

Feuille de route

La mission sur les hypertrucages générés ou manipulés par l'IA et ses enjeux pour les secteurs culturels et créatifs a conclu la première phase de ses travaux par la remise d'un rapport d'étape présenté en séance plénière du CSPLA le 9 juillet 2026 et livrant un diagnostic approfondi des problématiques juridiques soulevées.

La présente feuille de route a pour objet de définir le parcours d'une deuxième étape de la mission, pour aller, en concertation avec l'ensemble des acteurs intéressés, vers des propositions destinées à répondre aux enjeux soulevés lors de la première phase. Une présentation des pistes retenues sera effectuée en séance plénière du CSPLA en décembre 2026 et une remise du rapport final à la fin du premier semestre 2027.

Ces propositions répondront aux différentes dimensions du sujet décrites par le rapport d'étape. Elles viseront d'abord les hypertrucages, catégorie particulière de contenus générés par IA, tels que définis par le Règlement européen sur l'Intelligence artificielle (RIA) 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 ». Aux côtés des hypertrucages tels que le RIA les définit et les appréhende, deux autres catégories de contenus seront également prises en compte, parce qu'elles tangent le champ de cette définition et interrogent les mécanismes de régulation mis en place. Il s'agit d'une part des textes et d'autre part des autres contenus générés par IA qui peuvent satisfaire une partie des critères de la définition, sans y répondre entièrement ; ainsi, les « créations fantômes » sans avoir une ressemblance directe avec une personne, un lieu, une entité, un événement ou un objet existant *stricto sensu*, n'en présentent pas moins une illusion de réalité ; ces créations ne cherchent pas à s'appuyer sur une ressemblance avérée avec l'existant mais simplement à être perçues comme authentiques par le public.

Sur l'ensemble de ce champ, cinq problématiques orienteront les travaux : l'appréhension fine d'une nouvelle économie qui se déploie et qui ne profite que très partiellement aux créateurs et aux industries culturelles (1), l'identification des failles d'un encadrement juridique disparate et insuffisamment protecteur pour les personnes (2), la réponse au phénomène de confiance brisée du public (3), la prise en compte des difficultés croissantes à détecter et identifier les hypertrucages (4) et enfin la recherche d'une meilleure effectivité des droits, tant par la clarification du licite et de l'illicite que par l'amélioration des modes d'engagement de la responsabilité de acteurs.

1. PROBLEMATIQUE N°1 : UNE NOUVELLE ECONOMIE SE DEPLOIE, QUI NE PROFITE QUE TRES PARTIELLEMENT AUX INDUSTRIES CULTURELLES

Les hypertrucages s'intègrent dans une économie de la génération de contenus. La fabrique industrielle des hypertrucages se caractérise par une bicéphalie économique : deux marchés coexistent et se développent en parallèle, avec des dynamiques, des externalités et des modèles économiques très différents. Le premier est un marché d'entreprise à entreprise (B2B) en forte croissance. Le second est un marché grand public mêlant des usages anodins ou créatifs à des usages malveillants à grande échelle. En volume d'usages, le pôle grand public semble dominer. En valeur monétaire identifiable, le rapport s'inverse. L'émergence du marché des hypertrucages en tant que service (« deepfake as a service », DFAAS) repose sur des plateformes qui cherchent à gagner de l'argent en transformant des modèles complexes d'IA générative en services accessibles par crédit (« pay per use »), par abonnement, licences professionnelles ou selon une logique de monétisation indirecte via la publicité, démocratisant ainsi l'accès à des technologies de pointe auparavant exclusivement réservées aux très grosses entreprises. Dans leurs interactions avec les industries culturelles, certaines de ces sociétés se concentrent sur le marché grand public avec des interfaces simples et automatisées. D'autres se positionnent sur le marché professionnel en développant des services pour les studios, les plateformes ou les agences.

Les hypertrucages sont donc « fabriqués » pour de multiples usages. Contrairement aux représentations dominantes, ils ne sont cependant pas exclusivement associés à des usages malveillants, nocifs ou illégaux. De nombreux milieux professionnels de l'information et de la création ont en effet désormais recours à des exploitations positives des hypertrucages, pour des raisons essentiellement économiques, artistiques ou éditoriales. Les hypertrucages deviennent cependant problématiques dès lors qu'ils atteignent la réputation et les revenus individuels de personnes physiques et morales ou se substituent globalement, à des emplois,

fragilisant ainsi les modèles économiques des industries culturelles. Des effets d'éviction s'exercent par une concurrence à la fois par les prix et par les quantités ; la surabondance de contenus générés par les systèmes d'IA risque en effet de saturer le marché et, par conséquent, de réduire la découvrabilité des œuvres humaines et une partie des revenus qui auraient pu bénéficier aux créateurs humains.

Objectifs de travail

- Identifier les instruments de politiques publiques de nature à encourager la production de contenus humains plutôt que synthétiques (fiscalité, distribution d'aide, contribution, domaine public payant ...).
- Faire le point sur les initiatives pouvant encourager un meilleur partage de valeur issu de la fabrique des hypertrucages entre les acteurs issus de l'IA et ceux de la culture (par exemple, en incitant à la conclusion d'accords, y compris collectifs, protégeant les droits des acteurs des secteurs créatifs et culturels).

PROBLEMATIQUE N°2 : UN ENCADREMENT JURIDIQUE DISPARATE ET INSUFFISANT DE PROTECTION DES PERSONNES FACE AUX CONTENUS HYPERTRUQUES

La réponse apportée par le Règlement sur l'Intelligence artificielle (RIA) aux enjeux des hypertrucages repose sur la transparence, l'objectif étant de saisir les conséquences pouvant en résulter « *sur l'intégrité de l'écosystème informationnel et la confiance en celui-ci, ce qui pose de nouveaux risques de désinformation et de manipulation à grande échelle, de fraude, d'usurpation d'identité et de tromperie des consommateurs* » (RIA, Considérant 133). Les conséquences quant aux secteurs de la culture et de la création ne sont mentionnées qu'à la marge par le RIA. Pourtant, dans ces secteurs, les hypertrucages interviennent au cœur même de la production de valeur, c'est-à-dire dans la création, l'interprétation, la mise en scène et la circulation des œuvres, soulevant dès lors des enjeux distincts.

En se bornant à harmoniser les règles de transparence applicables aux hypertrucages, le législateur européen a fait le choix de renvoyer pour l'essentiel l'encadrement de ces contenus aux autres législations européennes et nationales susceptibles de s'appliquer. Une telle complémentarité est indispensable dans la mesure où l'obligation de transparence ne saurait épuiser l'ensemble des enjeux auxquels nos sociétés sont confrontées du fait de l'apparition de ces contenus, en particulier s'agissant de la protection des personnes. Traiter des hypertrucages suppose par conséquent de solliciter un ensemble de plusieurs normes ayant des finalités et des champs d'application différents. Or, une analyse élargie du cadre juridique **conduit à dresser le constat suivant : celui de l'existence d'un ensemble disparate de règles n'assurant pas une protection uniforme**. Cela résulte du fait que la plupart des textes applicables ont été conçus dans une perspective différente de celle des hypertrucages ; ils se recoupent en partie mais laissent subsister des doutes et des insuffisances au regard des enjeux identifiés dans le cadre des industries culturelles et créatives. Concernant plus particulièrement **la protection des personnes dont les caractéristiques sont susceptibles d'être reprises par les hypertrucages**, celle-ci **se caractérise par son caractère éparé et inégal**. Le développement des systèmes d'IA générative et la faculté de générer aisément ce type de contenus soulignent les limites d'une telle protection, qu'il s'agisse de celle de l'image, de la voix, du style ou encore de la réputation. Par ailleurs, si les outils contractuels constituent d'ores et déjà un cadre de réflexion pour traiter les enjeux des hypertrucages et sont porteurs d'importantes potentialités, les travaux conduits ont pu ici encore en identifier les insuffisances. Cela invite à prolonger la réflexion afin d'envisager les possibilités d'évolution, en tenant compte des enjeux contractuels.

Objectifs de travail :

- Identifier les moyens (suggestions en faveur d'une évolution jurisprudentielle, préconisation d'une modification du cadre législatif,...) qui permettraient de garantir une protection effective de certains droits de la personnalité, notamment du droit à l'image et du droit à la voix, y compris après le décès des personnes.
- Renforcer les garanties contractuelles en cas d'exploitation des attributs de la personnalité, en étudiant les conditions dans lesquelles des contrats ou clauses types, distincts des cessions de droits d'auteur, pourraient mieux encadrer la durée, l'objet et les usages autorisés de l'exploitation de la voix et de l'image.

PROBLEMATIQUE N° 3 : LA CONFIANCE BRISEE DU PUBLIC

Les limites cognitives des humains face aux contenus synthétiques brouillent leur perception. Les données empiriques convergentes des dernières années confirment en effet que la grande majorité des auditeurs et des spectateurs sont incapables de distinguer les contenus générés par intelligence artificielle de ceux produits par des humains. A la question de la perception se superpose celle de la réception individuelle. En la matière, on constate une tension structurelle entre normes déclarées et usages effectifs. D'un côté, les enquêtes d'opinion révèlent des attentes fortes en matière de transparence, et une méfiance persistante vis-à-vis de la qualité et de la légitimité des productions entièrement synthétiques. De l'autre, les comportements observés suggèrent, dans certains cas, une acceptation pragmatique, voire une valorisation conditionnelle, notamment lorsque les contenus suscitent la curiosité ou activent le transport narratif.

Sur le plan collectif, le danger principal des hypertrucages réside dans leur capacité à produire de l'incertitude et, par ce biais, à éroder la confiance générale du public par effet de contamination, dégradant ainsi la valeur perçue de l'ensemble des contenus y compris de ceux produits par des humains. Dès lors le coût social de la production d'un hypertrucage excède son coût privé, parce qu'une partie de ce coût social est répercutée sur des tiers sans compensation par celui qui est à l'origine du problème. Autrement dit les externalités négatives des hypertrucages ne sont pas internalisées par leurs producteurs qui en captent les bénéfices sans en supporter les coûts sociaux. Ainsi, sans signal fort qui internalise les coûts des effets négatifs des hypertrucages pour la société et réduise le coût asymétrique de la production face à la détection, les acteurs privés n'ont pas d'intérêt à agir.

L'étiquetage des contenus synthétiques constitue une première voie d'action. Le RIA impose en effet aux déployeurs d'IA une obligation de transparence spécifique aux hypertrucages via l'étiquetage. Dans le cas des contenus s'insérant dans « une œuvre ou un programme manifestement artistique, créatif, satirique, fictif ou analogue » l'ampleur de l'obligation est cependant limitée ; l'interprétation de cette limitation reste à préciser.

Objectifs de travail

- Accompagner la mise en œuvre du cadre européen en cours de déploiement, en identifiant les modèles d'étiquetage les mieux adaptés aux contenus artistiques et culturels, tout en tenant compte des exigences du RIA et du Code de bonnes pratiques sur la transparence des contenus générés par l'IA, des solutions techniques disponibles et des acteurs susceptibles de les mettre en œuvre, sans détériorer l'expérience du public et éviter le sur-étiquetage qui décrédibiliserait les contenus humains mêlés à des éléments synthétiques.
- Identifier les leviers permettant de favoriser la mise en œuvre effective du Code de bonnes pratiques par un nombre accru d'acteurs, en examinant par exemple les dispositifs de labellisation, de certification ou de présomption de conformité susceptibles d'accompagner les pratiques d'étiquetage.
- Définir des modalités concrètes et pratiques de sensibilisation du public à l'existence des hypertrucages.

PROBLEMATIQUE N° 4 : UNE DIFFICULTE CROISSANTE A DETECTER ET IDENTIFIER LES HYPERTRUCAGES

Sur le plan technique, on distingue les techniques de détection en aval des contenus synthétiques qui circulent, en l'absence de marquage (détection « forensique ») et les techniques visant à rendre détectables les contenus synthétiques dès la génération de contenus, par le marquage en amont.

En l'absence de marquage en amont, **les techniques de détection aval** vont comparer les propriétés statistiques des contenus humains et celles des contenus synthétiques avec pour objectif de repérer les artefacts caractéristiques laissés par le générateur. Par ailleurs, la distinction entre contenus humains et contenus synthétiques ne permet pas de repérer la ressemblance d'un contenu avec une personne existante ; d'autres techniques de détection d'identité, spécifiques aux hypertrucages, sont alors nécessaires. Détecter la présence d'une identité dans un contenu est aujourd'hui un problème en grande partie résolu, sur le plan technique – sans présager de sa prise en charge économique - à partir d'un corpus de référence authentifié. Au contraire, la détection, en l'absence de marquage en amont, reste peu fiable, avec des perspectives dégradées du fait de l'avancement des technologies, les modèles de génération progressant plus vite que les outils de détection.

Une autre famille de solutions techniques consiste à **rendre détectables les contenus synthétiques dès la génération de contenus, par le marquage en amont. Ce sont ces techniques de marquage que la législation européenne du RIA impose aux fournisseurs d'IA pour l'ensemble des contenus synthétiques et non pour les seuls hypertrucages.** Le marquage désigne l'insertion proactive d'une information signalant l'origine IA d'un contenu par le système qui le produit. La détection consistera alors à vérifier la présence ou l'intégrité de cette « marque » dans un contenu donné. Le marquage peut être appliqué pour un contenu synthétique, mais également pour un contenu humain, authentique. Une stratégie complémentaire consiste donc non pas à détecter le synthétique, mais à certifier l'authentique, en attachant une preuve de provenance au contenu dès sa création. Ces techniques rendent l'origine des contenus, humaine ou artificielle détectable par des outils automatisés.

Si le principe du marquage préventif en amont est posé, son efficacité dépendra de la standardisation des méthodes. Le risque est de voir émerger une multiplicité de systèmes distincts, développés par différents acteurs, sans interopérabilité ni cohérence. Ces techniques en amont nécessitent pour se développer une coopération entre les développeurs de solutions technologiques, les déployeurs d'outils, les diffuseurs de contenus culturels et les institutions publiques afin notamment de répartir la charge des investissements nécessaires. Les enjeux économiques de la répartition concrète de la charge de la détection dans toute la chaîne de valeur sont considérables. Parallèlement au développement du marché des hypertrucages un marché en miroir émerge en effet autour des outils défensifs. Ce marché reste pour l'heure majoritairement aval (réaction post-production) plutôt qu'amont (prévention à la génération).

Objectifs de travail

- Accompagner la mise en œuvre du cadre européen en cours de déploiement, en identifiant les conditions permettant de faciliter l'accès économique aux outils de détection d'identité en aval.
- Contribuer à renforcer la robustesse du marquage sur le plan technique et juridique en incitant les fournisseurs de SIA à prévoir des clauses contractuelles en ce sens et préciser les exigences de mise à disposition d'outils de détection du marquage.
- Accompagner les réflexions sur l'émergence d'un standard européen de provenance des contenus.

PROBLEMATIQUE N°5 : UNE APPROCHE PARTIELLE DE LA RESPONSABILISATION DES ACTEURS FACE AUX HYPERTRUCAGES

En miroir de l'encadrement juridique, se posent des enjeux de responsabilisation des acteurs de la chaîne de valeur autour de ces contenus. Cela requiert, de prime abord, de **préciser ce qui relève de l'hypertrucage admissible ou non** dès lors qu'il s'agit de la condition préalable à la détermination des responsabilités des différents acteurs. Il s'agit alors de déterminer, au regard des différentes législations applicables, dans quelle mesure l'illicéité de l'hypertrucage peut résulter, d'une part, du non-respect des exigences de transparence - ce qui demeure encore incertain - et, d'autre part, de la qualification d'un contenu comme étant illicite en tant que tel.

Il convient ensuite d'envisager comment se déploie la responsabilisation des acteurs. A cet égard, les textes européens et nationaux organisent un ensemble complexe de procédures destinées à responsabiliser les différents opérateurs des services numériques, à partir de l'idée fondamentale que, compte tenu des particularités de la société de l'information, les responsabilités ne peuvent se borner à être engagées à l'égard des seuls auteurs de l'illicéité ou du dommage, de sorte qu'il est nécessaire d'obtenir le concours de l'ensemble des opérateurs afin de prévenir et limiter autant que possible les effets de cette illégalité. La confrontation de cet ensemble aux hypertrucages et aux enjeux qu'ils soulèvent pour les secteurs de la culture et de la création met en évidence **des responsabilités très variables**. L'étude des multiples procédures existantes visant des opérateurs aux responsabilités diverses conduit à relever que **la protection peut s'avérer imparfaite selon le type d'enjeu soulevé**, en particulier s'agissant de la mise en œuvre de la procédure instaurée par l'article 6-3 de la LCEN lorsque les hypertrucages portent atteinte aux droits de la personnalité. Cela doit par conséquent conduire à envisager les pistes d'évolution pour protéger efficacement les droits des personnes face à ces contenus et responsabiliser pleinement l'ensemble des acteurs.

Objectifs de travail

- Approfondir la question de savoir si doivent être considérés illicites, au sens du DSA, les hypertrucages diffusés sans que les fournisseurs du système d'IA générative ou les déployeurs aient satisfait à leurs obligations tirées de l'article 50 du RIA.
- Envisager comment améliorer l'effectivité des procédures en particulier lorsque les droits liés à la vie privée, à la personnalité ou aux données personnelles sont remis en cause par un hypertrucage, notamment en engageant une réflexion sur l'adoption d'une disposition équivalente à celle de l'article L. 336-2 du Code de la propriété intellectuelle, afin de faciliter les actions engagées à l'encontre d'hypertrucages illicites.
- Proposer des clarifications quant à la définition des notions de « fournisseur » et de « déployeur » de systèmes d'IA, afin de déterminer plus facilement quel acteur est responsable et quelles obligations lui sont applicables au titre du RIA, du DSA et du Code de bonnes pratiques.