CDDH-IA(2024)09
08/11/2024

COMITÉ DIRECTEUR POUR LES DROITS HUMAINS
(CDDH)

_____

GROUPE DE RÉDACTION SUR LES DROITS HUMAINS ET L'INTELLIGENCE ARTIFICIELLE

(CDDH-IA)

Résumé de l'échange de vues avec des expert·e·s indépendant·e·s externes et des représentant·e·s du Comité sur l'intelligence artificielle (CAI), du Comité d'experts sur les implications de l'intelligence artificielle générative pour la liberté d'expression (MSI-AI), du Comité consultatif pour la Convention 108 (T-PD), et du Groupe de travail sur la cyberjustice et l'intelligence artificielle (CEPEJ-GT-CYBERJUST) (25 septembre 2024)

(Préparé par le Secrétariat)

1. INTRODUCTION

Le CDDH-IA a tenu un échange de vues, lors de sa première réunion, sur l'intersection de l'intelligence artificielle (IA) et des droits humains, en se concentrant sur les implications techniques, juridiques et éthiques des technologies de l'IA, ainsi que sur le travail de divers comités intergouvernementaux du Conseil de l'Europe sur les droits humains et l'IA.

Le CDDH-IA a organisé l'échange de vues avec :

- Marko GROBELNIK, Chercheur en IA et « Champion Digital » au AI Lab de l'Institut Jozef Stefan de Slovénie
- Alberto QUINTAVALLA, Professeur adjoint à l'université Erasmus de Rotterdam
- Miriam KULLMANN, Professeure de droit du travail et de droit de la sécurité sociale à la faculté de droit de l'université d'Utrecht et membre du Comité européen des droits sociaux.
- David LESLIE, Directeur de la recherche sur l'éthique et l'innovation responsable à l'Institut Alan Turing
- Thomas SCHNEIDER, ancien Président du Comité sur l'intelligence artificielle (CAI) du Conseil de l'Europe (2022-2024) et actuel vice-président du CAI
- Peter KIMPIÁN, Secrétaire du Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (T-PD)
- Andrin EICHIN, Président du Comité d'experts du Conseil de l'Europe sur les implications de l'intelligence artificielle générative sur la liberté d'expression (MSI-AI)
- Jan SPOENLE, Membre du groupe de travail sur la cyberjustice et l'intelligence artificielle (CEPEJ-GT-CYBERJUST) et contributeur à la Charte éthique européenne sur l'IA dans les systèmes judiciaires

2. RESUMÉ DES DISCUSSIONS

Principaux points soulevés par Marko GROBELNIK

- L'évolution de la définition et du cadre opérationnel des systèmes d'intelligence artificielle (IA) reflète le besoin croissant de clarté et de termes exploitables, les récentes mises à jour de l'OCDE ayant été largement adoptées dans les cadres politiques internationaux. Ces définitions décrivent les systèmes d'IA comme des cadres basés sur des machines capables de faire des prédictions, de formuler des recommandations ou de prendre des décisions qui influencent des environnements réels ou virtuels, et qui fonctionnent souvent de manière autonome.
- Les composants clés des systèmes d'IA comprennent une séquence commençant par l'entrée des données, progressant par la construction du modèle - souvent par le biais d'algorithmes d'apprentissage automatique - suivie par l'inférence et la sortie, qui influencent ensuite l'environnement externe. Au cœur des systèmes d'IA se trouve le modèle, une représentation comprimée et souvent opaque des données qui guide les réponses du système. Le processus d'inférence, par lequel l'IA interprète et applique le modèle, est crucial pour la logique opérationnelle du système et sa capacité à prendre des décisions.

- Les récentes mises à jour de la définition de l'OCDE intègrent des termes clés tels que « adaptabilité » et « autonomie », reflétant l'impact profond des systèmes d'IA génératifs et leur capacité à apprendre et à s'adapter à de nouvelles données et à de nouveaux environnements. Si ces développements représentent des avancées dans les capacités de l'IA, ils posent également des défis en matière de responsabilité et de robustesse des systèmes, en particulier lorsque les systèmes autonomes peuvent agir sans surveillance humaine immédiate.
- La nature « boîte noire » de l'IA pose d'importants problèmes de transparence, car les modèles avancés fonctionnent souvent au-delà de la compréhension totale des opérateurs, même experts. Cette opacité soulève d'importantes préoccupations en matière de responsabilité, d'autant plus que la vitesse, la complexité et l'autonomie de l'IA peuvent dépasser la capacité humaine à superviser efficacement. Il est essentiel de relever ces défis dans les cadres de l'IA pour aligner les développements technologiques sur les protections des droits humains.
- Les systèmes d'IA interagissent de plus en plus avec des phénomènes inconnus ou ininterprétables par l'homme, car les modèles d'apprentissage automatique découvrent des schémas et des idées qui vont au-delà des connaissances humaines actuelles. Si cette capacité représente une avancée technologique notable, elle souligne le besoin urgent de politiques solides qui empêchent les systèmes d'IA de porter atteinte par inadvertance aux droits fondamentaux ou aux normes sociétales établies.

Principaux points soulevés par Alberto QUINTAVALLA

- La relation entre l'IA et les droits humains est souvent envisagée sous l'angle de droits spécifiques tels que la protection de la vie privée et la non-discrimination, en raison de la nature de l'IA axée sur les données et de ses préjugés inhérents. Les risques pour la vie privée sont accrus dans les applications gouvernementales de l'IA, où la collecte de données personnelles est au cœur de fonctions telles que la police prédictive, la lutte contre le terrorisme et la surveillance, ce qui soulève d'importantes préoccupations en matière de protection de la vie privée. Bien que ces utilisations de l'IA comportent des risques pour la vie privée, les protections juridiques existantes, telles que la Convention 108+ du Conseil de l'Europe et divers cadres nationaux, offrent de solides garanties. Toutefois, à mesure que l'IA progresse, la vigilance est essentielle pour assurer une protection efficace de ces droits. De même, les droits à la non-discrimination sont remis en question par la susceptibilité de l'IA aux préjugés, en particulier dans les systèmes utilisés par les autorités publiques, comme le démontrent des cas tels que la police prédictive et la détection des fraudes à l'aide sociale. L'affaire néerlandaise « *Toeslagenaffaire* », dans laquelle les autorités fiscales ont utilisé des algorithmes qui signalaient de manière disproportionnée les ressortissants non néerlandais dans le cadre de la détection des fraudes aux allocations familiales, illustre bien la manière dont les préjugés intégrés dans l'IA peuvent conduire à des discriminations.
- Une approche fragmentaire de l'IA et des droits humains risque de négliger les impacts plus larges de l'IA, ce qui pourrait avoir des conséquences imprévues. En revanche, une approche systématique offre des avantages en reconnaissant l'indivisibilité, l'interdépendance et l'interconnexion de tous les droits humains. Cette approche garantit que les droits ne sont pas considérés isolément, mais dans un cadre qui reconnaît les impacts croisés. Une perspective plus large de l'impact de l'IA est importante car l'IA peut simultanément renforcer un droit tout en en compromettant un autre, ce qui nécessite un équilibre prudent entre les droits. En outre, une approche systématique permet d'identifier

les risques communs à tous les droits, ce qui donne aux décideurs politiques des indications précieuses pour créer des garanties permettant d'aborder et de prévenir les violations des droits humains associées à l'IA.

- L'influence de l'IA est à la fois positive et négative, créant une relation à double sens avec les droits. Par exemple, l'IA peut renforcer le droit à un procès équitable en améliorant l'accès aux ressources juridiques, en identifiant la jurisprudence et en prédisant les résultats, rendant ainsi la justice plus accessible. Cependant, elle peut également produire des résultats biaisés, violant ainsi les droits à un procès équitable et les protections contre la discrimination. La double nature de l'IA signifie qu'elle peut avoir un impact sur plusieurs droits à la fois ; par exemple, l'utilisation autoritaire de l'IA à des fins de surveillance peut porter atteinte à la vie privée, à la liberté de réunion et à la liberté d'expression. La Convention-cadre du Conseil de l'Europe souligne ce point en reconnaissant que l'IA peut faire progresser les droits humains tout en présentant des risques pour la dignité et l'autonomie humaines.
- Cette relation à double tranchant forme un réseau de droits contradictoires ou qui se renforcent. Par exemple, l'« internet des corps » - des dispositifs tels que les montres et les pilules connectées qui suivent les données de santé - peut améliorer le droit à la santé en fournissant des informations continues sur la santé, mais risque d'empiéter sur les droits à la protection de la vie privée. Les individus peuvent vouloir révoquer leur consentement à la collecte de données, mais cela pourrait avoir un impact sur leur santé. En outre, si ces dispositifs renforcent les droits en matière de santé, ils peuvent entrer en conflit avec les droits en matière d'environnement en raison de leur empreinte écologique élevée, ce qui met en évidence les compromis entre la santé et la protection de l'environnement. Le réseau complexe de l'impact de l'IA sur les droits humains présente des défis réglementaires importants. Les décideurs politiques et les juges doivent prendre en compte les différents compromis en matière de droits humains découlant de l'IA, ce qui rend le principe de proportionnalité de plus en plus pertinent. Bien que parfois controversé, ce principe fournit un moyen structuré d'équilibrer les droits concurrents, permettant des solutions pratiques dans les litiges.
- Les entreprises privées ont joué un rôle central dans le secteur de l'IA en étant les acteurs clés de la phase de développement et, parfois, de la phase de déploiement. Les gouvernements peuvent s'appuyer sur des acteurs non étatiques pour entreprendre certaines activités gouvernementales, notamment dans le secteur de la justice pénale. En outre, les acteurs non étatiques exercent eux-mêmes des activités de gouvernance pour combler les lacunes en la matière, par exemple dans le cadre du développement de villes intelligentes. C'est dans ce contexte que le droit relatif aux droits humains devrait développer des moyens de gérer le rôle des entreprises privées. L'effet horizontal indirect des droits humains et les lignes directrices existantes des Nations unies en matière de droits humains ont jusqu'à présent été utilisés pour atténuer ce problème. En vertu de l'effet horizontal indirect des droits humains, les États ont l'obligation positive de protéger les individus contre les ingérences préjudiciables d'acteurs non étatiques. Cette responsabilité appelle des mesures plus efficaces pour protéger les individus des préjudices potentiels causés par des entités privées. En outre, des instruments juridiques non contraignants, tels que les lignes directrices sur la responsabilité des entreprises, ont été introduits pour encourager le respect des droits humains. Toutefois, ces approches sont relativement nouvelles et leur impact réel reste à évaluer.
- L'interaction entre l'IA et les droits humains s'étend à des questions émergentes telles que la protection de l'environnement, l'IA ayant des effets positifs et négatifs sur

l'environnement. Par exemple, si l'IA peut contribuer à prévenir les activités illégales dans les zones protégées, elle est également très gourmande en énergie et en eau, ce qui suscite des inquiétudes quant à son empreinte écologique. Il est difficile de quantifier l'impact environnemental de l'IA, car il n'existe pas de consensus sur la mesure de son empreinte carbone. En outre, le Conseil de l'Europe ne reconnaît pas de droit autonome à un environnement sain, ce qui complique les efforts visant à porter en justice les dommages environnementaux liés à l'IA. Les plaintes doivent souvent être fondées sur des droits établis tels que la protection de la vie privée, ce qui fait qu'il est difficile pour les individus de satisfaire à la « condition de victime ». En outre, les instruments juridiques non contraignants relatifs aux droits humains visent généralement les États et non les entreprises privées, ce qui limite leur efficacité dans la lutte contre les incidences environnementales des technologies de l'IA.

Principaux points soulevés par Miriam KULLMANN

- Les autorités publiques et les entités privées ont de plus en plus recours à des systèmes de prise de décision automatisée, qu'elles appliquent à des tâches telles que l'octroi de prestations sociales, l'embauche et l'évaluation de la solvabilité. Si ces systèmes peuvent améliorer l'efficacité, on ne peut ignorer que l'automatisation de la prise de décision, même si un être humain intervient à un moment ou à un autre du processus, peut violer les droits économiques et sociaux dans la mesure où les processus de prise de décision automatisés traitent des données à caractère personnel. Les données personnelles pertinentes sont, par exemple, l'âge, le sexe, la religion ou les convictions, les opinions politiques, l'appartenance à un syndicat, le lieu de vie et/ou de travail, la situation familiale et/ou matrimoniale, la situation de revenu et le fait que la personne ait ou non besoin de la sécurité sociale ou de prestations sociales, ou l'état de santé d'une personne. Ces systèmes sont limités par la qualité de leurs données et des algorithmes qui les guident, ce qui entraîne souvent des conséquences négatives pour les groupes marginalisés.
- La numérisation a transformé le recrutement, les systèmes automatisés étant largement utilisés pour cibler et sélectionner les candidats, bien que les risques de partialité persistent. Ces systèmes contrôlent également la productivité, influençant les promotions, les salaires ou les licenciements. Dans le domaine de la protection sociale, la prise de décision automatisée est appliquée à la gestion des prestations, comme on l'a vu dans le scandale néerlandais des prestations de garde d'enfants (*kinderopvangtoeslagaffaire*), où des algorithmes ont repéré des personnes sur la base de profils tels que la double nationalité, ce qui a entraîné des sanctions injustifiées. Des problèmes similaires se posent au Danemark, où de nombreuses données sont recoupées pour détecter les fraudes aux prestations sociales, ce qui soulève des questions d'équité. L'automatisation affecte également l'accès au logement et aux soins de santé, les algorithmes déterminant l'éligibilité aux prêts hypothécaires, aux logements sociaux et même aux traitements médicaux. Lorsqu'ils sont formés à partir de données biaisées, ces systèmes risquent d'aboutir à des généralisations néfastes et peuvent négliger les besoins individuels.
- La CSE consacre plusieurs droits relatifs à l'emploi, qui sont affectés par l'essor de la prise de décision automatisée. L'article 1 garantit le droit au travail, y compris la non-discrimination (lié à l'article 20, le droit à l'égalité de traitement en matière d'emploi), l'article 8 (protection de la maternité), l'article 15 (emploi des personnes physiquement ou mentalement diminuées), l'article 23 (protection sociale des travailleurs âgés) et l'article 27 (égalité de traitement pour les travailleurs ayant des responsabilités familiales). La discrimination fondée sur des caractéristiques telles que le sexe ou le handicap est

interdite en matière de recrutement, de conditions d'emploi et de promotion, tandis que l'article 18 garantit aux ressortissants étrangers l'égalité d'accès au travail. Les systèmes automatisés de prise de décision affectant des facteurs tels que les heures de travail, la productivité contrôlée ou la rémunération doivent respecter les limites qui protègent la santé, la sécurité et la non-discrimination (article 4 et article 3). Lorsque l'IA est utilisée pour le suivi des performances professionnelles ou les licenciements, un préavis et des motifs valables sont requis (article 24). Les travailleurs doivent bénéficier d'un environnement sûr, y compris de droits tels que la déconnexion numérique, empêchant toute pénalisation en cas d'absence de travail en dehors des heures normales. L'article 5 de la CSE garantit la liberté d'association, qui peut être affectée si les systèmes d'IA prennent en compte l'appartenance à un syndicat, ce qui pourrait dissuader les travailleurs d'y adhérer. L'article 6 garantit les droits de négociation collective, encourageant la représentation dans les discussions sur les rôles des systèmes automatisés sur le lieu de travail. Ce droit est étroitement lié aux droits relatifs à la consultation, aux conditions et à la dignité sur le lieu de travail (articles 21, 22 et 26).

- En ce qui concerne la santé et la sécurité sociale, la CSE garantit le droit à la sécurité sociale pour les travailleurs salariés et indépendants et les personnes à leur charge en vertu de l'article 12, paragraphe 1, qui couvre des domaines tels que les soins médicaux, le chômage, les prestations familiales et les prestations de maternité. L'accès à la sécurité sociale doit être exempt de toute discrimination, comme le prévoient les articles 20 et E, ce qui garantit l'égalité de traitement dans des domaines tels que le calcul des prestations et l'éligibilité. Les autorités publiques qui recourent à la prise de décision automatisée pour la sécurité sociale ou l'assistance médicale (article 13, paragraphe 1) doivent veiller à ce que ces systèmes soient non discriminatoires, car l'éligibilité ne devrait dépendre que du besoin. L'accès aux services de protection sociale exige également des pratiques non discriminatoires (articles 14 et E).

- En ce qui concerne le logement, l'article 31 de la CSE stipule que toute personne a droit à un logement adéquat, une attention particulière étant accordée aux groupes vulnérables tels que les personnes à faible revenu, les parents isolés et les personnes handicapées. Lorsque des systèmes de prise de décision automatisés sont utilisés par des autorités publiques ou des entités privées pour attribuer des logements ou des prêts hypothécaires, l'accès doit être non discriminatoire, comme l'exige l'article E. Les États sont également tenus de prévenir et de réduire le nombre de sans-abri, en veillant à ce que les systèmes automatisés n'excluent pas ou ne discriminent pas les personnes qui risquent de perdre leur logement.

Principaux points soulevés par David LESLIE

- La croissance rapide de l'IA a remodelé l'environnement numérique, en intégrant des systèmes pilotés par l'IA à la fois dans l'infrastructure publique et dans la sphère privée. Cette transformation permet aux systèmes algorithmiques d'influencer l'identité personnelle et les interactions sociétales par le biais de prises de décision ciblées et de la curation de contenu. Des pratiques telles que l'exploration de l'attention, le classement par pertinence et la prédiction des tendances orientent le comportement des utilisateurs et affectent la manière dont les individus s'engagent dans l'information, ce qui soulève d'importantes préoccupations en matière de droits humains. En orientant les expériences numériques, ces technologies peuvent porter atteinte à l'autonomie individuelle, à la vie privée et à la

liberté de pensée, car elles contrôlent de plus en plus l'accès à l'information et l'interprétation de celle-ci dans la sphère publique.

- Le rôle omniprésent de l'IA dans la manipulation des comportements s'étend au lieu de travail, où des systèmes de gestion algorithmiques surveillent la productivité, la satisfaction et les interactions des employés. Ces systèmes, bien qu'ils visent à optimiser l'efficacité, peuvent porter atteinte à la vie privée et altérer les comportements en raison de la surveillance constante qu'ils exercent. Le manque de confiance et de solidarité interpersonnelles induit par cette surveillance compromet la liberté d'expression et de réunion, essentielle pour les environnements de travail collaboratifs. Dans les domaines du travail, de l'éducation et de la justice pénale, ces technologies peuvent imposer une surveillance coercitive qui limite la communication ouverte et décourage l'action et la participation démocratiques.
- Les technologies d'IA générative intensifient ces risques en permettant la production à grande échelle de désinformation, de contenus synthétiques et de « *deep fakes* » (ou « hypertrucage »), qui peuvent dégrader la qualité et la fiabilité des informations dans la sphère publique numérique. Cet afflux de contenus manipulés menace la confiance sociale qui est à la base des sociétés démocratiques, car il devient de plus en plus difficile pour les individus de discerner les faits de la fabrication. La pollution du discours public par des informations erronées générées par l'IA constitue une menace sérieuse pour l'intégrité de l'engagement démocratique et de la responsabilité civique.
- L'ensemble de ces évolutions met en évidence un défi éthique majeur : l'influence de l'IA sur les droits humains, la démocratie et l'État de droit est considérable, et son expansion incontrôlée risque de saper les fondements mêmes de ces institutions. Consciente du potentiel de transformation de l'IA, la société doit activement façonner son déploiement par le biais d'une gouvernance démocratique et d'une prospective éthique, en veillant à ce que le progrès technologique s'aligne sur les valeurs de justice, d'équité et de respect des droits humains.

Principaux points soulevés par Thomas SCHNEIDER

- L'IA, en tant que technologie de rupture, s'inscrit dans la lignée des changements technologiques passés, comme le passage de l'automatisation physique à l'automatisation cognitive. Contrairement aux moteurs physiques, l'IA fonctionne avec des données dématérialisées qui peuvent être reproduites et déployées à l'échelle mondiale, ce qui nécessite une gouvernance adaptative. M. Schneider a souligné que la convention ne visait pas à remplacer les normes existantes mais à les compléter, en mettant l'accent sur le respect des droits humains, de la démocratie et de l'État de droit tout en promouvant l'innovation. L'approche vise à prendre en compte les caractéristiques uniques de l'IA grâce à un cadre de haut niveau, à l'épreuve du temps, qui reste adaptable aux progrès technologiques.

- Le champ d'application de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l’homme, la démocratie et l'État de droit (Convention-cadre) s'étend aux secteurs public et privé, avec une certaine souplesse pour l'application nationale, mais tient tous les acteurs responsables de la protection des droits. Des exceptions limitées s'appliquent, telles que la sécurité nationale et certaines activités de recherche. La Convention-cadre applique une approche différenciée, en évaluant les risques en fonction du contexte et de la gravité de l'impact, afin de garantir que les

mesures réglementaires sont proportionnelles à l'impact potentiel de la technologie sur les droits.

- La Convention-cadre énonce des principes fondamentaux tels que la dignité humaine, l'autonomie individuelle, la transparence et la surveillance, l'obligation de rendre compte et la responsabilité, l'égalité et la protection de la vie privée. Les garanties procédurales prévoient une documentation permettant aux individus de contester les décisions prises par l'IA, conformément à l'engagement de transparence et d'accès aux voies de recours. En ce qui concerne les besoins réglementaires futurs, les cadres juridiques conventionnels pourraient être trop lents pour répondre à l'évolution rapide des technologies de l'IA. Des formes de gouvernance plus agiles, adaptables et adaptées aux technologies du XXIe siècle sont nécessaires. La Convention-cadre, qui attend maintenant d'être ratifiée, marque une étape importante, mais il convient de souligner la nécessité d'un dialogue continu, de normes techniques et d'une coopération internationale pour créer une vision commune de l'IA qui s'aligne sur les droits humains et les normes démocratiques au niveau mondial.

Principaux points soulevés par Andrin EICHIN

- Le Comité d'experts sur les implications de l'intelligence artificielle générative pour la liberté d'expression (MSI-AI) prépare un projet de lignes directrices sur les incidences de l'intelligence artificielle générative sur la liberté d'expression. Le travail a commencé en 2024, avec la première réunion visant à explorer les défis uniques de l'IA générative. Une première version des lignes directrices est prévue pour la fin de l'année 2024, le document final devant être achevé en 2025.

- Le processus consiste à déterminer si et comment l'IA générative a un impact unique sur la liberté d'expression par rapport à d'autres technologies, à évaluer si ces impacts nécessitent une intervention immédiate et à déterminer les conditions nécessaires pour sauvegarder la liberté d'expression dans une société de plus en plus façonnée par l'IA générative. Pour guider cette analyse, le comité utilise une approche par « *tech-stack* » ou « stack technique », examinant les risques à chaque étape du cycle de vie de l'IA générative - formation au modèle fondamental, mise au point du modèle et interaction avec l'utilisateur.

- La MSI-AI a indiqué que l'analyse devait porter à la fois sur les risques et les avantages de l'IA générative. L'IA générative présente des avantages, tels que l'aide à l'apprentissage personnalisée, l'accès aux ressources de santé mentale par l'intermédiaire d'avatars et des outils pour lutter contre la désinformation. Cependant, elle présente également des risques, notamment des résultats biaisés, des « hallucinations » (génération d'informations incorrectes) et un potentiel de manipulation par le biais d'un contenu hyper-personnalisé. Ces questions ont une incidence directe sur la liberté d'expression et soulignent la nécessité de mettre en place des garde-fous, notamment en ce qui concerne la transparence des contenus générés par l'IA. De telles mesures sont cruciales pour éviter toute utilisation abusive dans des contextes politiques, où des outils tels que le clonage de voix et la manipulation d'images pourraient influencer indûment l'opinion publique.

- Le projet de lignes directrices vise également à établir des responsabilités claires et des mécanismes de contrôle pour les acteurs publics et privés, en soulignant l'importance de la transparence, de l'équité et de la responsabilité à tous les stades du développement et du déploiement de l'IA générative. Il vise également à fournir des orientations pratiques aux utilisateurs, en mettant l'accent sur la sensibilisation du public à l'impact potentiel de l'IA générative sur la liberté d'expression et en dotant les individus d'outils leur permettant de gérer ces risques. Cette approche reflète un engagement plus large en faveur de la protection des valeurs démocratiques et des droits humains à mesure que les technologies de l'IA continuent de progresser.

Principaux points soulevés par Peter KIMPIAN

- La Convention 108 est le seul accord multilatéral juridiquement contraignant sur la protection de la vie privée et des données, couvrant à la fois les secteurs public et privé, avec 55 États membres et des observateurs supplémentaires dans le monde entier. La convention fournit un cadre conçu pour sauvegarder les droits individuels à la vie privée et à la protection des données, dans le but de protéger la dignité humaine à l'ère numérique. Elle intègre des principes de protection des données tels que la nécessité, la proportionnalité, la transparence et le droit des individus à comprendre et à contester les décisions prises par le traitement automatisé des données, un aspect essentiel également pour réguler l'impact sociétal de l'IA.
- Les flux de données transfrontaliers posent des défis importants, avec un besoin pressant de normes internationales harmonisées pour empêcher le contournement des lois sur la protection des données, en particulier compte tenu de la prévalence du partage de données basé sur l'IA. La coopération entre les autorités chargées de la protection des données est essentielle pour atteindre les objectifs de la Convention 108(+) dans les contextes transfrontaliers. En vertu du traité, cette coopération est obligatoire et favorise une approche coordonnée de la protection des données qui s'aligne sur les normes mondiales de gouvernance des données.
- Les lignes directrices sur l'intelligence artificielle et la protection des données, élaborées par le Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (TP-D), mettent l'accent sur des politiques d'IA qui protègent la vie privée, préviennent la discrimination et défendent les valeurs démocratiques. Les lignes directrices sont structurées de manière à offrir des conseils aux décideurs politiques, ainsi qu'aux développeurs, fabricants et fournisseurs de services d'IA. Ces lignes directrices prônent le respect de la vie privée dès la conception dans le développement de l'IA, en intégrant une solide protection des données afin de minimiser les risques liés à la prise de décision automatisée et au profilage. Elles mettent également l'accent sur la transparence, en encourageant des explications claires sur les décisions prises en matière d'IA afin de renforcer la confiance du public et de permettre la participation démocratique.
- Un autre outil important est la Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme, qui vise à prévenir les abus de ces systèmes et à assurer une protection efficace des droits humains, de l'État de droit et de la démocratie. Elle fournit des orientations sur les obligations des États et les responsabilités des entreprises, en mettant l'accent sur des droits tels que le procès équitable, la protection de la vie privée et des données, la liberté de pensée, de conscience et de religion, les libertés d'expression et de réunion, l'égalité de traitement, ainsi que les droits économiques et sociaux. Mettant à jour une

recommandation antérieure pour l'aligner sur la Convention 108+ modernisée, elle souligne les préoccupations relatives au profilage, à la transparence et à la nécessité de protéger les groupes vulnérables. La recommandation appelle également à une approche juridiquement contraignante de la "prise en compte dès la conception".

Principaux points soulevés par Jan SPOENLE

- La Commission pour l'efficacité de la justice (CEPEJ) vise à renforcer l'efficacité et le fonctionnement du système judiciaire et à permettre une meilleure mise en œuvre des instruments juridiques internationaux. Le travail de la CEPEJ soutient l'utilisation de la technologie pour rationaliser les processus judiciaires, améliorer la formation juridique et gérer l'administration des tribunaux, en s'alignant sur l'équité et l'accessibilité de la justice. Au fil du temps, le lien étroit entre la protection de l'environnement et les droits sociaux est devenu de plus en plus évident, la dégradation de l'environnement affectant de manière significative divers droits sociaux, tels que le droit à la santé et le droit à des conditions de travail sûres et saines.
- En réponse aux défis éthiques et juridiques posés par l'IA dans les systèmes judiciaires, la CEPEJ a adopté en 2018 la Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires et leur environnement. La Charte répond aux préoccupations concernant les biais de l'IA, ainsi qu'aux craintes concernant la prédiction par l'IA des décisions de justice et la nature de « boîte noire » des algorithmes. Il s'agit du premier instrument européen à énoncer cinq principes fondamentaux pour l'utilisation de l'IA dans les contextes judiciaires : (1) le respect des droits fondamentaux, en veillant à ce que les outils d'IA soient conformes aux normes en matière de droits humains ; (2) la prévention de la discrimination ; (3) la qualité et la sécurité, en exigeant un traitement des données certifié et multidisciplinaire ; (4) la transparence, l'impartialité et l'équité, en exigeant des processus d'IA clairs et vérifiables ; et (5) le contrôle de l'utilisateur, en veillant à ce que les utilisateurs restent des acteurs éclairés et maîtres de leurs choix.
- Pour faciliter l'application pratique, la CEPEJ a introduit un outil d'évaluation en 2023. Cet outil vise à compléter la Convention-cadre et aide les organes judiciaires à mettre en œuvre les principes de la Charte éthique en proposant 29 questions directrices sur l'utilisation éthique de l'IA. Les questions sont destinées aux fonctionnaires judiciaires et aux responsables informatiques afin d'évaluer les risques potentiels, tels que le profilage des juges ou des jurés, avant de mettre en œuvre des systèmes d'IA dans le système judiciaire.
- Pour soutenir davantage l'utilisation éthique de l'IA dans la justice, la CEPEJ a mis en place un groupe de travail permanent sur la cyberjustice et l'IA, un conseil consultatif sur l'IA et le réseau européen de cyberjustice, qui offre une collaboration transfrontalière et des webinaires entre les experts judiciaires. La CEPEJ a également amélioré ses outils de suivi avec un nouveau chapitre sur les TIC dans son cadre d'évaluation et a lancé un centre de ressources sur la cyberjustice pour suivre les applications de l'IA dans les systèmes judiciaires européens. En 2023, elle a publié une note d'orientation sur l'IA générative à l'intention des professionnels de la justice, qui donne un aperçu de l'utilisation responsable de l'IA dans les contextes de travail.

Discussion

- Un membre du CDDH-IA s'est inquiété de l'évolution de la définition de l'OCDE des systèmes d'IA et du problème de la « boîte noire ». Marko GROBELNIK, dans sa réponse, a précisé que la définition n'abordait pas la question de l'explicabilité des modèles d'IA. En ce qui concerne les problèmes d'interprétabilité de l'IA, il a été expliqué que certains phénomènes de l'IA sont observables mais ne sont pas entièrement compris. Cette question, où l'IA opère dans son propre « langage » au-delà de la compréhension humaine, souligne la nécessité de cadres réglementaires nuancés qui peuvent tenir compte des résultats inconnus ou imprévisibles de l'IA avancée.
- Des questions ont également été posées sur la difficulté d'identifier les principaux domaines concernant les droits économiques et sociaux. Miriam KULLMANN a expliqué qu'outre l'identification du champ d'application matériel des droits, une solution pourrait consister à identifier les principaux domaines thématiques et à essayer de rassembler les différents droits humains plutôt que de les mettre en exergue séparément.
- Une question a été posée sur le rôle potentiel du droit au développement dans le contexte de l'IA. Alberto QUINTAVALLA a souligné le rôle des entreprises et la question des droits de propriété intellectuelle (PI) et de l'accès aux sources ouvertes, se demandant dans quelle mesure les protections de la PI pouvaient être adaptées pour promouvoir des avantages technologiques plus larges. L'une des approches suggérées est l'utilisation de licences obligatoires pour équilibrer les obligations internationales et les objectifs de développement. La réponse a souligné l'importance d'un équilibre prudent par le biais, par exemple, du principe de proportionnalité, où le droit au développement, et éventuellement le droit au progrès scientifique, sont mis en balance avec d'autres droits concurrents tels que le droit émergent à un environnement sain. Enfin, la discussion a souligné la nécessité d'adopter des approches réglementaires souples pour répondre à la diversité des applications et des impacts de l'IA dans les différentes régions.
- La discussion a également porté sur l'impact de l'IA sur les droits de l'enfant, en soulignant les initiatives en cours pour rendre les plateformes numériques plus sûres pour les mineurs. David LESLIE a rappelé les travaux du Conseil de l'Europe sur la façon dont les applications de médias sociaux et le temps d'écran intensif en bas âge peuvent affecter le développement des enfants. Il a souligné l'importance de fonder la conception adaptée à l'âge sur les principes des droits humains, insistant sur le fait que les effets plus larges et cumulatifs de l'IA sur les enfants devraient être évalués dans le cadre des droits humains.
- Un membre du CDDH-IA a posé une question sur le potentiel de l'IA à interpréter et à comprendre le langage qu'elle génère et sur le fait que l'IA pourrait commettre des abus de manière indépendante, sans intervention humaine. En réponse, Marko GROBELNIK a expliqué que les machines perçoivent le monde comme un réseau de concepts interconnectés. Il a noté que s'il est possible de décoder le « langage » de l'IA dans une certaine mesure, cela nécessite d'importantes ressources informatiques et reste imparfait. David LESLIE a mis en garde contre l'anthropomorphisation de l'IA. Il a insisté sur une perspective « déflationniste », affirmant que les systèmes d'IA à grande échelle sont fondamentalement des outils statistiques qui traitent des modèles dans des ensembles de données massifs, y compris des structures syntaxiques qui ressemblent au langage humain. Cette capacité n'a toutefois rien de mystérieux ; elle résulte simplement du traitement de grandes quantités de données. Ces systèmes, a-t-il affirmé, sont des outils quantitatifs, dépourvus de langage indépendant, de conscience de soi ou de conscience, et ne devraient pas être considérés comme des entités autonomes capables d'abuser intentionnellement de leurs capacités.

- Une question a été posée sur la relation potentielle entre le Manuel du CDDH sur l'IA et les droits humains et l'étude de l'impact sur les droits de l'homme, la démocratie et l'État de droit (HUDERIA), en particulier en ce qui concerne leurs publics cibles. En réponse, un expert a souligné que la définition du public cible est cruciale. Pour HUDERIA, l'un des principaux défis consiste à déterminer le langage approprié, la profondeur technique et l'emplacement des messages clés. Si le Manuel vise à aider les non-juristes à comprendre les implications de l'IA en matière de droits humains, il devrait s'aligner sur l'HUDERIA sur le fond tout en garantissant un langage accessible. Lors de la discussion sur la valeur ajoutée du Manuel par rapport aux travaux antérieurs du Comité ad hoc sur l'intelligence artificielle (CAHAI) et du Comité sur l'intelligence artificielle (CAI), Thomas SCHNEIDER a suggéré que, à l'instar du Guide des droits de l'homme pour les utilisateurs d'Internet, le Manuel pourrait aider les lecteurs à comprendre leurs droits dans le contexte de l'IA et donner un aperçu des recours disponibles. L'identification des droits à elle seule est insuffisante ; le Manuel devrait mettre l'accent sur l'accessibilité des recours et s'attaquer au problème de l'imprécision des voies de recours.