



El Reglament d'IA de la UE no aconsegueix establir un “estàndard d’or” per als DDHH

Dimecres 3 d'abril de 2024

Durant els últims tres anys, un ampli grup d'organitzacions de drets humans i de justícia global hem treballat en coalició per exigir que la intel·ligència artificial (IA) funcioni per a les persones i doni prioritat a la protecció dels drets humans fonamentals. Per això, vam presentar la [nostra visió col·lectiva](#) per a un enfocament on l'expressió "IA centrada en l'ésser humà" no sigui només una paraula de moda, on les [persones en moviment siguin tractades amb dignitat](#), i on les legisladores siguin prou valentes per establir línies vermelles contra els [usos inacceptables](#) dels sistemes d'IA.

Després d'un [procés de negociació esgotador](#), es preveu que les institucions de la UE adoptin definitivament el Reglament d'IA l'abril de 2024. Però mentre celebren, nosaltres prenem una postura molt més crítica, destacant les moltes oportunitats

perdues per assegurar que els nostres drets a la privacitat, la igualtat, a la no discriminació, a la presumpció d'innocència i molts altres, així com llibertats estiguin protegits en matèria d'IA. Aquí teniu el nostre resum de com la llei final es compara amb les nostres demandes col·lectives.

Si us plau, tingueu en compte que aquesta anàlisi es basa en la [versió més recent disponible del text del Reglament d'IA](#), datada del 6 de març de 2024. Encara podrien haver-hi petits canvis abans de l'adopció final de la llei.

En primer lloc, vam reclamar a les legisladores de la UE a empoderar les persones afectades mantenint un marc de responsabilitat, transparència, accessibilitat i reparació. Com ho han fet?

S'han superat algunes barreres d'accessibilitat, però cal fer més:

L'article 16 (ja) del Reglament d'IA compleix amb la nostra petició de protegir el dret a l'accessibilitat en certa manera, ja que afirma que els sistemes d'IA d'alt risc han de complir amb els requisits d'accessibilitat. No obstant això, encara creiem que la garantia s'hauria d'estendre per aplicar-se també als sistemes d'IA de risc baix i mitjà, per tal de garantir que les necessitats de les persones amb discapacitats siguin centrals en el desenvolupament de tots els sistemes d'IA que podrien impactar-les.

Més transparència sobre certs desplegaments d'IA, però grans buits per al sector privat i les agències i cossos de seguretat:

El Reglament d'IA estableix una base de dades de la UE accessible al públic per proporcionar transparència sobre els sistemes d'IA que representen major risc per als drets o la seguretat de les persones. Encara que originalment només els proveïdors de sistemes d'IA d'alt risc estaven subjectes a aquests requisits de transparència, vam persuadir amb èxit els responsables de la presa de decisions per a incloure que els implementadors del sistema d'IA – aquells que realment utilitzen el sistema – també estiguessin subjectes a obligacions de transparència.

Aquests proveïdors i implementadors estaran subjectes a obligacions de transparència quan posin en el mercat o utilitzin sistemes d'IA en àrees d'alt risc – com en les àrees d'ocupació i educació – tal com designa l'Annex III.

Es requerirà que els proveïdors registrin el seu sistema d'alt risc a la base de dades i que introdueixin informació sobre aquest, com la descripció del seu propòsit previst, la descripció concisa de la informació utilitzada pel sistema i la seva lògica operativa. Els implementadors de sistemes d'IA d'alt risc que siguin autoritats públiques – o aquells que actuïn en nom seu – estaran obligats a registrar l'ús del sistema. Es requerirà que introdueixin informació a la base de dades com un resum dels resultats d'una avaluació d'impacte en els drets fonamentals (FRIA en anglès) i un resum de l'avaluació d'impacte sobre la protecció de dades. No obstant això, no es requerirà que els desplejadors de sistemes d'IA d'alt risc en l'àrea del sector privat registrin l'ús de sistemes d'alt risc – un altre tema crític.

La principal mancança de la base de dades de la UE és que els negociadors van acordar una excepció per a les autoritats i forces d'aplicació de la llei, migració, asil i control de fronteres. Se sol·licitarà als proveïdors i desplejadors de sistemes d'alt risc en aquestes àrees que registrin només una quantitat limitada d'informació, i només en una secció de la base de dades no accessible al públic. Certes peces d'informació importants, com les dades d'entrenament utilitzades, no es revelaran en absolut. Això impedirà que les persones afectades, la societat civil, periodistes, organitzacions de vigilància i acadèmiques exerceixin escrutini públic en aquestes àrees d'alt risc que són propenses a la violació dels drets fonamentals i els facin responsables.

Les avaluacions d'impacte en els drets fonamentals estan incloses, però persisteixen les nostres preocupacions sobre la seva utilitat:

Vam convèncer amb èxit a les institucions de la UE de la necessitat d'avaluacions d'impacte en els drets fonamentals (FRIAs). No obstant això, basant-nos en el text final del Reglament, tenim dubtes sobre si realment evitaran violacions dels drets humans i seran útils com a eines significatives de responsabilitat.

Veiem tres mancances principals:

Falta d'avaluació significativa i d'obligació de prevenir impactes negatius: encara que les noves normes requereixen que els implementadors de sistemes d'IA d'alt risc llistin els riscos cap a les persones, no hi ha una obligació explícita d'avaluar si aquests riscos són acceptables en l'àmbit legal dels drets fonamentals, ni de prevenir-los sempre que sigui possible. Lamentablement, els implementadors només han d'especificar quines mesures es prendran un cop els riscos es materialitzin, probablement un cop ja s'hagi comès la vulneració.

No hi ha obligació de participació de col·lectius i agents interessats: el requisit de fer participar agents externs, inclosa la societat civil i les persones afectades per la IA, en el procés d'avaluació també s'ha suprimit de l'article en les últimes etapes de

les negociacions. Això significa que les organitzacions de la societat civil no tindran una manera directa i legalment vinculant de contribuir a les avaluacions d'impacte.

Excepcions de transparència per a les autoritats d'aplicació de la llei i migració: mentre que en principi, els implementadors de sistemes d'IA d'alt risc hauran de publicar el resum dels resultats de les avaluacions, aquest no serà el cas per a les autoritats d'aplicació de la llei i migració. El públic ni tan sols tindrà accés a la mera informació sobre el fet que una autoritat utilitzi un sistema d'IA d'alt risc. En canvi, tota la informació relacionada amb l'ús de la IA en l'aplicació de la llei i migració només s'inclourà en una base de dades no pública, limitant severament la supervisió pública constructiva i l'escrutini democràtic. Això és un aspecte molt preocupant, ja que, indiscutiblement, els riscos per als drets humans, l'espai cívic i l'estat de dret són els més greus en aquestes dues àrees. A més, mentre que els implementadors estan obligats a notificar a l'autoritat de vigilància del mercat pertinent el resultat de les seves avaluacions, hi ha una excepció per complir amb aquesta obligació de notificar per "raons excepcionals de seguretat pública". Aquesta excusa sovint s'utilitzarà com a justificació per continuar amb activitats policials i de gestió de fronteres desproporcionades.

Pel que fa al dret a la queixa i reparació, hi ha algunes solucions, però no hi ha un reconeixement clar de la "persona afectada":

La societat civil ha defensat mecanismes robustos de drets i reparació per a individus i grups afectats per sistemes d'IA d'alt risc. Hem exigint la creació d'una nova secció titulada "Drets de les Persones Afectades", que delimitaria drets i remeis específics per a individus impactats per sistemes d'IA. No obstant això, la secció no s'ha creat i tenim un capítol de "remeis" que inclou només algunes de les nostres demandes.

Aquest capítol de remeis inclou el dret a presentar queixes davant una autoritat de vigilància del mercat, però aquest dret està mancat de força, ja que no està clar com aquestes autoritats podran fer complir efectivament el reglament i responsabilitzar els infractors. De manera similar, es reconeix el dret a una explicació dels processos de presa de decisions individuals, particularment per a sistemes d'IA llistats com d'alt risc, però la redacció final planteja preguntes sobre la practicitat i accessibilitat d'obtenir explicacions significatives per part dels implementadors. A més, l'eficàcia d'aquests mecanismes en la pràctica roman incerta, donada l'absència de disposicions com el dret a la representació de persones naturals, o la capacitat per a organitzacions d'interès públic de presentar queixes davant autoritats supervidores nacionals.

El Reglament permet un doble estàndard quan es tracta dels drets humans de les persones a fora de la UE:

El Reglament d'IA no compleix amb la demanda de la societat civil d'assegurar que els proveïdors d'IA establerts a la UE, els sistemes dels quals impacten persones a fora de la UE, estiguin subjectes als mateixos requisits que aquells dins de la UE. El Reglament no impedeix que les empreses establertes a la UE exportin sistemes d'IA que estan prohibits a la UE, creant així un risc enorme de violar els drets de les persones en països no membres de la UE mitjançant tecnologies fetes a la UE que són fonamentalment incompatibles amb els drets humans. A més, el Reglament no exigeix que els sistemes d'alt risc exportats segueixin els requisits tècnics, de transparència o altres salvaguardes altrament requerides quan els sistemes d'IA estan destinats a ser utilitzats dins de la UE, posant en risc de nou la violació dels drets de les persones fora de la UE per tecnologies fetes a la UE.

En segon lloc, vam instar les legisladores de la UE a limitar la vigilància nociva i discriminatòria per part de les autoritats de seguretat nacional, aplicació de la llei i migració. Com ho han resultat?

L'exempció general per a la seguretat nacional provoca el risc d'infringir altres normes:

El Reglament d'IA i les seves mesures de seguretat no s'aplicaran als sistemes d'IA si es desenvolupen o utilitzen únicament amb el propòsit de seguretat nacional, independentment de si això és portat a terme per una autoritat pública o una empresa privada. Aquesta exempció introdueix un buit significatiu que exclourà automàticament certs sistemes d'IA de l'escrutini i limitarà l'aplicabilitat de les salvaguardes dels drets humans previstes al Reglament.

En termes pràctics, significarà que els governs podrien invocar la seguretat nacional per introduir sistemes de vigilància biomètrica en massa, sense haver d'aplicar cap salvaguarda prevista al Reglament, ni avaluar d'impacte en els drets fonamentals i sense assegurar que el sistema d'IA compleixi amb estàndards tècnics alts i no discrimini contra certs grups.

Una exempció tan àmplia no està justificada segons els tractats de la UE i contradiu la jurisprudència establerta del Tribunal de Justícia de la Unió Europea. Encara que la

seguretat nacional pot ser un motiu justificat per a excepcions del Reglament, en realitat això ha de ser avaluat cas per cas, d'acord amb la Carta de Drets Fonamentals de la UE. El text adoptat, però, fa, en gran manera, de la seguretat nacional una zona lliure de drets digitals. Ens preocupa la falta de procediments clars en l'àmbit nacional per verificar si l'amenaça de seguretat nacional invocada pel govern és realment legítima i prou greu per justificar l'ús del sistema i si el sistema es desenvolupa i utilitza amb respecte pels drets fonamentals. La UE també ha establert un precedent preocupant en l'àmbit regional i global; ja que ara s'han introduït també exempcions àmplies per a la seguretat nacional en la recentment adoptada Convenció del Consell d'Europa sobre l'IA.

La policia predictiva, el reconeixement facial públic en remot i en temps real, la categorització biomètrica i el reconeixement d'emocions només estan parcialment prohibits, legitimant aquestes pràctiques perilloses:

Vam demanar prohibicions completes contra qualsevol ús de la IA que no sigui compatible amb els drets i les llibertats - com la pretesa 'lectura de pensaments' per part de la IA, els sistemes de vigilància biomètrica que ens tracten com a codis de barres ambulants, o els algorismes utilitzats per decidir si som innocents o culpables. Tots aquests exemples ara estan parcialment prohibits al Reglament de la IA, la qual cosa és un senyal important que la UE està preparada per establir límits clars contra usos de l'IA inacceptables perjudicials.

Al mateix temps, totes aquestes prohibicions contenen buits significatius i decebedors, fet que provoca que possiblement no aconseguiran el seu potencial complet. En alguns casos, aquests buits podrien tenir l'efecte contrari del que una prohibició hauria de tenir: donen el senyal que algunes formes de vigilància biomètrica en massa i de discriminació alimentada per la IA són legítimes a la UE, el que establirà un precedent global perillós. Per exemple, el fet que els sistemes de reconeixement d'emocions i de categorització biomètrica estiguin prohibits en el lloc de treball i en entorns educatius, però estiguin permesos quan els utilitzen autoritats i cossos d'aplicació de la llei i de migració, sembla assenyalar que la voluntat de la UE és de provar els sistemes de vigilància més abusius i intrusius contra les persones i col·lectius més vulnerabilitzats de la societat.

A més, pel que fa al reconeixement facial remot en l'àmbit públic en temps real, el Reglament obre el camí per legalitzar alguns usos específics per primera vegada a la UE - malgrat la [nostra anàlisi](#) que il·lustra com tots els usos en espais públics d'aquests sistemes constitueixen una violació inacceptable dels drets i llibertats de tothom.

Les serioses vulneracions que provoca el reconeixement facial retrospectiu són en gran part ignorades:

Pel que fa al reconeixement facial retrospectiu, aquesta pràctica no està prohibida en absolut pel Reglament d'IA. [Com hem explicat](#), l'ús del reconeixement facial retrospectiu (post) i altres sistemes de vigilància biomètrica (anomenats 'identificació biomètrica remota', o 'RBI' en el text) són tan invasius i vulneradors de drets com els sistemes en temps real (en viu). No obstant això, el Reglament d'IA comet un gran error en afirmar que el temps addicional per a usos retrospectius mitigarà possibles danys.

Encara que diverses legisladores han argumentat que van aconseguir inserir salvaguardes, la nostra anàlisi és que les salvaguardes no són prou significatives i podrien ser fàcilment evadides per la policia. En una part del text, la salvaguarda pretesa fins i tot suggereix que la simple sospita de qualsevol delictes hagi tingut lloc seria suficient per justificar l'ús d'un sistema RBI posterior - un llindar més baix del que actualment ens beneficia sota la llei de protecció de dades de la UE.

Les persones en moviment no tenen els mateixos drets que la resta, amb regles més febles - i a vegades absents - sobre l'ús de la IA a les fronteres i en context de migració:

En la seva versió final, el Reglament de la UE estableix un [precedent perillós per a l'ús de tecnologia de vigilància contra migrants](#), persones en moviment i grups vulnerabilitzats i dels marges. La legislació desenvolupa un marc legal separat per a l'ús de la IA per autoritats de control de migracions, per permetre la prova i l'ús de tecnologies de vigilància perilloses a les fronteres de la UE i de manera desproporcionada contra persones racialitzades.

Cap de les prohibicions s'aplica de manera significativa al context de migració, i les obligacions de transparència presenten exempcions ad hoc per a autoritats de migració, permetent-les actuar amb impunitat i fora de l'escrutini públic.

La llista de sistemes d'alt risc no captura els molts sistemes d'IA utilitzats en el context de migració, ja que exclou sistemes perillosos com els sistemes d'identificació biomètrica no remota, escàners d'empremtes dactilars o eines de predicció utilitzades per predir, intercedir i restringir la migració.

Finalment, els sistemes d'IA utilitzats com a part de bases de dades de migració a gran escala de la UE (p. ex., Eurodac, el Sistema d'Informació de Schengen i ETIAS) no hauran de complir amb el Reglament fins al 2030, la qual cosa dona molt de temps per normalitzar l'ús de tecnologia de vigilància.

En tercer lloc, vam instar les legisladores de la UE a resistir el lobby de les Big Tech i a incorporar els impactes ambientals. Com ha quedat aquest punt?

El marc de classificació de riscos s'ha convertit en un exercici d'autoregulació:

Inicialment, tots els casos d'ús inclosos a la llista d'aplicacions d'alt risc haurien hagut de seguir obligacions específiques. No obstant això, com a resultat del fort lobby de la indústria, ara els proveïdors de sistemes d'alt risc podran decidir si els seus sistemes són d'alt risc o no, ja que s'ha afegit un "filtre" addicional a aquest sistema de classificació.

Els proveïdors encara hauran de registrar documentació suficient a la base de dades pública per explicar per què no consideren que el seu sistema sigui d'alt risc. No obstant això, aquesta obligació no s'aplicarà quan estiguin proporcionant sistemes a autoritats d'aplicació de la llei i autoritats de migració. Això obrirà el camí per a l'adquisició lliure i desregulada de sistemes de vigilància en els contextos policials i de fronteres.

El Reglament només fa un primer pas provisional per abordar els impactes ambientals de la IA:

Tenim serioses preocupacions sobre com l'ús exponencial dels sistemes d'IA pot tenir impactes greus sobre el medi ambient, tenint en compte el consum de recursos, la mineria extractiva i el processament intensiu d'energia. Avui en dia, la informació sobre els impactes ambientals de la IA és un secret corporatiu molt ben guardat. Això fa difícil avaluar els danys ambientals de la IA i desenvolupar solucions polítiques per reduir les emissions de carboni i altres impactes negatius.

El primer esborrany de Reglament d'IA va ignorar completament aquests riscos, malgrat que la societat civil i les investigadores van alertar repetidament sobre la necessitat de fer transparent el consum d'energia dels sistemes d'IA. Per abordar aquest problema, el Reglament d'IA ara requereix que els proveïdors de models de GPAI (IA de propòsit general) que s'entrenen amb grans quantitats de dades i consumeixen molta electricitat documentin el seu consum d'energia. Ara la Comissió té la tasca de desenvolupar una metodologia adequada per mesurar el consum d'energia d'una manera comparable i verificable.

El Reglament d'IA també requereix que es creïn procediments estandarditzats d'informació i documentació per assegurar l'ús eficient dels recursos per alguns

sistemes d'IA. Aquests procediments haurien d'ajudar a reduir el consum d'energia i altres recursos dels sistemes d'IA considerats d'alt risc durant el seu cicle de vida. Aquests estàndards també tenen com a objectiu promoure el desenvolupament eficient en energia de models d'IA de propòsit general.

Aquests estàndards d'informació són un primer pas crucial per proporcionar transparència bàsica sobre alguns impactes ecològics de la IA, principalment pel que fa a l'ús d'energia. Però només poden servir com a punt de partida per a enfocaments polítics més complets que aborden tots els danys ambientals al llarg del procés de producció de la IA, com l'aigua i els minerals. No podem confiar en l'autoregulació, donada l'evolució ràpida de la crisi climàtica.

Què ve ara en el marc del Reglament de la IA?

L'any vinent serà decisiu pel Reglament de la IA de la UE, amb diferents institucions, legisladors nacionals i fins i tot representants d'empreses establint estàndards, publicant directrius interpretatives i impulsant la implementació del Reglament a través dels països membres de la UE. Algunes parts de la llei - les prohibicions - podrien esdevenir operatives tan aviat com aquest novembre. Per tant, és vital que els grups de la societat civil estiguem inclosos a la taula de decisions, i que aquest treball no es faci en entorns opacs i darrere de portes tancades.

Instem a les persones legisladores de tot el món que també estan considerant introduir normes horitzontals sobre la IA a aprendre dels molts errors de la UE esbossats en aquesta anàlisi. Un conjunt significatiu de proteccions ha d'assegurar que les normes d'IA realment funcionin per a individus, comunitats, la societat, l'estat de dret i el planeta.

Mentre aquest llarg capítol de legislació ara s'acosta al seu final, el pròxim capítol, el de la implementació - i intentar treure el màxim partit d'aquesta Regulació com sigui possible - només està començant. Com a grup, estem redactant una guia d'implementació per a la societat civil, que esperem que arribi més endavant durant aquest mateix any. Volem expressar el nostre agraïment a tot el grup nucli d'incidència, que ha treballat incansablement durant més de tres anys per analitzar, incidir i mobilitzar-se al voltant del Reglament d'IA de la UE. En particular, agraïm la feina, dedicació i visió de Sarah Chander, de l'Equinox Racial Justice Institute, pel seu lideratge durant els últims tres anys.

Per aprendre més sobre les opinions i anàlisis de la nostra coalició sobre el Reglament d'IA final, consulteu els següents recursos:

Access Now, 'The EU AI Act: a failure for human rights, a victory for industry and law enforcement':

<https://www.accessnow.org/press-release/ai-act-failure-for-human-rights-victory-for-industry-and-law-enforcement/>

Amnesty International: 'EU: Artificial Intelligence rulebook fails to stop proliferation of abusive technologies':

<https://www.amnesty.org/en/latest/news/2024/03/eu-artificial-intelligence-rulebook-fails-to-stop-proliferation-of-abusive-technologies/>

European Disability Forum: 'AI Act agreement: partial win on accessibility':

<https://www.edf-feph.org/ai-act-agreement-partial-win-on-accessibility/>

#ProtectNotSurveil: A dangerous precedent: how the EU AI Act fails migrants and people on the move:

<https://www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move/>

Bits of Freedom, 'De AI-verordening is er, maar wij zijn sceptisch':

<https://www.bitsoffreedom.nl/2023/12/09/de-ai-verordening-is-er-maar-wij-zijn-sceptisch/>

AlgorithmWatch, 'EU Parliament votes on AI Act; member states will have to plug surveillance loopholes': <https://algorithmwatch.org/en/eu-parliament-votes-on-ai-act/>

ARTICLE 19, 'EU: AI Act passed in Parliament fails to ban harmful biometric technologies':

<https://www.article19.org/resources/eu-ai-act-passed-in-parliament-fails-to-ban-harmful-biometric-technologies/>

Persones col·laboradores en aquesta anàlisi conjunta:

Ella Jakubowska, European Digital Rights (EDRi);

Kave Noori, European Disability Forum (EDF);

Mher Hakobyan, Amnesty International;

Karolina Iwańska, European Center for Not-for-profit Law (ECNL);

Kilian Vieth-Ditlmann, AlgorithmWatch;

Nikolett Aszodi, AlgorithmWatch;

Judith Membrives Llorens, Lafede.cat / Algorights;

Caterina Rodelli, Access Now;

Daniel Leufer, Access Now;

Nadia Benaissa, Bits of Freedom;

Ilaria Fevola, ARTICLE 19;

Amb el treball, suport i col·laboració de totes les organitzacions de la coalició.