EAPC - Gestió de la informació: transparència i protecció de dades

La transparència exigible a la intel·ligència artificial

04-09-2023

En una societat democràticament avançada, esdevé necessari convenir els requeriments, les regles, els criteris i els principis que han de regir l'ús dels algoritmes en la presa de decisions. Aquesta necessitat esdevé especialment rellevant, òbviament, quan es tracta de la presa de decisions per part de les institucions públiques, atesos els interessos que representen i l'afectació que les decisions públiques comporten als drets i deures de la ciutadania.

Quan parlem de regular l'ús de sistemes d'intel·ligència artificial (IA), ens referim no només a la necessitat de, primerament, decidir fins a quin punt volem delegar-los les nostres decisions –és a dir, aclarir en quins casos, i en quina mesura, farem ús de la IA per prendre decisions–, sinó també, alhora, a la conveniència de concretar quines mesures de control i informació hem de garantir i assumir a l'hora d'emprar la IA en els procediments de presa de decisions, control i informació que esdevenen cabdals per tal de minimitzar i gestionar els riscos i biaixos inherents als sistemes automatitzats de decisió; perquè, en paraules de Rabelais, "el món digital sense consciència no és més que una ruïna per a l'ànima". L'emergència de la IA, el seu potencial i utilitats indiscutibles i, al mateix temps, els perills i temors que innegablement suscita, proliferen en un context de força inseguretat jurídica i manca de consens pel que fa al seu ús. Tot i així, per moments s'hi estan assolint grans fites, amb l'objectiu d'establir-ne un control i seguiment òptims.

En aquest sentit, cal destacar els cinc principis fonamentals que, d'acord amb La déclaration de Montréal pour un développement responsable de l'intelligence artificielle, de la Universitat de Mont-real (2018), han de regir l'ús dels algoritmes en la presa de decisions:

1. Transparència: les decisions que empren IA han d'estar justificades i ser intel·ligibles per a tothom, de tal manera que esdevé fonamental exigir la mateixa difusió i coneixement de la justificació i l'explicació dels factors que motiven la decisió respecte de la que requerim a l'ésser humà.
2. Equitat: els sistemes automatitzats han de ser dissenyats de manera que no generin ni incentivin la generació de tractes discriminatoris entre les persones.
3. Seguretat: els sistemes intel·ligents han de satisfer criteris de fiabilitat, seguretat i integritat.

4. Responsabilitat: els éssers humans ha de ser responsables en el desenvolupament i la utilització de les màquines intel·ligents, que en cap cas els poden eximir de les conseqüències exigibles un cop presa una decisió.
5. Intimitat i protecció de la vida privada de les persones, que cal preservar, amb caràcter fonamental, també en l'ús de la IA.

Des del punt de vista normatiu, convé fer una especial referència a la proposta de Reglament del Parlament europeu i del Consell pel qual s'estableixen normes harmonitzades en matèria d'intel·ligència artificial (llei d'intel·ligència artificial –LIA–), pel qual la Comissió proposa un marc reglamentari sobre l'ús de sistemes d'IA amb els objectius específics de garantir que els sistemes d'IA introduïts i usats al mercat de la Unió Europea siguin segurs i respectin la legislació vigent en matèria de drets fonamentals i valors de la Unió, entre d'altres. Aquesta proposta és part integrant d'un paquet de mesures que tracten els problemes derivats del desenvolupament i l'ús de la IA, i són objecte d'anàlisi detallada al Llibre blanc sobre IA de la Comissió Europea, de febrer de 2020.

L'objectiu de reglamentar la IA de manera harmonitzada va néixer el 25 d'abril de 2018 amb l'establiment per la Comissió Europea de l'estratègia sobre IA –COM(2018) 237–, juntament amb la creació d'un grup d'experts i la publicació la Comunicació COM(2019) 168, que acollia els set requisits essencials considerats a les directrius emeses pel grup d'experts referit:

1. Acció i supervisió humanes
2. Solidesa tècnica i seguretat
3. Gestió de la privadesa i de les dades
4. Transparència
5. Diversitat, no-discriminació i equitat
6. Benestar social i mediambiental
7. Rendició de comptes.

Les mesures i cauteles que imposa el marc normatiu harmonitzat han de ser en tot cas proporcionades per assolir els objectius indicats, en la mesura que imposa càrregues només quan és probable que un sistema d'IA comporti riscos per als drets fonamentals i la seguretat.

En aquesta línia, el principi de transparència algorítmica, que respon a la necessitat d'explicabilitat pública davant la ciutadania –usuaris però també auditors i autoritats supervisores– a mode d'eina per abordar l'opacitat dels sistemes basats en algoritmes, ha de regir sempre de forma proporcional, és a dir, depenent de l'impacte i del risc de l'ús dels sistemes d'IA en cada cas, de manera que, com més risc, més transparència exigible.

La transparència esdevé, doncs, un requeriment cabdal en l'establiment i l'ús dels sistemes d'IA, perquè lluny de presentar-se com humans davant dels usuaris, cal garantir que aquests tenen dret a saber que estan interactuant amb un sistema d'IA, el qual, per tant, ha de ser identificable com a tal. Alhora, cal oferir a l'usuari la possibilitat de decidir si prefereix interactuar amb un sistema d'IA o amb un altre mitjà alternatiu, si escau, per tal de garantir el respecte dels drets fonamentals.

Aquesta **obligació bàsica de transparència en els sistemes d'IA**, consistent a haver d'informar que un determinat contingut s'ha generat per mitjans automatitzats, és exigida a la proposta de norma europea (LIA) en nombrosos casos, fins i tot davant sistemes d'IA que no són catalogats d'alt risc, sempre que, però, aquests sistemes (i) es destinin a interactuar amb persones físiques; (ii) s'utilitzin per detectar o reconèixer emocions; (iii) s'emprin per determinar l'associació a

categories –socials– concretes a partir de dades biomètriques, o (iv) generin o manipulin continguts d'imatges, arxius d'àudio o vídeo, de tal manera que s'assemblin notablement a persones, objectes, llocs o altres entitats o successos existents i pugui induir erròniament una persona a pensar que són autèntics o verídics (ultrafalsificació) –art. 52 de la LIA. És el cas, per exemple, de sistemes d'IA que poden comportar riscos específics de suplantació o falsificació, independentment, doncs, de si són classificats com d'alt risc o no.

Es tracta de garantir en tots aquests casos que les persones usuàries siguin informades en tot moment de la circumstància d'estar interactuant amb un sistema de IA. Però d'aquests deures generals d'informació s'exceptuen els sistemes d'IA autoritzats per llei per a finalitats de detecció, prevenció, investigació o enjudiciament d'infraccions penals.

En el cas particular dels **sistemes d'IA classificats com d'alt risc** per a la salut i la seguretat o els drets fonamentals de les persones –art. 6 i annexos II i III de la LIA–, la necessitat de transparentar-ne informació s'incrementa exponencialment, atesa la seva especial singularitat, per la qual cosa la LIA exigeix en aquests casos obligacions superiors i addicionals de transparència:

- Primerament, preveu un deure dels proveïdors de sistemes d'IA d'alt risc de registrar aquests sistemes en una base de dades de la UE que gestionarà la Comissió (art. 60).
- Paral·lelament es requereix, en el seu disseny i desenvolupament, un nivell d'accés a informació suficient perquè els usuaris interpretin i usin correctament la seva informació de sortida –generada–, de tal manera que els sistemes han d'anar acompanyats d'informació concisa, completa, correcta, clara, accessible i entenedora per als usuaris sobre les instruccions d'ús, en un format digital o d'un altre tipus adequat. L'article 13 en concreta el contingut a publicar. Prenent com a base aquest precepte, entenem que la informació a difondre en relació amb els sistemes d'alt risc hauria de comprendre els extrems següents:

Dades del proveïdor

S'ha d'informar de la identitat i les dades de contacte del proveïdor i, si escau, del seu representant autoritzat.

Es tracta de poder conèixer l'empresa que dissenya i desenvolupa el sistema i el posa en el mercat per usar-lo, és a dir, un "punt de contacte" o responsable del sistema a qui poder adreçar-se.

Informació sobre el funcionament del sistema d'IA d'alt risc

Sobre les característiques, capacitats i limitacions del seu funcionament, en particular, caldria fer-hi constar les informacions següents:

- 1. La finalitat prevista de l'algoritme

Esdevé necessari descriure per a què s'ha dissenyat una eina d'IA i per a què serveix, és a dir, les seves finalitats generals i concretes. Precisament aquestes finalitats previstes són les que en determinen el risc associat.

En aquest context, la doctrina aposta per entendre que entre les exigències de la transparència cal que les administracions donin explicacions sobre com i amb quins criteris s'adopten decisions sobre l'ús dels sistemes algorítmics i sobre per què la seva implementació constitueix la millor alternativa enfront d'altres possibles, no algorítmiques.

- **2. Les especificacions relatives a les dades d'entrada, o qualsevol altra informació pertinent en relació amb els conjunts de dades d'entrenament, validació i prova, tenint en compte la finalitat prevista del sistema d'IA**

Conèixer la procedència o la font, l'abast i les característiques d'aquest tipus de dades emprades en la configuració, l'aprenentatge i la presa de decisions, constitueix un element clau, en la mesura que incideixen en la qualitat del sistema i permeten controlar els biaixos, els errors o les discriminacions.

- **3. Les especificacions tècniques del sistema d'IA**

La proposta de reglamentació europea es limita a exigir un deure de facilitar informació sobre les garanties de seguretat i equitat del sistema d'IA, és a dir, "el nivell de precisió, solidesa i ciberseguretat" i sobre "les circumstàncies conegudes o previsibles que podrien afectar aquest nivell esperat o donar lloc a riscos per a la salut i la seguretat o els drets fonamentals".

Més enllà del tenor literal de la proposta normativa europea, caldria fer públics els següents continguts d'informacions:

- Descripció del disseny i funcionament del sistema d'IA.
- Nom comercial del sistema, la seva comercialització, dades del certificat emès i la declaració UE de conformitat.
- Instruccions d'ús electròniques.
- Determinació del nivell de risc.
- Informació sobre el tipus de models algorítmics i algoritmes emprats (mòduls, grau de maduresa), és a dir, de l'arquitectura del sistema o funcionament intern.
- Codi font, programa o programari que acompanya el sistema d'IA i la documentació tècnica relacionada. Cal dir que l'accés a aquestes dades no comporta necessàriament, de fet, més transparència ni que sigui més adequada, per la qual cosa la seva difusió hauria d'anar acompanyada d'una descripció de l'algoritme en llenguatge natural i comprensible.

Tota aquesta informació tècnica acostuma a quedar lluny de la comprensió per part del públic en general, per la qual cosa s'adreça especialment als usuaris del sistema i a les autoritats supervisores.

- **4. El funcionament del sistema pel que fa a les persones o grups de persones en relació amb els quals es vol utilitzar el sistema**

En aquest extrem, la UNESCO, a la Recomanació 39, es refereix a la necessitat d'informar sobre els "factors que influeixen en una predicció o decisió específica", és a dir, de les regles i instruccions en què es basa l'algoritme, el seu tractament i el seu impacte. Es tracta d'informar, doncs, de quina manera i en quina mesura l'eina d'IA s'integra en el procés de presa de decisió i, per tant, pot afectar les persones destinatàries.

Informació sobre les mesures de vigilància i supervisió humana

La LIA es refereix al deure d'informar sobre "les mesures tècniques establertes per facilitar la interpretació de la informació de sortida dels sistemes d'IA per part dels usuaris", de tal manera que els usuaris puguin entendre les capacitats i limitacions del sistema.

En qualsevol cas, cal tant mostrar com i de quina manera l'ésser humà revisa o verifica la decisió automatitzada i, per tant, supervisa l'ús de l'eina d'IA, com també motivar les decisions públiques que se separin del criteri proposat per un sistema intel·ligent.

Informació sobre les mesures de manteniment i cura necessàries per garantir el funcionament correcte del sistema d'IA d'alt risc, incloent-hi la vida útil prevista del sistema i l'actualització del programari
La norma europea proposada inclou informació sobre l'anàlisi i les accions que s'han implementat per a la mitigació de riscos i perills, i sobre sistemes d'avaluació, d'anàlisi d'impacte, manteniment i revisió. Al·ludeix també a donar explicació dels "canvis en el sistema d'IA d'alt risc i el seu funcionament predeterminats pel proveïdor en el moment d'efectuar l'avaluació de la conformitat inicial", si escau.

En aquest sentit, convindria facilitar informació sobre els estudis d'impacte efectuats, a través d'un enllaç a l'avaluació o d'un resum amb el detall de la descripció i la data (Recomanació de la UNESCO, 2021), així com acreditar, a través de l'accés a informes i dictamen emesos, que els sistemes d'IA han superat l'auditoria que en verifiqui un funcionament adequat.

Pel que fa al sector públic, és pràcticament unànime el reconeixement de la dificultat de conèixer els sistemes d'IA que estan utilitzant les autoritats públiques per assistir les decisions públiques o donar-hi suport, especialment en fases prèvies, de tràmit i preparatòries, la qual cosa fa pràcticament impossible poder conèixer en quina mesura els algoritmes determinen o condicionen les decisions i polítiques públiques. Prova d'això és que les lleis de transparència, estatal o catalana, ometen cap referència a aquest respecte.

Conscient d'aquesta mancança i de la necessitat de donar una clara traçabilitat i transparència a l'ús de sistemes algorítmics en qualsevol actuació administrativa, l'article 13 del Reial decret 203/2021, de 30 de març, pel qual s'aprova el Reglament d'actuació i funcionament del sector públic per mitjans electrònics, en desenvolupament de l'article 41 de la Llei 40/2015, d'1 d'octubre, de règim jurídic del sector públic, preveu, tot i que només en l'àmbit estatal, que la determinació d'una actuació administrativa com a automatitzada s'autoritza per resolució del titular de l'òrgan administratiu competent per raó de la matèria o de l'òrgan executiu competent de l'organisme o entitat de dret públic, que s'ha de publicar a la seu electrònica o a la seu electrònica associada.

Cal posar especialment en relleu la previsió normativa pionera a l'Estat espanyol que ha comportat a aquest respecte l'article 16.1.*l* de la Llei 1/2022, de 13 d'abril, de Transparència i Bon Govern de la Comunitat Valenciana, que exigeix a les administracions valencianes fer publicitat activa, d'acord amb els principis de transparència i explicabilitat, de la relació de sistemes algorítmics o d'IA que tinguin impacte en els procediments administratius o la prestació dels serveis públics. Entre les informacions que ha de contenir la relació a publicar, el precepte exigeix fer-hi constar una descripció comprensible del seu disseny i funcionament, el nivell de risc que impliquen i el punt de contacte al qual poder adreçar-se en cada cas.

A la Generalitat de Catalunya cal destacar la creació d'un registre d'eines d'IA en matèria de serveis assistencials a l'Observatori d'IA en Salut, que ha permès la identificació de més de 100 algorismes d'IA del sistema sanitari català (SISCAT) i els centres de recerca de Catalunya, en el marc del Programa per a la promoció i desenvolupament de la IA al sistema de salut (Resolució SLT/954/2023, de 19 de març). Així és vol millorar el servei al ciutadà amb l'obtenció de més eines de suport a la prevenció, la predicció, el diagnòstic i el tractament, garantint l'equitat a tot el

país, la transparència, la seguretat i l'ètica en el tractament de les seves dades. Un dels seus objectius consisteix precisament a assegurar que les solucions desenvolupades es trobin disponibles per a tothom.

Per la seva banda, el Departament de Governança Pública i Autogovern del País Basc, per mandat parlamentari, està treballant en l'elaboració d'un manifest ètic de la dada per garantir un ús intel·ligent, ètic, segur, respectuós i no discriminatori de les dades i els algoritmes que gestiona l'Administració, per la qual cosa sembla que molt probablement disposarà en breu del primer registre d'algorismes de l'Estat a escala autonòmica.

En definitiva, resta molt camí a fer, especialment pel que fa a l'ús dels sistemes automatitzats per part de les administracions públiques en exercici de llurs funcions que, atesos els perills potencials, requereixen més garanties de control, explicabilitat i transparència i que siguin més exigents.

Òliver Garcia Muñoz

Subdirector general de Transparència i Grups d'Interès, de la DG de Bon Govern, Innovació i Qualitat Democràtiques, del Dept. de la Presidència