EduTechCluster

Avaluació sobre l'ús ètic de dades i sistemes d'intel·ligència artificial

Jornades: Impuls de la IA en l'educació
Reptes i propostes

Jornada Girona
20 d'octubre de 2023

EIAC
Observatori d'Ètica en Intel·ligència Artificial de Catalunya

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.

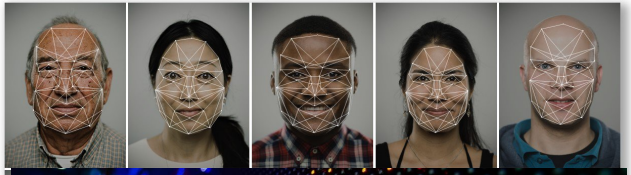
CATALONIA.AI inclou un eix sobre "Ètica i Societat" per promoure el desenvolupament d'una IA ètica.

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

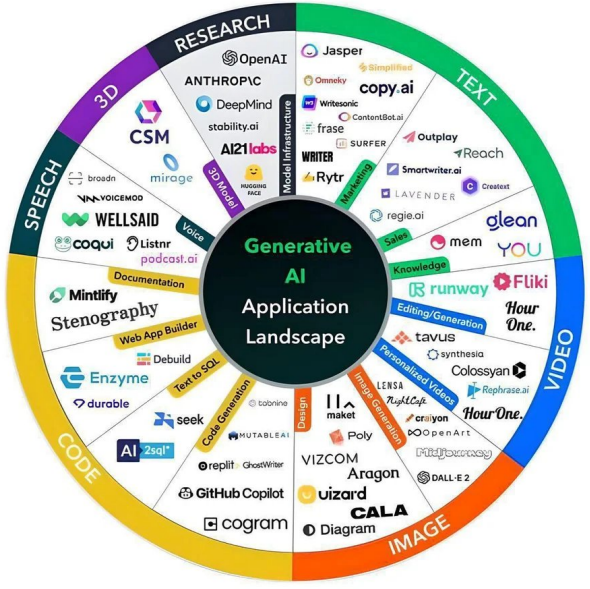
Objectiu general

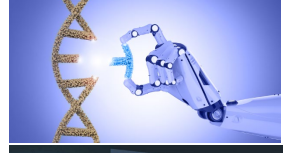
Estudiar les ***conseqüències*** ètiques, socials i legals, així com els riscos i oportunitats de la implantació de la IA en la vida diària a Catalunya, des d'una òptica plenament transversal.

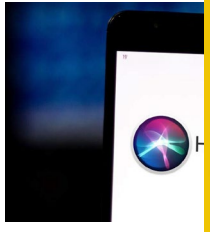

MEDIONmobile
5:31 PM
Netflix
Einstellungen
Fotos
Wattpad
Facebook
129
Kamera
Twitter
App Store
Mail
3



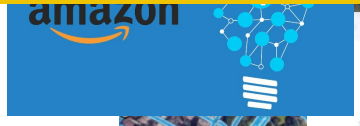

AI
in Education


Generative AI Application Landscape
RESEARCH
TEXT
VIDEO
IMAGE
CODE
SPEECH
3D
OpenAI
ANTHROPIC
DeepMind
stability.ai
AI21labs
CSM
mirage
broadn
VOICEMOD
WELLSAID
coqui
Listnr
podcast.ai
Mintlify
Stenography
Debuild
Enzyme
durable
seek
AI2sql
replit
GhostWriter
GitHub Copilot
cogram
Jasper
Simplified
Omneky
copy.ai
Writesonic
ContentBot.ai
frase
SURFER
Outplay
Reach
WRITER
Rytr
Smartwriter.ai
LAVENDER
Creatext
regie.ai
glean
mem
YOU
runway
Fliki
Hour One.
tavus
synthesia
Colossyan
Rephrase.ai
LENSA
NightCafe
craiyon
OpenArt
Midjourney
DALL·E 2
maket
Poly
VIZCOM
Aragon
uizard
CALA
Diagram
tabnine
MUTABLE.AI
Model Infrastructure
3D Model
Marketing
Sales
Knowledge
Editing/Generation
Personalized Videos
Image Generation
Design
Code Generation
Text to SQL
Web App Builder
Documentation
Voice

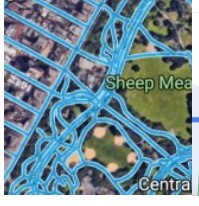


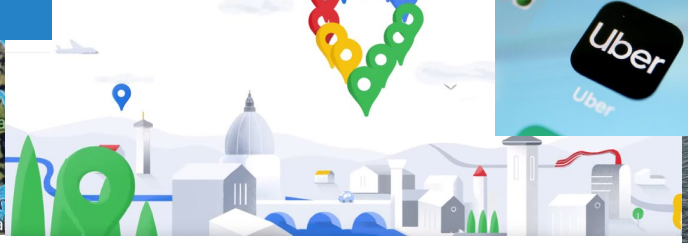


Great La


amazon


JUST ASK
amazon alexa
amazon


Sheep Mea
Centra



Uber
Uber

Principis i limitacions

Proliferació de principis per l'ús ètic de la IA

Adopció de principis per l'ús ètic de la IA **segons context**

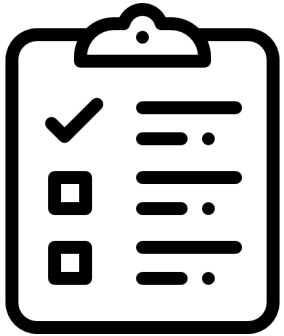
Avaluació de principis per l'ús ètic de la IA **segons eines disponibles**

Toolkits i checklists

- Existeixen diversos recursos, especialment des del món acadèmic i d'algunes organitzacions sense ànim de lucre, que es poden dividir en ***toolkits*** i ***checklists***.

- Mentre els primers són eines específiques per avaluar de manera tècnica aspectes determinats (e.g. seguretat, privacitat), els segons són **eines de verificació per determinar si un projecte o organització s'adequa o no a diferents usos ètics de les dades i sistemes d'IA.**

Els checklists d'autoavaluació I



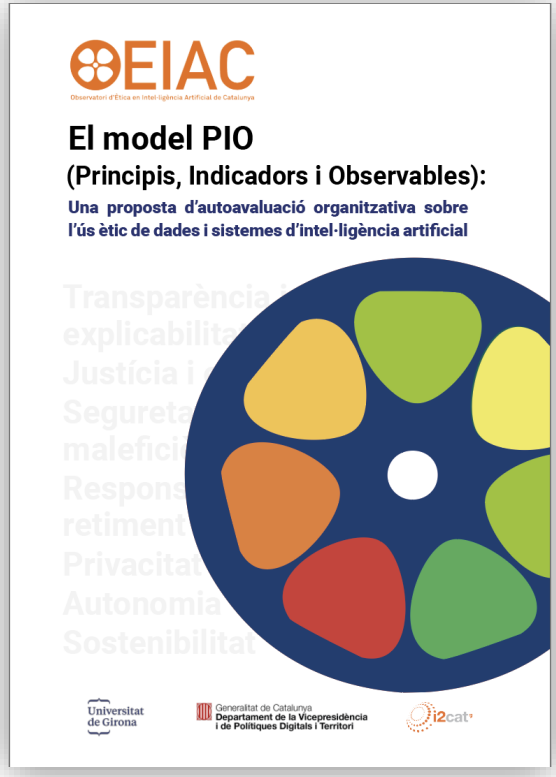
- Els checklists són útils **per observar i avaluar** l'adequació organitzativa envers els usos ètics de dades i sistemes d'IA.
- D'una manera holística: considerant **diferents principis** i **diferents etapes** del cicle de vida d'un sistema d'IA.
- Aspiracionals però, molt sovint, esdevenen **operacionals en accions concretes.**

Els checklists d'autoavaluació II

Diferents **organismes internacionals** com la Comissió Europea, UNESCO, la OECD o l'Ada Lovelace Institute **han publicat guies i checklists d'adequació ètica** per a sistemes de IA basant-se en diferents principis ètics per aplicar-los als sectors públic i privat.

Ethics Guidelines for Trustworthy AI (European Commission, 2019), *Draft text of the Recommendation on the Ethics of Artificial Intelligence* (UNESCO, 2021), *Declaració de Barcelona* (IIIA CSIC, 2017), *AI Ethics Impact Group: From Principles to Practice* (AIEI Group, 2020), *Technical methods for regulatory inspection of algorithmic systems in social media platforms* (Ada Lovelace Institute, 2021), *AI systems classification framework at the OECD* (OECD, 2022), *AI and Data Protection Risk Mitigation And Management Toolkit* (ICO, 2021)

Desenvolupament de la primera proposta i metodologia d'avaluació PIO



NIVELL ABSTRACTE

Revisió teòrica

Revisió teòrica de més de 150 fonts bibliogràfiques sobre metodologies de checklists i principis ètics

NIVELL FORMAL

Adopció dels principis

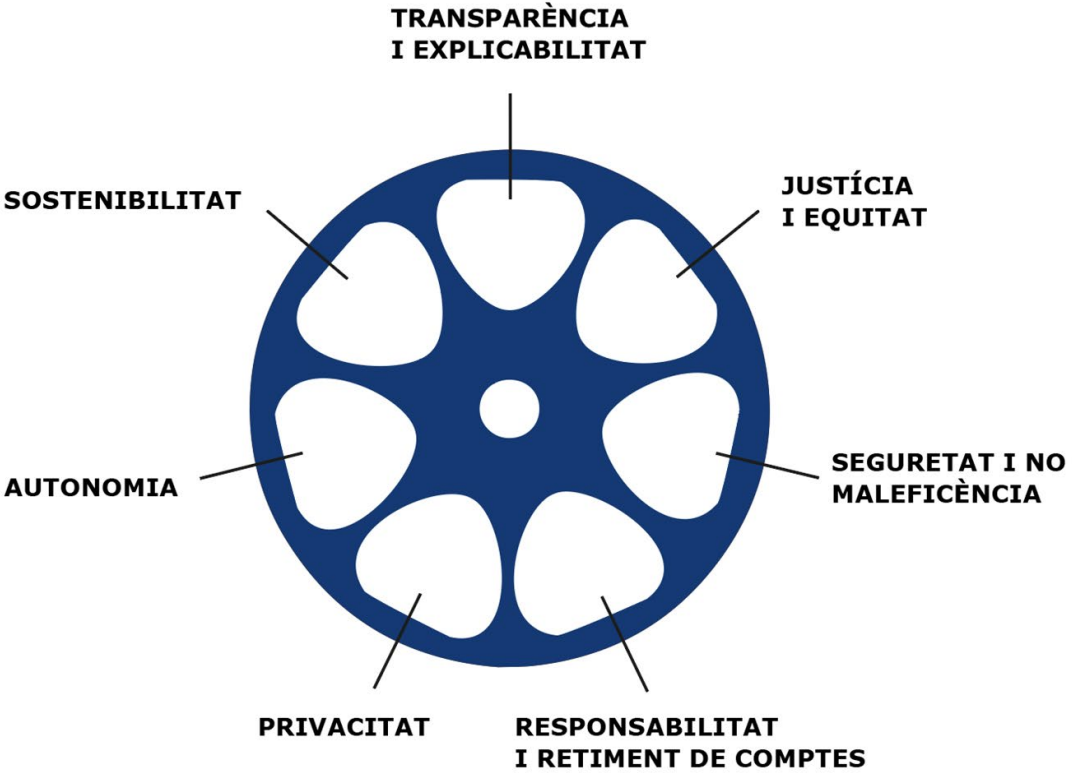
Anàlisi de documents, models, toolkits i projectes de llei relacionats amb l'ús de la IA, principalment en el context europeu.

NIVELL PRÀCTIC

Aplicació pràctica

Desenvolupament d'un model propi d'aplicació d'auditoria ètica

Adopció de 7 principis



Transparència i explicabilitat

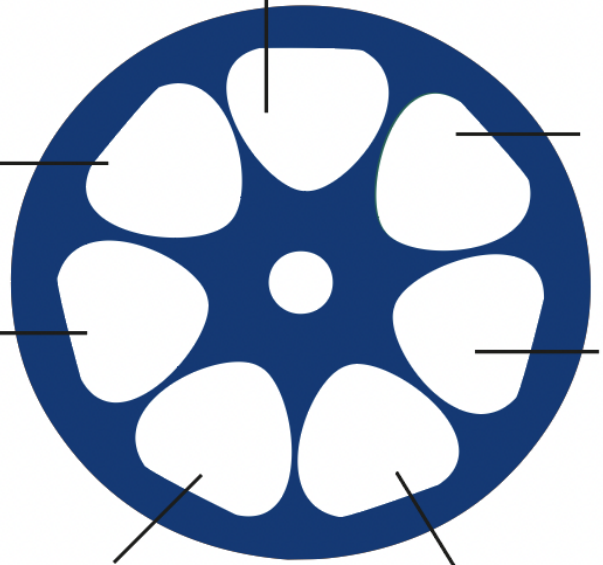
1. Anàlisi preliminar de l'ús sistema
2. Anàlisi técnic del sistema
3. Explicabilitat
4. Precisió del sistema
5. Traçabilitat
6. Protocols de comunicació

Sostenibilitat

1. Desenvolupament sostenible
2. Impacte mediambiental

Justícia i equitat

1. Inclusió i equitat
2. Identificació de biaxos



Autonomia

1. Impacte en les capacitats dels usuaris
2. Percepció del sistema

Seguretat i no maleficència

1. Funcionament del sistema
2. Traçabilitat per la gestió del risc
3. Mecanismes de seguretat
4. Protecció enfront d'atacs

Privacitat

1. Protecció de la privacitat
2. Governança de dades

Responsabilitat i rendició de comptes

1. Control
2. Rendició de comptes
3. Auditabilitat
4. Figures de protecció

El model PIO

Model d'autoavaluació PIO

Presentem el model d'autoavaluació PIO (Principis, Indicadors i Observables) per avançar en l'avaluació ètica de dades i sistemes d'intel·ligència artificial a través d'un formulari de verificació o checklist

Iniciar Formulari

Objectius

Objectius del model PIO

Sensibilitzar als diferents agents de la quàdruple hèlix que utilitzen dades i sistemes d'intel·ligència artificial, sobre la importància d'adoptar principis ètics fonamentals per minimitzar riscos coneguts i desconeguts i maximitzar oportunitats.

Identificar accions adequades o inadequades a través d'una proposta d'autoavaluació basada en principis, indicadors i observables per valoritzar, recomanar i avançar en l'ús ètic de dades i sistemes d'intel·ligència artificial.

Principis, Indicadors, Observables

P

Principis
7 principis ètics

I

Indicadors
70 preguntes

O

Observables
Respostes

Senzillesa

El model PIO es construeix al voltant de la senzillesa i d'una pregunta clau i efectiva que qualsevol persona que desenvolupa, gestiona o dirigeix un projecte d'IA pot entendre fàcilment: **Ho hem fet?**

Aplicable a tot el cicle de vida: disseny i modelització; desenvolupament i validació, desplegament, seguiment i perfeccionament

Com funciona I

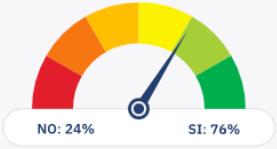
FASE 1

Presentació del principi a través de diversos indicadors

FASE 2

Recollida de totes les respostes observables

FASE 3



Avaluació basada en el percentatge de preguntes recollides

FASE 4

Visualització del resultat a través de la insígnia OEIAC

Com funciona II

Tipus d'avaluació

Només podràs sol·licitar el feedback de l'OEIAC en el cas que hagis justificat totes les teves respostes. En el cas que prefereixis fer una avaluació ràpida o no estiguis segur, podràs canviar d'opinió en qualsevol moment mentre realitzes l'avaluació i justificar les respostes que ja hagis contestat per a obtenir el feedback de l'OEIAC.

Tria el tipus d'autoavaluació *

- ○ Desitjo fer l'avaluació ràpida
- ○ Desitjo fer l'avaluació completa

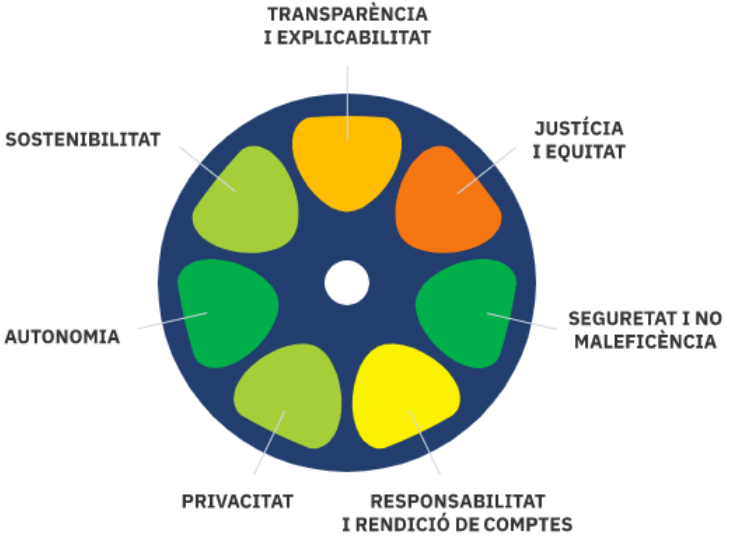
Declaració de responsabilitat *

- ☐ Afirmo haver llegit i accepto la Declaració de Responsabilitat

Següent | Save Draft

Com funciona III

PRINCIPI	RESPOSTES
TRANSPARÈNCIA I EXPLICABILITAT PREGUNTES CONTESTADES: 15 DE 16	27% SÍ / 73% NO
JUSTÍCIA I EQUITAT PREGUNTES CONTESTADES: 8 DE 10	13% SÍ / 87% NO
SEGURETAT I NO MELEFICÈNCIA PREGUNTES CONTESTADES: 11 DE 11	82% SÍ / 18% NO
RESPONSABILITAT I RETIMENT DE COMPTES PREGUNTES CONTESTADES: 7 DE 9	43% SÍ / 57% NO
PRIVACITAT PREGUNTES CONTESTADES: 9 DE 10	78% SÍ / 22% NO
AUTONOMIA PREGUNTES CONTESTADES: 5 DE 6	100% SÍ / 0% NO
SOSTENIBILITAT PREGUNTES CONTESTADES: 8 DE 8	75% SÍ / 25% NO



Descarregar insígnia

Descarregar resultats

A qui va dirigit

A totes les organitzacions públiques o privades que tinguin un interès en el disseny, desenvolupament o desplegament de tecnologies d'IA, i a totes les persones que estiguin utilitzant, comprant o siguin destinatàries

Quins són els riscos i beneficis?
Qui es veu afectat per elles i com?
De quina manera es pot millorar el benestar de les persones amb aquestes tecnologies?

Accés i registre

Accedir

Email

Contrasenya

Recordar-me

Accedir

Registrar-me | Has perdut la contrasenya?

Encara no estàs registrat?

Per poder accedir al nostre model d'autoavaluació has d'estar registrat

Registrar-me

Registre

Nom*

Cognoms*

Nom de l'organització*

Adreça electrònica de l'organització*

Aquesta direcció de correu serà el seu nom d'usuari

Contrasenya*

Longitud mínima de 8 caràcters. La contrasenya ha de tenir una força mínima de Medium

Indicador de força

Repetir contrasenya*

Accepto la Política de Privadesa i la Política de Protecció de Dades *

Registrar-me

Sumari

Avaluem

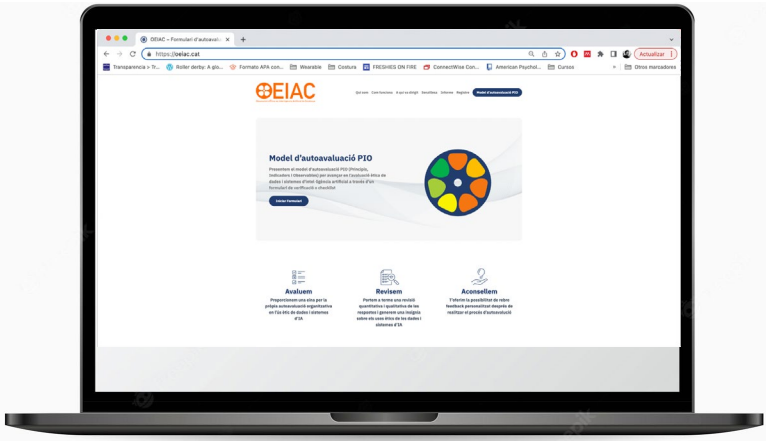
Proporcionem una eina per la pròpia autoavaluació organitzativa en l'ús ètic de dades i sistemes d'IA

Revisem

Portem a terme una revisió quantitativa i qualitativa de les respostes i generem una insígnia sobre els usos ètics de les dades i sistemes d'IA

Aconsellem

T'oferim la possibilitat de rebre feedback personalitzat després de realitzar el procés d'autoavaluació



Nº Registres:

+300

https://oeiac.cat/

report BENIFEI, TUDORACHE A9-0188/2023

Commission proposal

620

\+ 499

— 028

○ 093

ARTIFICIAL INTELLIGENCE

In NYC, companies will have to prove their AI hiring software isn't sexist or racist

AI-infused hiring programs have drawn scrutiny, most notably over whether they end up exhibiting biases based on the data they're trained on.

July 5, 2023, 3:00 PM CEST

By Kevin Collier

EduTechCluster

@OEIAC_UdG @albert_sabater

Jornades: Impuls de la IA en l'educació
Reptes i propostes

Jornada Girona
20 d'octubre de 2023

OEIAC
Observatori d'Ètica en Intel·ligència Artificial de Catalunya

Universitat de Girona