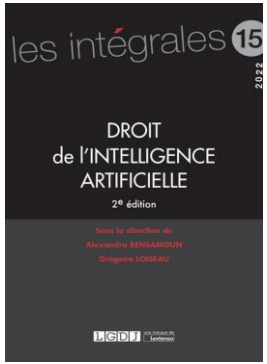
QUEL ENCADREMENT JURIDIQUE POUR LES IA MANIPULATRICES ?

Alexandra BENSAMOUN
Professeure de droit
Université Paris-Saclay
Membre de la Commission interministérielle de l'IA



CORPUS NORMATIF DE LA MANIPULATION, ETC.

- Publicité
- Double clic/contrat électronique
- Droit de rétractation
- Pratiques commerciales déloyales
- Droit des données à caractère personnel
- DSA
- DMA…

- Charte des droits fondamentaux de l'UE

AUGMENTATION DU RISQUE

Ex. : bulles de filtre informationnelles et désinformation

Raisons ? Ampleur et efficacité

- Moyens importants et asymétrie
- Personnalisation accrue

UNION EUROPEENNE

AI Act ou RIA, proposition de règlement, 21 avril 2021

Dernier trilogue décembre 2023 (tendu…)

Vote COREPER janvier 2024

Vote en plénière du PE 13 mars 2024

SYSTÈME D'IA (SIA)

Définition, art. 3, 1) :

« un système automatisé conçu pour fonctionner à différents niveaux d'autonomie, qui peut faire preuve d'une capacité d'adaptation après son déploiement et qui, pour des objectifs explicites ou implicites, déduit, à partir des données d'entrée qu'il reçoit, la manière de générer des résultats tels que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels »

- Neutralité technologique
- Convergence des autorités intergouvernementales

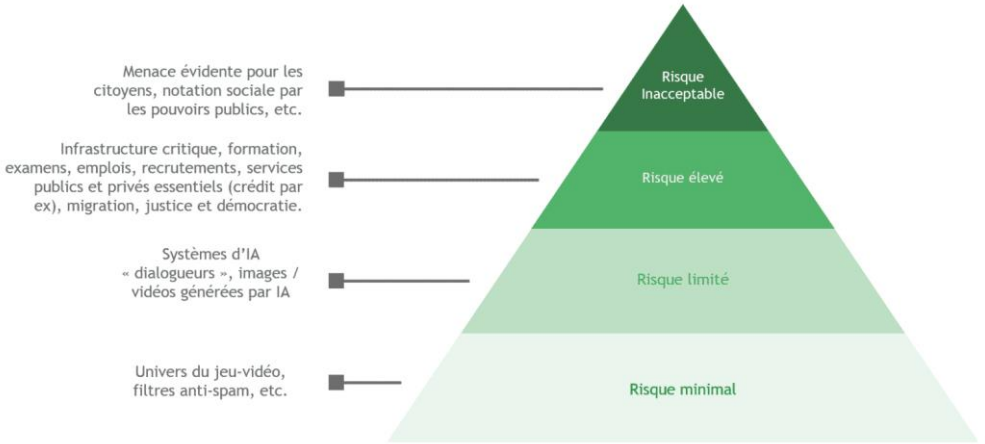
PYRAMIDE DES RISQUES

Approche fondée sur des cas d'usage et sur l'appréciation des risques liés

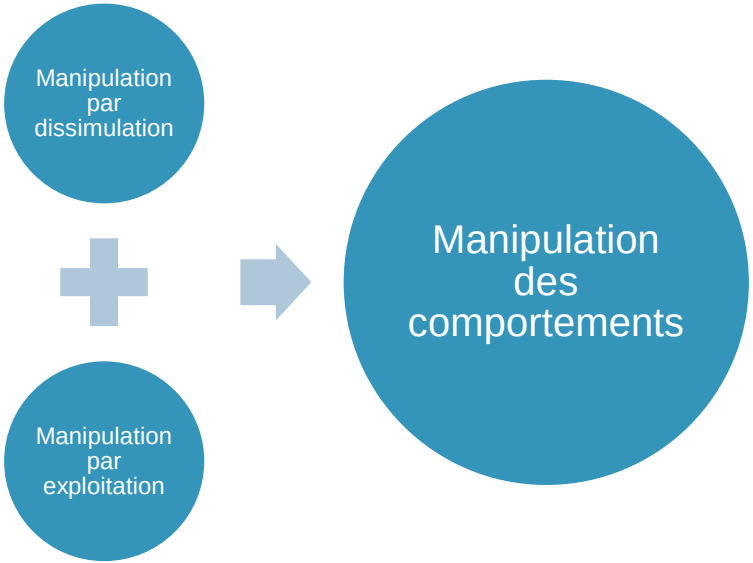
Règles procédurales **modulées** en fonction du **risque**

Gradation normative :

- interdiction légale (risque inacceptable)
- obligations renforcées et une procédure de mise en conformité (haut risque)
- obligations de transparence et d'information (risque modéré)
- simple recommandation de droit souple/codes de conduite (risque faible ou nul)



DEUX FORMES DE MANIPULATION DES COMPORTEMENTS



CINQ COMPORTEMENTS MANIPULATOIRES

Manipulation par dissimulation

- Interdiction des techniques subliminales
- Obligations de transparence

Manipulation par exploitation

- Interdiction/vulnérabilités cognitives
- Interdiction ou encadrement/vulnérabilités émotionnelles

DISSIMULATION : LES TECHNIQUES SUBLIMINALES

Art. 5, §1 a) :

Interdiction de « la mise sur le marché, la mise en service ou l'utilisation d'un système d'IA qui a recours à des techniques subliminales, au-dessous du seuil de conscience d'une personne, ou à des techniques délibérément manipulatrices ou trompeuses, avec pour objectif ou effet d'altérer substantiellement le comportement d'une personne ou d'un groupe de personnes en portant considérablement atteinte à leur capacité à prendre une décision éclairée, amenant ainsi la personne à prendre une décision qu'elle n'aurait pas prise autrement, d'une manière qui cause ou est susceptible de causer un préjudice important à cette personne, à une autre personne ou à un groupe de personnes »

DISSIMULATION : LES OBLIGATIONS DE TRANSPARENCE

Art. 50 :

1. Les fournisseurs veillent à ce que les systèmes d'IA destinés à interagir directement avec des personnes physiques soient conçus et développés de manière à ce que les personnes physiques concernées soient informées qu'elles interagissent avec un système d'IA, sauf si cela ressort clairement du point de vue d'une personne physique normalement informée et raisonnablement attentive et avisée, compte tenu des circonstances et du contexte d'utilisation. Cette obligation ne s'applique pas aux systèmes d'IA dont la loi autorise l'utilisation à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière, sous réserve de garanties appropriées pour les droits et libertés des tiers, sauf si ces systèmes sont mis à la disposition du public pour permettre le signalement d'une infraction pénale.

2. Les fournisseurs de systèmes d'IA, y compris de systèmes d'IA à usage général, qui génèrent des contenus de synthèse de type audio, image, vidéo ou texte, veillent à ce que les résultats produits par les systèmes d'IA soient marqués dans un format lisible par machine et identifiables comme ayant été générés ou manipulés par une IA. Les fournisseurs veillent à ce que leurs solutions techniques soient aussi efficaces, interopérables, solides et fiables que la technologie le permet, compte tenu des spécificités et des limites des différents types de contenus, des coûts de mise en œuvre et de l'état de la technique généralement reconnu, comme cela peut ressortir des normes techniques pertinentes. Cette obligation ne s'applique pas dans la mesure où les systèmes d'IA remplissent une fonction d'assistance pour la mise en forme standard ou ne modifient pas de manière substantielle les données d'entrée fournies par le déployeur ou leur sémantique, ou lorsque leur utilisation est autorisée par la loi à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière.

3. Les déployeurs d'un système de reconnaissance des émotions ou d'un système de catégorisation biométrique informent les personnes physiques qui y sont exposées du fonctionnement du système et traitent les données à caractère personnel conformément au règlement (UE) 2016/679, au règlement (UE) 2018/1725 et à la directive (UE) 2016/680, selon le cas. Cette obligation ne s'applique pas aux systèmes d'IA utilisés pour la catégorisation biométrique et la reconnaissance des émotions dont la loi autorise l'utilisation à des fins de prévention ou de détection des infractions pénales ou d'enquêtes en la matière, sous réserve de garanties appropriées pour les droits et libertés des tiers, et dans le respect du droit de l'Union.

4. Les déployeurs d'un système d'IA qui génère ou manipule des images ou des contenus audio ou vidéo constituant un hypertrucage indiquent que les contenus ont été générés ou manipulés par une IA. Cette obligation ne s'applique pas lorsque l'utilisation est autorisée par la loi à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière. Lorsque le contenu fait partie d'une œuvre ou d'un programme manifestement artistique, créatif, satirique, fictif ou analogue, les obligations de transparence énoncées au présent paragraphe se limitent à la divulgation de l'existence de tels contenus générés ou manipulés d'une manière appropriée qui n'entrave pas l'affichage ou la jouissance de l'œuvre.

Les déployeurs d'un système d'IA qui génère ou manipule des textes publiés dans le but d'informer le public sur des questions d'intérêt public indiquent que le texte a été généré ou manipulé par une IA. Cette obligation ne s'applique pas lorsque l'utilisation est autorisée par la loi à des fins de prévention ou de détection des infractions pénales, d'enquêtes ou de poursuites en la matière, ou lorsque le contenu généré par l'IA a fait l'objet d'un processus d'examen humain ou de contrôle éditorial et lorsqu'une personne physique ou morale assume la responsabilité éditoriale de la publication du contenu.

EXPLOITATION : LES VULNERABILITES COGNITIVES

Art. 5, §1 b) :

Interdiction de « la mise sur le marché, la mise en service ou l'utilisation d'un système d'IA qui exploite les éventuelles vulnérabilités dues à l'âge, au handicap ou à la situation sociale ou économique spécifique d'une personne ou d'un groupe de personnes donné avec pour objectif ou effet d'altérer substantiellement le comportement de cette personne ou d'un membre de ce groupe d'une manière qui cause ou est raisonnablement susceptible de causer un préjudice important à cette personne ou à un tiers »

EXPLOITATION : LES VULNERABILITES EMOTIONNELLES

Interdiction :

Article 5, § 1 f) : « la mise sur le marché, la mise en service à cette fin spécifique ou l'utilisation de systèmes d'IA pour inférer les émotions d'une personne physique sur le lieu de travail et dans les établissements d'enseignement, sauf lorsque l'utilisation du système d'IA est destinée à être mise en place ou mise sur le marché pour des raisons médicales ou de sécurité »

Encadrement strict (haut risque) :

Art. 6, § 2 et Annexe III, 1, c) :

« Biométrie, dans la mesure où leur utilisation est autorisée par la législation nationale ou de l'Union applicable : (…) systèmes d'IA destinés à être utilisés pour la reconnaissance des émotions »