DRETS HUMANS I SOCIETAT DIGITAL

Intel·ligència artificial i drets humans: es necessiten mútuament?

Karlos A. Castilla Juárez



Agents de la policia de Bòsnia, a l'estació de tren de Bihać, fan tornar un grup de migrants cap a Sarajevo amb autobús, el 10 de novembre de 2018. Fotografia: Alessio Mamo

En parlar de la relació entre drets humans i intel·ligència artificial (IA), el més comú actualment és preguntar-se com la IA afecta els drets humans. De fet, hi ha un grup concret de drets humans que constantment s'assenyala com el més exposat a aquestes afectacions; per tant, la seva protecció en la governança i la regulació de la intel·ligència artificial es considera prioritària. És el cas de la privacitat i la protecció de dades, la llibertat d'expressió i l'accés a la informació, la igualtat i la no-discriminació.

Que aquests drets humans i no d'altres estiguin constantment al centre d'atenció sembla normal si s'entén que, per al seu disseny, desenvolupament i desplegament, la intel·ligència artificial necessita dades, que són la matèria primera que alimenta els algoritmes amb què es posa en marxa qualsevol IA. És per això que conèixer la manera com s'obtenen aquestes dades (si és amb autorització o sense), com s'emmagatzemen i com s'utilitzen, quina part s'utilitza o s'exclou, quines no es recullen, com es difonen o comparteixen, etc., són aspectes de gran importància. Al final, les dades són informació rellevant sobre la vida i les activitats que desenvolupem les persones.

Davant d'això, el que menys ens hem preguntat és si els drets humans necessiten la intel·ligència artificial. És a dir, si cal disposar de la IA per exercir alguns o tots els nostres drets humans, perquè és aquí on el nombre de drets humans susceptibles d'estar afectats per la intel·ligència artificial s'ampliaria, ja que no només estaríem parlant de temes entorn de les dades, sinó que es podria incidir en el gaudiment, exercici i garantia d'altres drets.

Així, per exemple, necessitem la intel·ligència artificial per garantir el dret a l'habitatge? Necessitem la IA per exercir la llibertat de circulació? Necessitem la IA per exercir i garantir l'accés a la justícia? La resposta a aquestes preguntes, si es pensa en la pretesa universalitat dels drets humans, no és senzilla i difícilment pot ser absoluta en sentit positiu o negatiu, per la qual cosa és important reflexionar entorn d'això.

Quina és la relació entre drets humans i intel·ligència artificial?

L'any 2023 bé podem afirmar que la relació entre drets humans i intel·ligència artificial és més d'un "veïnatge llunyà" que d'un "compromís mutu". Això és així perquè encara que drets humans i IA viuen al mateix món i es coneixen en termes genèrics, han interactuat poc de manera propera, malgrat els múltiples esforços, especialment des d'organismes internacionals de drets humans [1], perquè es coneguin millor i col·laborin mútuament.

Els qui treballem en l'àmbit dels drets humans, amb el pas dels anys —i en veure com la IA arribava a més àmbits de l'exercici, gaudiment i garantia d'aquests drets— ens hi hem anat apropant, moltes vegades sense entendre en detall i en profunditat tot el que implica. D'altra banda, els qui treballen en el sector de la intel·ligència artificial s'han interessat poc en els drets humans i han preferit mirar a l'ètica quan se'ls han qüestionat els efectes negatius que pot tenir la IA en la vida de les persones.

Per tot plegat, podem afirmar que la relació entre la IA i els drets humans encara està en construcció, malgrat la gran rellevància que té: pràcticament en tots els àmbits on s'està implementant la intel·ligència artificial hi ha un dret humà vinculat, directament o indirectament. Sí, aquesta afirmació no és exagerada. El problema és que moltes vegades no veiem aquesta vinculació perquè perdem de vista que, per exemple, en pretendre utilitzar un transport públic on hi ha aplicacions d'IA es podria estar incidint en l'exercici i garantia dels nostres drets humans a la llibertat de circulació, privacitat, no-discriminació o reconeixement de la personalitat jurídica, entre d'altres. I així en moltes més activitats quotidianes.

> La relació entre la intel·ligència artificial i els drets humans encara està en construcció, malgrat la gran rellevància que té: pràcticament en tots els àmbits on s'està implementant la intel·ligència artificial hi ha un dret humà vinculat

En aquest sentit, la relació entre els drets humans i la intel·ligència artificial és molt important i ha de ser prioritària. Cal que hi hagi més interacció entre la comunitat de drets humans i de la IA perquè el desenvolupament tecnològic no es creï sense la comunitat de drets humans. Especialment perquè l'ètica a la qual s'apel·la com a forma d'autoregulació des de la intel·ligència artificial (empreses, centres de recerca, acadèmies de ciències, així com alguns estats, organitzacions internacionals i associacions de la societat civil) és insuficient per prevenir i evitar els possibles efectes negatius de la IA en les persones.

Així, resulta evident que en la relació entre la IA i els drets humans queda molt per fer i que, en el futur proper, hauria de ser una relació de compromís mutu per tot el que s'explicarà a continuació.

Drets humans en l'àmbit de la intel·ligència artificial: on queda l'ètica?

Al món, qui més IA produeix són les empreses, entre les quals destaquen Google, Amazon, Apple, Microsoft, Meta, Baidu, Alibaba i Tencent. Això significa que el gruix de la intel·ligència artificial s'està produint als Estats Units i a la Xina, o sota interessos econòmics de persones amb aquestes nacionalitats. Per aquest motiu sembla evident que són aquestes empreses les que han anat establint a la pràctica la seva pròpia regulació i el marc d'actuació vinculat amb la IA. Això, sens dubte, influeix també en petites empreses i en tota la generació (o mercat) de la IA. Per aquest motiu, justament, des de molts àmbits vinculats amb sectors dominants de la IA (empreses, centres de recerca, acadèmies especialitzades i associacions vinculades amb tots els àmbits anteriors) es proposa insistentment l'ètica com a marc òptim de regulació que pot garantir drets i no posa en risc la llibertat d'empresa, la propietat intel·lectual i, en general, els seus legítims interessos econòmics, d'innovació o de negoci.

Si no fos així, no s'entendria per què Europa —on el poder econòmic i polític dels qui produeixen IA és encara menor— és l'única regió del món on, l'any 2023, s'està avançant en una regulació de la IA amb una mirada més pròxima als drets humans, abandonant les simples regulacions ètiques. És cert que hi ha obstacles, i que certs aspectes requeriran anàlisis més completes i ajustaments, especialment si un dels objectius és la protecció dels drets fonamentals de les persones.

En aquest context, l'ètica en la intel·ligència artificial que s'impulsa i es vol establir com a marc de regulació i prevenció d'afectacions a les persones es mostra com a canviant, mal·leable, parcial i incerta, i permet la impunitat en casos d'incompliment. Això és especialment així quan són les grans empreses, amb els seus interessos econòmics o de negoci, les que finalment estableixen aquesta ètica o les que en proposen els paràmetres i àmbits d'aplicació.

Davant d'això, tenim els drets humans, que, en primer lloc, ens permeten pressuposar que s'han resolt un cert conjunt de debats filosòfics per concretar-los i identificar-los. En segon lloc, representen acords mínims sobre valors universals (encara que no ho siguin tots) en

forma de drets i llibertats, per la qual cosa sabem amb molta certesa quins drets hi ha [2], quin abast tenen, on estan reconeguts i qui els ha de respectar, protegir i garantir.

A més, el suport mundial a aquests drets és bastant ampli: hi ha aspectes comuns en normes nacionals i internacionals de moltes regions del món. Així, aquestes empreses (i d'altres que també produeixen IA), com qualsevol altra persona, tenen responsabilitats davant els drets humans dels altres a partir del contingut que estableixen les lleis a escala nacional en matèria civil, laboral, administrativa i penal, i també amb les derivades de normes nacionals i internacionals de drets humans, encara que això es negui constantment i s'eludeixi per part de tota mena d'empreses; tant és així que hi va haver la necessitat de desenvolupar els principis rectors sobre les empreses i els drets humans de les Nacions Unides com un intent de reafirmar aquest conjunt de mínims a observar.

Així doncs, els drets humans en l'àmbit de la intel·ligència artificial han de tenir, en primer lloc, una incidència principal en tot allò relacionat amb les dades que utilitza la IA per als drets que es podrien veure afectats (la privacitat i la protecció de dades, la llibertat d'expressió i l'accés a la informació, la igualtat i la no-discriminació). En segon lloc, també han de tenir influència en els àmbits en què pot ser aplicada la IA, és a dir, respecte a les preguntes i respostes que els diferents algoritmes donaran com a resultat, ja que és aquí on es pot incidir en molts més drets.

Si ens fixem en la primera perspectiva vinculada amb les dades, per exemple, sempre serà millor aplicar els estàndards internacionals de drets humans que s'han anat creant, desenvolupant i actualitzant respecte al dret a la vida privada, ja que tenen uns paràmetres clars quan es tracta de persones públiques i persones comunes, i tenen definida quina part d'aquest dret no es pot transgredir en cap cas. Aplicar aquests estàndards és preferible a deixar que una empresa o persona valori o avaluï èticament què és el que es considera "èticament adequat" que se sàpiga, es divulgui o s'utilitzi de la vida privada d'una persona o d'un conjunt de persones sobre els quals s'han recollit una sèrie de dades amb el seu consentiment total o sense.

Des de la segona perspectiva, i com ja s'intenta fer en la proposta de regulació europea, és necessari definir els àmbits en què els drets humans es troben en risc evident i, per tant, també els paràmetres sobre els quals s'ha de limitar l'ús i el desenvolupament de la intel·ligència artificial. En el cas d'Europa, per exemple, a partir de l'establiment de "nivells de risc" [3], tots els sistemes d'IA que suposin una amenaça clara per a la seguretat, els mitjans de vida i els drets de les persones es consideren un risc inacceptable: des de la puntuació social per part dels governs fins a les joguines que fan ús d'assistents de veu, les quals fomenten comportaments perillosos. Fer valoracions ètiques *ad hoc* d'aquestes qüestions no garanteix uns mínims de certesa i seguretat jurídica, per més que aquestes valoracions es facin amb bona voluntat.

Així, a la Unió Europea, a partir de l'establiment d'àmbits d'alt risc, risc limitat i risc nul, s'ha proposat que es prohibeixi l'ús de la intel·ligència artificial en sistemes de categorització biomètrica que facin servir característiques personals sensibles com el gènere, la raça, l'ètnia, la religió o l'orientació política. Així mateix, estarien prohibits els

sistemes predictius de vigilància per avaluar el risc d'una persona o grup de persones de cometre un delicte o ofensa, així com els sistemes de reconeixement d'emocions per part tant de la policia com dels agents de fronteres, o en llocs de treball i escoles. De la mateixa manera, es limitarà el rastreig aleatori de dades biomètriques de xarxes socials o càmeres de vigilància per crear o ampliar bases de dades de reconeixement facial. Amb tot plegat, sembla clar que hi ha àmbits dels drets humans on clarament la IA ha d'estar vetada, per més que èticament algú pogués considerar admissible que hi entrés.

> L'ètica com a forma d'autoregulació de la intel·ligència artificial és insuficient per prevenir els efectes negatius en les persones

Des d'una perspectiva empresarial, de negoci, estrictament econòmica i mercantilista capitalista, l'establiment d'àmbits prohibits per a la IA es pot considerar una intromissió greu. D'aquí ve, en part, la insistència per apostar per les regulacions ètiques. Però des de la perspectiva de la protecció dels drets humans, és evident que hi ha uns mínims d'aquests drets i llibertats que han de ser innegociables, de la mateixa manera que també hi ha límits per no vulnerar els drets humans en molts altres àmbits al "món analògic" en què hem viscut fins ara.

La intel·ligència artificial en l'àmbit dels drets humans

En el sector de la intel·ligència artificial, els espais en què cal tenir en compte els drets humans semblen clars. Ara, però, és necessari analitzar una perspectiva menys explorada: en l'àmbit dels drets humans, cal que la IA hi sigui present? O, dit d'una altra manera, és necessària la IA per a l'exercici d'alguns o de tots els nostres drets humans?

Si prenem en consideració el fet que, a l'origen, els drets humans no van necessitar la intel·ligència artificial per al seu exercici, garantia i protecció, l'any 2023, i de cara al futur, no sembla imprescindible l'ús d'aquesta tecnologia per a aquestes finalitats, tret que la intel·ligència artificial permeti assegurar la pretesa universalitat dels drets humans que, 75 anys després de l'aprovació de la Declaració Universal dels Drets Humans, continua sent un tema pendent a totes les regions del món.

En tot cas, determinar si la intel·ligència artificial és necessària per a l'exercici, garantia i protecció dels drets humans no pot partir d'una conclusió general per a tots els drets, sinó que requereix l'anàlisi de cadascun dels drets reconeguts en els àmbits nacional i internacional per determinar, a partir dels seus elements i caracterització, si la IA és necessària o útil. Fins i tot requereix una anàlisi més particular del lloc o el país en qüestió, ja que hi ha llocs al món on l'accés a internet no està garantit, amb la qual cosa pensar en la IA seria multiplicar les desigualtats [4].

L'anàlisi particular dels drets és molt important perquè, per exemple, no és el mateix l'exercici del dret a la protesta pacífica que l'exercici del dret a l'habitatge. Protestar ho

podríem fer amb impactes idèntics als del món analògic o, fins i tot, es podria pensar, en alguns casos, en efectes més focalitzats, si es fa en entorns virtuals o digitals, mentre es compleixi amb les finalitats del dret. Però sempre seria insuficient disposar d'un habitatge en aquests mateixos entorns si en la realitat analògica no se'n té o si ni tan sols s'hi pot accedir perquè no podem utilitzar l'aplicació que ens ho permetria, o perquè l'aplicació esmentada ens exclou sense cap justificació raonable per les característiques i dades amb les quals funciona l'algoritme.

En aquest sentit, és rellevant assenyalar que el fet que no existeixi una intel·ligència artificial determinada per a l'exercici, la garantia i la protecció d'un dret específic no s'ha de considerar *per se* com un aspecte negatiu. Hem d'assumir i entendre que no tots els drets humans necessiten l'ús de la IA per ser efectivament exercits, garantits i protegits. Més aviat el que passa és que si al món analògic hi ha persones que estan sent excloses o no poden exercir un dret determinat, amb molta probabilitat també serà així amb l'ús d'aplicacions d'intel·ligència artificial.

Fixem-nos, per exemple, en l'accés a la justícia. Al món real analògic, actualment moltes persones no tenen garantit de manera efectiva aquest dret per múltiples raons, fins i tot en països anomenats desenvolupats [5], per la qual cosa la implementació d'una "justícia digital" o de "tribunals automatitzats" no és en si mateixa una solució a aquest problema. Encara menys si l'accés a aquestes aplicacions digitals ha de fer front a les mateixes barreres d'accés que la justícia analògica [6]. Per tant, molt possiblement tan sols és una bona opció per als qui ja accedien a la "justícia analògica".

Però, fins i tot així, tampoc suposa necessàriament una solució per als qui ja accedien a la justícia si aquesta "justícia digital" és incapaç d'assegurar les garanties judicials que es requereixen en tot procés (la imparcialitat del jutge, la publicitat del procés, la possibilitat d'assistència d'un advocat, la prohibició de dilacions indegudes i la utilització dels mitjans de prova disponibles, entre d'altres). Així doncs, hem d'anar amb compte amb allò que moltes vegades es promou com a "modernitat", perquè no necessàriament implica accedir als drets humans, exercir-los i respectar-los de manera efectiva.

> És important decidir l'ús de la intel·ligència artificial pensant en els drets humans. Si només pensem en la modernitat, hi ha moltes probabilitats que simplement s'estiguin creant formes modernes d'exclusió

Davant d'aquesta situació, és important decidir l'ús de la intel·ligència artificial pensant en els drets humans, per valorar si la IA aporta o no alguna cosa millor al que ja hi ha al món analògic per al seu exercici, garantia i protecció. És a dir, cal no tenir por de dir que no a una aparent modernitat que no aportarà res de positiu als drets humans, i encara menys quan qui l'ofereix són empreses a qui importa el seu negoci i no la universalitat dels drets humans.

En l'exemple anterior, hi pot haver aspectes concrets d'aquest accés a la justícia o tutela judicial en els quals la IA resulti útil, com podria ser, per exemple, la gestió d'expedients, l'accés a aquests expedients, la gestió d'informació o la solució de problemes simples [7]. Però això no significa que sigui útil per a tot el que implica i ha de contenir el dret.

Tampoc no podem esperar que tota decisió respecte a un dret humà recaigui en la intel·ligència artificial, ja que, per més dades que hi hagi, la intervenció humana sempre serà rellevant en l'assignació individual de drets i llibertats. Per tot plegat, sembla innegable que la IA pot ajudar amb consells, suggeriments i paràmetres a l'hora de tenir una millor comprensió de situacions o aconseguir informació rellevant, però la decisió final no ha de dependre exclusivament d'aquesta tecnologia.

Per determinar si necessitem la IA per a l'exercici, garantia i protecció d'un dret específic o d'una llibertat concreta, és recomanable plantejar-se, almenys, les preguntes següents:

- Què aporta la IA al *dret X* que vagi més enllà del seu exercici, garantia o protecció en la realitat analògica?
- Quin nivell d'accés efectiu té en aquests moments el *dret X* en la realitat analògica?
- L'ús de la IA millora l'accés o exercici del *dret X* i no perpetua les exclusions existents (o, fins i tot, en crea de noves)?
- La IA és capaç de garantir tots els elements i caracteritzacions que se li han donat al *dret X* en els estàndards de drets humans?
- Quin és el nivell de decisió que es delegarà a la intel·ligència artificial per a l'exercici, garantia o protecció del *dret X* i, per tant, quin és el nivell de decisió que tindran les persones?

Si abans de decidir que el dret a la salut, el dret al treball, el dret a l'educació, la llibertat d'expressió, el dret d'accés a la informació, el dret a la nacionalitat o la llibertat de circulació s'exerciran, es garantiran o es protegiran per mitjà de la intel·ligència artificial ens plantegem les preguntes anteriors, possiblement es podran reduir o evitar noves afectacions als drets humans. Altrament, si només pensem en la "modernitat", hi ha moltes probabilitats que simplement s'estiguin creant formes "modernes" d'exclusió i vulneració de drets humans.

Conclusió

Respecte als drets humans, el que ens pot preocupar de qualsevol tecnologia, no només de la intel·ligència artificial, és, d'una banda, el procés de creació i disseny de tecnologies i, de l'altra, els usos, aplicacions, creacions, efectes o resultats de les tecnologies que excloguin, restringeixin, discriminin, anul·lin, menyscabin, impedeixin o redueixin el reconeixement,

gaudiment o exercici, en condicions d'igualtat, dels drets humans i les llibertats fonamentals [8].

Des d'aquesta perspectiva, és important no perdre de vista els drets humans en la intel·ligència artificial, en la mateixa mesura que la IA en l'àmbit dels drets humans, ja que en totes dues direccions es produeixen incidències, encara que no totes siguin sempre negatives.

Si l'objectiu és salvaguardar els drets i llibertats de les persones, la realitat actual ens mostra que els drets humans són més necessaris en l'àmbit de la intel·ligència artificial que no pas la IA en l'àmbit dels drets humans.

REFERÈNCIES I NOTES

1 — Vegeu, per exemple: Comissària de Drets Humans del Consell d'Europa (2023). *Human rights by design future-proofing human rights protection in the era of AI*. Estrasburg: Consell d'Europa.
2 — Risse, Mathias (2018). "Human Rights and Artificial Intelligence: An Urgently Needed Agenda". *HKS Faculty Research Working Paper Series.* RWP18-015, p. 10.
3 — Comissió Europea (2020*)*. *White Paper on Artificial Intelligence: a European approach to excellence and trust*. Brussel·les.
4 — Castilla, Karlos (2022). *Cuatro ángulos de análisis de la igualdad y la no discriminación en la inteligencia artificial*, Barcelona: Institut de Drets Humans de Catalunya, p. 6.
5 — Deseau, Arnaud; Levai, Adam; Schmiegelow, Michèle (2019). "Access to Justice and Economic Development: Evidence from an International Panel Dataset". *LIDAM Discussion Papers IRES*, 2019009. Lovaina: Universitat Catòlica de Lovaina, Institut de Recherches Economiques et Sociales (IRES).
6 — Castilla, Karlos (2012). *Acceso efectivo a la justicia: elementos y caracterización*. Mèxic: Porrua
7 — Reiling, Dory (2020). "Courts and Artificial Intelligence". *International Journal for Court Administration 8.* 11(2)*,* p. 10.
8 — Pont, Anna; Passera, Agostina; Castilla, Karlos (2022). "Análisis introductorio"*. Impactos de las nuevas tecnologías en los derechos humanos*. Barcelona: Institut de Drets Humans de Catalunya, p. 11.



Karlos A. Castilla Juárez
Karlos A. Castilla Juárez és advocat per la Universitat Nacional Autònoma de Mèxic i doctor en Dret per la Universitat Pompeu Fabra. És responsable de recerca de l'Institut de Drets Humans de Catalunya, membre de l'Observatori de Dret Públic de la Universitat de Barcelona i professor associat de Dret Internacional dels Drets Humans a la Universitat Pompeu Fabra. La seva activitat investigadora s'ha desenvolupat en l'àmbit del dret internacional (sistemes internacionals de drets humans i litigi internacional), del dret constitucional (control de convencionalitat i sistemes d'accés a la justícia), drets de les persones migrants (detenció migratòria, expulsió, migració i canvi climàtic) i temàtiques específiques de drets humans (racisme, igualtat, no-discriminació, interseccionalitat, llibertat d'expressió i justícia transicional).