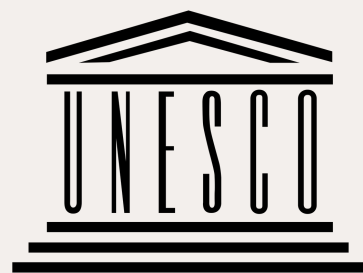


UNESCO: RECOMANACIONS D'ÈTICA EN IA



Han passat gairebé dos anys i un llarg procés de diàleg i negociació per tal que la UNESCO hagi elaborat el primer instrument global d'establiment de normes sobre l'ètica de la intel·ligència artificial (IA) en forma de recomanacions. A continuació t'expliquem els principals objectius, valors fonamentals i principis bàsics adquirits per part de 193 Estats.

Objectius principals

01

Proporcionar un marc universal de valors, principis i accions per orientar als Estats en la formulació de les seves lleis, polítiques o altres instruments relatius a la IA;

**02**

Orientar les accions de les persones, els grups, les comunitats, les institucions i les empreses a fi d'assegurar la incorporació de l'ètica en totes les etapes del cicle de vida dels sistemes de IA;

**04**

Fomentar el diàleg multidisciplinari i pluralista entre múltiples parts interessades sobre qüestions ètiques relacionades amb els sistemes de IA;

03

Promoure el respecte de la dignitat humana i la igualtat de gènere, i protegir els drets humans, les llibertats fonamentals i el medi ambient i els ecosistemes en totes les etapes del cicle de vida dels sistemes de IA;

**05**

Promoure l'accés equitatiu als avenços i els coneixements en l'àmbit de la IA i l'aprofitament compartit dels beneficis.



UNESCO: RECOMANACIONS D'ÈTICA EN IA



Valors Fonamentals



1. Respecte, protecció i promoció de la dignitat humana, els drets humans i les llibertats fonamentals

Cap ésser humà hauria de sofrir danys físics, econòmics, socials, polítics o mentals durant cap etapa del cycle de vida dels sistemes d'IA. Les persones mai haurien de ser cosificadas i els drets humans i llibertats fonamentals han de ser respectats, protegits i promoguts.

2. Prosperitat del medi ambient i els ecosistemes

La prosperitat del medi ambient i els ecosistemes hauria de ser reconeguda i promoguda al llarg del cycle de vida dels sistemes d'IA. Tots els actors involucrats haurien de reduir l'impacte ambiental dels sistemes de IA.

3. Diversitat i inclusió

El respecte, la protecció i la promoció de la diversitat i inclusió haurien de garantir-se al llarg del cycle de vida dels sistemes d'IA. A més, caldria esforçar-se per pal·liar la falta d'infraestructura, educació i competències tecnològiques necessàries.

4. Viure en harmonia i pau

Els actors de la IA haurien d'exercir una funció propícia per a la vida harmoniosa i pacífica, garantint un futur interconnectat i una coexistència harmoniosa entre éssers humans, no humans i el medi natural en benefici de tots.



UNESCO: RECOMANACIONS D'ÈTICA EN IA



Principis Bàsics



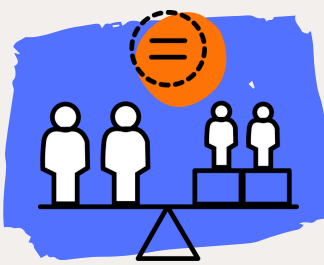
1. Proporcionalitat i innocuïtat

L'elecció d'un mètode d'IA hauria de justificar-se de les maneres següents: a) el mètode d'IA triat hauria de ser convenient i proporcional per assolir un objectiu legítim determinat; b) el mètode d'IA triat no hauria de repercutir negativament en els valors fonamentals; c) el mètode d'IA hauria de ser adequat al context i basar-se en fonaments científics rigorosos.



2. Seguretat i protecció

Els danys no desitjats (riscos de seguretat) i les vulnerabilitats als atacs (riscos de protecció) haurien d'evitar-se. La seguretat i la protecció de la IA serà possible gràcies al desenvolupament de marcs d'accés a les dades que siguin sostenibles, respectin la privacitat i fomentin un millor entrenament dels models d'IA utilitzant dades de qualitat.



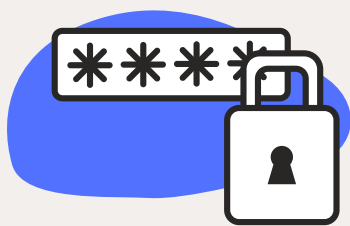
3. Equitat i no discriminació

Els actors de la IA haurien de promoure la justícia social i respectar l'equitat. L'equitat suposa compartir els beneficis de les tecnologies de la IA als plànols local, nacional i internacional, tenint en compte les necessitats específiques dels diferents grups d'edat, sistemes culturals, diferents grups lingüístics, persones amb discapacitat, nenes i dones, i poblacions desfavorides, marginades i vulnerables.



4. Sostenibilitat

L'aparició de les tecnologies de la IA pot beneficiar els objectius de sostenibilitat o dificultar-ne la realització, depenent de la manera com s'apliquin als països amb diferents nivells de desenvolupament. Per tant, s'hauria de dur a terme una avaluació continuada dels efectes socials, culturals, econòmics i ambientals de les tecnologies de la IA amb ple coneixement de les repercussions d'aquestes tecnologies en la sostenibilitat.



5. Privacitat

La privacitat ha de ser respectada, protegida i promoguda al llarg del cicle de vida dels sistemes de IA, tant a nivell personal com col·lectiu. Els organismes reguladors haurien d'establir marcs de protecció de dades i mecanismes de governança adequats. Els sistemes algorítmics requereixen d'avaluacions exhaustives d'impacte en la privacitat, que incloguin també consideracions socials i ètiques en la seva utilització.



6. Supervisió i decisió humanes

Ha de ser sempre possible atribuir la responsabilitat ètica i jurídica, a qualsevol etapa del cicle de vida dels sistemes d'IA, a persones físiques o entitats jurídiques existents. La supervisió humana es refereix, per tant, no sols a la supervisió humana individual, sinó també a la supervisió pública, segons correspongui.



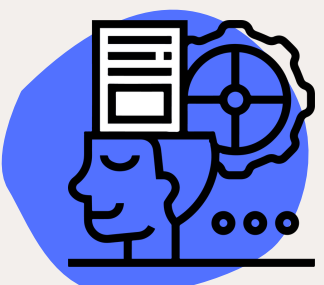
7. Transparència i explicabilitat

La transparència i l'explicabilitat estan estretament relacionades amb les mesures adequades de responsabilitat i rendició de comptes, així com amb la fiabilitat dels sistemes d'IA. Si bé cal fer tot el possible per augmentar la transparència i l'explicabilitat dels sistemes de IA, el grau de transparència i explicabilitat hauria de ser sempre adequat al context. Cal trobar l'equilibri just entre la transparència i l'explicabilitat i altres principis com la seguretat i la protecció. Les persones tenen dret a saber quan es pren una decisió sobre la base d'algorismes de IA i exigir o sol·licitar explicacions i informació a empreses del sector privat o institucions del sector públic.



8. Responsabilitat i rendició de comptes

Haurien d'elaborar-se mecanismes adequats de supervisió, d'avaluació de l'impacte i de diligència deguda per tal de garantir la rendició de comptes al llarg del cicle de vida dels sistemes d'IA. La responsabilitat ètica i l'obligació de rendir comptes sempre haurien de ser atribuïbles, en última instància, als actors de la IA.



9. Sensibilització i alfabetització

La sensibilització i la comprensió del públic respecte de les tecnologies de la IA i el valor de les dades haurien de promoure's mitjançant una educació oberta i accessible. L'aprenentatge sobre l'impacte dels sistemes de IA hauria d'incloure l'aprenentatge sobre els drets humans, a través d'ells i per a ells, el que significa que l'enfocament i la comprensió dels sistemes d'IA s'hauria de basar en l'impacte d'aquests sistemes pels drets humans i per l'accés a aquests drets.



10. Governança i col·laboració adaptatives i de múltiples parts interessades

En la utilització de dades hauria de respectar-se el dret internacional i la sobirania. Cal adoptar normes obertes i garantir la interoperabilitat per a facilitar la col·laboració. Han d'adoptar-se mesures per a tenir en compte els canvis en les tecnologies i l'aparició de nous grups de parts interessades, i permetre una intervenció significativa dels grups, les comunitats i les persones marginades.

Font: UNESCO 2021