



TERCER EJERCICIO: TRADUCCIÓN DIRECTA

TEXTO PARA SU TRADUCCIÓN DEL INGLÉS AL ESPAÑOL

The Ethics of Artificial Intelligence (AI)

This Recommendation approaches AI ethics as a systematic normative reflection, based on a holistic, comprehensive, multicultural and evolving framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies on human beings, societies and the environment and ecosystems, and offers them a basis to accept or reject AI technologies. It considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and as rooted in the ethics of science and technology.

AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations.

Ethical questions regarding AI systems pertain to all stages of the AI system life cycle, understood here to range from research, design and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly and termination.

AI systems raise new types of ethical issues that include, but are not limited to, their impact on decision-making, employment and labour, social interaction, health care, education, media, access to information, digital divide, personal data and consumer protection, environment, democracy, rule of law, security and policing, dual use, and human rights and fundamental freedoms, including freedom of expression, privacy and non-discrimination. Furthermore, new ethical challenges are created by the potential of AI algorithms to reproduce and reinforce existing biases, and thus to exacerbate already existing forms of discrimination, prejudice and stereotyping. Some of these issues are related to the capacity of AI systems to perform tasks which previously only living beings could do, and which were in some cases even limited to human beings only. These characteristics give AI systems a profound, new role in human practices and society, as well as in their relationship with the environment and ecosystems, creating a new context for children and young people to grow up in, develop an understanding of the world and themselves, critically understand media and information, and learn to make decisions. In the long term, AI systems could challenge humans' special sense of experience and



agency, raising additional concerns about, inter alia, human self-understanding, social, cultural and environmental interaction, autonomy, agency, worth and dignity.

It should be recognized that AI technologies do not necessarily, per se, ensure human and environmental and ecosystem flourishing. Furthermore, none of the processes related to the AI system life cycle shall exceed what is necessary to achieve legitimate aims or objectives and should be appropriate to the context. In the event of possible occurrence of any harm to human beings, human rights and fundamental freedoms, communities and society at large or the environment and ecosystems, the implementation of procedures for risk assessment and the adoption of measures in order to preclude the occurrence of such harm should be ensured.

The choice to use AI systems and which AI method to use should be justified in the following ways: (a) the AI method chosen should be appropriate and proportional to achieve a given legitimate aim; (b) the AI method chosen should not infringe upon the foundational values, in particular, its use must not violate or abuse human rights; and (c) the AI method should be appropriate to the context and should be based on rigorous scientific foundations. In scenarios where decisions are understood to have an impact that is irreversible or difficult to reverse or may involve life and death decisions, final human determination should apply. In particular, AI systems should not be used for social scoring or mass surveillance purposes.

AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems. Effective remedy should be available against discrimination and biased algorithmic determination.

Furthermore, digital and knowledge divides within and between countries need to be addressed throughout an AI system life cycle, including in terms of access and quality of access to technology and data, in accordance with relevant national, regional and international legal frameworks, as well as in terms of connectivity, knowledge and skills and meaningful participation of the affected communities, such that every person is treated equitably.

Governments should adopt a regulatory framework that sets out a procedure, particularly for public authorities, to carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges. The assessment should also establish appropriate oversight mechanisms, including auditability, traceability and explainability, which enable the assessment of algorithms, data and design processes, as well as include external review of AI systems. Ethical impact assessments should be transparent and open to the public, where appropriate. Such assessments should also be multidisciplinary, multi-stakeholder, multicultural, pluralistic and inclusive. The public authorities should be required to monitor the AI systems implemented and/or deployed by those authorities by introducing appropriate mechanisms and tools.



TEXTO PARA SU TRADUCCIÓN DEL FRANCÉS AL ESPAÑOL

L'éthique de l'intelligence artificielle (IA)

La présente Recommandation aborde l'éthique de l'IA en tant que réflexion normative systématique, basée sur un cadre global, détaillé, multiculturel et évolutif de valeurs, de principes et d'actions interdépendants de nature à orienter les sociétés pour qu'elles apportent des réponses responsables aux conséquences connues et inconnues des impacts des technologies de l'IA sur les êtres humains, les sociétés, l'environnement et les écosystèmes, et leur offre une base pour accepter ou rejeter les technologies de l'IA. Elle considère l'éthique comme une base dynamique pour l'évaluation et l'orientation normatives des technologies de l'IA, en faisant de la dignité humaine, du bien-être et de la prévention des dommages des repères et en s'appuyant sur l'éthique des sciences et des technologies.

Les systèmes d'IA sont des technologies de traitement des informations qui intègrent des modèles et des algorithmes, lesquels génèrent une capacité d'apprentissage et d'exécution de tâches cognitives conduisant à des résultats tels que l'anticipation et la prise de décisions dans des environnements matériels et virtuels. Les systèmes d'IA sont conçus pour fonctionner avec différents degrés d'autonomie, au moyen de la modélisation et la représentation des connaissances, de l'exploitation des données et du calcul de corrélations.

Les questions éthiques concernant les systèmes d'IA se rapportent à toutes les étapes du cycle de vie de ces systèmes, compris ici comme allant de la recherche, la conception et le développement au déploiement et à l'utilisation, et incluant la maintenance, l'exploitation, la commercialisation, le financement, le suivi et l'évaluation, la validation, la fin de l'utilisation, le démontage et la mise hors service.

Les systèmes d'IA soulèvent de nouveaux types de questions éthiques qui comprennent, sans s'y limiter, leur impact sur la prise de décisions, l'emploi et le travail, les interactions sociales, les soins de santé, l'éducation, les médias, l'accès à l'information, la fracture numérique, la protection des consommateurs et des données personnelles, l'environnement, la démocratie, l'état de droit, la sécurité et la police, le double usage et les droits de l'homme et les libertés fondamentales, y compris la liberté d'expression, la protection de la vie privée et la non-discrimination. En outre, de nouveaux défis éthiques sont générés par la reproduction et le renforcement potentiels des biais existants par les algorithmes d'IA, ce qui exacerbe des formes déjà existantes de discriminations, de préjugés et de stéréotypes. Certaines de ces questions ont trait à la capacité des systèmes d'IA d'effectuer des tâches qu'auparavant seuls des êtres vivants pouvaient réaliser, parfois même uniquement des êtres humains. Ces caractéristiques confèrent aux systèmes d'IA un rôle déterminant et inédit dans les pratiques et les sociétés humaines, ainsi que dans leur relation avec l'environnement et les écosystèmes, créant un nouveau contexte dans lequel les enfants et les jeunes vont grandir, développer une compréhension du monde et d'eux-mêmes, comprendre de manière critique les



médias et l'information, et apprendre à prendre des décisions. Les systèmes d'IA pourraient, à long terme, disputer aux êtres humains le sentiment d'expérience et la capacité d'agir qui leur sont propres, ce qui susciterait de nouvelles inquiétudes quant à la compréhension qu'ont les êtres humains d'eux-mêmes, leurs interactions sociales, culturelles et environnementales, leur autonomie, leur capacité d'agir, leur valeur et leur dignité, entre autres.

Il faut reconnaître que les technologies de l'IA ne garantissent pas nécessairement, en soi, l'épanouissement des êtres humains ni le maintien d'un environnement et d'écosystèmes prospères. Qui plus est, aucun des processus liés au cycle de vie des systèmes d'IA ne devrait aller au-delà de ce qui est nécessaire pour atteindre des buts ou objectifs légitimes, et ces processus devraient être adaptés au contexte dans lequel ils interviennent. S'il s'avérait qu'un préjudice quelconque puisse affecter les êtres humains, les droits de l'homme et les libertés fondamentales, les communautés et la société en général, ou l'environnement et les écosystèmes, il conviendrait d'assurer la mise en œuvre de procédures d'évaluation des risques et l'adoption de mesures visant à empêcher qu'un tel préjudice ne survienne.

La décision de recourir à des systèmes d'IA et le choix de la méthode d'IA devraient être justifiés des manières suivantes: (a) la méthode d'IA retenue devrait être adaptée et proportionnée pour atteindre un objectif légitime; (b) la méthode d'IA retenue ne devrait pas porter atteinte aux valeurs fondamentales– en particulier, son utilisation ne devrait pas constituer une violation ou un abus des droits de l'homme; et (c) la méthode d'IA retenue devrait être adaptée au contexte et reposer sur des bases scientifiques rigoureuses. Dans les scénarios où les décisions sont considérées comme ayant un impact irréversible ou difficile à renverser, ou qui pourraient impliquer des décisions de vie et de mort, la décision finale devrait être prise par l'homme. En particulier, les systèmes d'IA ne devraient pas être utilisés à des fins de notation sociale ou de surveillance de masse.

Les acteurs de l'IA doivent faire tout ce qui est raisonnablement possible pour réduire au maximum et éviter de renforcer ou de perpétuer des applications et des résultats discriminatoires ou biaisés, tout au long du cycle de vie des systèmes d'IA, afin de garantir l'équité de ces systèmes. Un recours efficace devrait être disponible contre la discrimination et la détermination algorithmique biaisée.

En outre, les fractures numérique et cognitive, au sein des pays et entre eux, doivent être traitées tout au long du cycle de vie des systèmes d'IA, notamment en termes d'accès et de qualité d'accès aux technologies et aux données, conformément aux cadres juridiques nationaux, régionaux et internationaux pertinents, ainsi qu'en termes de connectivité, de connaissances et de compétences et de participation effective des communautés concernées, de sorte que chaque personne soit traitée équitablement.

Les gouvernements devraient adopter un cadre réglementaire qui définisse une procédure permettant en particulier aux autorités publiques de mener à bien des évaluations sur l'impact des systèmes d'IA afin d'anticiper les répercussions, d'atténuer les risques, d'éviter les conséquences préjudiciables, de faciliter la participation des citoyens et de faire face aux défis sociétaux. L'évaluation devrait



MINISTERIO
DE CIENCIA
E INNOVACIÓN



Tribunal de oposiciones nº 26

FINANCIACIÓN, EVALUACIÓN Y
SEGUIMIENTO ECONÓMICO Y
CIENTÍFICO-TÉCNICO DE
ACTIVIDADES DE I+D+I

également établir des mécanismes de supervision adaptés, notamment les principes de vérifiabilité, de traçabilité et d'explicabilité, permettant d'évaluer les algorithmes, les données et les processus de conception, et inclure un examen externe des systèmes d'IA. Les évaluations de l'impact éthique devraient être transparentes et ouvertes au public, selon le cas. Elles devraient par ailleurs être multidisciplinaires, multipartites, multiculturelles, pluralistes et inclusives. Les autorités publiques devraient être tenues de surveiller les systèmes d'IA qu'elles mettent en œuvre et/ou déploient, en établissant des mécanismes et des outils appropriés.