

Le droit international de la culture au défi des médias synthétiques

International cultural law to the test of synthetic media

William LETRÔNE¹

¹ CNRS, Laboratoire DCS, Université de Nantes, France, William.Letrone@univ-nantes.fr

RÉSUMÉ. Les médias synthétiques produits par intelligence artificielle générative (IAG) ont rapidement inondé le Web, faisant planer un risque de préjudices culturels auquel il convient de remédier. Or le sujet des préjudices culturels pouvant découler de la dissémination de médias synthétiques n'a été que partiellement traité dans la littérature juridique. Ce travail vise à combler cette lacune en proposant une exploration des implications juridiques des médias synthétiques, dès lors qu'ils rendent plus probable la survenance de préjudices culturels. A cette fin, le concept-clé de préjudice culturel ainsi que le rôle occupé par les droits culturels et le principe de diversité culturelle dans sa détermination feront l'objet d'une première analyse. Seront ensuite abordés les manières dont les médias synthétiques peuvent entraîner des préjudices culturels et les conséquences juridiques potentielles. Il sera ainsi démontré que bien que le droit international prévoie des mécanismes permettant la prévention des préjudices culturels, peu de moyens permettent une réelle prise en compte des spécificités des médias synthétiques culturellement préjudiciables.

ABSTRACT. Synthetic media produced by generative artificial intelligence (GAI) tools are flooding the Web, causing risks of cultural harm that should be addressed. Yet, the question of cultural harm resulting from the dissemination of synthetic media has only been partially addressed in the legal literature. This article aims to fill this gap, by exploring the legal implications of culturally harmful synthetic media. To that end, this article analyzes the key concept of cultural harm and the role played by cultural rights and the principle of cultural diversity in its characterization. The ways in which synthetic media can cause cultural harm and the potential legal consequences are then discussed. It will be shown that, while international law provides mechanisms for preventing cultural harm, there are few means to really take into account the specificities of culturally harmful synthetic media.

MOTS-CLÉS. Contenus Synthétiques, Préjudice Culturel, Droit International, Désordres Informationnels, Droits Fondamentaux, Liberté D'expression.

KEYWORDS. Synthetic Content, Cultural Harm, International Law, Information Disorders, Human rights, Free Speech.

1. Introduction

La recommandation sur l'éthique de l'intelligence artificielle adoptée par l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) le 23 novembre 2021 encourage les États membres à:

« (...) examiner et à traiter les répercussions des systèmes d'IA sur la culture ». [UNE 21, P.34, §95].

Citant l'exemple des applications de traitement de langage, l'UNESCO alerte notamment sur le risque de disparition de dialectes peu ou pas représentés résultant de l'absence de nuances langagières dans les sorties de ces systèmes, [UNE 21, P.34, §95], et souligne l'importance de l'accès à une pluralité d'expressions culturelles dans le contexte numérique, et de la protection de la propriété intellectuelle des créateurs dont les œuvres sont utilisées pour développer des intelligences artificielles (IA) [UNE 21, P.35, §99].

Les exigences contenues dans les lignes directrices en matière d'éthique pour une IA digne de confiance présentées en 2019 par le groupe d'experts de haut niveau sur l'IA réuni par la Commission européenne demandent quant à elles, à ce que les systèmes d'IA soient mis au point de manière à

respecter, protéger et servir le sentiment d'identité personnel et culturel des êtres humains [GRO 19, P.13].

Ces recommandations et lignes directrices trouvent un écho tout particulier dans le contexte des intelligences artificielles génératives (IAG), ces outils bouleversant l'environnement numérique culturel, entendu ici largement, comme l'ensemble des contenus informationnels ayant une valeur culturelle, ainsi que les acteurs, outils et méthodes de distribution de ces contenus.

Révélée au grand public en 2018, l'IAG, est un type d'IA spécialisé dans la génération d'information reposant sur une classe d'algorithme appelés réseaux antagonistes génératifs (GAN). La Commission nationale informatique et liberté (CNIL) définit l'IAG comme « un système capable de créer du texte, des images ou d'autres contenus (musique, vidéo, voix, etc.) à partir d'une instruction d'un utilisateur humain» [CNI 23]. Un système d'IAG comprend généralement une interface prenant la forme d'un agent conversationnel avec lequel l'utilisateur va interagir afin de faire générer du contenu (ou média) dit « synthétique » au modèle sous-jacent.

Le ton assuré employé dans leurs sorties textuelles et le réalisme de leurs sorties visuelles font qu'il est aisément de leur faire confiance [WAC 24]. Pour autant, les IAG ne sont pas infaillibles. De par leur facilité d'accès et la vraisemblance de leurs productions, elles occupent un rôle croissant dans la formation des désordres informationnels -entendus comme des phénomènes numériques liés à l'expression pouvant déboucher sur des préjuges individuels et collectifs divers [SPI 24], [BTE 23], [SED 21], [GOL 23], [JAI 24].

Par essence, un système d'IA est une représentation d'un corpus symbolique qui contient par définition des biais. Puisqu'un système d'IA est dépourvu d'autonomie personnelle et de sens moral, seuls ses développeurs, et dans une moindre mesure, les utilisateurs, sont en capacité de prévenir les potentielles externalités négatives induites par les biais contenus dans les données qui ont été fournies lors de la phase d'apprentissage [DEM 17, p.31]. Or, plus la confiance accordée aux systèmes d'IAG est grande, plus ceux-ci sont en mesure de façonner l'imaginaire collectif au travers des contenus synthétiques qu'ils génèrent. Dans un contexte où le risque d'anthropomorphisation de ces outils est particulièrement élevé [JI 24], des préjuges liés à l'expression peuvent survenir, lesquels sont à même d'affecter l'environnement numérique culturel, et au-delà, de manière durable.

A l'heure où les États investissent massivement dans l'IAG, une réflexion sur l'impact juridique des contenus synthétiques générés par IAG doit être conduite de façon plus approfondie, car à ce jour, le sujet des préjuges culturels suscités par l'IAG n'a été que peu traité dans la littérature juridique. Ce travail propose donc de combler cette lacune, en explorant l'étendue des implications juridiques des contenus synthétiques, dès lors qu'ils rendent plus probable la survenance de préjuges culturels. Ce faisant, cet article reviendra d'abord sur le concept-clé de préjudice culturel, ainsi que le rôle occupé par les droits culturels et le principe de diversité culturelle dans la détermination d'un tel préjudice (2.), avant d'aborder plus en détails les manières dont les médias synthétiques peuvent entraîner des préjuges culturels ainsi que les conséquences juridiques d'une telle qualification (3.).

2. La caractérisation des préjuges culturels suscités par l'IAG

Caractériser les préjuges culturels suscités par l'IAG implique, dans un premier temps, à ce que soit définie la notion centrale de préjudice culturel (2.1) En tant que grand principe du droit international de la culture, le principe de diversité culturelle sera également exploré (2.2).

2.1. Le préjudice culturel en droit

2.1.1. Tentative de définition du préjudice culturel

Avant de définir la notion de préjudice culturel, il paraît opportun de commencer par définir la notion de préjudice telle que comprise en droit. Ainsi, en droit de la responsabilité civile, le préjudice désigne une atteinte subie par une personne par le fait d'un tiers, ou d'une chose. Le terme de « dommage » est parfois confondu avec celui de préjudice. Néanmoins, ces deux notions se rapportent à des choses différentes, [POU 05]. En effet là où le dommage désigne un fait objectif, le préjudice se rattache plutôt aux conséquences juridiques du dommage, sous forme de lésion d'un droit subjectif. [POU 05, P.203].

Un préjudice peut être individuel ou collectif, comme dans le cas des préjudices écologique ou, justement, culturels [BOS 22]. Puisqu'il s'apprécie en référence à un droit, le préjudice peut prendre de multiples formes. Il peut donc être moral, matériel, corporel, ou culturel, selon la nature de l'intérêt protégé. Ainsi, le préjudice culturel s'attache aux conséquences juridiques résultant d'un fait impactant la culture, au sens large. Le caractère collectif du préjudice culturel trouve sa source dans le fait que les individus tirent collectivement un bénéfice de la culture et de ses composantes. Il convient dès lors de définir le concept de « Culture ».

Bien que le concept de culture soit difficile à définir [RAU 11], une définition généralement acceptée est celle proposée dans la Déclaration universelle de l'UNESCO sur la diversité culturelle (Déclaration UNESCO), où la culture, au sens large, est définie comme :

« (...) l'ensemble des traits distinctifs spirituels et matériels, intellectuels et affectifs qui caractérisent une société ou un groupe social et qu'elle englobe, outre les arts et les lettres, les modes de vie, les façons de vivre ensemble, les systèmes de valeurs, les traditions et les croyances» [UNE 01].

La définition du terme « culture » étant particulièrement large, la notion de préjudice culturel s'en retrouve d'autant plus difficile à définir précisément. Ainsi, le préjudice culturel ;

« est protéiforme [et] se concentre sur la question de l'équilibre et du lien entre nature et culture, c'est-à-dire sur les rapports qu'entretient l'homme avec son environnement (créations culturelles, représentations culturelles, savoir-faire» [LAF 12].

Le préjudice culturel semble donc pouvoir émaner de comportements extrêmement variés, tant que ceux-ci affectent l'ensemble de ce qui constitue l'identité culturelle d'une société ou d'un groupe social. On parle alors de « patrimoine culturel ». Dans le cadre des médias synthétiques générés par IAG, les préjudice culturels semblent principalement se rattacher aux représentations véhiculées des cultures et des éléments qui les composent. Ces représentations peuvent avoir pour objet des éléments du patrimoine culturel matériel et immatériel. La reproduction de stéréotypes [GHO 24], la distorsion de faits historiques [WAC 24], et l'omission de subtilités culturelles [RET 24], [CHO 23] peuvent avoir des répercussions directes sur le patrimoine culturel immatériel, qui comprend ;

« les pratiques, représentations, expressions, connaissances et savoir-faire - ainsi que les instruments, objets, artefacts et espaces culturels qui leur sont associés - que les communautés, les groupes et, le cas échéant, les individus reconnaissent comme faisant partie de leur patrimoine culturel. » [UNE 03, ARTICLE 2].

Reste que le statut juridique du préjudice culturel demeure inachevé en droit français [BOS 22]. Ce constat ne doit pas pour autant nous décourager dans cette analyse. En effet, des cadres internationaux permettent déjà de discerner la forme que peut revêtir le préjudice culturel en droit. Le contexte de ce travail doit ainsi être perçu comme une opportunité d'avancer dans le sens d'une consolidation du statut juridique des atteintes aux intérêts culturels, lesquels peuvent être abordés au travers du prisme des droits culturels tels que reconnus en droit international.

2.1.2. Le rôle des droits culturels dans l'appréciation des préjugés culturels

Si les préjugés culturels se rapportent aux phénomènes liés à l'expression qui résultent de la méconnaissance des droits culturels, il convient alors d'explorer l'étendue des droits culturels, qui sont entendus par Meyer-Bisch et Bidault comme :

« les droits et libertés pour une personne, seule ou en commun, de choisir et d'exprimer son identité et d'accéder aux références culturelles comme à autant de ressources qui sont nécessaires à son processus d'identification, de communication et de création. » [MEY 98, P.23].

Développés tardivement, les droits culturels n'en demeurent pas moins des droits fondamentaux à part entière [MEY 08]. De la même manière que les droits économiques, sociaux, civils et politiques, les droits culturels sont directement connectés à la dignité humaine, axiome de tout droit fondamental [COM 09, P.1], [GRO 07]. La plupart se présentent d'ailleurs comme les émanations de la dimension culturelle de ces droits fondamentaux. En tant que tels, ils se trouvent régulièrement intégrés dans les dispositions des instruments de protection des droits de l'Homme, ce qui a pu participer à leur conférer une importance moindre. Les droits culturels sont donc longtemps restés dispersés dans plusieurs instruments juridiques [SHA 10]. Par exemple, les articles premiers du Pacte international relatif aux droits économiques sociaux et culturels (PIDESC) et du Pacte international relatif aux droits civils et politiques de 1966 (PIDCP), deux instruments majeurs ratifiés par la quasi-totalité des Etats, disposent tous deux;

« Tous les peuples ont le droit de disposer d'eux-mêmes. En vertu de ce droit, ils déterminent librement leur statut politique et assurent librement leur développement économique, social et culturel. » [AGN 66, ARTICLE 1§1], [AGN 66, ARTICLE 1§1].

Les articles 27 du PIDCP et de la Déclaration universelle des droits de l'Homme de 1948, (DUDH) viennent quant à eux respectivement établir une obligation de protection des droits des minorités ethniques, notamment dans le cadre du choix de leur culture, ainsi que le droit de toute personne de prendre part à la vie culturelle de la communauté. Ce dernier est d'ailleurs reproduit à l'article 15 du PIDESC. Compte-tenu de leurs liens étroits avec les droits fondamentaux traditionnels, les droits culturels peuvent également être déduits des dispositions d'instruments plus spécifiques, comme la Convention relative aux droits de l'enfant de 1989 et la Convention internationale sur l'élimination de toutes les formes de discrimination raciale de 1965.

Afin de clarifier leur contenu et d'en faciliter l'application, un groupe d'experts internationaux, le « Groupe de Fribourg », a entrepris de codifier les droits culturels au sein d'un unique instrument: la Déclaration de Fribourg de 2007 [GRO 07]. Bien que non-normative, la Déclaration de Fribourg rassemble des droits dont l'existence ne prête plus vraiment à débat [SHA 10]. Extraits de leur cadres conceptuels originels, les droits culturels gagnent ainsi en autonomie [MEY 98], [MEY 08].

La Déclaration de Fribourg énumère six droits, qui ont respectivement trait à l'identité et au patrimoine culturels, la référence à des communautés culturelles, l'accès et la participation à la vie culturelle, l'éducation et la formation, la communication et l'information, et la coopération culturelle, chacun comprenant plusieurs facettes. Comme pour les autres droits fondamentaux, les États sont tenus de respecter, de protéger et d'assurer les droits culturels [MEY 98].

2.2. Le préjudice culturel à la lumière du principe de diversité culturelle

2.2.1. Définition et fonction du principe de diversité culturelle

Les cultures constituent des témoins de l'histoire, elles sont les reflets des sociétés qui les ont forgées et doivent être protégées dans leur diversité. La paix mondiale repose d'ailleurs à bien des égards sur le respect et la préservation de cultures diverses. Pour ces raisons, plusieurs instruments juridiques mandatent les États à assurer la diversité culturelle [VAR 23]. Ce concept doit impérativement servir de

boussole aux gouvernements nationaux dans la lutte contre la détérioration de l'environnement numérique culturel pouvant résulter de l'usage massif et non régulé des systèmes d'IAG.

Le principe d'accès à la diversité culturelle s'impose comme élément central de la sauvegarde de la culture dans le contexte de l'IAG en ce qu'il incarne l'opposé de l'homogénéité culturelle, un paradigme particulièrement craint par les défenseurs de la cause culturelle [OTA 18 P.15.]. La diversité culturelle prend appui sur les droits culturels, dont la garantie effective est indispensable pour sa propre réalisation [UNE 01], [SHA 10, P.12]. Ces derniers perdant toute leur essence sans diversité culturelle, la sauvegarde de la diversité culturelle va de pair avec la garantie des droits culturels [SHA 10, P 11].

La diversité culturelle est donc un principe englobant, développé en réaction aux troubles liés à la mondialisation et la libéralisation subséquente des échanges commerciaux. C'est avec la Déclaration UNESCO que la diversité culturelle en tant que point focal du droit international de la culture, prend de l'ampleur sur le plan normatif [RAU 11]. La Déclaration UNESCO promeut la diversité culturelle en tant que patrimoine commun de l'humanité, et dispose en son article second, que la diversité des cultures ;

« (...) s'incarne dans l'originalité et la pluralité des identités qui caractérisent les groupes et les sociétés composant l'humanité » [UNE 01].

Il faudra néanmoins attendre l'adoption de la Convention de l'UNESCO sur la protection et la promotion de la diversité des expressions culturelles adoptée en 2005 (Convention UNESCO) pour que la diversité culturelle profite d'une protection juridique effective en droit international, une première [RAU 11], [KOL 05]. En effet, contrairement à la Déclaration, la Convention UNESCO est juridiquement contraignante. Entrée en vigueur en Mars 2007, elle constitue un instrument majeur du droit international de la culture [VAR 23]. Proposée à l'origine par la France et le Canada, la Convention UNESCO a pour objet la sauvegarde de l'originalité et de la pluralité des identités culturelles. Pour y parvenir, le texte instaure des obligations positives de moyen et de résultat à destination des Etats parties, dont la méconnaissance peut engendrer des préjudices culturels graves susceptibles d'engager leur responsabilité internationale. Parmi les nombreux objectifs de la Convention UNESCO figurent notamment la création des conditions permettant aux cultures de s'épanouir et d'interagir (Article 1 (b)), la stimulation de l'interculturalité (Article 1 (d)), et le respect de la diversité des expressions culturelles (Article 1 (e)).

À ce jour la Convention UNESCO rassemble 160 États Parties, dont les membres de l'Union européenne et de nombreux d'états du Moyen-Orient. Sa large reconnaissance lui confère néanmoins une autorité forte, si bien qu'elle pourrait conduire à la formation d'une norme coutumière allant dans le sens de la protection du principe de diversité culturelle en droit international, qui s'imposerait alors à tous les Etats, à l'exception des objecteurs persistants [OTA 18].

Une série de directives opérationnelles est venue éclairer l'application de la Convention UNESCO. Les Directives opérationnelles sur la mise en œuvre de la Convention dans l'environnement numérique de 2017 visent à guider les États dans la mise en œuvre de la Convention UNESCO dans l'environnement numérique [VAR 23, P.8].

2.2.2. L'accès équitable à une variété de cultures

Il est un aspect de la diversité culturelle qui présente un intérêt certain dans le cadre de cette étude ; celui de l'accès équitable à une diversité de cultures. L'article 6 de la Déclaration UNESCO vient en effet affirmer l'importance de ce principe ;

« Tout en assurant la libre circulation des idées par le mot et par l'image, il faut veiller à ce que toutes les cultures puissent s'exprimer et se faire connaître. La liberté d'expression, le pluralisme des médias, le multilinguisme, l'égalité d'accès aux expressions artistiques, au savoir scientifique et technologique - y compris sous la forme numérique - et la possibilité,

pour toutes les cultures, d'être présentes dans les moyens d'expression et de diffusion, sont les garants de la diversité culturelle. » [UN 01, ARTICLE 6].

L'accès équitable à une variété de cultures est érigé en tant que principe directeur de la Convention UNESCO;

« L'accès équitable à une gamme riche et diversifiée d'expressions culturelles provenant du monde entier et l'accès des cultures aux moyens d'expression et de diffusion constituent des éléments importants pour mettre en valeur la diversité culturelle et encourager la compréhension mutuelle. » [UNE 05, ARTICLE 7].

Le Comité des droits économiques, sociaux et culturels le définit à son tour, dans son Observation générale n°21 comme;

« le droit de chacun de rechercher, de recevoir et de partager des informations sur toutes les manifestations culturelles dans la langue de son choix et l'accès des communautés aux moyens d'expression et de diffusion » [COM 09, P.5, §16].

L'accès équitable à une variété de cultures et le principe de diversité culturelle sont intrinsèquement liés, en ce qu'il nécessite d'abord à ce que la diversité culturelle soit assurée [UNE 05, ARTICLE 2, PRINCIPE 7]. Le droit d'accès à une diversité culturelle constitue alors un prérequis pour la jouissance des autres droits culturels, particulièrement ceux relatifs à la formation de l'identité culturelle [VAR 23, P.4], [COM 09, P.5§16].

2.2.3. Diversité culturelle dans le cadre de l'IAG

La mise en œuvre des principes d'accès à une diversité culturelle a donné lieu à l'adoption de politiques nationales et régionales variées, visant à promouvoir le contenu culturel national ainsi que les cultures autochtones [VAR 23], [BRIN 18]. Dans ce cadre, le rôle des médias traditionnels a longtemps été souligné [BRIN 18]. En ce sens, la Directive européenne relative à la fourniture de services de médias audiovisuels de 2018 impose par exemple, à ce qu'une part minimale de 30% du catalogue des fournisseurs de service de médias audiovisuels soit occupée par des œuvres européennes. D'autres politiques nationales vont plus loin, certains États comme la France ayant prévu des quotas bien plus élevés [VAR 23]. Pour autant, l'implémentation du principe d'accès à une diversité culturelle dans l'espace numérique se heurte à plusieurs difficultés.

Il a d'abord été remarqué que les stratégies nationales visant l'accès à une diversité culturelle ont tardé à se moderniser pour s'adapter aux mutations de l'écosystème culturel suscité par l'essor de l'espace numérique [BRIN 18, P.64]. Les habitudes de consommation des contenus culturels ont évolué, avec une préférence croissante pour les grandes plateformes de streaming vidéo. Cela engendre un environnement défavorable pour les médias traditionnels, qui doivent innover pour conserver leur audience. Par ailleurs, la domination des larges plateformes américaines sur le marché de la consommation de contenu numérique, aussi bien sur le sol américain qu'à l'international, tend à fragiliser la diversité culturelle, [VAR 23, P.14]. La prédominance des représentations d'expression culturelle anglophone dans l'environnement numérique fait ainsi craindre un appauvrissement culturel, où le référentiel de la majorité entraîne un retrait de la diversité dans l'espace numérique [VAR 23, P.14], [SAN 14, P.728]

Le principe de diversité culturelle résonne aujourd'hui de manière particulièrement significative dans le contexte des IAG, ces outils pouvant à la fois assumer le rôle de moteur de recherche, et de créateur de contenus. Ainsi la Résolution A/78/L.49 adoptée par l'Assemblée générale de l'ONU en mars 2024 encourage les États membres, à promouvoir ;

« des systèmes d'intelligence artificielle qui favorisent, protègent et préservent la diversité linguistique et culturelle, en tenant compte du multilinguisme dans leurs données

d'apprentissage et tout au long de leur cycle de vie, en particulier pour les grands modèles de langage » [AGN 24, p.7].

3. Caractérisation et conséquences juridiques de la caractérisation des préjudices culturels suscités par l'IAG

Après avoir défini les notions clés nécessaires à l'étude des impacts culturels de l'IAG, cette recherche se poursuit par une exploration plus concrète de la manière dont ces outils peuvent porter atteinte aux droits culturels (3.1), avant d'examiner les conséquences de la caractérisation juridique des préjudices culturels en question (3.2).

3.1. Les contenus synthétiques, vecteurs de préjudices culturels

3.1.1. Médias synthétiques et dynamiques des espaces numériques

Les contenus synthétiques peuvent être vecteurs majeurs de préjudices culturels, et affecter le patrimoine culturel immatériel. Selon la définition de l'institut national américain des normes et de la technologie, (NIST), un « contenu synthétique » ou « contenu de synthèse » désigne un média tel qu'une image, une vidéo, un extrait audio, ou un texte, qui a été manipulé ou généré de façon significative par un algorithme [NIS 24]. L'hypertrucage, (ou *deepfake*), est une sous-catégorie de contenu de synthèse, que le règlement européen sur l'IA (RIA) définit comme ;

« une image ou un contenu audio ou vidéo généré ou manipulé par l'IA, présentant une ressemblance avec des personnes, des objets, des lieux, des entités ou événements existants et pouvant être perçu à tort par une personne comme authentiques ou véridiques » [EU 24].

De la même manière que les contenus créés directement par l'humain, les contenus synthétiques ne sont pas fondamentalement problématiques, mais peuvent le devenir en fonction de leurs caractéristiques et du contexte de leur dissémination [SPI 24], [CAL 24], [NIS 23], [FIS 24]. La gravité d'un préjudice culturel résultant de l'exposition à un média, qu'il soit synthétique ou authentique, sera ainsi proportionnelle aux niveaux de manipulation exercés sur l'audience, aux caractéristiques exhibées par le média, aux proportions du phénomène informationnel global dans lequel s'inscrit le média, ainsi qu'aux caractéristiques des audiences exposées [NIS 24].

Les utilisateurs se tournent de plus en plus vers l'IAG pour la création de contenu et la recherche d'information [KNI 24], [BIC 24], [SIN 24], un constat qui peut aisément s'expliquer par la versatilité de ces outils, leur niveau de personnalisation, et leur facilité d'accès et d'utilisation. En tant que nouvel intermédiaire clé dans le processus de création et de restitution de contenu, l'IAG s'imbrique parfaitement dans le modèle économique du Web, où l'attention des utilisateurs, ou « temps de cerveau humain disponible », tout comme ses données, se monnayent. La création de contenu, qu'il prenne la forme d'images, de vidéos ou de texte, constitue une des pierres angulaires de cet écosystème.

Cependant, une combinaison de facteurs tend à fragiliser l'intégrité de l'information dans l'espace numérique. Autant les politiques de modération et de recommandation des contenus des plateformes [CEY 23] que le besoin de validation des individus [BAL 23], et la recherche d'originalité [VOS 18] induisent un paradigme où le contenu mis en avant sur les plateformes l'est avant tout grâce aux interactions qu'il génère plutôt que sa qualité. À cela s'ajoute la difficulté de trouver un équilibre satisfaisant entre la régulation d'espaces numériques désordonnés et structurellement décentralisés et le respect des droits fondamentaux. Dans cet environnement ouvert mettant en relation des cultures très diverses [SAN 14, p.728], les désordres informationnels peuvent proliférer.

3.1.2. Le préjudice culturel des médias synthétiques utilisés dans les désordres informationnels

Un rapport du Conseil de l'Europe sur les désordres informationnels les décompose en trois catégories : la désinformation, la mésinformation, et la mal-information [WAR 17]. La première

catégorie englobe les fausses informations délibérément créées et partagées afin de porter préjudice à une personne, un groupe, une organisation ou un pays. La seconde concerne les fausses informations qui ne sont pas créées ni partagées avec l'intention de causer un préjudice. La dernière catégorie englobe l'information véridique, utilisée afin de causer un préjudice à une personne, organisation, ou pays. L'IAG peut être impliquée dans la survenance de chacun de ces phénomènes.

Les désordres informationnels sont pris très au sérieux par les autorités. En France, en 2021, le président de la République a créé une commission diverse, la «Commission Bronner», afin de travailler sur l'atténuation des désordres informationnels. Cette initiative venait compléter toute une série de mesure visant à assurer la qualité de l'information accessible en ligne et de lutter contre les ingérences étrangères tout en garantissant la diversité de l'information. La lutte contre les désordres informationnels occupe jusqu'aux grandes instances supranationales puisque le maintien de l'intégrité de l'information en ligne a fait l'objet de rapports, notamment du côté du Conseil de l'Europe [WAR 17], qui a établi deux comités d'experts afin de travailler sur le sujet. La sauvegarde de l'intégrité de l'information en ligne constitue également un axe de travail à part entière pour l'ONU.

Les systèmes d'IAG actuels ne maîtrisent ni la signification des données qui leurs ont été fournies, ni le sens de leurs propres productions [HAR 23, P.2]. En l'état actuel de la technologie, un système d'IA n'est pas en mesure de palier de lui-même aux problèmes de diversité affectant ses données d'entraînement ainsi que ses productions, puisqu'il ne maîtrise que les liaisons statistiques entre unités sémiotiques dans un corpus donné. Il incombe donc aux développeurs de limiter les transgressions éthiques des systèmes, par exemple, en concevant des solutions techniques pour aligner ces outils sur un système de règles particulier. En l'absence de garde-fous adéquats, les IAG peuvent être utilisées à des fins de discours de haine dans le cadre de désordres informationnels, et engendrer des préjudices culturels graves.

Les préjudices se rattachant aux désordres informationnels sont variés, et peuvent s'apprécier à la lumière des intérêts individuels et collectifs expressément reconnus et protégés en droit international. Par exemple, les désordres informationnels sont néfastes pour l'intégrité mentale des audiences, dont la protection est assurée par l'article 3 de la Charte des droits fondamentaux de l'Union européenne, et, de manière implicite, par l'article 9 du PIDCP. La liberté de conscience et de religion, le droit à la liberté d'opinion et d'expression, le droit à la vie privée, à la réputation, à l'image, le droit à la santé, le droit à la participation politique, à la non-discrimination, et à l'autodétermination interne peuvent être atteints également [KHA 21 P.8.], [AGN 66], [JON 19], [DOU 21, P.185], [LET 22], [SCH 20], [STO 24].

Par-dessous tout, les désordres informationnels peuvent être discriminants et/ou entraîner des conséquences discriminantes pour des individus ou groupes d'individus spécifiques. Ce peut être le cas, lorsque des campagnes de désinformation véhiculent des clichés socioculturels et des stéréotypes afin de produire de l'instabilité, en aggravant des fractures sociétales préexistantes [RID 17]. Plusieurs études ont notamment permis de mettre la lumière sur la façon dont les IAG pouvaient être utilisées pour véhiculer des discours de haine [CEN 24], [MCG 20]. Un rapport du *Center on Extremism* rapportait ainsi que l'IAG permettait de camoufler des messages racistes et antisémites à l'intérieur de médias synthétiques audios ou visuels de façon à ce que les contenus échappent à la modération [CEN 24].

Dans certains cas, ces systèmes reproduisent d'eux-mêmes des biais socio-culturels répréhensibles [GHO 24], tiennent des propos racistes [GAS 24], ou refusent catégoriquement d'aborder des sujets censurés dans leur juridiction [YAN 25]. Ce dernier cas s'est d'ailleurs présenté tout récemment avec l'IAG chinoise *Deepseek*, qui décline toute requête faisant mention de la situation politique de Taïwan, ou des manifestations de Tian'anmen [YAN 25]. Le manque d'inclusivité dans les bases de données ainsi que les biais des développeurs et les politiques de régulation peuvent avoir des conséquences indirectes néfastes sur les populations marginalisées, en perpétuant les clichés sans que l'utilisateur n'ait cherché à orienter le modèle vers une réponse particulière, [IRC 24], [GIL 24], [SHW 24], un problème qui accablait déjà les IA prédictives [DEM 17, P.31].

Les droits culturels sont particulièrement vulnérables aux désordres informationnels vecteurs de discriminations. La représentation et le partage de contenus discriminants peuvent entraîner des dommages réputationnels, des atteintes au droit à l'image, au principe de non-discrimination, aux libertés d'opinion et d'expression, et, bien sûr, aux droits culturels [SPI 24], [RUG 24], [HOF 24]. L'article 3 de la Déclaration de Fribourg sur identité et patrimoine culturels, et l'article 4 sur la référence à des communautés culturelles disposent ainsi ;

« Toute personne, aussi bien seule qu'en commun, a le droit: a. de choisir et de voir respecter son identité culturelle dans la diversité de ses modes d'expression ; ce droit s'exerce dans la connexion notamment des libertés de pensée, de conscience, de religion, d'opinion et d'expression ; b. de connaître et de voir respecter sa propre culture ainsi que les cultures qui, dans leurs diversités, constituent le patrimoine commun de l'humanité; (...). » [GRO 07].

et ;

« b. Nul ne peut se voir imposer la mention d'une référence ou être assimilé à une communauté culturelle contre son gré. » [GRO 07].

Ces dispositions viennent ainsi protéger le processus de formation de l'identité culturelle de l'individu contre les désordres informationnels véhiculant des contenus discriminants. Il en découle que toute représentation culturelle ne peut être imposée, ni à une personne, ni à un groupe de personnes, qu'elle soit biaisée, discriminatoire, ou simplement en contradiction avec ses valeurs.

3.1.3. *Au-delà des désordres informationnels ; le préjudice culturel des expressions culturelles synthétiques*

La Déclaration de Fribourg vient également reconnaître un droit « d'accès et participation à la vie culturelle », qu'elle développe en son Article 5 de la sorte :

« a.Toute personne, aussi bien seule qu'en commun, a le droit d'accéder et de participer librement, sans considération de frontières, à la vie culturelle (...) b. Ce droit comprend notamment : (...) ; la liberté d'exercer (...) ses propres pratiques culturelles et de poursuivre un mode de vie associé à la valorisation de ses ressources culturelles, notamment dans le domaine de l'utilisation, de la production et de la diffusion de biens et de services ; la liberté de développer et de partager des connaissances, des expressions culturelles, de conduire des recherches et de participer aux différentes formes de création ainsi qu'à leurs bienfaits ; le droit à la protection des intérêts moraux et matériels liés aux œuvres qui sont le fruit de son activité culturelle. » [GRO 07].

Ainsi, les créations artistiques peuvent constituer des expressions culturelles par lesquelles leurs auteurs viennent participer à la vie culturelle. En principe, la création artistique doit être libre et soutenue par les pouvoirs publics. En effet, dès lors qu'elles « incarnent ou transmettent des expressions culturelles » indépendamment de leur valeur commerciale, les créations artistiques peuvent être qualifiées de « biens culturels » selon les termes de l'Article 4 de la Convention UNESCO, qui encourage les États à adopter des politiques et mesures afin de promouvoir les expressions culturelles, notamment à soutenir les artistes dans leurs efforts pour créer des biens culturels. Dès lors il apparaît tout à fait pertinent d'analyser en priorité l'impact des contenus synthétiques sur le droit d'auteur, puisque ce dernier vient justement protéger les auteurs d'œuvres artistiques.

La question de la protection de la propriété intellectuelle se trouve au cœur d'une activité contentieuse particulièrement riche autour des contenus synthétiques. Il est en effet possible de faire générer à une IAG des copies d'œuvres normalement protégées par le droit d'auteur, ou d'instruire le modèle à reproduire le style d'un artiste humain dans ses sorties. Ces usages peuvent conduire à des violations du droit de la propriété intellectuelle, qui protège les œuvres artistiques, compromettant par la même, le droit d'accès et participation à la vie culturelle tel que reconnu à l'article 5 de la Déclaration de Fribourg.

Les questions de propriété intellectuelle sont intrinsèquement liées aux modalités de constitution des bases de données d'entraînement des systèmes d'IAG. Le risque trouve sa source dans le fait que ces outils sont également entraînés sur des créations numériques ou numérisées protégées par le droit d'auteur. On parlera alors de « données culturelles ». Même les corpus préformés tels que *LAION-5B*, *The Pile* ou *CommonCrawl* contiennent d'importantes quantités d'œuvres protégées, et il n'existe pas à ce jour d'outils techniques et règlementaires permettant de contrôler efficacement la manière dont les acteurs de l'IA s'emparent et utilisent les données culturelles. En conséquence, des frictions avec le droit de propriété artistique peuvent se réaliser en amont du déploiement du système, au stade de la constitution du corpus d'entraînement [DES 24]. La présence d'œuvres artistiques dans ce dernier engendre également un risque de reproduction illégale en sortie.

La reproduction partielle ou complète d'une œuvre protégée en sortie de système peut constituer une violation du droit d'auteur, bien que des exceptions puissent s'appliquer en fonction des juridictions [COU 22, p. 861]. L'affaire *NewYork Times contre Microsoft Corporation & OpenAI* illustre bien cette difficulté, le journal reprochant aux deux géants numériques d'avoir permis à leur modèle d'IAG de reproduire Verbatim ses articles et de les résumer [NEW 23]. A l'inverse, le fait qu'il soit possible pour un utilisateur de reproduire, sur demande, le style d'un artiste en sortie n'entraîne pas en principe de violation des droits de propriété mais pourrait s'apprécier à la lumière du droit de la concurrence, qui demeure également pertinente pour les cas de reproduction d'œuvres. En effet, les capacités de reproduction d'un style ou d'une œuvre par l'IA font peser le risque d'une situation de concurrence déloyale, où un système va de façon fortuite, ou voulue, se substituer à un service déjà existant, entraînant un manque à gagner pour l'auteur du contenu préalablement capté par moissonnage.

L'IAG possède un fort potentiel de création et de divertissement en ce qu'elle offre la possibilité aux utilisateurs de donner vie à leur imagination. En ce sens, l'IAG peut aussi servir de moyen d'expression artistique alimentant le patrimoine culturel. Se pose alors la question de la possibilité d'accorder une protection à du contenu artistique généré en tout ou partie par l'IAG au titre de la propriété intellectuelle ; A qui du développeur, de l'utilisateur, des créateurs dont les œuvres ont servi à l'entraînement du système, ou du système lui-même, devrait revenir la propriété d'une œuvre synthétique ? Cette dernière possibilité peut d'emblée être évacuée, puisque les IAG n'ont pas la capacité à être titulaire de droits, ou « personnalité juridique » [BUR 23]. Par ailleurs, bien que cela ne soit pas formulé explicitement, la plupart des instruments internationaux pertinents laisse entendre que la qualité d'« auteur » n'est en principe attribuable qu'aux humains [FRI 24].

Pour ce qui est de l'utilisateur, la *Beijing Internet Court* a tranché en faveur de l'obtention de droit de propriété intellectuelle sur une œuvre originale entièrement générée par IAG [LIV 23]. Dans cette affaire, les juges chinois ont estimé que l'investissement de l'utilisateur instructeur était suffisant pour le considérer comme auteur d'une œuvre authentique pouvant bénéficier de la protection juridique. Tout récemment, le Bureau des droits d'auteur aux États-Unis a admis la possibilité d'octroyer les droits sur une œuvre synthétique à l'utilisateur si celui-ci justifiait d'« éléments expressifs suffisants » [BUR 25]. Bien que la majorité des juridictions nationales n'aient pas encore tranché la question, la tendance serait à l'admission de l'octroi de la propriété à l'utilisateur si celui-ci justifie d'un investissement substantiel dans la génération de l'œuvre. Les critères de détermination de l'investissement demeurent peu clairs à ce jour. En ce qui concerne la possibilité d'attribuer des droits de propriété sur les générations des modèles aux développeurs desdits modèles, la situation n'est pas non plus réglée. Toutefois, l'argument de l'octroi de droits de propriété aux développeurs peine à convaincre, ceux-ci n'étant engagés dans aucune démarche créatrice au sens artistique du terme.

D'une part, l'impossibilité de voir leurs productions synthétiques un jour protégées par le droit d'auteur peut décourager certains artistes. D'autre part, la crainte que leurs œuvres soient utilisées pour l'entraînement de modèles d'IAG dissuade les artistes de partager leurs créations avec le grand public. En définitive, l'indétermination de la question de l'octroi de la protection intellectuelle aux créations synthétiques, ajoutée à la dépréciation de la valeur d'une œuvre artistique engendrée par des pratiques

de moissonnage peu scrupuleuses des intérêts des artistes, participent à créer un environnement défavorable à la création et la distribution d'expressions culturelles.

Or la Convention UNESCO impose aux États de s'efforcer ;

« de créer sur leur territoire un environnement encourageant les individus et les groupes sociaux à créer, produire, diffuser et distribuer leurs propres expressions culturelles (...) » [UNE 05].

3.1.4. *L'appauprissement culturel dans l'espace numérique*

Les IAG tendent à reproduire les écueils de représentativité qui accablaient déjà le Web. Ces outils sont basés sur une logique de fonctionnement probabiliste. C'est-à-dire qu'au moment de la génération, un modèle d'IAG calcule la probabilité qu'un mot ou pixel en suive un autre en fonction de ses données d'entraînement et des instructions plus ou moins précises de l'utilisateur. Il peut aussi utiliser le contexte de production c'est à dire la suite d'instruction précédentes et les diverses versions de la sortie ou un document fourni directement par l'utilisateur. Or, les grands corpus préformés servant pour l'entraînement des modèles d'IAG sont rarement contrôlés pour leur manque de diversité [BAA 24].

Si toute représentation biaisée prise indépendamment n'est pas nécessairement problématique, la reproduction de biais culturels et l'omission répétée de spécificités culturelles font peser un risque de disparition de certaines cultures (ou homogénéisation culturelle). En ce sens, plusieurs autorités, dont la Sous-Directrice générale pour les sciences sociales et humaines à l'UNESCO ont mis en garde contre la sous-représentation de certaines cultures dans les sorties des IAG;

« Un manque de diversité est évident, pas seulement sein des équipes, mais aussi dans les données utilisées et les algorithmes, résultant directement en la sous-représentation des communautés minoritaires. »¹ [RAM 24].

A long terme, le risque est d'assister à un appauvrissement culturel généralisé, où certaines cultures manquant de structures de conservation et de canaux de dissémination adéquats, disparaissent sous le poids d'une *Western Glaze* devenue omniprésente dans la sphère informationnelle ;

« (...) de tels biais cognitifs à l'échelle individuelle peuvent s'accumuler au fil du temps et façonner le système culturel au sens large. »² [TAO 24].

Ce phénomène participerait à la tendance déjà amplement documentée d'homogénéisation des cultures, caractérisée par l'alignement des valeurs derrière celles occidentales [GHO 24], [GIL 24], [SHW 24]. Or ;

« La pauvreté culturelle est la base des autres dimensions de la pauvreté ; elle empêche de sortir de l'enchaînement des précarités et fait obstacle à tout développement individuel et collectif. » [MEY 08].

Plus l'IAG prendra de place dans le processus de dissémination de l'information, plus sa capacité à altérer le processus par lequel les individus forgent leur propre identité sera forte. A l'échelle individuelle, l'exposition systématique à des représentations culturelles stéréotypées ou erronées ne permettra pas aux utilisateurs d'acquérir une compétence interculturelle suffisante pour comprendre et s'adapter à des groupes sociaux différents des leurs [SAX 24], [MEY 98, p.50]. À l'échelle collective, le risque est que

¹ Traduit de l'original: “A lack of diversity is evident not only within the teams but also in the data used and the algorithms, directly resulting in poor representation of minority communities.”

² Traduit de l'original: “Such small individual-level cognitive biases can accumulate over time to shape the broader cultural system”.

les systèmes développés par l'Occident servent d'assises pour l'idéologie capitaliste occidentale, dont la projection est d'ores et déjà assurée dans la sphère informationnelle. Ce phénomène pourrait susciter de la résistance, voire du défi. La majorité des IAG étant accessibles dans des pays pouvant présenter des identités culturelles extrêmement éloignées les unes des autres, ce paradigme est propice aux instabilités géopolitiques.

A long terme, ces tendances auront des répercussions sur la diversité culturelle ainsi qu'une variété de droits culturels, dont le droit de connaître et de voir respecter sa propre culture ainsi que les autres cultures, le droit à la non-imposition d'une référence, et le droit à une information libre et pluraliste, qui sont reconnus respectivement, aux articles 3, 4 et 7 de la Déclaration de Fribourg.

3.2. Les conséquences juridiques de la violation des droits culturels

3.2.1. La question de la réparation des préjudices culturels immatériels

Les acteurs privés, principaux auteurs des préjudices culturels, ne sont qu'indirectement tenus au respect des droits fondamentaux. Il revient en effet aux États de s'assurer que les droits culturels des populations placées sous leur juridiction sont protégés, en édictant des lois protectrices, en finançant les projets culturels, et plus généralement, en adoptant des politiques afin de valoriser les cultures autochtones sur leurs sols, en application des recommandations de l'UNESCO.

La qualification d'un préjudice peut engendrer des conséquences juridiques pour son auteur et l'autorité qui a fait preuve de négligence dans la prévention de celui-ci. Ces effets se mesurent à la fois à la lumière des dispositions pertinentes du droit national et des engagements internationaux de l'Etat dans lequel le préjudice culturel est survenu. Sur le plan procédural, les juridictions nationales sont tenues par le droit international d'accueillir les demandes en réparation des victimes de violation des droits de l'Homme. Les juridictions internationales compétentes peuvent être saisies après l'épuisement des voies de recours internes.

Dans la pratique, et plus encore dans le contexte des médias synthétiques culturellement préjudiciables, la situation n'est pas si claire. D'abord, le préjudice culturel tel qu'il survient dans le contexte des IAG est rarement circonscrit au seul territoire des auteurs dudit préjudice. Cela pose un problème de responsabilité, puisqu'un Etat n'est en principe pas tenu de réparer un dommage survenu en dehors de sa juridiction, même si celui-ci résulte de sa propre négligence.

Ensuite, la Déclaration de Fribourg n'est pas juridiquement contraignante, et la plupart des dispositions des Conventions pertinentes sont non prescriptives, c'est-à-dire qu'elles laissent une large marge de manœuvre aux États Parties pour atteindre leurs objectifs, pouvant conduire à des degrés de protection variables selon les juridictions [VAR 19]. Bien sûr, il demeure possible pour les parties de dénoncer la Convention UNESCO si l'une ou plusieurs d'entre elles fait preuve de mauvaise volonté dans leur application, entraînant ainsi leur retrait des obligations qui en découle. Toutefois, la dénonciation pourrait avoir des effets contre-productifs en ce qui concerne la sauvegarde de la diversité culturelle à l'échelle internationale. Se pose, en outre, la question de l'adhésion des Etats auxdites Conventions, car bien que la Convention UNESCO rassemble un grand nombre d'Etats, les Etats-Unis, où sont basés les principaux acteurs de l'IAG, n'y ont pas adhéré, s'étant systématiquement opposés au projet. En l'état, la Convention ne leur est donc pas opposable.

Enfin, le sujet de la quantification du dommage résultant de la dissémination d'un média synthétique doit également être mentionné. Les tribunaux *ad hoc* tels que le tribunal pénal international pour l'ex-Yougoslavie (TPIY) et le tribunal de Nuremberg ainsi que la Cour pénale internationale (CPI), ont déjà eu à se prononcer sur la question de la réparation du préjudice culturel collectif et individuel résultant de la détresse morale ressentie par un ou plusieurs individus concomitamment à la destruction d'éléments du patrimoine culturel physique. Dans l'affaire *Le Procureur c. Ahmad Al Faqi Al Mahdi*, qui impliquait la destruction de biens culturels par des djihadistes au Mali, la Chambre de première instance de la CPI a par exemple accordé l'octroi de réparations collectives,

«pour la douleur mentale/l'angoisse et pour la perturbation de sa culture subies par l'ensemble de la communauté de Tombouctou » [LEP 17, P.37].

Le préjudice moral lié à la perturbation de la culture a également fait l'objet d'évaluation par les Cours internationales des droits de l'Homme. Dans *Yakye Axa Indigenous Community c. Paraguay*, la Cour interaméricaine des droits de l'Homme (CIADH) a par exemple établi que le déplacement forcé de populations autochtones hors de leurs terres ancestrales avait altéré leur identité culturelle. Dans *Plan de Sánchez Massacre c. Guatemala*, la CIADH a également reconnu le vide culturel causé par le meurtre des principaux transmetteurs des savoirs et pratiques culturels d'une communauté.

Le préjudice culturel immatériel tel qu'il survient à la suite de la dissémination d'un média synthétique se démarque cependant, puisqu'il ne découle ni de la destruction d'un élément physique du patrimoine culturel, ni d'un déplacement forcé ou du massacre d'une population. Par ailleurs, il n'apparaît pas de façon immédiate, mais survient au terme de l'accumulation d'expressions et de médias culturellement préjudiciables dans la sphère informationnelle et en dehors.

Le dommage s'en retrouve d'autant plus ardu à quantifier qu'il n'est pas toujours possible d'accéder à l'historique des échanges entre l'utilisateur et le système, hormis par le biais d'accès ciblés, encadrés par le droit et nécessitant la coopération de la société détentrice du système, mais qui ne pourront être complètement représentatifs. En outre, l'opacité des systèmes d'IAG complique la détection de la négligence dans leur conception. Bout-à-bout, ces facteurs rendent plus complexe l'évaluation de la réparation ainsi que la détermination de l'intérêt à agir des justiciables lésés [BOS 22].

Bien que le texte final n'y fasse plus référence, il est intéressant de noter que les auteurs de la Déclaration de Fribourg ont évoqué la possibilité d'un droit de réponse des individus lésés par la représentation incorrecte des expressions culturelles aux informations erronées, qui pourrait dès lors, être assuré par les États par l'intermédiaire de canaux médiatiques dédiés ;

« Cette compréhension contemporaine des libertés d'information et de communication, prenant en compte la diversité culturelle, amène à reconnaître un droit de réponse aux personnes et aux groupes qui considèrent que des informations erronées sur les cultures ont été transmises. Comme cela est reconnu notamment aux articles 19 et 20 du PDCP, la liberté d'expression peut être restreinte à certaines conditions très encadrées. » [MEY 08, P.91].

D'une manière générale, le droit international de la culture manque d'outils pour assurer sa pleine effectivité, et le préjudice culturel est encore mal défini sur le plan national [BOS 22]. Les cadres de protection portant sur certains droits connexes comme la liberté d'expression ou la non-discrimination, permettent toutefois de protéger les droits culturels de façon indirecte.

3.2.2. L'enjeu de la liberté d'expression

3.2.2.1. Fonction et reconnaissance du droit à la liberté d'expression

Si la sauvegarde de la diversité et des droits culturels doit être assurée, cet impératif ne saurait être invoqué en justification de violations commises à l'encontre d'autres droits fondamentaux [BEN 18, P.9], [COM 09, P.6], [UNE 05, ARTICLE 2], [MEY 08]. En d'autres termes, l'objectif de prévention des préjudices culturels ne peut justifier à lui seul l'application de restrictions extensives sur le processus de génération et l'acte de disséminer du contenu synthétique, qui empièterait alors sur la liberté d'expression des utilisateurs.

En effet, les utilisateurs des systèmes d'IAG jouissent des garanties légales au titre des règles relatives à la protection de la liberté d'expression. A l'instar de nombreuses expressions constituant des désordres informationnels, les expressions susceptibles de causer des préjudices culturels sont dans leur majorité protégées en droit international au titre de la liberté d'expression, qui a pour fonction de protéger les personnes lorsqu'elles s'expriment. La liberté d'expression constitue un droit fondamental majeur, dont

la garantie permet l'effectivité d'une multitude d'autres droits fondamentaux, dont les droits culturels [COM 11], [MEY 98, p.97]. Le droit à la liberté d'expression présente un volet actif ; la liberté d'informer et de participer librement au débat démocratique, et un volet passif, qui prend la forme d'un droit à être informé et à recevoir de l'information [CAS 20].

Le droit à la liberté d'expression fût d'abord consacré dans la DUDH, avant d'être reproduit dans le PIDCP, ce dernier proclamant en son Article 19, que :

« Toute personne a droit à la liberté d'expression; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix. » [AGN 66].

Aujourd'hui, la liberté d'expression est reconnue sur tous les continents, étant promulguée, entre autres, à l'article 10 de la Convention de sauvegarde des droits de l'Homme et des libertés fondamentales du conseil de l'Europe, à l'article 13 de la Convention américaine relative aux droits de l'Homme, à l'article 9 de la Charte africaine des droits de l'Homme et des peuples, et au point 23 de la Déclaration des droits de l'Homme de l' Association des nations de l'Asie du Sud-Est (ASEAN). Il est aussi largement établi que la liberté d'expression s'applique en ligne, comme hors ligne [AGN 16].

Par ailleurs, la liberté d'expression joue un rôle prépondérant dans l'affirmation des droits culturels. L'article 7 de la Déclaration de Fribourg dispose notamment que ;

« (...) toute personne, seule ou en commun, a droit à une information libre et pluraliste qui contribue au plein développement de son identité culturelle. » [GRO 07].

3.2.2. *Liberté d'expression et médias synthétiques*

Un utilisateur bénéficie de la protection offerte par la liberté d'expression, aussi bien lorsqu'il interroge une IAG pour glaner des informations que lorsqu'il entreprend d'en disséminer les sorties, même lorsque celles-ci s'avèrent de piètre qualité, ou « profondément offensante » [COM 11, p.3], et même lorsque sa contribution dans le processus de génération est dérisoire, entendue comme une faible quantité d'instructions saisies en amont de la phase de génération [DEV 22]. Des productions synthétiques à caractère artistiques pourraient également se voir protégées par la liberté de création artistique découlant elle-même de la liberté d'expression [COU 22, p. 861], [MEY 98, p.97].

La liberté d'expression de l'utilisateur et par extension, de son éventuelle audience, le protège contre toutes restrictions abusives visant à altérer les mécanismes de génération du système d'IAG, puisque la liberté d'expression se décline aussi en un droit d'accès à l'information. En revanche, la liberté d'expression des développeurs d'IA ne saurait être opposée à la mise en œuvre par une autorité publique, de restrictions visant à limiter les possibilités de sorties de ces systèmes, car les sorties des modèles d'IAG ne sont en principe pas couverts par la liberté d'expression des développeurs. Pareillement, les choix de modération du processus de synthétisation qui intervient dans la génération de sorties ne relèvent pas de la liberté d'expression des développeurs d'IA, ni du système lui-même, ce dernier étant dépourvu de personnalité juridique.

3.2.2.2. *Liberté d'expression et accès à l'IAG*

En tant que moyen d'expression, et d'accès à l'information, les systèmes d'IAG intéressent la liberté d'expression en ses deux volets [CAL 24], [DEV 22]. Dans les cas où leur utilisation n'est pas cantonnée à de la récupération d'information, les IAG se posent en facilitatrices d'expression, permettant d'assister l'utilisateur dans la formulation et la structuration de ses idées, se substituant même parfois au processus créatif dans les cas où l'investissement de l'utilisateur est maigre. Ce constat a des conséquences sur l'étendue de la protection offerte par la liberté d'expression en l'espèce. Land explique en effet que ;

« Parce que des limites sur les moyens affectent nécessairement l'expression, protéger les moyens d'expression peut également servir à protéger l'expression. »³ [LAN 13].

Dans son observation générale no 34 de 2011, le Comité des droits de l'Homme de l'ONU affirme de façon similaire que la liberté d'expression :

« (...) protège toutes les formes d'expression et les moyens de les diffuser. (...) Sont visées aussi toutes les formes de médias audiovisuels ainsi que les modes d'expression électroniques et l'Internet. » [COM 11, p.3].

Ainsi, se pose la question de savoir, si par analogie aux autres moyens d'expression et d'accès à l'information, l'accès et l'utilisation des modèles d'IAG tombe dans le champ d'application du droit à la liberté d'expression ? A ce jour, il n'existe pas de droit dérivé de la liberté d'expression venant spécifiquement protéger l'accès aux systèmes d'IAG, bien que des garanties nouvelles dans le domaine de la communication soient régulièrement consacrées en droit positif [HOT 10], et que l'IAG s'impose graduellement comme un moyen crédible d'expression et d'accès à l'information. Par ailleurs, un tel droit dérivé ne répondrait pas à une nécessité impérieuse, puisque des alternatives équivalentes voire supérieures à l'IAG existent, que ce soit en matière de facilitation d'expression, ou d'accès à l'information. Néanmoins, et comme il a été souligné dans la littérature ;

« L'idée n'est pas de dire que l'IAG est nécessaire pour réaliser les valeurs d'expression libre. (...) Mais plutôt que tant que l'IAG peut être utilisée de façon responsable pour réaliser ces intérêts, il apparaît à première vue objectionable d'en nier aux citoyens l'accès en tant qu'outil. »⁴ [FIS 24].

3.2.2.3. La pertinence relative des limites de la liberté d'expression

La liberté d'expression ne protège pas *toutes* les expressions, puisqu'une protection totale ouvrirait la voie à des conflits de normes, où l'expression est utilisée à des fins manifestement contraires à d'autres droits tels que le droit à la vie privée et à la réputation. Par ailleurs, des différences culturelles, historiques, idéologiques et religieuses peuvent conduire à des différences dans l'étendue des garanties assurées au titre de la liberté d'expression selon les juridictions [SAN 13, p.714]. Cependant, le droit international positif autorise la prise de mesures de restrictions de l'expression lorsque celle-ci contreviendrait de manière ostentatoire à d'autres valeurs [AGN 66, ARTICLE 19§2]. Par exemple, la Cour européenne des droits de l'Homme a considéré qu'un article encourageant le recours à la violence ne bénéficiait pas de la protection offerte par la liberté d'expression compte tenu des risques pour la sécurité nationale et l'intégrité territoriale [CED 99].

Les éventuelles interventions dans la liberté d'expression, comme pour les autres droits fondamentaux, doivent cependant satisfaire les trois grands critères cumulatifs que sont la légalité, la nécessité et la proportionnalité. En d'autres termes, la liberté d'expression n'admet de restrictions que lorsque celles-ci sont expressément prévues par la loi, nécessaires dans une société démocratique, et proportionnelles aux intérêts protégés.

Le Comité des droits de l'Homme est venu quelque peu préciser des motifs admissibles de limitation, soulignant à plusieurs reprises que la liberté d'expression doit rester la norme. Ainsi, des interférences dans la liberté d'expression déployées dans le but de protéger les électeurs contre la coercition et

³ Traduit de l'original : “Because limits on the means necessarily affect expression, protecting the means of expression can serve to protect the expression as well.”

⁴ Traduit de l'original : “The point is not that generative AI is somehow necessary to realising free speech values—clearly it was possible beforehand. (...) Rather, the point is that insofar as generative AI can be used responsibly to further these interests, it is *prima facie* objectionable to deny citizens access to it as a tool.”

l'intimidation qui les empêcheraient d'exprimer leur choix politiques en période électorale sont admissibles. A l'inverse, des mesures qui sont de nature à restreindre le débat politique sont inadmissibles [COM 11, P.8]. L'intégralité de ces critères s'applique, de fait, aux communications utilisant du contenu synthétique.

Qu'en est-il des médias synthétiques à l'origine de préjudices culturels? A l'instar de nombreuses expressions produisant des désordres informationnels, les expressions susceptibles de provoquer des préjudices culturels sont pour la plupart protégées en droit international. Le caractère artificiel d'un contenu ne saurait d'ailleurs, à lui seul, constituer un motif de censure [FIS 24];

« Tout contenu synthétique n'est pas trompeur ou préjudiciable. Au contraire, la plupart ont une réelle plus-value, ce qui en fait de prime abord des cibles illégitimes pour la suppression de la sphère publique. »⁵ [DEV 22].

Dans la même veine, l'argument du niveau de réalisme élevé de certains contenus synthétiques ne saurait justifier la prise de mesures de censure à leur encontre. Enfin, dans la continuité des débats sur la lutte contre la désinformation, il est également proscrit de censurer des contenus pour la seule raison que ceux-ci seraient faux ou trompeurs, au risque d'empêter sur la liberté d'expression des utilisateurs ;

« Selon un principe de non-ingérence, l'Etat ne peut intervenir sur le contenu de l'information car cela signifierait la porte ouverte à la censure et à la propagande. » [MEY 98, P.97].

Il en va différemment pour ce qui est des contenus synthétiques discriminants, lesquels peuvent constituer des discours de haine prohibés. Reste que la définition du discours de haine prohibé n'est pas fixée dans le droit international, ce qui laisse une certaine marge d'appréciation aux États. Il en résulte que la qualification de discours de haine ne présuppose pas automatiquement une réponse pénale, ni même une réponse en matière civile. A l'origine, la Convention internationale sur l'élimination de toutes les formes de discrimination raciale oblige les États :

a) A déclarer délits punissables par la loi toute diffusion d'idées fondées sur la supériorité ou la haine raciale, toute incitation à la discrimination raciale, ainsi que tous actes de violence, ou provocation à de tels actes, dirigés contre toute race ou tout groupe de personnes d'une autre couleur ou d'une autre origine ethnique [AGN 65, ARTICLE 4§A].

Le PIDCP a quant à lui semblé souscrire à une définition plus restrictive du discours de haine prohibé, conduisant des experts du Haut-Commissariat aux droits de l'Homme à élaborer une grille d'évaluation afin de clarifier le statut des discours de haine sous la forme du Plan d'action de Rabat en 2012. Le seuil de reconnaissance d'un discours de haine appelant à une réponse pénale est maintenu à un niveau élevé, au travers du critère de l'incitation. Concernant les discours de haine constituant un délit, la détermination se fera au cas par cas, bien que la plupart des médias synthétiques susceptibles de déboucher sur un préjudice culturel demeurent moralement condamnables sans être pour autant illégaux ou punissables [WAC 24]. Le terme « *careless speech* » a d'ailleurs été introduit dans la littérature juridique, afin de caractériser les expressions productrices de désordres informationnels et de préjudices culturels qui échappent à la qualification juridique [WAC 24].

Quoiqu'il en soit, il ne serait pas opportun de réprimer la reproduction de stéréotypes, lorsque ceux-ci ne sont pas outrageusement discriminatoires, même si certaines représentations s'avèrent offensantes pour certains groupes [COM 11, P.3]. Pareillement, et à l'exception des cas graves de négationisme, le révisionisme issu de la représentation d'évènements fictifs dans un média synthétique doit constituer un

⁵ Traduit de l'original: “Not all synthetic content is wrongfully deceptive or otherwise harmful. On the contrary, much of it arguably has genuine value, making it a *prima facie* illegitimate target for removal from the public sphere.”

cas de « *careless speech*. » Compte-tenu du risque d'instrumentalisation de la cause culturelle par certains acteurs, il s'avère peu opportun de permettre aux autorités nationales d'établir elles-mêmes leurs critères d'intervention sur les modèles, leurs productions, ainsi que sur leur dissémination par les tiers, lorsque ces mesures viendraient directement affecter la liberté d'expression. L'accent doit être mis sur la prévention, ce qui implique la valorisation de toutes les identités culturelles, tel que préconisé par l'intégralité des instruments rencontrés jusque là.

4. Remarques conclusives

Il est attendu que les États adoptent des mesures concrètes afin de mitiger la survenance de préjudice culturels dans le contexte de l'IAG, puisque ces outils sont fort susceptibles de porter atteinte aux droits culturels. Les politiques de régulations sur l'IAG gagneraient d'ailleurs en légitimité si elles prenaient pleinement appui sur les dispositions du droit international de la culture. En effet, les bénéfices à long terme d'une IAG respectueuse de la diversité culturelle et des droits culturels sont considérables, favorisant la valorisation des expressions culturelles et stimulant l'écosystème culturel des États.

Reste que des doutes subsistent quant à la meilleure façon de réaliser les droits culturels et, *in fine*, d'atteindre les objectifs de diversité culturelle prônés par la Convention UNESCO. Sur le plan artistique, la question de la valorisation de la créativité humaine doit être envisagée à la lumière des possibilités d'utilisations innovantes de l'IAG. Sur le plan du droit, le statut juridique du préjudice culturel doit être consolidé.

Enfin, plusieurs questions se posent sur le plan technique ; comment s'assurer de la fidélité des représentations culturelles produites par l'IAG [GIL 24] ? Comment accroître la diversité dans les IAG tout en préservant la vie privée [SPI 24] et sans nuire à la pertinence de leurs sorties [GIL 24] ? Cette dernière interrogation touche à un paradoxe complexe : plus un système est conçu pour refléter fidèlement une culture, plus sa capacité à intégrer et représenter d'autres cultures s'en trouve limitée. Un second paradoxe mérite d'être mentionné en conclusion ; celui de l'encadrement de l'accès aux sources culturelles par les acteurs du monde de l'IA eux-mêmes, qui devra nécessairement l'objet d'une pondération au regard du rôle de garants et de diffuseurs de la culture que ces firmes sont contraintes d'assumer en raison de la place occupée par leurs outils. Ces problématiques requièrent à ce que juristes, acteurs du monde de la culture, et informaticiens travaillent de concert dans le cadre de collaborations pluridisciplinaires. Le droit international de la culture en ressortira grandement enrichi.

Bibliographie

- [AGN 66] Assemblée Générale des Nations Unies (AGNU), Pacte international relative aux droits civils et politiques, *Treaty Series*, vol. 999, 1966.
- [AGN 16] AGNU, 32/... La promotion, la protection et l'exercice des droits de l'Homme sur Internet, Résolution A/HRC/32/L.20, 2016.
- [BAA 24] BAACK, S., MOZILLA INSIGHTS, Training Data for the Price of a Sandwich; Common Crawl's Impact on Generative AI, Rapport, Mozilla, 2024.
- [BAL 23] BALLARA, N., « The Power of Social Validation: A Literature Review on How Likes, Comments, and Shares Shape User Behavior on Social Media », *International Journal of Research Publication and Reviews* n°4, p. 3355-3367, 2023.
- [LIV 23] *Li v. Liu*, Beijing Internet Court, Case Number Jing 0491 Min Chu No. 11279, 2023.
- [BEN 18] BENOUNE, K., Universalité, diversité culturelle et droits culturels, AGN, A/73/227, 2018.
- [BIC 24] BICK, A., BLANDIN, A., DEMING, D., The Rapid Adoption of Generative AI, National Bureau of Economic Research, Working Paper, 2024.
- [BOS 22] BOSC, L., « Le préjudice culturel, d'une nouveauté à une perspective de la responsabilité civile », *Les Cahiers Portalis*, n°9(1), p. 85-98, 2022.

[BTE 23] B-TECH, Taxonomy of Human Rights Risks Connected to Generative AI, Supplement to B-Tech's Foundational Paper on the Responsible Development and Deployment of Generative AI, United Nations Human Rights Office of the High Commissioner, 2023.

[BUR 25] BUREAU DES DROITS D'AUTEUR, Copyright and Artificial Intelligence ; Part 2: Copyrightability, rapport du responsable de l'enregistrement des droits d'auteur, 2025.

[BUR 23] BUREAU DES DROITS D'AUTEUR, « Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence », *16190 Federal Register*, n°88(51) Rules and Regulations 37 CFR PART 202, 2023.

[CHO 23] CHOUDHURY, M., « Generative AI has a language problem », *Nature Human Behaviour*, n°7, p. 1802–1803, 2023.

[CAL 24] CALVET-BADEMUNT J., and MCHANGAMA J., Freedom of Expression in Generative AI: A Snapshot of Content Policies, rapport, The Future of Free Speech, 2024.

[CEN 24] CENTER ON EXTREMISM, “Propaganda for Fun:” How Extremists Use GAI to Camouflage Hate, 2024.

[CNI 23] CNIL, Intelligence artificielle : le plan d'action de la CNIL, 2023.

[COM 11] COMITE DES DROITS DE L'HOMME, Observation générale No. 34, Article 19: Liberté d'opinion et liberté d'expression, CCPR/C/GC/34, 2011.

[COM 09] COMITE DES DROITS ECONOMIQUES, SOCIAUX ET CULTURELS, Observation générale No.21: Droit de chacun de participer à la vie culturelle (Article 15, paragraphe 1.a), 2009.

[DEM 17] DEMIAUX, V., ABDALLAH, Y., « Comment Permettre à l'Homme de Garder la Main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle », synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la loi pour une république numérique, 2017.

[DES 24] DE SILVA J. T., DE LETONA, L., « The Right to Scrape Data on the Internet: From the US Case hiQLabs, Inc. v. LinkedIn Corp to the ChatGPT Scraping Cases: Differences Between US and EU Law », *Global Privacy Law Review*, n°1, p. 5-22, 2024.

[EU 24] Règlement (EU) 2024/1689 du Parlement européen et du Conseil du 13 Juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle. (RIA)

[FIS 24] FISCHER, S.A., HOWARD, J.W. & KIRA, B. « Moderating Synthetic Content: The Challenge of Generative AI», *Philosophy & Technology*, n°37(133), p.1-20, 2024.

[FRI 24] FRITZ, J., “The Notion of ‘authorship’ under EU law-who can be an author and what makes one an author? An analysis of the Legislative framework and case law” *Journal of Intellectual Property Law & Practice*, n°19:7, p. 552-556, 2024.

[GIL 24] GILLESPIE, T., « Generative AI and the politics of visibility », *Big Data & Society* n°11:2, p.1-14, 2024.

[GOL 23] GOLDSTEIN J.A. et al., Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations, rapport, Internet Observatory Publications, 2023.

[GRO 19] GROUPE D'EXPERTS INDEPENDANTS DE HAUT NIVEAU SUR L'IA, Lignes directrices en matière d'éthiques pour une IA digne de confiance, 2019.

[HAR 23] HARRER, S., « Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine », *The Lancet*, n°90., p.1-12, 2023.

[JAI 24] JAIDKA K., et al., « Misinformation, Disinformation, and Generative AI: Implications for Perception and Policy »*Digital Government: Research and Practice*, p.1-16, 2024.

[JI 24] JI, J., « Demystify ChatGPT: Anthropomorphism and generative AI, GRACE: Global Review of AI Community Ethics», *AI in Education, Culture, Finance, and War*, n°2(1), P.1-11, 2024.

[JON 19] JONES, K., Online Disinformation and Political Discourse Applying a Human Rights Framework, Papier de Recherche, Chatham House, 2019.

[KHA 21] KHAN I., Désinformation et liberté d'opinion et d'expression, Conseil des droits de l'Homme, A/HRC/47/25, 2021.

[KNI 24] KNIBBS, K., Yes, That Viral LinkedIn Post You Read Was Probably AI-Generated, *Wired*, 2024.

[LAF 12] LAFARGUE R. « Le préjudice culturel né du dommage environnemental : par-delà nature et culture, un préjudice écologique spécifique », dans NEYRET, L., et MARTIN, G. J (dir.), *Nomenclature des préjuges environnementaux*, LGDJ, Paris, 2012.

[LAN 13] LAND M., « Toward an International Law of the Internet », *Harvard International Law Journal*, n°54(2), (2013). p. 393-458, 2013.

[LEP 17] Le Procureur c. Ahmad Al Faqi Al Mahdi, Chambre de Première Instance, Cour Pénale Internationale, Ordonnance de réparation, N° : ICC-01/12-01/15, 2017.

- [MCG 20] MCGUFFIE, K., NEWHOUSE, A., « The Radicalization Risks of GPT-3 and Advanced Neural Language Models», *Arxiv*, 2020.
- [MEY 08] MEYER-BISH, P., « Les droits culturels Enfin sur le devant de la scène ? » *L'Observatoire*, N°33(1), p.9-13, 2008.
- [NEW 23] NewYork Times v. Microsoft Corporation & OpenAI Cour de District pour le district Sud de New York, 2023.
- [NIS 23] NIST AI, Reducing Risks Posed by Synthetic Content An Overview of Technical Approaches to Digital Content Transparency, NIST AI 100-41, Draft for public comment, 2023.
- [OTA 18] OTASEVIC, I., l'émergence d'une norme non écrite en matière de protection de la diversité culturelle en droit international, Thèse de doctorat, Université Laval, 2018.
- [POU 05] POUPARD Myriam. « La distinction entre le dommage et le préjudice » *Revue juridique de l'Ouest*, n°2, p. 187-233, 2005.
- [RAM 24] RAMOS G. in BALCERZAK M., and KAPELANSKA-PREGOWSKA J., (ed.) *Artificial Intelligence and International Human Rights Law Developing Standards for a Changing World*, Edward Elgar Publishing, Cheltenham, UK, 2024.
- [RAU 11] RAUTENBACH, C., « Cultural Diversity and International Law », dans SCHWABACH, A., et COCKFIELD, A., J., (ed), *International Law and Institutions in Encyclopedia of Life Support Systems (EOLSS)*, EOLSS & UNESCO, 2011.
- [RET 24] RETTBERG, J., W., « How Generative AI Endangers Cultural Narratives », *Issues in Science and Technology*, n°40(2), p.77-79, 2024.
- [RID 17] RID, R., Disinformation; A primer in Russian active measures and influence campaigns, Hearings before the select committee on intelligence, United States Senate, 115th congress, 2017.
- [RUG 24] RUGHINIS R., et al., « Generative AI and Social Engines of Hate », dans LIEBOWITZ J., (Ed.), *Regulating Hate Speech Created by GenAI*, Taylor & Francis, Oxford, 2024.
- [SAN 13] SANGSUVAN, K., « Balancing Freedom of Speech on the Internet under International Law », *North Carolina Journal of International Law*, n°39(701), p.723, 2013.
- [SED 21] SEDOVA et al., AI and the Future of Disinformation Campaigns Part 2: A Threat Model, Policy Brief, Center for Security and Emerging Technology (CSET), 2021.
- [SHA 10] SHAHEED, F., Rapport de l'Expert indépendante dans le domaine des droits culturels, Mme Farida Shaheed, soumis en application de la résolution 10/23 du Conseil des droits de l'Homme, A/HRC/14/36, 2010.
- [SHW 24] SHWARTZ V., « Artificial intelligence needs to be trained on culturally diverse datasets to avoid bias », TheConversation, 2024.
- [SPI 24] SPIVACK J., Synthetic Content; Exploring the Risks, Technical Approaches, and Regulatory Responses, Rapport, Future of Privacy Forum, 2024.
- [UNE 01] UNESCO, Déclaration universelle de l'UNESCO sur la diversité culturelle. 20e séance plénière de la 31e session de la Conférence générale, 2021.
- [UNE 21] UNESCO, Recommandation sur l'éthique de l'intelligence artificielle, 2021
- [UNE 03] UNESCO, Convention pour la sauvegarde du patrimoine culturel immatériel, 2003.
- [VAR 23] VARIN C., GUEVREMONT, V., OTASEVIC I., « L'accès à une diversité d'expressions culturelles dans l'environnement numérique : un droit fondamental en quête de reconnaissance et de protection » *ADI/ILA 2023 150ème anniversaire – Laboratoire d'idées*.
- [VAN 22] VAN DER SLOOT, B. & Wagenveld, Y., « Deepfakes: regulatory challenges for the synthetic society », *Computer Law & Security Review*, n°46, 2022.
- [VAR 19] VARIN, C. & GUEVREMONT, V., ‘La Convention sur la diversité des expressions culturelles quinze ans après : une mise en oeuvre effective d’un instrument juridique faiblement contraignant », *Revue québécoise de droit international*, n°32(2), p. 1-27, 2019.
- [VOS 18] VOSOUGHI, S., DEB R., et SINAN, A, « The Spread of True and False News Online », n°359(6380) *Science*, p.1146-1151, 2018.
- [WAC 24] WACHTER S., MITTELSTADT B., et RUSSELL C., “Do large language models have a legal duty to tell the truth?” *Royal Society Open Science*, n°11(8), p.1-38, 2024.
- [WAR 17] WARDLE, C., & DERAKHSHAN, H., INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making, Rapport, Conseil de l'Europe, 2017.
- [YAN 25] YANG, Z., « Here's How DeepSeek Censorship Actually Works—and How to Get Around It », WIRED, 2025.