Législation sur l’intelligence artificielle

2021/0106(COD) - 13/03/2024 - Texte adopté du Parlement, 1ère lecture/lecture unique

Le Parlement européen a adopté par 523 voix pour, 46 contre et 49 abstentions, une résolution législative sur la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l’intelligence artificielle (législation sur l’intelligence artificielle) et modifiant certains actes législatifs de l’Union.

La position du Parlement européen arrêtée en première lecture dans le cadre de la procédure législative ordinaire modifie la proposition comme suit:

Objet

L’objectif du règlement est d'améliorer le fonctionnement du marché intérieur et de promouvoir l'adoption de **l'intelligence artificielle (IA) axée sur le facteur humain et digne de confiance**, tout en garantissant un niveau élevé de protection de la santé, de la sécurité et des droits fondamentaux consacrés dans la charte des droits fondamentaux, notamment la démocratie, l'État de droit et la protection de l'environnement, contre les effets néfastes des systèmes d'intelligence artificielle dans l'Union, ainsi que de soutenir l'innovation.

Le règlement ne s'appliquera pas aux systèmes d'IA ou aux modèles d'IA spécifiquement développés et mis en service uniquement à des fins de recherche et développement scientifiques, ni aux résultats qu'ils génèrent. Des «**bacs à sable réglementaires**» et des essais en conditions réelles devront être mis en place par les autorités nationales et mis à disposition des PME et des start-ups pour développer et tester des IA innovantes avant leur mise sur le marché.

Le règlement s'appliquera aux systèmes d'IA publiés dans le cadre de licences libres et ouvertes, sauf s'ils sont mis sur le marché ou mis en service en tant que systèmes d'IA à haut risque.

Maîtrise de l’IA

Les fournisseurs de systèmes d'IA devront prendre des mesures pour garantir un niveau suffisant de maîtrise de l'IA pour leur personnel et les autres personnes s'occupant du fonctionnement et de l'utilisation des systèmes d'IA pour leur compte, en prenant en considération leurs connaissances techniques, leur expérience, leur éducation et leur formation, ainsi que le contexte dans lequel les systèmes d'IA sont destinés à être utilisés, et en tenant compte des personnes ou des groupes de personnes à l'égard desquels les systèmes d'IA sont destinés à être utilisés.

Applications interdites en matière d’IA

Les nouvelles règles interdisent certaines applications fondées sur l’IA telles que :

- les systèmes qui ont recours à des **techniques subliminales** ou à des techniques délibérément manipulatrices ou trompeuses, avec pour objectif d’altérer substantiellement le comportement d’une personne en portant considérablement atteinte à la capacité de la personne à prendre une décision éclairée, l’amenant ainsi à prendre une décision qu’elle n’aurait pas prise autrement;

- les systèmes qui exploitent les éventuelles **vulnérabilités dues à l'âge, au handicap ou à la situation sociale ou économique** spécifique d'une personne avec pour objectif ou effet d'altérer substantiellement le comportement de cette personne;

- les systèmes qui sont utilisés pour la **notation sociale** (classifiant les personnes en fonction de leur comportement social ou de leurs caractéristiques personnelles);

- les système pour mener des **évaluations des risques des personnes physiques** visant à évaluer ou à prédire la probabilité qu'une personne physique commette une infraction pénale, uniquement sur la base du **profilage** d'une personne physique ou de l'évaluation de ses traits de personnalités ou caractéristiques;

- les systèmes qui créent ou développent des **bases de données de reconnaissance faciale** par le moissonnage non ciblé d'images faciales provenant de l'internet ou de la vidéosurveillance;

- les systèmes de **reconnaissance des 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);

- les systèmes de **catégorisation biométrique** qui catégorisent individuellement les personnes physiques sur la base de leurs données biométriques afin d'arriver à des déductions concernant leur race, leurs opinions politiques, leur affiliation à une organisation syndicale, leurs convictions religieuses ou philosophiques, leur vie sexuelle ou leur orientation sexuelle;

- l'utilisation de **systèmes d'identification biométrique à distance en temps réel dans des espaces accessibles au public à des fins répressives**, sauf si cette utilisation est strictement nécessaire pour i) la recherche ciblée de victimes spécifiques d'enlèvement, de la traite et de l'exploitation sexuelle d'êtres humains, ainsi que la recherche de personnes disparues; ii) la prévention d'une menace réelle d'attaque terroriste; ii) l'identification d'une personne soupçonnée d'avoir commis une infraction pénale, aux fins de mener une enquête pénale, d'engager des poursuites ou d'exécuter une sanction pénale pour des infractions passibles d'une mesure de sûreté privatives de liberté d'une durée maximale d'au moins quatre ans.

L'utilisation des systèmes d'identification biométrique «en temps réel» ne sera autorisée que si l'autorité répressive a réalisé une **analyse d'impact sur les droits fondamentaux**. En outre, leur utilisation sera limitée dans le temps et géographiquement et soumise à une autorisation judiciaire ou administrative préalable spécifique. Aucune décision produisant des effets juridiques défavorables à l'égard d'une personne ne pourra être prise sur la seule base du produit du système d'identification biométrique à distance en temps réel.

Obligations pour les systèmes à haut risque

Le règlement prévoit des obligations strictes pour les systèmes d'IA à haut risque (en raison de leur préjudice potentiel important pour la santé, la sécurité, les droits fondamentaux, l'environnement, la démocratie et l'État de droit).

Ont été ajoutés à la liste des systèmes à haut risque, en particulier les systèmes destinés à être utilisés:

- en tant que composants de sécurité dans la gestion et l'exploitation **d'infrastructures numériques critiques**, du trafic routier ou de la fourniture d'eau, de gaz, de chauffage ou d'électricité;

- pour déterminer l'accès, l'admission ou l'affectation de personnes physiques à des établissements **d'enseignement et de formation professionnelle**, à tous les niveaux;

- pour le recrutement ou la sélection de personnes physiques, en particulier pour publier des **offres d'emploi ciblées**, analyser et filtrer les candidatures et évaluer les candidats;

- pour évaluer l'éligibilité des personnes physiques aux **prestations et services d'aide sociale essentiels**, y compris les services de soins de santé;

- pour l'évaluation des risques et la tarification en ce qui concerne les personnes physiques en matière **d'assurance-vie et d'assurance maladie**;

- dans le cadre de **la migration, de l'asile et de la gestion des contrôles aux frontières**, aux fins de la détection, de la reconnaissance ou de l'identification des personnes physiques;

- pour **influencer le résultat d'une élection** ou d'un référendum ou le comportement électoral de personnes physiques dans l'exercice de leur vote.

Ces systèmes devront faire l'objet d'une évaluation et d'une réduction des risques, être assortis de registres d'utilisation, être transparents et précis et être soumis à une supervision humaine. Les citoyens auront le droit de déposer des **plaintes** concernant les systèmes d'IA et de recevoir des explications sur les décisions basées sur des systèmes d'IA à haut risque qui ont une incidence sur leurs droits.

Systèmes d'IA à usage général

Les systèmes d'IA générative basés sur des modèles tels que ChatGPT devront respecter les exigences en matière de **transparence**. Les fournisseurs de modèles d'IA à usage général devront respecter la législation européenne sur les droits d'auteurs et mettre à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle d'IA à usage général. Les systèmes d'IA à usage général plus puissants présentant un risque systémique seront soumis à des exigences supplémentaires. De plus, les images et les contenus audio et vidéo artificiels ou manipulés (hypertrucages ou «deep fakes ») devront être clairement signalés comme tels.