvres sont diffusées en dehors du site de l'auteur, notamment sur les réseaux sociaux, sur les sites de musées ou sur les sites de presse, de sorte que celui-ci n'a que rarement la main sur la configuration du site de diffusion.



**C'est pour ce cas de figure, notamment, que l'ADAGP a procédé à une déclaration générale d'*opt-out* pour les œuvres des artistes dont elle gère le droit de reproduction et qu'elle intègre désormais dans ses contrats avec les diffuseurs des clauses interdisant la fouille de données ([consulter la déclaration](#)).**

Les auteurs peuvent également intervenir directement sur les images, en utilisant les métadonnées ou en recourant à des techniques permettant de perturber le fonctionnement des IA.

# J'interviens sur les métadonnées

Le standard IPTC (*International Press Telecommunications Council*) est une norme qui définit comment des métadonnées peuvent être ajoutées à des fichiers images.

Ceci permet d'ajouter des informations à un fichier image telles que le titre, l'auteur de l'image, le copyright, des informations relatives au droit d'auteur, etc., de façon structurée afin qu'elles puissent être lues par la plupart des logiciels et machines (par exemple Google Image affiche désormais les informations trouvées dans les champs IPTC des images indexées).

En octobre 2023, la nouvelle version de la norme IPTC 2023.1 a adopté l'ajout de 2 nouveaux champs dans les métadonnées IPTC :

- **Data Mining** permet de préciser ce que les robots ont le droit de faire avec l'image, parmi les valeurs suivantes :

DMI-UNSPECIFIED	Non précisé, pas d'interdiction
DMI-ALLOWED	Autorisé
DMI-PROHIBITED-AIMLTRAINING	Interdit pour les moteurs IA
DMI-PROHIBITED-GENAIMLTRAINING	Interdit pour les moteurs d'IA générative
DMI-PROHIBITED-EXCEPTSEARCHENGINEINDEXING	Interdit, sauf pour les moteurs de recherche
DMI-PROHIBITED	Interdit
DMI-PROHIBITED-SEECONSTRAINT	Interdit, voir les infos dans le champ <i>other constraints</i>
DMI-PROHIBITED-SEEEMBEDDEDRIGHTSEXPR	Interdit, voir les infos complémentaires dans le champ <i>Embedded Encoded Rights Expression</i>
DMI-PROHIBITED-SEELINKEDRIGHTSEXPR	Interdit, voir les infos complémentaires dans le champ <i>Linked Encoded Rights Expression</i>

- **Other Constraints** est un champ de texte libre qui permet de donner des informations complémentaires en langage compréhensibles par un humain, par exemple : « L'entraînement des moteurs d'IA générative est autorisé seulement dans un cadre éducatif ».

L'intérêt d'intégrer les informations d'*opt-out* dans le fichier image publié est que comme ces métadonnées sont encapsulées dans le fichier avec l'image, si le fichier est copié et publié sur un autre site, ces informations d'*opt-out* restent présentes dans le fichier, contrairement à toutes les directives vues précédemment qui ne s'appliquent qu'aux fichiers trouvés sur le site ou sur les pages web où sont placées ces directives.

## Ajout des données IPTC aux fichiers images

La plupart des logiciels de traitement d'images tels que Photoshop permettent de modifier les données IPTC d'un fichier image, éventuellement avec un traitement automatique de lots d'images.



**Pour en savoir plus sur les champs IPTC relatifs à l'*opt-out* (en anglais)**

<https://iptc.org/news/exclude-images-from-generative-ai-iptc-photo-metadata-standard-2023-1/>

# J'utilise des outils visant à perturber le fonctionnement des IA

## Je veux protéger mon style : Glaze

Glaze est un logiciel gratuit développé par une équipe de chercheurs de l'université de Chicago qui ne cherche pas à empêcher que les robots puissent récupérer une image, mais à faire en sorte que l'image récupérée soit mal analysée en y ajoutant des altérations presque imperceptibles à l'œil nu mais qui vont tromper le moteur d'apprentissage de l'intelligence artificielle et le conduire à générer des images de mauvaise qualité, en particulier dans le cas de la génération d'images IA "à la manière de".

Glaze fonctionne comme un logiciel de traitement d'images qui va appliquer des déformations calculées spécifiquement pour chaque image. Il convient donc de "*glazifier*" les images avant de les publier ou de les distribuer.

NB : Glaze a fait ses preuves avec les IA génératives actuelles mais ne garantit pas son efficacité avec les futur moteurs d'IA à venir.

**Présentation de Glaze par Frédéric Maupomé, auteur de bande dessinée, adhérent de l'ADAGP et président d'honneur de la Ligue des auteurs professionnels :**



(Vidéo disponible sur la chaîne YouTube de l'ADAGP)



**Pour en savoir plus sur Glaze (en anglais)**

<https://glaze.cs.uchicago.edu/>

## Je veux empoisonner les IA : NightShade

NightShade est un outil développé par la même équipe de chercheurs que Glaze dont la première version vient d'être livrée. NightShade repose sur le même principe que Glaze, introduire des déformations dans l'image, mais avec un objectif clairement plus offensif (et des déformations plus visibles également) puisqu'il cherche à tromper les moteurs d'IA en empoisonnant leur apprentissage, ou, *a minima*, à rendre le coût d'apprentissage beaucoup plus élevé.

**Présentation de NightShade par Frédéric Maupomé, auteur de bande dessinée, adhérent de l'ADAGP et président d'honneur de la Ligue des auteurs professionnels :**



(Vidéo disponible sur la chaîne YouTube de l'ADAGP)



**Pour en savoir plus sur NightShade (en anglais)**  
<https://nightshade.cs.uchicago.edu/whatis.html>

# Autres ressources



## Outils

### Have I been trained?

HavelBeenTrained est un outil permettant de rechercher si votre travail a été utilisé dans le jeu de données LAION 5B pour entraîner les générateurs d'images. Il permet de retirer les images des IA respectant des moteurs le principe de désinscription en vous permettant d'introduire vos œuvres à un registre « Do not train Registry ».



**En savoir plus sur HavelBeenTrained (en anglais)**

<https://haveibeen trained.com/>

### Dark Visitors

Ce site recense les robots d'exploration utilisés par les IA et vous permet d'actualiser la liste des robots interdits de fouille sur votre site. Vous pouvez choisir de recevoir des alertes par courriel pour être notifié des nouveaux robots recensés.



**En savoir plus sur Dark Visitors (en anglais)**

<https://darkvisitors.com/>



## Textes légaux

Article L. 122-5-3 du code de la propriété intellectuelle

[https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000044363192](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000044363192)

Article R. 122-28 du code de la propriété intellectuelle

[https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000045960687](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000045960687)

Directive 2019/790 sur le droit d'auteur et les droits voisins dans le marché unique numérique (DAMUN)

<https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32019L0790>

# En savoir plus

Les bureaux de l'ADAGP sont ouverts  
du lundi au vendredi de 9 h15 à 13 h  
et de 14 h15 à 18 h.

Pour en savoir plus sur les actions  
de défense des droits d'auteur,  
les évènements organisés autour  
de l'actualité des arts visuels, les appels  
à candidature de l'action culturelle,  
les avantages adhérents, etc.

↘ [adagp.fr](https://adagp.fr)

Pour suivre l'ADAGP sur les réseaux sociaux :



[@adagp.fr](https://adagp.fr) | [adagp\\_](https://adagp.fr) | [@Adagp\\_](https://adagp.fr) | [ADAGP\\_](https://adagp.fr)

[#PourledroitdesArtistes](https://adagp.fr)

# @dagp

Pour le droit des artistes

11 rue Duguay-Trouin - 75006 PARIS

T+33 (0)1 43 59 09 79

[adagp@adagp.fr](mailto:adagp@adagp.fr)

Société civile à capital variable

RCS Paris D 339 330 722

Février 2024