IA: Riscos i resposta jurídica

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Generalitat de Catalunya

OEIAC

Universitat de Girona

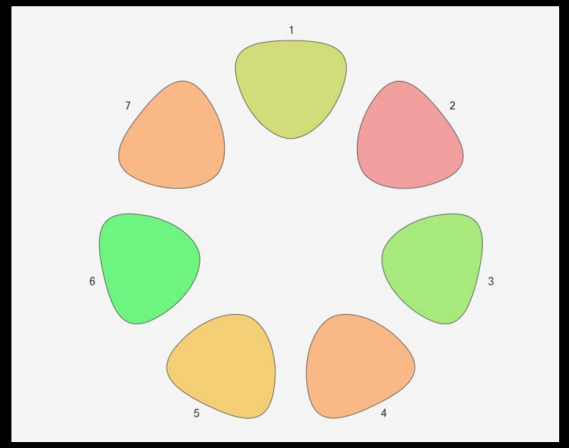
L'OEIAC dins l'estratègia CATALONIA.AI

La Generalitat de Catalunya impulsa l'Estratègia d'Intel·ligència Artificial (IA) de Catalunya que, amb el nom de **CATALONIA.AI**, realitzarà un programa d'actuacions per enfortir l'ecosistema català en IA.

L'Estratègia CATALONIA.AI coincideix amb el **"Pla Estratègic UdG2030: la Suma d'Intel·ligències"** de la Universitat de Girona.

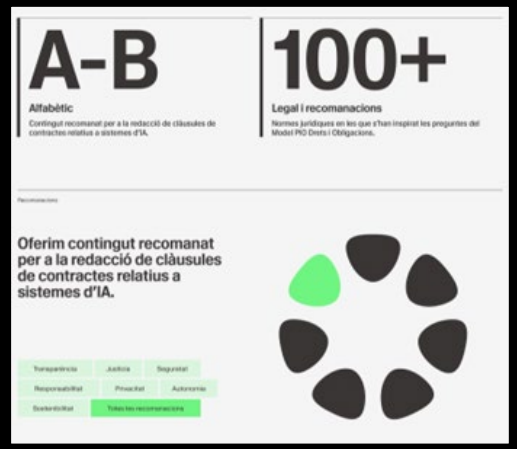
Estudiar les ***conseqüències*** ètiques, socials i legals, així com els riscos i oportunitats de la implantació de la IA en la vida diària a Catalunya, des d'una òptica plenament transversal.

Principals activitats



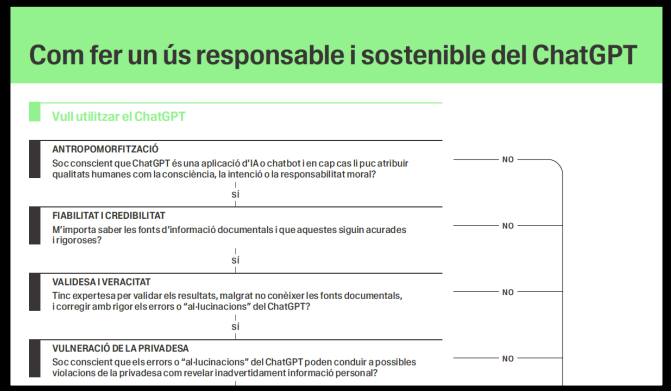
Avaluació de princpis ètics

Eines d'avaluació sobre usos ètics de la IA per garantir que el desenvolupament, la implantació i l'ús dels sistemes d'IA compleixen els requisits per a una IA responsable i sostenible.



Informes i guies sobre usos ètics

Informes i guies per promoure les bones pràctiques i ús de la IA de confiança entre les organitzacions públiques i privades de l'ecosistema català.



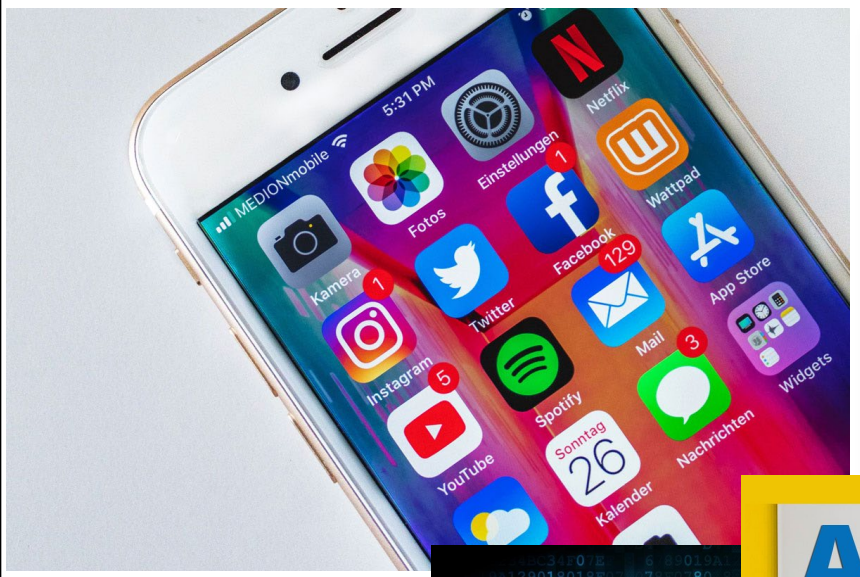
Recursos pedagògics

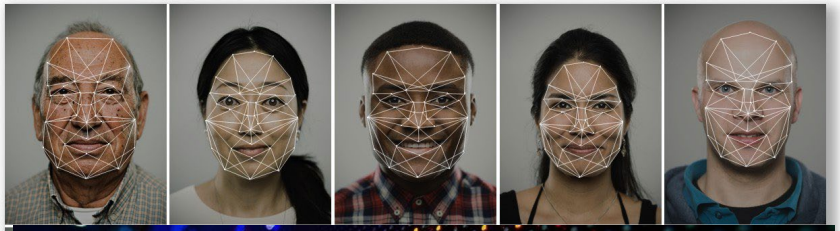
Materials dissenyats per a l'educació a IA per facilitar la comprensió dels usos ètics de la IA amb recomanacions pràctiques per utilitzar la IA de manera responsible i sostenible.



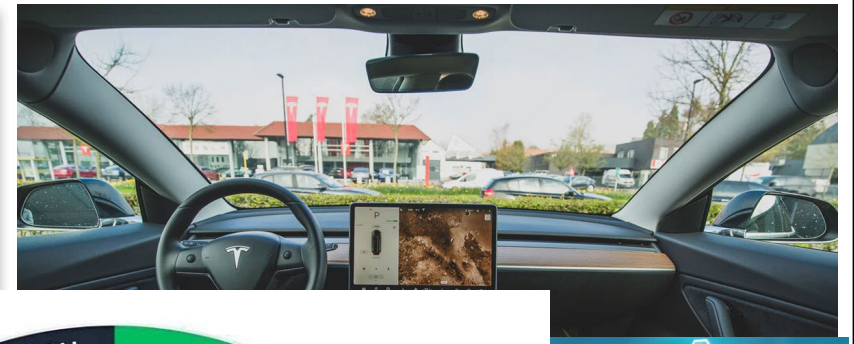
Cicle Seminaris

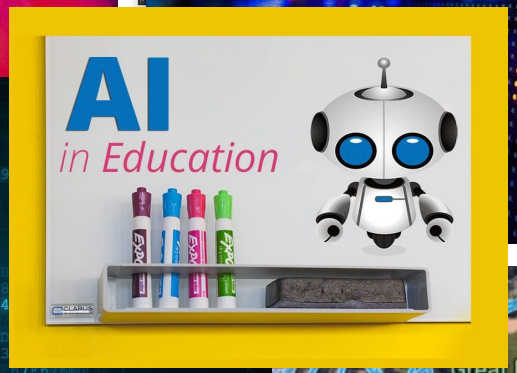
Una sèrie de seminaris per a la difusió, transferència de coneixements i desenvolupament de capacitats per al públic en general i també per a la capacitació sociotècnica a IA.

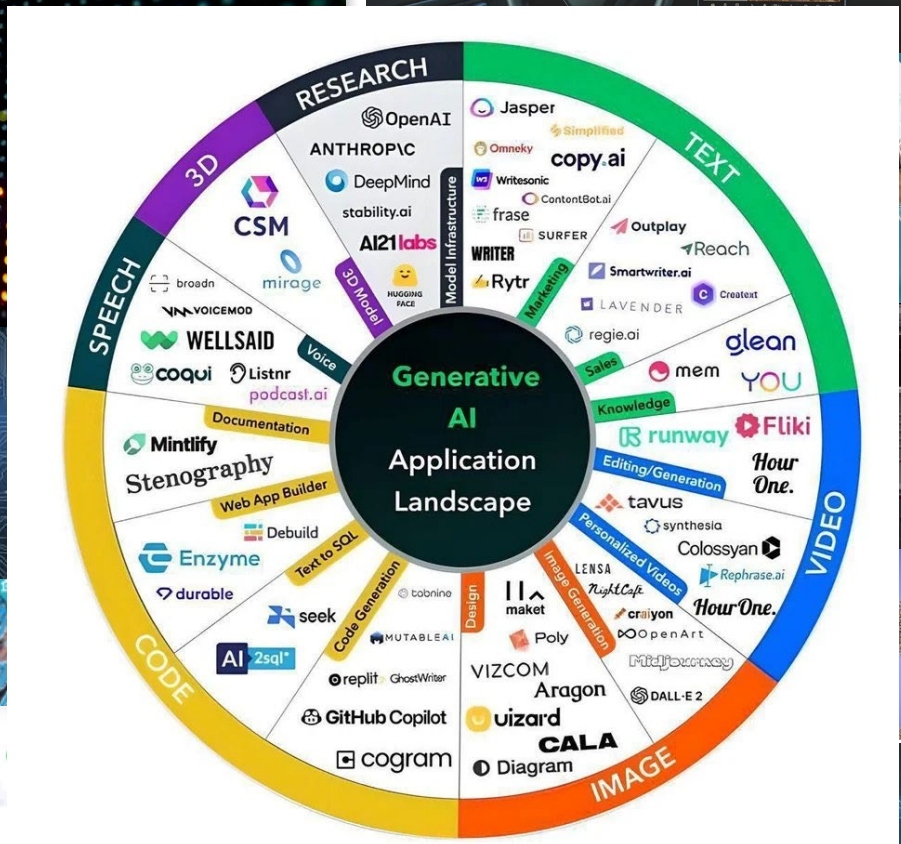

MEDIONmobile
5:31 PM
Netflix
Einstellungen
Fotos
Wattpad
Facebook
Kamera
App Store
Twitter
Mail
Instagram
Spotify
Widgets
YouTube
Sonntag
26
Kalender
Nachrichten



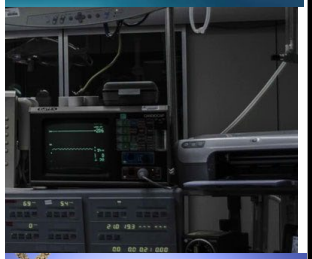


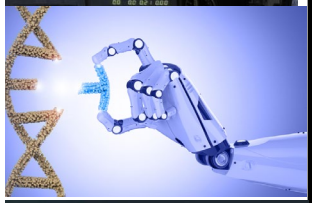


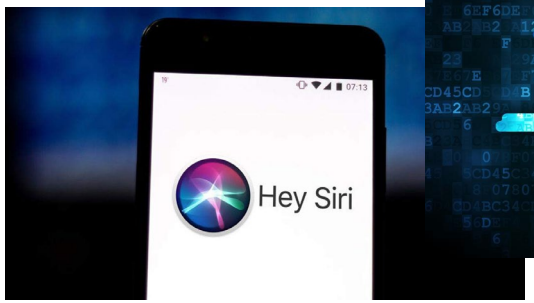

AI
in Education

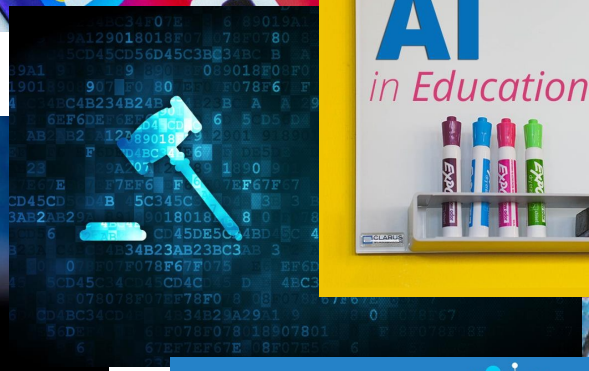

Generative AI Application Landscape
RESEARCH
TEXT
VIDEO
IMAGE
CODE
SPEECH
3D
OpenAI
ANTHROP\C
DeepMind
stability.ai
AI21labs
HUGGING FACE
CSM
mirage
broadn
VOICEMOD
WELLSAID
coqui
Listnr
podcast.ai
Model Infrastructure
3D Model
Voice
Jasper
Simplified
Omneky
copy.ai
Writesonic
ContentBot.ai
frase
SURFER
WRITER
Rytr
Outplay
Reach
Smartwriter.ai
LAVENDER
Creatext
regie.ai
glean
mem
YOU
Marketing
Sales
Knowledge
runway
Fliki
Hour One.
Editing/Generation
tavus
synthesia
Colossyan
Personalized Videos
Rephrase.ai
Hour One.
LENSA
NightCafe
craiyon
OpenArt
Midjourney
DALL·E 2
Image Generation
Design
maket
Poly
VIZCOM
Aragon
uizard
CALA
Diagram
tabnine
MUTABLE.AI
replit
GhostWriter
GitHub Copilot
cogram
Code Generation
seek
AI 2sql
Text to SQL
Debuild
Enzyme
durable
Web App Builder
Documentation
Mintlify
Stenography





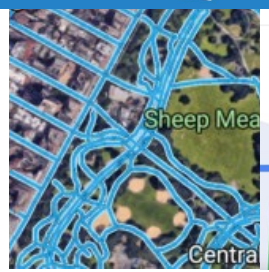


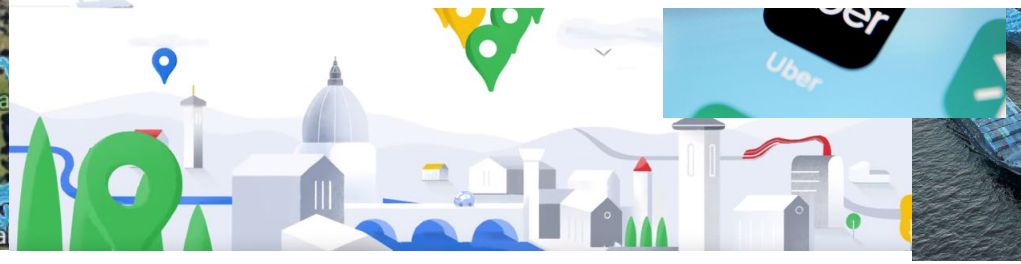

Hey Siri




amazon


JUST ASK
amazon alexa
amazon

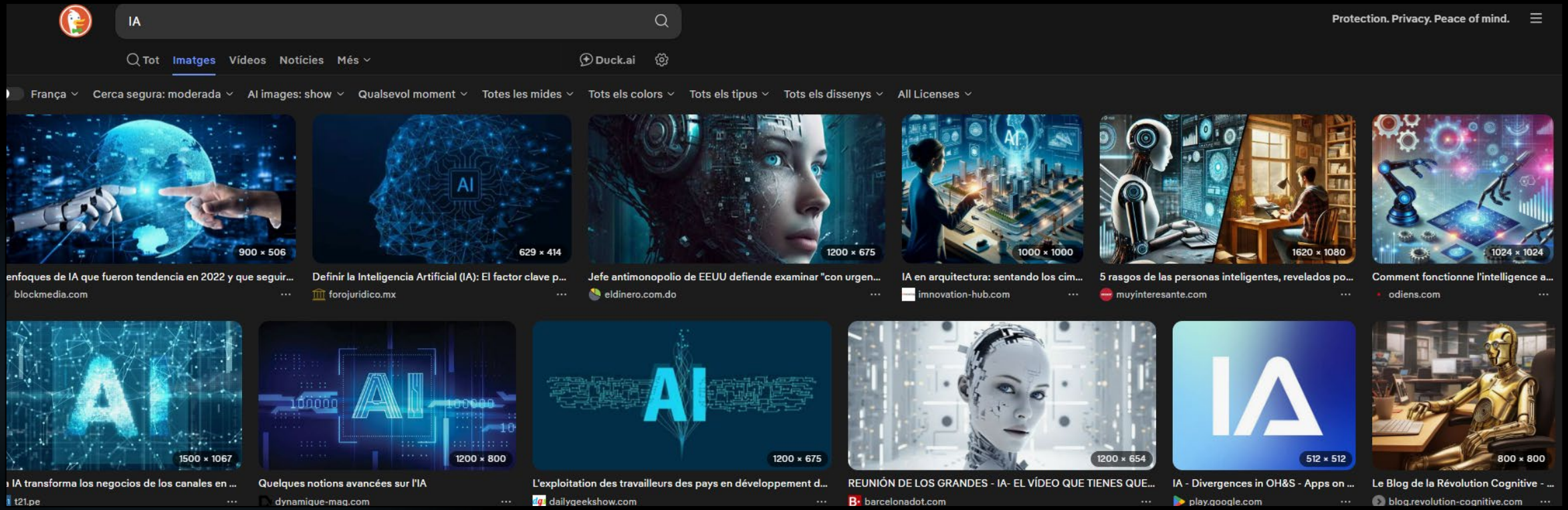

Sheep Mea
Centra



Uber






IA
Protection. Privacy. Peace of mind.
Tot
Imatges
Vídeos
Notícies
Més
Duck.ai
França
Cerca segura: moderada
AI images: show
Qualsevol moment
Totes les mides
Tots els colors
Tots els tipus
Tots els dissenys
All Licenses
900 × 506
enfoques de IA que fueron tendencia en 2022 y que seguir...
blockmedia.com
629 × 414
Definir la Inteligencia Artificial (IA): El factor clave p...
forojuridico.mx
1200 × 675
Jefe antimonopolio de EEUU defiende examinar "con urgen...
eldinero.com.do
1000 × 1000
IA en arquitectura: sentando los cim...
imnovation-hub.com
1620 × 1080
5 rasgos de las personas inteligentes, revelados po...
muyinteresante.com
1024 × 1024
Comment fonctionne l'intelligence a...
odiens.com
1500 × 1067
IA transforma los negocios de los canales en ...
t21.pe
1200 × 800
Quelques notions avancées sur l'IA
dynamique-mag.com
1200 × 675
L'exploitation des travailleurs des pays en développement d...
dailygeekshow.com
1200 × 654
REUNIÓN DE LOS GRANDES - IA- EL VÍDEO QUE TIENES QUE...
barcelonadot.com
512 × 512
IA - Divergences in OH&S - Apps on ...
play.google.com
800 × 800
Le Blog de la Révolution Cognitive - ...
blog.revolution-cognitive.com

La imatge de la IA

- Descriure els sistemes d'IA amb **agencia independent** de la supervisió humana o donant entendre que aviat podrien fer-ho no és ètic.

- Assumir que els algorismes d'IA **aprenen de la mateixa manera** que ho fan els humans és problemàtic ja que donen credibilitat a les afirmacions que la IA és "sentinent".

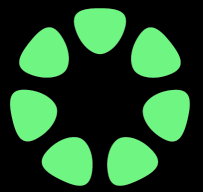
- Per exemple, són habituals i errònies les comparacions d'algoritmes d'aprenentatge profund amb el **funcionament del cervell humà**.



La bona notícia és que està perfectament sa, la mala notícia és que el meu algoritme preveu que es morirà al cap de dos dies.

Image credit to Timo Elliot

Sembla obvi però no ho és



- **No poden respondre preguntes** les respostes de les quals no estiguin d'alguna forma en el seu conjunt de formació.
- **No poden resoldre problemes** sobre els quals no han estat entrenats. Això no tan sols inclou les dades sinó l'arquitectura.
- **No poden adquirir noves habilitats** sense la nostra ajuda. Això inclou des de la recollida d'informació a l'etiquetatge i anàlisi.
- **No poden inventar coses noves**. Necessiten les dades, la programació i l'arquitectura.

Implementació

Les dades d'IA plantegen diferents **reptes d'implementació**:

- **Cost**: els dispositius necessaris per obtenir-les poden ser prohibitius.
- **Emmagatzematge i anàlisi**: les grans mides dels arxius fan que no encaixin en les infraestructures de dades que tenim a disposició.
- **Etiquetatge no expert**: mentre les etiquetes poden ser assignades per persones no expertes, hi ha el risc que això comporti una manca de precisió.

Biaxos

L'ús i implantació de sistemes d'IA pot **reforçar biaixos**:

- **Biaixos estadístics**: l'ús de dades massives no garanteix que existeixi representativitat de diferents grups de població, sent la subrepresentació i sobrerepresentació un problema.
- **Biaixos cognitius**: el disseny i opcions de sistemes i/o l'etiquetatge de les dades poden estar subjectes a biaixos que tenen lloc en la presa de decisions de grups de persones amb característiques molt semblants .
- **Categories d'entrenament**: aquestes són difícils de definir i enfosqueixen la diversitat poblacional, creant biaixos estadístics i cognitius.

Privacitat

L'ús i implantació de sistemes d'IA planteja **reptes de privacitat**:

- **Dades sensibles**: totes les dades recollides i utilitzades requereixen de consentiment previ i calen mesures d'anonimització.
- **Pirateig**: hi ha preocupacions sobre el robatori de dades sensibles i altres conductes indegudes que requereixen de mesures específiques.
- **Reutilització de dades**: sempre desitjable per la minimització de dades però no sempre possible, fet que suposa cercar equilibris i interessos.

Sostenibilitat

Les preocupacions se centren en 3 tipus d'impacte de la IA:

1. **La uberització del treball al voltant de la IA:**

   Preocupacions relacionades amb el creixement d'una nova "underclass".

2. **Una demanda exponencial d'energia eléctrica:**

   Preocupacions relacionades amb l'origen (energies fòssils i nuclears).

3. **Una violació de drets d'autoria a gran escala:**

   Preocupacions relacionades amb la compensació i innovación.

El repte de la confiança

Els sistemes d'IA plantegen diferents **reptes de confiança**:

- **Transparència i explicació**: un repte tècnic actual ja que molts sistemes d'IA funcionen com a "caixes negres" ininterpretables.
- **Fiabilitat**: per quantificar la fiabilitat dels sistemes d'IA calen més estudis replicables, verificació de les dades i models entrenats a nivell públic.
- **Comoditat**: els dispositius han de ser còmodes i fàcils d'integrar a la pràctica de negoci, amb instruccions fàcils de llegir i explicant com i quan s'han d'utilitzar.

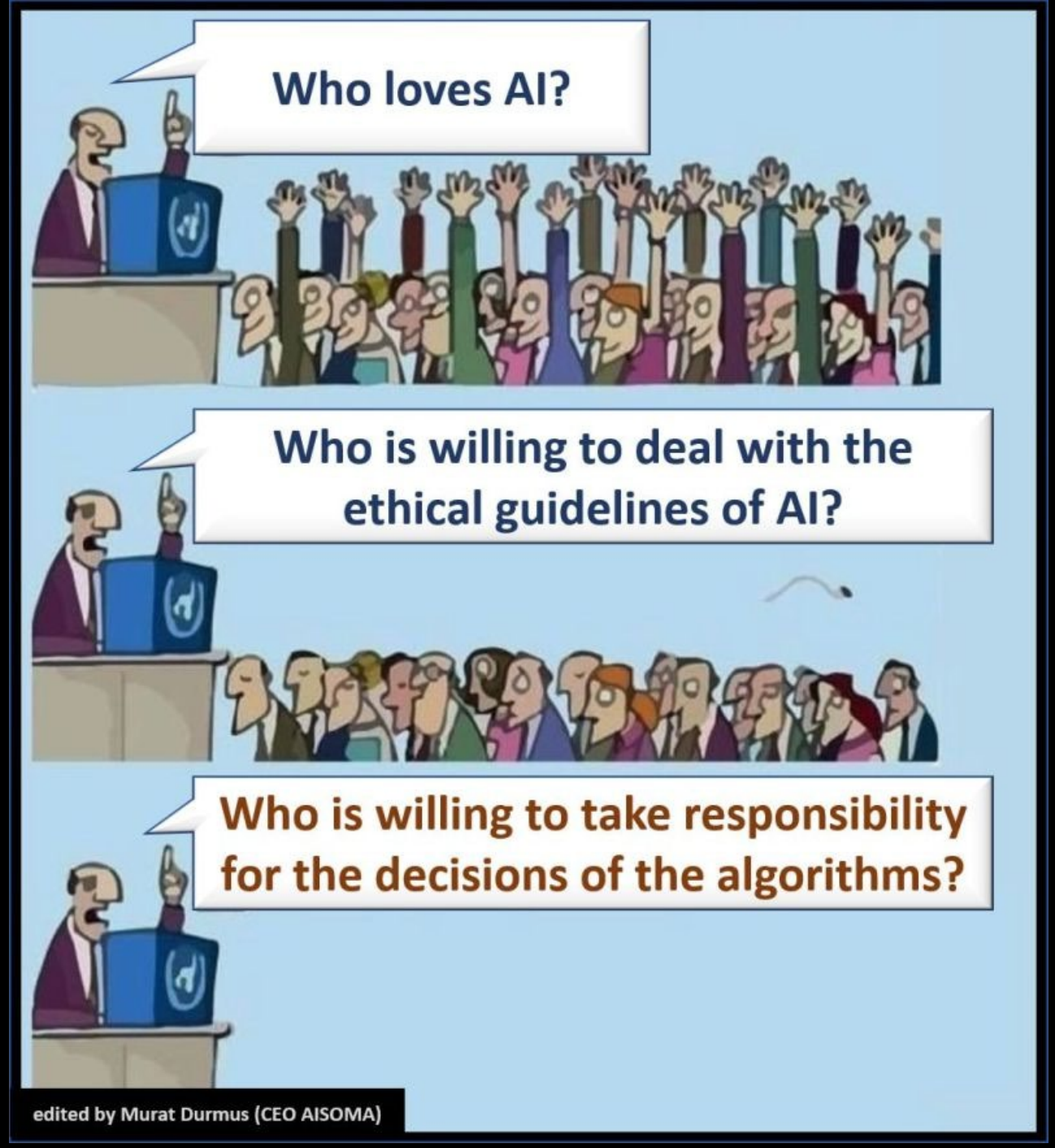
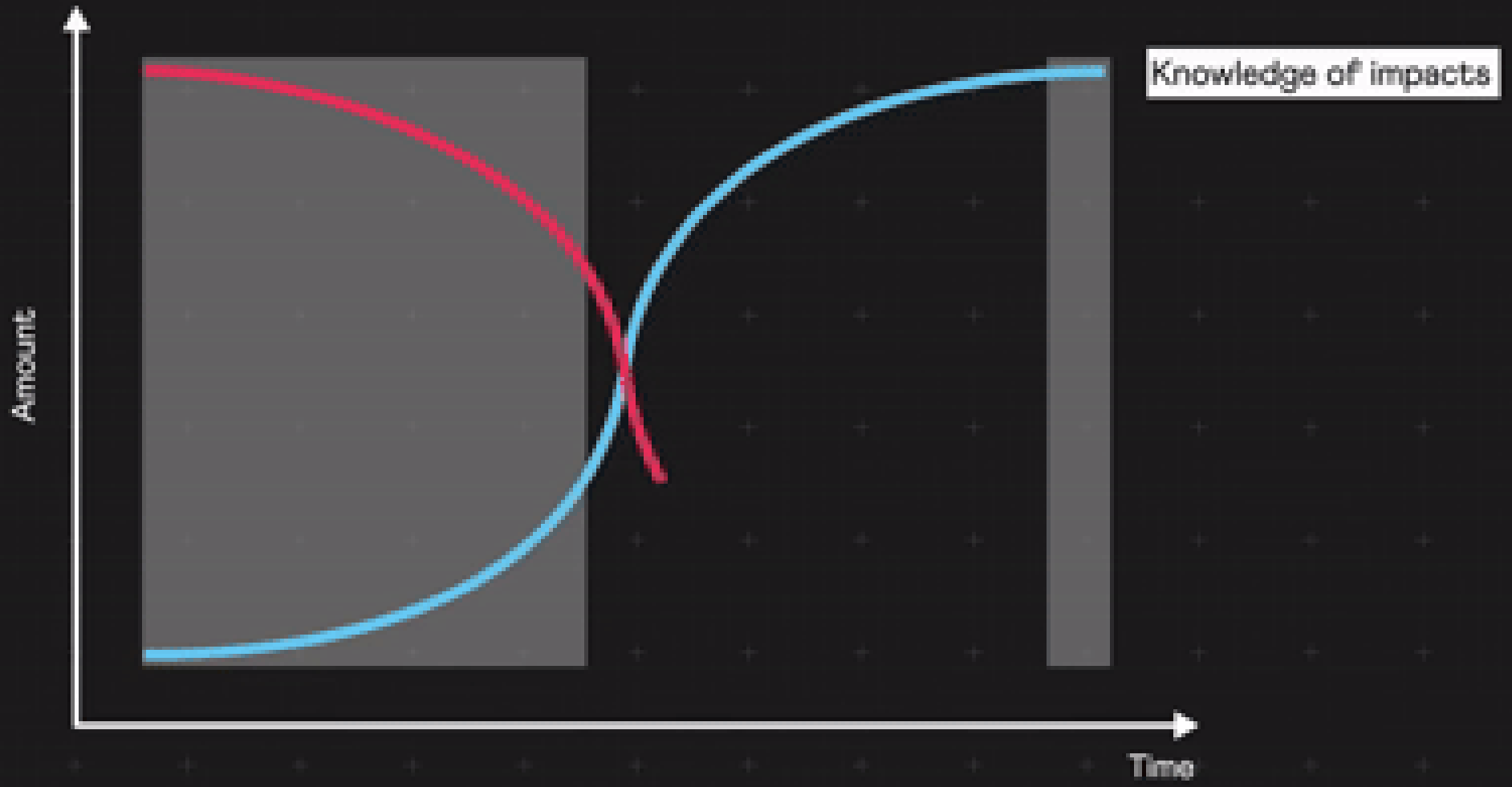


Image credit to Murat Durmus

Preguntes inicials

Sabeu quin és el problema que esteu intentant resoldre amb l'ús d'un sistema d'IA?
Us heu plantejat si és necessari utilitzar un sistema d'IA per resoldre aquest problema?
Heu considerat els beneficis que aportarà l'ús d'un sistema d'IA i si aquests superen els riscos que planteja?
Sabeu explicar de quina manera el sistema d'IA resoldrà el problema o objectiu concret d'una manera comprensible?
Heu establert mecanismes per evitar que el sistema d'IA no s'exploti de manera maliciosa?

The Collingridge Dilemma - When Do You Regulate Technology?

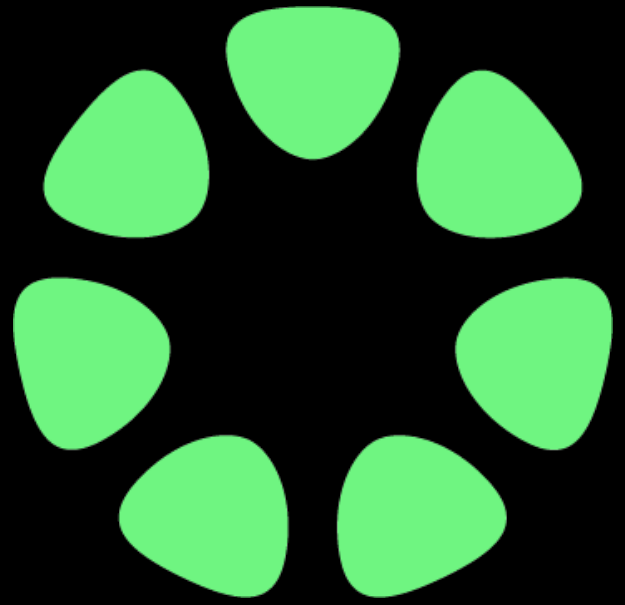


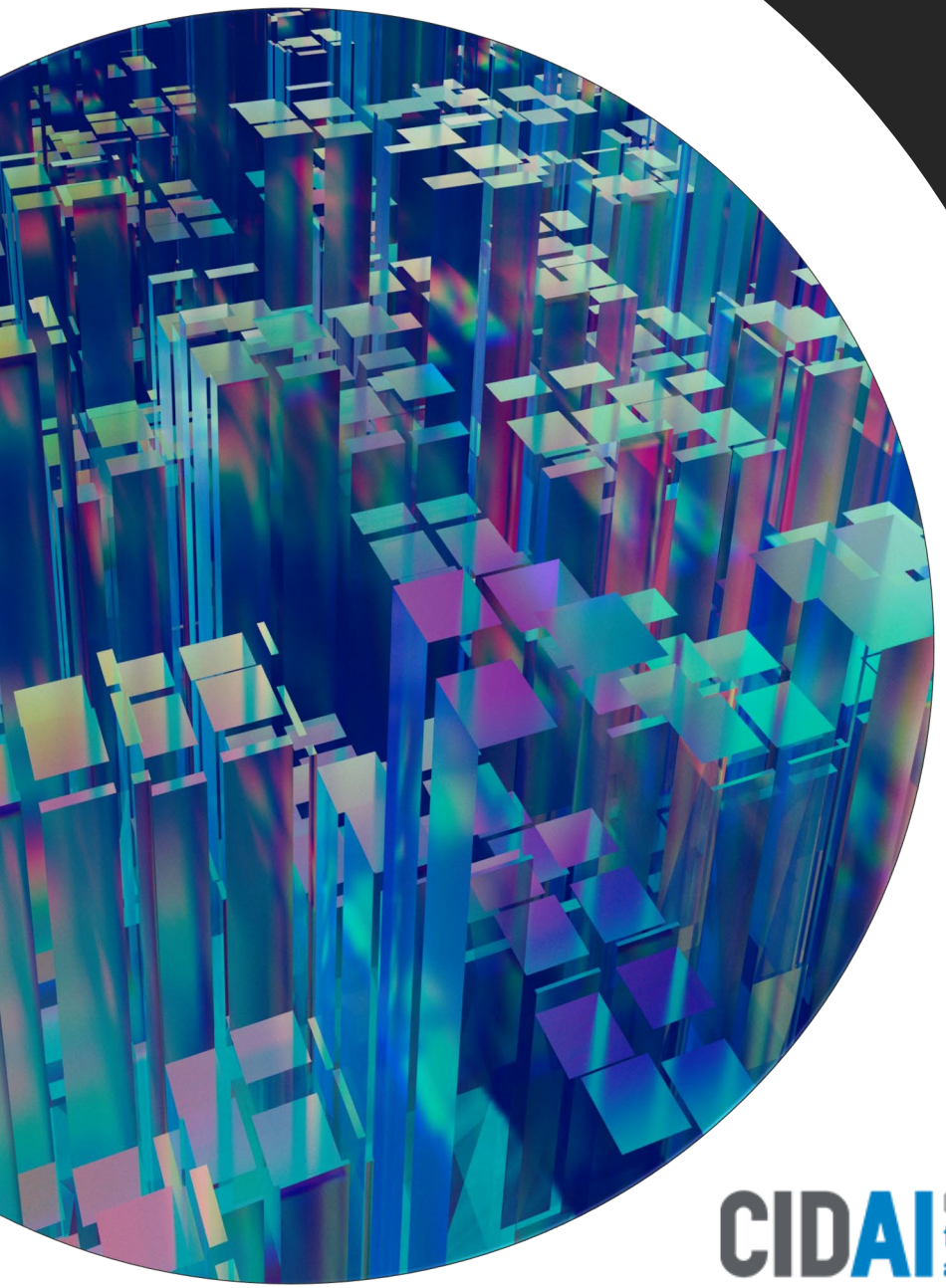
GIFRUN.COM

Present i futur

No perdre les característiques claus que ens defineixen com a humans:

- **L'autonomia**: la nostra capacitat de **prendre decisions** informades i sense restriccions.
- **L'agència**: la capacitat dels individus per **actuar de manera independent**.
- **L'empatia**: la capacitat **d'entendre i relacionar-nos** amb el món que ens envolta.
- **L'ètica**: valors compartits que ens ajuden a distingir el que **està bé** del que **està malament**.





Marcs legals i compliment (RIA)

Implicació per les empreses

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Generalitat de Catalunya

OEIAC

Universitat de Girona

Característiques generals del RIA

El RIA

1. **Primera regulació** que estableix un marc normatiu per a l´ús de la IA.

2. **Perspectiva de mercat**, no de drets.

3. Aproximació basada en els **riscos** → Desacords respecte als usos inclosos en cada categoria.

4. Diferents **obligacions segons** les categories de risc i el rol respecte al desenvolupament, comercialització i desplegament del sistema.

Objecte – Article 1 RIA

L'objectiu del present Reglament és millorar el funcionament del mercat interior i promoure l'adopció de la intel·ligència artificial (IA) centrada en l'ésser humà i digna de confiança, garantint alhora un alt nivell de protecció de la salut, la seguretat, els drets fonamentals consagrats en la Carta, inclosa la democràcia, l'Estat de Dret i la protecció del medi ambient, contra els efectes nocius dels sistemes d' IA a la Unió i donant suport a la innovació.

Àmbit d'aplicació – Art. 2 RIA(I)

Subjectes:

- **Proveïdors** de sistemes comercialitzats o posats en servei dins la UE, o el producte del qual s'utilitzi a la UE, així com els seus **representants autoritzats**.
- **Desplegadors** de sistemes establerts a la UE, o el producte del qual s'utilitzi a la UE.
- **Importadors** i **distribuïdors** de sistemes.
- **Representant autoritzat** d'un sistema d'IA.

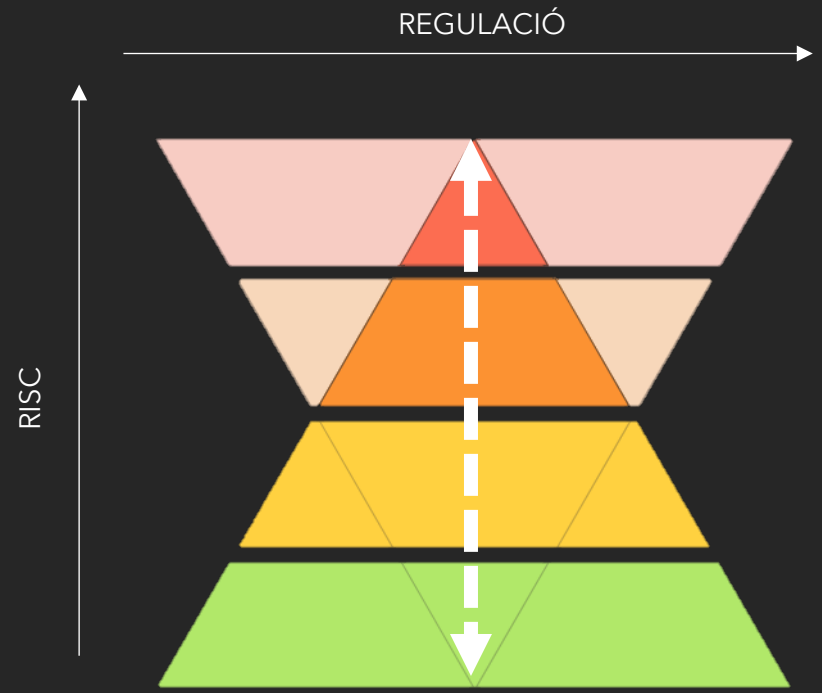
Àmbit d'aplicació – Art. 2 RIA (II)

Àmbti terriotiral:

- Tot sistema comercialitzat i posat en servei dins de la UE, o el producte del qual s'utilitzarà a la UE.
- És indiferent on es trobi establert el subjecte.
- Per tal d'introduir els seu producte a la UE, els objectes no comunitaris han de complir amb els requisits de la RIA.

Categories de risc

Aproximació segons el risc



Risc Inacceptable
e.g. Puntuació social
— **Prohibits**

Alt risc
e.g. Classificació CVs
— **Permès** amb obligacions estrictes i la realització d'una avaluació de compliment

Risc limitat
e.g. Chatbots
— **Permès** amb obligacions de transparència associades

Risc mínim
e.g. Classificador de mails
— **Permès** amb restriccions mínimes i s'anima a seguir codis de conducta

Risc inacceptable – Art. 5 RIA

TÈCNIQUES SUBLIMINARS

ALTERAR EL COMPORTAMENT

CLASSIFICACIÓ SOCIAL

ELABORACIÓ DE PERFILS

IMATGES NO SELECTIVES

EDUCACIÓ O LLOC DE FEINA

CATEGORITZACIÓ BIOMÈTRICA

BIOMETRIA AMB FINS POLICIALS

Risc alt – Art. 6 RIA

Apartat 1r: Annex I

PRODUCTES REGULATS PER LA UE

SOTMESOS A *AVALUACIÓ DE CONFORMITAT* PER TERCERS

Apartat 2n: Annex III

BIOMETRIA

INFRAESTRUCTURES CRÍTIQUES

EDUCACIÓ

GESTIÓ DE PERSONES TREBLLADORES

SERVEIS ESSENCIALS

COMPLIMENT DE LA LLEI

MIGRACIÓ I ASÍL

JUSTÍCIA I PROCESSOS DEMOCRÀTICS

Risc alt – Excepcions i revisió

Article 6.3: Un sistema no serà d'alt risc quan no existeixi un risc per a la salut, la seguretat o els DDFF:

- Tasca de procediment limitat.
- Previsió o millora d'una activitat humana anterior.
- Identificació de patrons i desviació de patrons de pressa de decisions.
- Tasca preparatòria de l'avaluació d'un sistema d'alt risc.

Article 7: La Comissió podrà afegir, modificar i suprimir els casos de l'annex III

Risc limitat – Articles 50–52 RIA

IA GENERATIVA

RECONEIXEMENT D'EMOCIONS

MODELS D'IA DE PROPÒSIT GENERAL (GPAI)

INTERACCIÓ AMB PERSONES

CATEGORITZACIÓ BIOMÈTRICA

Risc limitat – Art. 50



Obligacions de transparència:

- Informar a les persones quan **interactuïn** o estiguin **exposades** a una IA.
- Informar quan un **contingut** ha estat generat per una IA.

Excepcions:

- Sistemes autoritzats legalment per a detectar, prevenir o investigar delictes.

Risc limitat – GPAI



- **Model d'IA d'ús general (GPAI):** capacitat de servir per a una varietat de propòsits, tant per a l'ús directe com la integració en altres sistemes d'IA.

- **GPAI de risc sistèmic:** alta capacitat d'impacte a:
  - ✓ Seguretat
  - ✓ Salut
  - ✓ Drets fonamentals
  - ✓ Democràcia
  - ✓ Economia
  - ✓ Societat en general

  → Obligacions addicionals.

Risc mínim



Categoria residual.

EXEMPLE:

- Un sistema d'IA que organitza el catàleg d'una biblioteca segons quins llibres són més sol·licitats.

Exemples de sistemes

Sistema d’IA d’identificació biomètrica per accedir a les oficines públiques.

Sistema d’IA que genera correus electrònics i textos de resposta.

Sistema d’IA entrenat amb les dades històriques de l’administració per a preveure les franges horàries de les cites prèvies.

Sistema de propòsit general (GPAI) per a preveure aglomeracions i manifestacions segons les dades de les xxss.

Sistema d’IA d’identificació biomètrica per accedir a les oficines públiques: ALT RISC

Sistema d’IA que genera correus electrònics i textos de resposta: RISC LIMITAT

Sistema d’IA entrenat amb les dades històriques de l’administració per a preveure les franges horàries de les cites prèvies: RISC MÍNIM

Sistema de propòsit general (GPAI) per a preveure aglomeracions i manifestacions segons les dades de les xxss: RISC LIMITAT

Persones proveïdores dels sistemes d'IA

Proveïdor:

- **Persona física o jurídica**, **administració pública** o **altre organisme** que desenvolupa o ha desenvolupat un sistema d'IA i el comercialitza o posa en servei amb el seu propi nom o marca comercial, ja sigui a títol onerós o gratuït.

- Les seves obligacions es recullen en **els arts. 16, 50, 53 i 55 del RIA**.

- **Qui** ho és?

  i. La persona que posa **el seu nom o marca** en un producte.

  ii. La persona que el modifica **substancialment.**

  → **Proveïdor d'alt risc:**

  a. La persona que modifica **la finalitat** d'un sistema de manera que el converteix en alt risc.

  b. El productor d'un producte de l'Annex I RIA quan l'incorpora com a component de seguretat en la posada al mercat conjuntament o sota el nom del fabricant.

Avaluació de conformitat

Obligatòria per als sistemes d'alt risc (arts. 16 i 43 del RIA).

3 tipus d'avaluacions:

1. **Procediment d'avaluació intern (**Annex VI) – el propi proveïdor.
2. **Procediment d'avaluació extern (**Annex VII) – organisme encarregat.
3. **Procediment d'avaluació específic per a alguns sistemes** – Annex I, secció A.

Avaluació d'impacte sobre DDFF

Es contempla a l'artícle 27 del RIA.

És obligatòria per:

- Els organismes públics o els organismes privats que **presten serveis públics** en tots els casos de **l'annex III – alt risc** (excepte infraestructures crítiques).
- Avaluar la **solvència** de les persones físiques o establir la seva puntuació creditícia.
- Avaluar els riscos i la fixació de preus en relació amb **assegurances de vida i salut**.

Altres obligacions:

- Sistemes d'IA de Alt risc.
- Sistemes d'IA de Risc limitat.
- Sistemes d'IA de propòsit general.
- Sistemes d'IA de propòsit general de risc sistèmic.

Sistemes d'alt risc – Art. 16 RIA (I):

- Sistema de **gestió de riscos** (Art.9).
- **Governança** de dades (Art.10).
- Documentació **tècnica** (Art.11).
- Els **registres** automàtics (Art.12).
- **Transparència** i informació (Art.13).
- Mesures de **supervisió** humana. (Art.14).
- **Requisits** de precisió, robustesa i ciberseguretat (Art.15).

Sistemes d'alt risc – Art. 16 RIA (II):

- Informar de les **dades** de contacte i llur identificació (en el sistema o la documentació que l'acompanya) (Art.16).
- Incloure un sistema de **gestió de qualitat** (Art.17).
- Conservar i mantenir a disposició de les autoritats competents la documentació del sistema **durant 10 anys** (Art.18).
- Generació de **registres** automàtics (Art.19).

Sistemes d'alt risc – Art. 16 RIA (III):

- **Avaluació de conformitat** del sistema i la declaració europea de **responsabilitat** (Art.43).
- L'obligació de **notificar** la informació als organismes responsables (Art.45).
- El certificat **"CE"** (compleix els estàndards de la UE) (Art.48).
- **Registre** del proveïdor i llur producte (Art.49).

Sistemes de risc limitat – Art. 50 RIA:

- Garantir que les persones exposades a sistemes que interactuïn directament amb persones físiques siguin **informades** d'aquesta exposició.

- Garantir que els resultats d'un sistema IAG siguin **marcats** com a generats o manipulats artificialment.

Sistemes de propòsit general (GPAI)– Art. 53 RIA (I):

- Elaborar i mantenir actualitzada la **documentació tècnica** del model.
- Demostrar **el compliment** del RIA quan les autoritats competents ho sol·licitin. Es presumirà la conformitat quan el proveïdor compleixi amb les normes harmonitzades existents.
- **Lliurar** la informació i documentació necessàries a altres proveïdors que vulguin integrar el model al seu sistema.

Sistemes de propòsit general (GPAI)– Art. 53 RIA (II):

- Establir una política per complir la legislació europea en matèria de **drets d'autor i similars**.
- Elaborar i posar a la disposició del públic **un resum suficientment detallat** sobre el contingut utilitzat per a l'entrenament del model.
- **Cooperar** amb la Comissió i les autoritats competents en l'exercici de les seves competències i facultats.

Sistemes de propòsit general (GPAI) de risc sistèmic– Art. 55 RIA:

- **Avaluar** el model amb protocols i eines normalitzades.
- Avaluar i mitigar els **possibles riscos sistèmics** derivat: desenvolupament; comercialització; o ús del sistema, incloses les seves fonts.
- Tenir un registre, documentar i comunicar a les autoritats competents, la informació sobre **incidents greus** i les possibles mesures correctives.
- Garantir un **nivell de ciberseguretat** adequat per al model i la seva infraestructura física.

Codi de bones pràctiques (I)

Documentació complementaria per als proveïdors de sistemes d'IA de propòsit general en matèria de:

- Transparència.
- Drets d'autoria.
- Seguretat i protecció.

Obligacions dels proveïdors de GPAI– Capítol 2 Codi de bones pràctiques: Transparència (II)

- Elaborar i mantenir la documentació de models actualitzada a partir del formulari de documentació del model de l'Oficina de la IA.
- Proporcionar informació rellevant als proveïdors intermedis i a la oficina de la IA, incloent la publicació de la seva informació de contacte a la seva pàgina web o, si no en tenen, altres medis apropiats.
- Garantir la qualitat, la integritat i la seguretat de la informació.

Obligacions dels proveïdors de GPAI– Capítol 2 Codi de bones pràctiques: Drets d'autoria (III)

- Redactar, actualitzar i implementar una política de drets d'autoria.
- Reproduir i extreure només contingut protegit per drets d'autor accessible legalment en rastrejar internet.
- Identificar i complir les reserves de drets a l'hora de rastrejar internet.
- Mitigar el risc de reproduir il·legalment materials protegit per drets d'autoria.
- Designar una via de contacte i possibilitar la presentació de queixes.

Obligacions dels proveïdors de GPAI de risc sistèmic– Capítol 3 Codi de bones pràctiques: Seguretat i protecció (IV)

Creació del marc	• Ha de contenir una descripció dels elements descrits al compromís 1 del capítol 3 del codi de bones pràctiques per a proveïdors de GPAI, incloent els processos i mesures d'avaluació i mitigació de riscos adoptades. • El marc ha d'estar complet com a mínim 4 setmanes després de notificar a la comissió en virtut de l'art. 52.1 RIA, i 2 setmanes abans d'introduir el sistema al mercat.
Implementació del marc	• Implementar el marc de manera continua durant tot el cicle de vida del model. • Completar el procés complet d' avaluació i mitigació del risc sistèmic abans de compilar l' informe del model i cada actualització del mateix. Avaluar i mitigar el risc sistèmic de forma contínua, incloent avaluacions del model menys rigoroses en els punts d' activació adequats.
Actualització del marc	• Actualitzar el marc segons correspongui. • Realitzar una avaluació del Marc quan existeixin motius raonables per creure que l'adequació i/o el compliment del Marc s'han vist o es veuran substancialment soscavats, o cada 12 mesos a partir de la comercialització del model, cosa que passi primer.

Desplegadors dels sistemes d'IA

Desplegador:

- Persona física o jurídica, administració pública o altre organisme que **utilitza un sistema d'IA sota la seva autoritat**.

  Excepte: ús en una activitat personal no professional.

- Les seves obligacions es recullen en els arts. 26 i 50 del RIA.

Obligacions dels desplegadors de sistemes d'alt risc – Art. 26 RIA (I):

- Adoptar **mesures tècniques i organitzatives** per garantir que els sistemes s'usen d'acord amb les instruccions d'ús.
- Establir la tasca de **supervisió** a persones formades i competents.
- Assegurar que les **dades d'entrada** siguin rellevants i representatives.
- **Supervisar** el funcionament del sistema i informar als altres operadors i a les autoritats competents si identifiquin un risc (salut, la seguretat i els DDFF).

Obligacions dels desplegadors de sistemes d'alt risc – Art. 26 RIA (II):

- En cas **d'incident greu:** informar als altres operadors i les autoritats competents. Si cal, iniciar el procediment de notificació.
- Conservar els **registres** generats automàticament.
- Portar a terme una **avaluació d'impacte** sobre la protecció de **dades**.
- **Informar:**

Als representants dels treballadors i als treballadors afectats per un sistema usat en el lloc de treball.

Les persones físiques **afectades** per les decisions del sistema.

Obligacions dels desplegadors de sistemes de risc limitat – Art. 50 RIA:

- Informar a les **persones exposades** a un sistema de reconeixement d'emocions o categorització biomètrica.
- Identificar les imatges, àudios o vídeos que **constitueixin una manipulació profunda** (*deepfake*) com a generats o manipulats artificialment.
- Informar si **un text** amb la finalitat d'informar sobre qüestions d'interès públic ha estat generat o manipulat artificialment.

Altres subjectes dels sistemes d'IA

Representant autoritzat:

- Persona física o jurídica situada o establerta a la UE que ha acceptat un **mandat escrit d'un proveïdor d'un sistema d'IA** per complir i dur a terme en nom seu les obligacions i els procediments establerts al RIA.

- Aquesta figura és **obligatòria** en els casos en els que el proveïdor està situat en un territori fora dels estats membres de la UE.

- Les seves obligacions es contemplen als arts. 22 i 54 RIA.

Importador:

- Persona física o jurídica situada o establerta a la UE que **comercialitza** un sistema d'IA que porta el nom o la marca comercial d'una persona física o jurídica **establerta en un tercer país**.

- Les seves obligacions es contemplen a l’art. 23 del RIA.

Distribuïdor:

- Persona física o jurídica de la **cadena de subministrament**, diferent del proveïdor o de l'importador, que posa un sistema d'IA **a disposició** del mercat de la UE.

- Les seves obligacions es contemplen a l’art. del 24 RIA.

Obligacions (I):

Les obligacions d'aquests tres subjectes es poden resumir en les següents:

- Verificar que el sistema porta el **marcatge CE** i va acompanyat de la **declaració UE de conformitat** i les **instruccions d'ús**.
- Verificar que la resta d'operadors, inclòs el proveïdors, **compleixen amb les seves obligacions**.
- **Informar** a les autoritats competents, i a la resta d'operadors, sobre riscos per a la salut, seguretat o DDFF derivats del sistema.

Obligacions (II):

- **No posar en perill** la conformitat del sistema respecte el RIA.

En cas de falta de conformitat, no comercialitzar-ho i adoptar les mesures correctores necessàries.

- **Cooperar** amb les autoritats competents en el que sol·licitin.

- El representant autoritzar: també haurà de complir amb la resta d'obligacions incloses al seu mandat, el qual haurà de rescindir en cas d'incompliment per part del proveïdor, i amb les obligacions de registre del sistema.

Terminis d'aplicació

Terminis d'aplicació – Article 113 RIA

- **1 d'agost del 2024:** entrada en vigor
- **2 de febrer del 2025:** aplicació dels capítols I (disposicions generals) i II practiques prohibides).
- **2 d'agost del 2025: :** aplicació dels capítols III Secció 4 (autoritats notificadores), V (GPAI), VII (governança), XII (confidencialitat i sancions) i l'article 78 (confidencialitat), amb l'excepció de l'article 101 (multes per a les proveïdores dels GPAI).
- **2 d'agost del 2026:** s'aplicarà la resta de la Llei d'IA, exceptuant l'article 6.1,
- **2 d'agost del 2027:** s'aplicarà l'article 6.1 (sistemes d'alt risc de l'annex I)

Sancions

Sancions – Capítol XI RIA

- **Art. 99: Sancions**

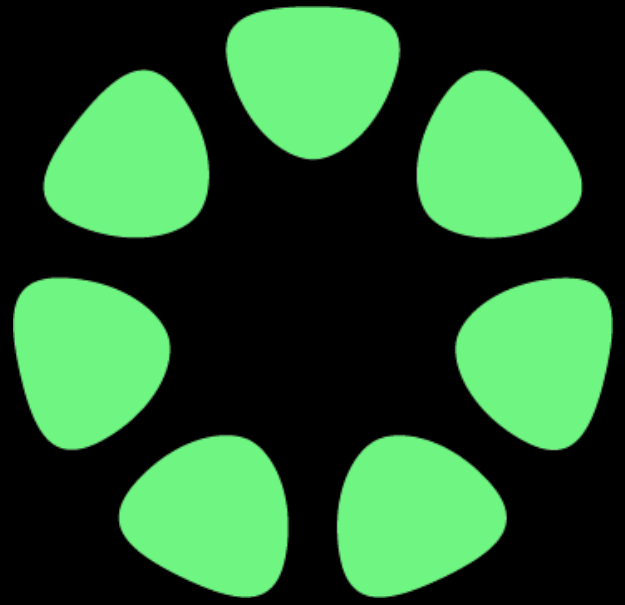
a) Prohibició de les pràctiques de la IA – fins a 35 milions € / empresa fins al 7% de llur facturació.
b) Altres pràctiques - fins a 15 milions € / empresa fins al 3%.
c) Facilitar informació incorrecte o enganyosa – fins a 7,5 milions € / empresa fins al 1%.
d) PiME – multes anteriors seran proporcionals.

- **Art. 100: Multes a institucions, òrgans i organismes de la UE**

a) Ponderació segons les característiques de l'òrgan.
b) Prohibició de pràctiques de la IA – fins a 1.500 mil €
c) Altres pràctiques – fins a 750 mil €

- **Art. 101: Multes a proveïdors de models d'IA d'ús general**

Penalització general - fins a 15 milions € / empresa fins al 3%.



Per una IA responsable: Principis, Indicadors i Observables

Model PIO

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

Generalitat de Catalunya

OEIAC

Universitat de Girona

El Model PIO

Què és el Model PIO?

1. Un model d'autoavaluació integrat per preguntes.

2. És fonamenta en la normativa aplicable i recomanacions ètiques i legals.

3. Ajuda les persones usuàries a identificar si el seu sistema compleix amb els principis legals i ètics.

Principis, Indicadors, Observables

P	I	O
Principis	**Indicadors**	**Observables**
7 principis	+130 preguntes	Resultats globals Matriu de riscos Punts de millora

Els 7 principis

Sostenibilitat:
Impacte mediambiental i social així com governança.

Autonomia:
Control de l'usuari, llibertat d'elecció i consentiment.

Privacitat:
Protecció de dades, llibertat i confiança.

Seguretat i no maleficència:
Evitar danys previsibles, no intencionats i mecanismes de seguretat.

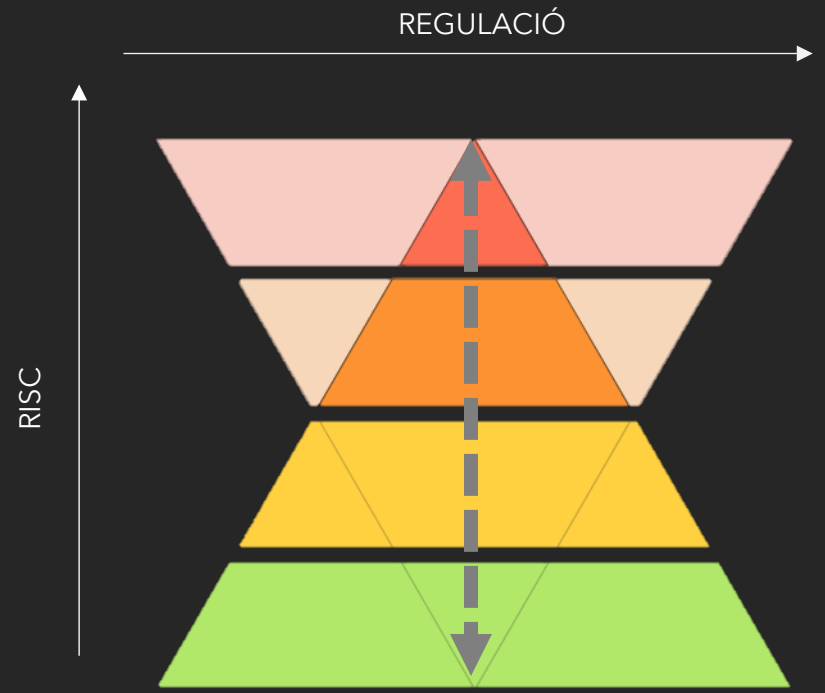


Transparència:
Explicació, interpretació i comunicació correcta.

Justícia i equitat:
Tracte just i imparcial, prevenció, seguiment i mitigació de biaixos i discriminacions.

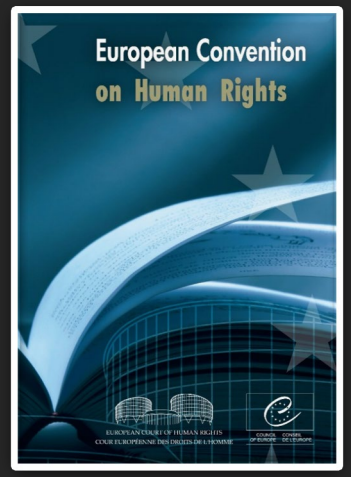
Responsabilitat i retiment de comptes:
Integritat en les accions i decisions i atribució de responsabilitat.

Avaluació adaptada a les categories de risc

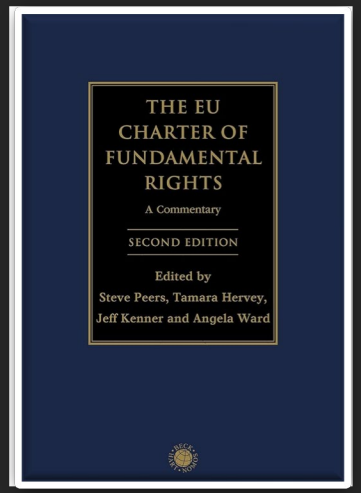


Risc Inacceptable e.g. Puntuació social	**Prohibits**
Alt risc e.g. Classificació CVs	**Permès** amb obligacions estrictes i la realització d'una avaluació de compliment
Risc limitat e.g. Chatbots	**Permès** amb obligacions de transparència associades
Risc mínim e.g. Videojocs	**Permès** amb restriccions mínimes i s'anima a seguir codis de conducta

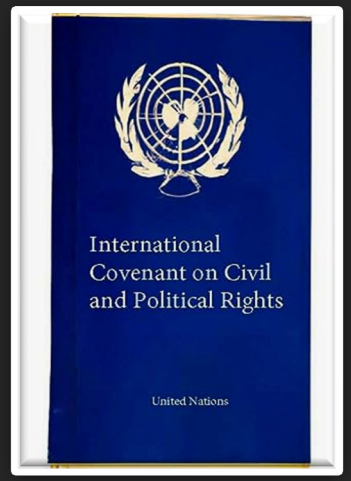
Per què obligacions i drets?



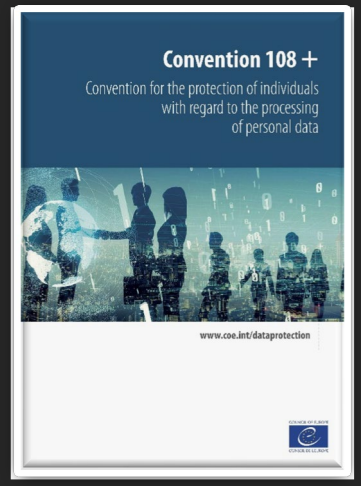










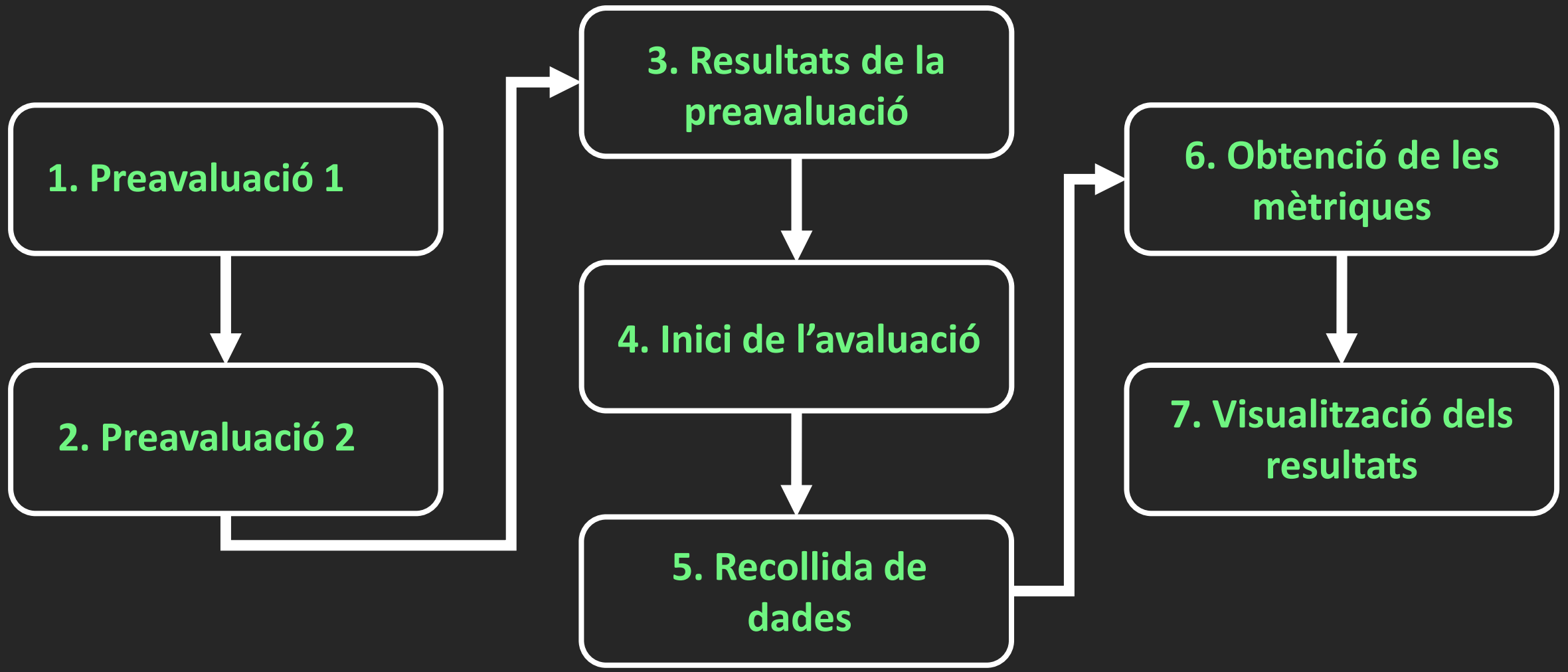


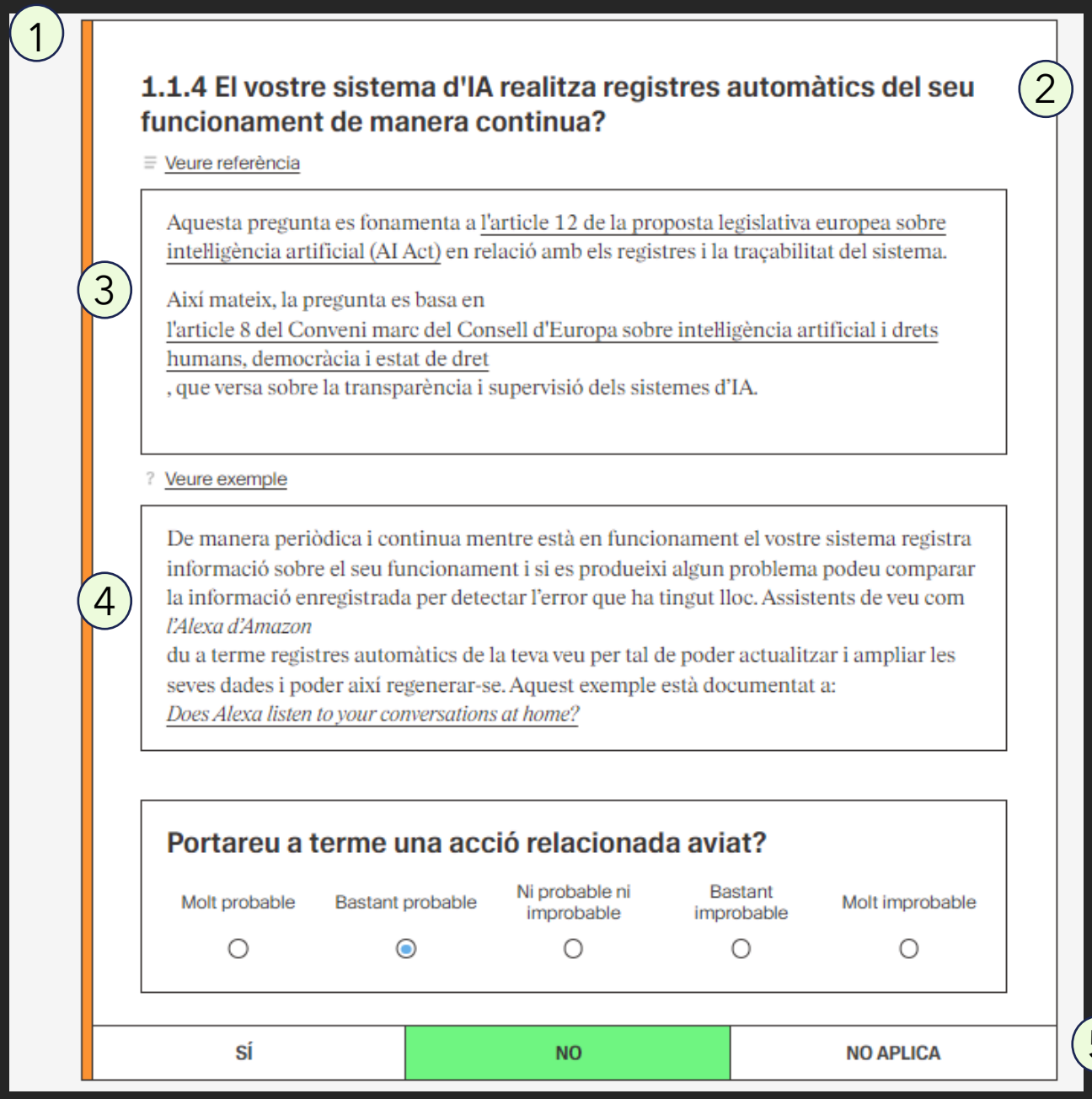


La protecció de dades més enllà del RIA

- Reglament general de protecció de dades.
- Llei europea de governança de dades.
- Conveni 108+ del Consell d'Europa sobre la protecció dels individus respecte el processament de dades personals.

Funcionament


1. Preavaluació 1
2. Preavaluació 2
3. Resultats de la preavaluació
4. Inici de l'avaluació
5. Recollida de dades
6. Obtenció de les mètriques
7. Visualització dels resultats



1- Franja de color.

2- Pregunta i numeral.

3- Referència legal.

4- Exemple.

5- Resposta.

2.2.2 Heu establert mecanismes que permetin mitigar els possibles biaixos de classificació en el vostre sistema d'IA, siguin per categories canviants o inadequades segons el context sociocultural?

Veure referència

Aquesta pregunta fa referència als articles 9, 10.3, 15.3, 53, 55.1 i 95 del Reglament 2024/1689 del Parlament i del Consell sobre intel·ligència artificial (AI Act), relatius a les mesures de gestió de riscos, la gestió de dades, la resiliència, robustesa dels sistemes d'IA, les obligacions de les persones proveïdores de sistemes de propòsit general i la promoció de codis de conducta d'aplicació voluntària en matèria de grups vulnerables.

Així mateix, la pregunta es basa en l'article 4.2 de la proposta de directiva del Parlament Europeu i del Consell relativa a normes de responsabilitat civil extracontractual a la intel·ligència artificial (AI Liability Act), relativa a la presumpció de culpa dels danys provocats per una IA.

A més, la pregunta es fonamenta en els articles 10, 12 i 16 del Conveni marc del Consell d'Europa sobre intel·ligència artificial i drets humans, democràcia i estat de dret, que versen sobre la igualtat i la no discriminació, la fiabilitat dels sistemes d'IA i el marc de gestió de riscos i impactes. Juntament amb el dret humà a la no discriminació, consagrat a l'article 14 del Conveni Europeu de Drets Humans, l'article 21 de la Carta de Drets Fonamentals de la UE, l'article 26 del Pacte Internacional de Drets Civils i Polítics i l'article 7 de la Declaració Universal dels Drets Humans.

La pregunta també està relacionada amb el principi *Self-identification* del document A Human Rights-Based Approach To Data de l'ACNUDH l'apartat de *Diversity, Non-discrimination and Fairness* d'ALTAI.

? Veure exemple

Per tal de garantir la mitigació de biaixos es pot recopilar dades de manera equilibrada i una de les tècniques és el submostreig aleatori o *Random Undersampling*. Consisteix en reduir la mida de la classe de majoritària i així s'iguala amb la classe minoritària. En el següent enllaç hi trobareu informació: https://www.sciencedirect.com/science/article/pii/S1877050919313456 i

https://ieeexplore.ieee.org/document/9078901.

SÍ	NO	NO APLICA

5.1.2 Heu establert mecanismes de protecció de dades per tal que les dades de caràcter personal subministrades per la persona interessada no siguin venudes, llogades ni cedides, sense el seu consentiment?

Veure referència

Aquesta pregunta fa referència als articles 10.5, 17.1 50.3 i 55.1 del Reglament 2024/1689 del Parlament i del Consell sobre intel·ligència artificial (AI Act), que versen sobre la governança de dades i el sistema de gestió de qualitat dels sistemes d'alt risc, així com les obligacions dels sistemes de reconeixement d'emocions, categorització biomètrica i de propòsit general de risc sistemàtic.

La pregunta també esta relacionada amb els articles 6 i 7 del Reglament 2016/679, de 27 d'abril 2016 [Reglament general de protecció de dades] que versen sobre la licitud del tractament de dades i el consentiment dels afectats, i l'article 21 de la Regulació 2022/868, de 30 de Maig del 2022 [Llei Europea sobre la Governança de dades], sobre els requisits específics per a protegir els drets i els interessos de les persones interessades i titulars de dades perquè respecta a les seves dades.

Així mateix, la pregunta es fonamenta en l'article 11 del Conveni marc del Consell d'Europa sobre intel·ligència artificial i drets humans, democràcia i estat de dret, que versa sobre la privacitat i protecció de dades personals; l'article 12 de la Declaració Universal dels Drets Humans, l'article 8 del Conveni Europeu de Drets Humans, i l'article 17 del Pacte Internacional de Drets Civils i Polítics, sobre el Dret humà a la privacitat; els articles 7 i 8 de la Carta de Drets Fonamentals de la UE sobre la privacitat i la protecció de dades personals; i l'article 7 del Conveni 108 per a la protecció de les persones físiques pel que fa al tractament de dades personals, que versa sobre la seguretat de les dades.

En aquest sentit, la qüestió guarda relació amb el principi *Privacy* del document A Human Rights-Based Approach To Data de l'ACNUDH i les recomanacions contingudes al *toolkit* d'ICO sobre riscos i protecció de dades en la IA.

? Veure exemple

L'empresa *Meta* va ser multada el maig del 2023 per violar la Llei de Protecció de Dades incomplint amb la sol·licitud del consentiment de deure de protecció de les dades personals, tal com s'explica als següents articles: *This article is more than 4 months old Facebook owner Meta fined €1.2bn for mishandling user information, Meta Fined $1.3 Billion for Violating E.U. Data Privacy Rules* i *Meta slapped with record $1.3 billion EU fine over data privacy*.

SÍ	NO	NO APLICA

Els resultats

Mètriques (I): Resultats globals

Resultats globals

1 Transparència i explicabilitat
Preguntes contestades: 24 de 24 — 70% SÍ / 30% NO

2 Justícia i equitat
Preguntes contestades: 20 de 21 — 95% SÍ / 5% NO

3 Seguretat i no maleficència
Preguntes contestades: 20 de 20 — 100% SÍ / 0% NO

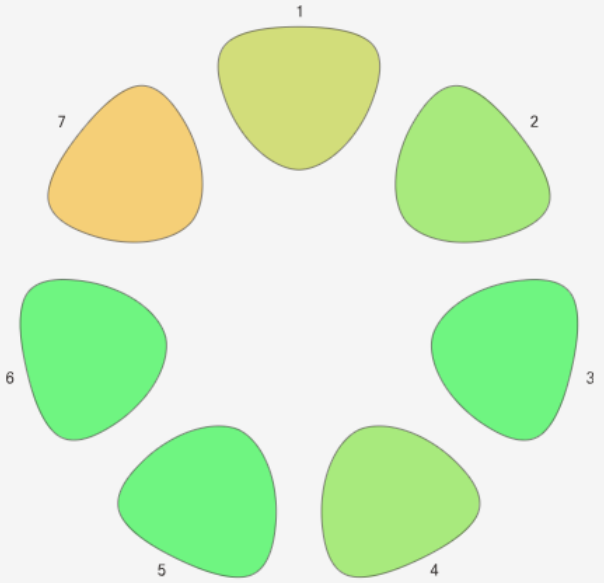
4 Responsabilitat i retiment de comptes
Preguntes contestades: 19 de 19 — 94% SÍ / 6% NO

5 Privacitat
Preguntes contestades: 19 de 19 — 100% SÍ / 0% NO

6 Autonomia
Preguntes contestades: 15 de 15 — 100% SÍ / 0% NO

7 Sostenibilitat
Preguntes contestades: 16 de 16 — 56% SÍ / 44% NO

0% 0-19% 20-39% 40-59% 60-79% 80-99%



Descarregar insígnia | Descarregar resultats

Mètriques (II): Matriu de riscos

Matriu de riscos

La següent matriu de riscos és una eina per obtenir una fotografia del nivell de risc segons la severitat (normativa o AI Act) i la probabilitat de portar a terme una acció al respecte. Els valors en cada casella representen les respostes negatives proporcionades segons la severitat i probabilitat esmentades. Aquesta informació també s'utilitza per als punts de millora.

	SEVERITAT SEGONS LA NORMATIVA O AI ACT		
PROBABILITAT DE PORTAR A TERME UNA ACCIÓ	Mínima	Limitada	Alta
Molt alta	1	0	4
Alta	2	0	4
Ni alta ni baixa	1	0	0
Baixa	1	1	2
Molt baixa	0	0	0

Mètriques (III): Punts de millora

Punts de millora

A continuació trobareu aquelles preguntes que heu contestat amb un "NO" i el seu grau de severitat. Els colors utilitzats en aquesta secció fan referència a la combinació donada entre els nivells de severitat (segons la normativa de la AI Act) i la probabilitat de que porteu a terme alguna acció aviat. Aquests apareixen en la mateixa taula anterior.

Molt baix | Baix | Baix/mitjà | Mitjà | Mitjà/alt | Alt

1 Transparència i explicabilitat

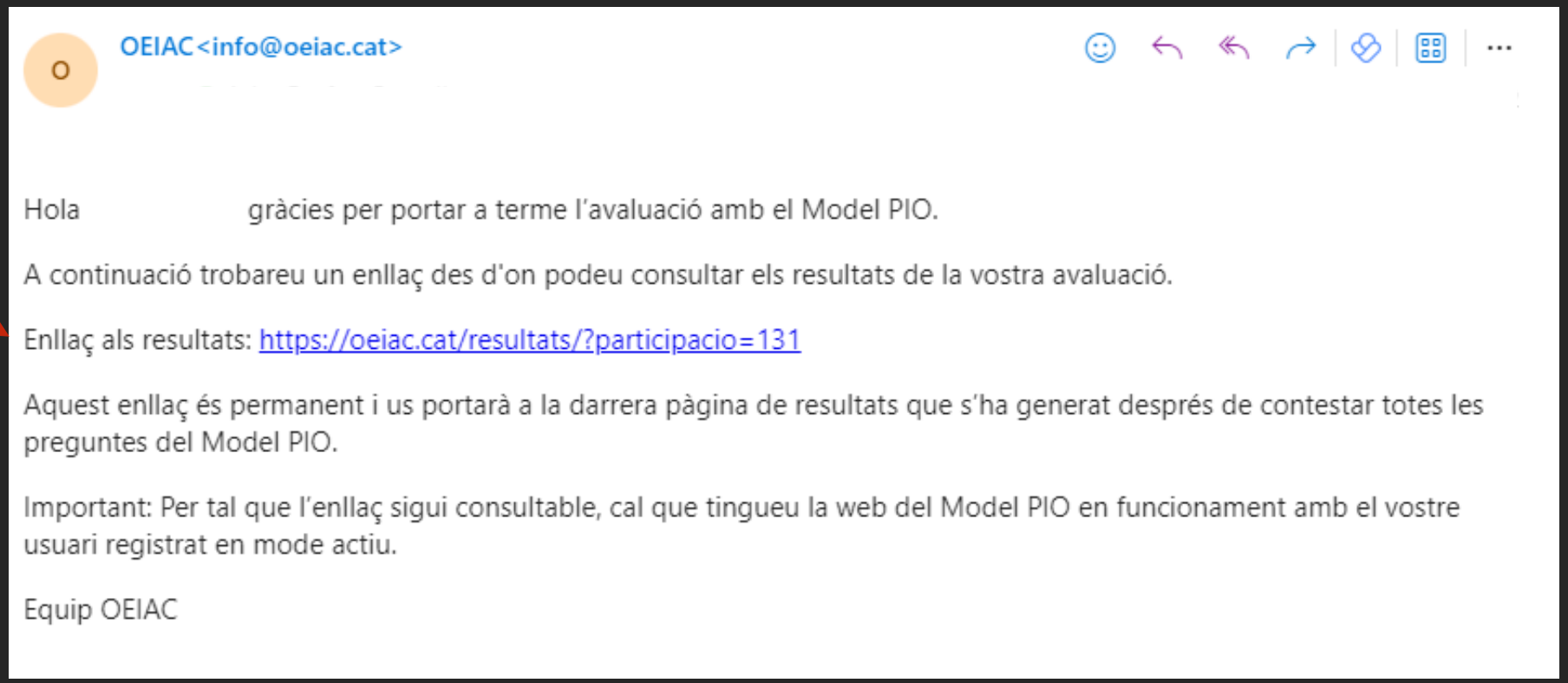
1.1.5 Heu facilitat o posat a disposició informació sobre la identitat i les dades de contacte de la persona proveïdora (o els seus representants)?	Aquesta pregunta es fonamenta en els articles 13.3, 22 i 54 del Reglament 2024/1689 del Parlament i del Consell sobre intel·ligència artificial (AI Act), que versen sobre la transparència i sobre les obligacions dels representants autoritzats del proveïdors de sistemes d'alt risc i els models de propòsit general. Així mateix, la pregunta es fonamenta en els articles 8 i 14 del Conveni marc del Consell d'Europa sobre intel·ligència artificial i drets humans, democràcia i estat de dret, que versen sobre la transparència i supervisió dels sistemes d'IA i els recursos per a les vulneracions dels drets humans derivades dels sistemes d'IA. Juntament amb el dret a la informació reconegut a l'article 19 del Pacte Internacional de Drets Civils i Polítics, l'article 11 de la Carta de Drets Fonamentals de la UE, l'article 10 del Conveni Europeu de Drets Humans i l'article 19 de la Declaració Universal dels Drets Humans.

Mètriques (III): Punts de millora

2 Justícia i equitat

2.1.1 Heu establert mesures d'inclusió i equitat com provar els resultats del vostre sistema d'IA per a diferents grups afectats?	Aquesta pregunta fa referència als articles 9, 10.3, 15.3, 27.1 i 95 del Reglament 2024/1689 del Parlament i del Consell sobre inteHigència artificial (AI Act), relatius a les mesures de gestió de riscos, la governança de dades, la resiliència i robustesa dels sistemes d'IA, l'avaluació d'impacte dels drets fonamentals per a sistemes d'IA d'alt risc i la promoció de codis de conducta d'aplicació voluntària per a persones o grups vulnerables. La pregunta també es relaciona amb els articles 10 i 16 del Conveni marc del Consell d'Europa sobre inteHigència artificial i drets humans, democràcia i estat de dret, que versen sobre la Igualtat i no discriminació i el marc de gestió de riscos i impactes. Juntament amb el dret humà la no discriminació, consagrat a l'article 14 del Conveni Europeu de Drets Humans, l'article 21 de la Carta de Drets Fonamentals de la UE, l'article 26 del Pacte Internacional de Drets Civils i Polítics i l'article 7 de la Declaració Universal dels Drets Humans. A més, la pregunta també està relacionada l'apartat *Specific safeguards to ensure non-discrimination when using* de l'informe *Getting The Future Right, Artificial Intelligence And Fundamental Rights* de la FRA. D'altra banda, la pregunta es fonamenta en l'apartat de *Fairness and non-discrimination* de les recomanacions sobre ètica en la InteHigència Artificial d'UNESCO, i està relacionada amb l'apartat *Fairness* del AI Ethics Assessment de GSMA.
2.1.2 Us heu assegurat que el vostre sistema d'IA no utilitza variables o "proxies" que poden ser injustament discriminatòries?	Aquesta pregunta fa referència als articles 9, 15.3 27.1, 55.1 i 95 del Reglament 2024/1689 del Parlament i del Consell sobre inteHigència artificial (AI Act), relatius a les mesures de gestió de riscos, la resiliència i robustesa dels sistemes d'IA, l'avaluació d'impacte dels drets fonamentals per a sistemes d'IA d'alt risc, les obligacions des proveïdors de sistemes de propòsit general de risc sistèmic i la promoció de codis de conducta d'aplicació voluntària per a persones o grups vulnerables. i la promoció de codis de conducta d'aplicació voluntària per a persones o grups vulnerables. A més, la pregunta es fonamenta en l'article 10 del Conveni marc del Consell d'Europa sobre inteHigència artificial i drets humans, democràcia i estat de dret, que versa sobre la Igualtat i no discriminació, així com el dret humà a la no discriminació, consagrat a l'article 14 del Conveni Europeu de Drets Humans, l'article 21 de la Carta de Drets Fonamentals de la UE, l'article 26 del Pacte Internacional de Drets Civils i Polítics i l'article 7 de la Declaració Universal dels Drets Humans. En aquest sentit, la pregunta també està relacionada amb l'apartat *Specific safeguards to ensure non-discrimination when using* de l'informe *Getting The Future Right, Artificial Intelligence And Fundamental Rights* de la FRA. Finalment, destaquem l'apartat de *Fairness and non-discrimination* de les recomanacions sobre ètica en la InteHigència Artificial d'UNESCO i l'apartat *Fairness* del *AI Ethics Assessment* de GSMA.
2.1.3 Heu avaluat si el vostre sistema d'IA s'adapta a un ampli espectre de preferències i habilitats individuals, incloent-hi usuaris que requereixen tecnologies d'assistència?	Aquesta pregunta fa referència a l'article 95 del Reglament 2024/1689 del Parlament i del Consell sobre inteHigència artificial (AI Act), relatiu a la promoció de codis de conducta d'aplicació voluntària en matèria d'accessibilitat. A més, la pregunta es fonamenta en l'article 10 del Conveni marc del Consell d'Europa sobre inteHigència artificial i drets humans, democràcia i estat de dret, que versa sobre la Igualtat i no discriminació, així com el dret humà a la no discriminació, consagrat a l'article 14 del Conveni Europeu de Drets Humans, l'article 21 de la Carta de Drets Fonamentals de la UE, i l'article 26 del Pacte Internacional de Drets Civils i Polítics. Destaquem també les recomanacions contingudes a l'apartat de *Diversity, Non-discrimination and Fairness: Accessibility and Universal Design* d'ALTAI, i està relacionada amb l'apartat *Fairness del AI Ethics Assessment de GSMA.*

Mètriques (IV)



O OEIAC<info@oeiac.cat>

Hola gràcies per portar a terme l'avaluació amb el Model PIO.

A continuació trobareu un enllaç des d'on podeu consultar els resultats de la vostra avaluació.

Enllaç als resultats: https://oeiac.cat/resultats/?participacio=131

Aquest enllaç és permanent i us portarà a la darrera pàgina de resultats que s'ha generat després de contestar totes les preguntes del Model PIO.

Important: Per tal que l'enllaç sigui consultable, cal que tingueu la web del Model PIO en funcionament amb el vostre usuari registrat en mode actiu.

Equip OEIAC

Altres recursos

Catàlegs: Exemples i Legal i recomanacions

A-B

Exemples

Exemples reals i hipotètics a les preguntes del Model PIO.

1 Transparència i explicabilitat +

2 Justícia i equitat +

3 Seguretat i no maleficència +

4 Responsabilitat i retiment de comptes ×

Control

	PREGUNTA MODEL PIO
Acomiadament Sr.Diallo	4.1.4
Formacions especialitzades	4.1.1
MasterCard	4.1.2
Persona responsable (H)	4.1.5
Protocol de monitoratge continu (H)	4.1.6
Revisió manual (H)	4.1.3

100+

Legal i recomanacions

Normes jurídiques en les que s'han inspirat les preguntes del Model PIO.

1 Transparència i explicabilitat +

2 Justícia i equitat +

3 Seguretat i no maleficència ×

1. Funcionament del sistema

	ARTICLE	PREGUNTA MODEL PIO
Conveni marc del Consell d'Europa sobre intel·ligència artificial i drets humans, democràcia i estat de dret	8 12 14 16	3.1.1 3.1.2 3.1.3 3.1.4 3.1.5
Proposta Directiva del Parlament Europeu del Consell relativa a normes de responsabilitat civil extracontractual a la intel·ligència artificial [AI Liability Act]	4	3.1.5
Reglament 2024/1689 del Parlament i del Consell sobre intel·ligència artificial [AI Act]	9 13 14 17 27 53 55 72	3.1.1 3.1.2 3.1.3 3.1.4 3.1.5
Assessment list for Trustworthy artificial Intelligence (ALTAI) for self assessment		3.1.1 3.1.4 3.1.5

2. Traçabilitat per a la gestió del risc

	ARTICLE	PREGUNTA MODEL PIO
Carta de Drets Fonamentals de la UE	1 3 7 8	3.2.1

Recomanacions de clàusules contractuals

Contractes

En aquest apartat trobareu contingut recomanat per a la redacció de clàusules de contractes relatius a sistemes d'IA. Aquest contingut organitzat a través dels principis del Model PIO, és rellevant per a molts casos d'ús i per determinar l'aplicació de la legislació europea sobre intel·ligència artificial (AI Act) així com el compliment dels més alts estàndards de responsabilitat, seguretat i respecte pels drets humans.

1 Transparència i explicabilitat +

2 Justícia i equitat +

3 Seguretat i no maleficència ×

→ 3.1. Funcionament del sistema

Nº	RECOMANACIÓ DE CLÀUSULA
3.1.1	El sistema anirà acompanyat de la següent informació: • El nivell de precisió i seguretat del sistema. • La mesura en la qual influeix en la presa de decisions i en el treball humà i el grau d'adequació del sistema

Avaluacions

Avaluacions

Avaluació de conformitat
Avaluació d'impacte sobre drets fonamentals
Obligacions de transparència

Avaluació de conformitat:

La AI Act exigeix que les persones proveïdores dels sistemes d'alt risc sotmetin la seva IA a una avaluació de conformitat (art.16 i desenvolupada a l'art. 43) i remet als annexos de la Llei per a la seva especificitat. Aquesta consisteix a demostrar que el sistema en qüestió compleix amb els requisits que li són imposats per la AI Act.

La regulació preveu diferents modalitats d'avaluació en funció del tipus de sistema d'alt risc.

A continuació, us proporcionem informació sobre les característiques principals de les avaluacions de conformitat previstes a la AI Act.

1. Procediment d'avaluació intern contemplat a l'annex VI de la AI Act.

Quins sistemes han de ser sotmesos a aquesta avaluació?

Subjectes

Guia d'actors subjectes al RIA

Les diferents persones relacionades amb el desenvolupament, comercialització i implementació d'un sistema d'IA no han de respondre de la mateixa manera a totes les obligacions contemplades al Reglament, sinó que el RIA assigna diferents obligacions segons el rol particular que juga cadascun d'aquests subjectes. Aquesta guia té l'objectiu de facilitar als operadors identificar quin és el seu rol en relació amb el RIA i sobre quins requisits se'ls pot fer responsables en cas d'incompliment.

- Proveïdor +
- Representant autoritzat +
- Importador +
- Distribuidors +
- Desplegador o responsable del desplegament +
- Responsabilitats al llarg de la cadena de valor +

Més recursos

La AI Act

Antecedents legislatius
Cronologia
Terminis d'aplicació
Altres dates rellevants
AI Liability Directive

Recursos

Avaluem
Acompanyem
Aconsellem
Formem

Models PIO sectorials

Models sectorials



Salut

- **Accés al Model Sectorial Salut**
- Catàleg Legal Salut
- Clàusules de recomanacions
- Informe



Educació

- **Accés al Model Sectorial Educació**
- Catàleg Legal Educació
- Clàusules de recomanacions
- Guies educació alumnat
- Guies educació professorat
- Informe



www.oeiac.cat
X: @OEIAC_UdG
arlet.brufau@udg.edu


TQRCG

