IA: reptes i oportunitats per als parlaments i els poders públics

Democràcia i intel·ligència artificial

La intel·ligència artificial en els parlaments i en el sector públic

Intel·ligència artificial i regulació

Elaborat per:

upf. Universitat Pompeu Fabra Barcelona

cap cit Consell Assessor del Parlament sobre Ciència i Tecnologia

IA: reptes i oportunitats per als parlaments i els poders públics

Democràcia i intel·ligència artificial

— Dra. Migle Laukyte, Departament de Dret
— Dra. Clara I. Velasco Rico, Departament de Dret
— Sr. Roger Cuartielles, Departament de Comunicació
— Dra. Júlia Vilasís-Pamos, Departament de Comunicació
— Dr. Marcel Mauri de los Rios, Departament de Comunicació

La intel·ligència artificial en els parlaments i en el sector públic

— Dra. Migle Laukyte, Departament de Dret
— Dra. Clara I. Velasco Rico, Departament de Dret

Intel·ligència artificial i regulació

— Sr. Aurelio Ruiz García, Departament d'Enginyeria
— Dra. Mireia Artigot Golobardes, Departament de Dret

upf. Universitat Pompeu Fabra Barcelona

Barcelona, maig de 2025

Prefaci

El Consell Assessor del Parlament sobre Ciència i Tecnologia (CAPCIT) va iniciar una tasca divulgativa sobre la intel·ligència artificial (IA) en el marc de la presidència de la Xarxa Parlamentària Europea d'Avaluació Tecnològica (EPTA), que el Parlament de Catalunya va exercir l'any 2023.

Arran d'aquella primera experiència, els membres de l'EPTA van acordar elaborar un segon informe específic sobre IA i democràcia amb l'objectiu d'informar els parlaments que formen part d'aquesta xarxa.

El CAPCIT, amb la voluntat d'anar més enllà de la participació feta en el marc de l'informe *Intel·ligència artificial i democràcia: Informe EPTA 2024*, va decidir elaborar-ne un d'específic per al Parlament de Catalunya, que tractés d'una manera més àmplia, i contextualitzada a Catalunya, les qüestions que s'aborden en el marc de l'Informe EPTA 2024.

Fruit d'aquesta iniciativa, s'ha concebut el present informe, format per tres informes autònoms però relacionats en la mesura que tots tres aborden una dimensió de les relacions entre la IA i la democràcia.

En primer lloc s'explora, sota el títol *Democràcia i intel·ligència artificial*, l'impacte que té la IA, i en concret la intel·ligència artificial generativa (IAG), en la democràcia per mitjà d'una reflexió sobre els mecanismes que permeten controlar l'ús de la IA per part dels poders públics legitimats democràticament. Aquesta reflexió s'acompanya de situacions reals d'aplicació de la IA a les darreres eleccions al Parlament de Catalunya.

En segon lloc, sota el títol *La intel·ligència artificial en els parlaments i en el sector públic*, s'exposa la situació actual de l'ús de la IA en les institucions pilars de l'estat democràtic, com ho són els parlaments, i també els governs per mitjà de la prestació de serveis públics. Així mateix, s'analitzen els possibles usos de la IAG al Parlament de Catalunya prenent com a referència altres experiències internacionals, i es fa una anàlisi dels riscos i les oportunitats d'aquests usos.

Finalment, en tercer lloc, sota el títol *Intel·ligència artificial i regulació*, es fa un breu repàs a la regulació de la IA a dia d'avui i s'emmarquen normativament les temàtiques abordades al llarg de l'informe, tenint en compte aspectes comuns a altres indústries, aspectes tècnics específics de la IA o aspectes derivats de la seva ràpida i massiva adopció.

Índex

Democràcia i intel·ligència artificial

La intel·ligència artificial en els parlaments i en el sector públic

Intel·ligència artificial i regulació

Democràcia i intel·ligència artificial

No cal recordar al lector l'abast que ha tingut la irrupció de la intel·ligència artificial (IA) i en particular la intel·ligència artificial generativa (IAG) en tots els àmbits de la vida pública, social i econòmica.

La infinitud d'impactes que ha produït l'eclosió de la IA en el conjunt de la societat, sobretot de la IA generativa cap a finals de l'any 2022, és inabastable en un document de les característiques d'aquest informe; per això, en aquest punt tenim un únic objectiu a tractar, i, per tant, ens centrarem en un tema molt específic. Concretament, s'hi analitza l'impacte que pot tenir la IA en la democràcia. Aquesta qüestió s'encara tant des d'un punt de vista teòric, reflexionant sobre els mecanismes que permeten controlar l'ús de la IA per part dels poders públics legitimats democràticament, en aplicació del concepte de sobirania digital, com des d'un punt de vista més pràctic o aplicat, en relació amb l'ús de la IA a les eleccions al Parlament.

Com hem avançat, l'informe està dividit en **dues parts**: la primera tracta el tema de la sobirania digital i del control democràtic d'aquesta nova tecnologia, i es centra en particular en les mesures i les respostes que han adoptat les institucions públiques per gestionar i canalitzar el tsunami provocat per la irrupció de la IA, sobretot de la IAG, i la segona es centra en el tema de les potencialitats de l'ús de la IA a les eleccions: tant en les campanyes electorals com els seus riscos i problemàtiques d'inclusió, competència digital i qualitat democràtica.

Amb l'ús del terme **intel·ligència artificial de propòsit general**, ens referim a un tipus d'IA, a aquells sistemes que es poden adaptar a multitud d'usos, incloent-hi aquells per als quals no van ser dissenyats. Aquests sistemes poden desenvolupar múltiples funcions i no estan especialitzats en desenvolupar tasques concretes. Aquests tipus de sistemes d'IA són capaços d'aprendre, raonar, entendre contextos complexos, resoldre problemes i adaptar-se a situacions noves, de manera similar a com ho faria una persona.

La **intel·ligència artificial generativa (IAG)** es pot entendre per IA generativa aquell tipus d'intel·ligència artificial amb capacitat per crear nous continguts com ara música, vídeos, text, àudio o imatges. Els models d'aquest tipus d'intel·ligència artificial aprenen dels patrons i l'estructura de les dades d'entrenament d'entrada i després generen noves dades que tenen característiques similars. La IA generativa utilitza un model d'aprenentatge automàtic per aprendre els patrons i les relacions d'un conjunt de dades de contingut creat per persones per posteriorment utilitzar patrons que ja coneix amb els quals generar contingut. La gran diferència entre la IA generativa i els tipus més comuns d'IA –com l'analítica o la discriminativa– és que es passa de les capacitats cognitives a l'àmbit de les capacitats creatives. Amb la IA generativa, la màquina produeix informació nova en comptes de limitar-se a reconèixer, analitzar o classificar contingut existent.

L'objectiu d'aquest informe és ajudar els lectors a copsar i entendre el potencial de la IA(G) per a la millora dels sistemes democràtics, sense perdre de vista els seus riscos, així com a identificar la gran varietat d'iniciatives i l'ampli ventall d'institucions i d'entitats que s'han involucrat per observar, gestionar, qüestionar i controlar el desenvolupament i el desplegament de la IA(G) a Catalunya. Aquest ecosistema ha de fornir-nos d'eines per prendre les mesures que s'escaiguin i siguin necessàries, davant de la possibilitat que, per exemple, per exemple, la IA(G) generi un perill cert per a l'avenç democràtic i la convivència pacífica de la societat catalana.

Nota sobre la terminologia: en aquest informe s'utilitzaran indistintament els termes *intel·ligència artificial* (IA) i *intel·ligència artificial generativa* (IAG). Certament, tot i que no són el mateix, sovint, també a la literatura científica, s'utilitza el terme *IA* com el terme paraigua que engloba els dos conceptes. Tanmateix, sempre que ens és possible, en el text s'especifica quan es tracta d'IA i quan d'IAG per no generar confusió.

Sobirania digital i control democràtic

En aquest primer apartat tractem el tema de la sobirania digital. Per *sobirania digital* hem d'entendre la capacitat que té un ens o una institució per protegir i incidir en l'ús i la gestió de les dades i de la informació que es generen en el seu àmbit d'actuació fruit de l'ús de les tecnologies per part dels ciutadans i de les empreses.

Tot i així, abans d'assenyalar ni que sigui de forma molt breu com reforça Catalunya la seva sobirania digital, cal tenir en compte, encara que sigui de manera sumària, quins són els riscos i els potencials que presenta la IA per als sistemes democràtics.

En aquest sentit, en l'informe de la UNESCO de 2024 titulat «IA i democràcia», elaborat per D. Innerarity, es posava el focus en els següents punts. En primer lloc, quant al coneixement que com a societat tenim sobre aquesta nova tecnologia, es constata que encara no se sap amb exactitud, i que, fins i tot, encara és d'hora per determinar-ho, si aquestes noves eines milloraran o faran impossible la democràcia. En segon lloc, l'autor posa de manifest que la digitalització està suposant una modificació tan radical dels espais públics que ens ha d'obligar a pensar, des de noves categories, sobre com es conjumina aquest desenvolupament tecnològic amb el desenvolupament i garantia de la democràcia. En tercer lloc, l'autor emfasitza que aquestes noves eines han tingut un fort impacte en termes de cohesió social, ja que semblen afavorir aquelles actituds més individualistes i polaritzades i semblen conduir a l'aparició de comunitats virtuals homogènies i tancades, la qual cosa soscava la cohesió social. En quart lloc, també s'hi destaca un nou element de poder en aquest nou estadi de desenvolupament tecnològic: les dades i el control que es pugui exercir sobre elles. En aquest sentit, cal analitzar les noves relacions de poder que genera la capacitat d'analitzar les dades i d'explotar-les. Només algunes institucions o ens poden aprofitar tot aquest potencial i això els permet seguir acaparant poder, obrint-se una escletxa respecte d'aquells altres que no tenen els mitjans per fer-ho. Finalment, l'informe recorda que les tecnologies de la IA no són neutrals i codifiquen els valors de les persones que les creen i de l'ecosistema d'on provenen.

Un cop sistematitzats aquests perills, en l'informe s'advoca per un sistema de governança que controli els riscos de l'automatització en els processos polítics. Com a primera proposta s'apunta la necessitat d'una governança ètica que alineï la IA amb els principis democràtics per mitjà de l'aprovació de normes i regulacions vinculants. També s'hi afegeix la necessitat que es garanteixi la transparència algorítmica a través d'auditories dels sistemes d'IA que permetin garantir que aquests actuen de manera legal i amb equitat.

Certament, les institucions democràtiques presenten diverses vulnerabilitats derivades de les noves amenaces digitals que rauen en l'ús de la IA. Per això De Filippi i Blázquez, en un treball recent titulat «Democracy Reloaded» (2025), aposten per la posada en marxa de noves polítiques públiques d'inversions en infraestructures d'IA, acompanyades de marcs reguladors globals i d'iniciatives de col·laboració publicoprivada que permetin provar les tecnologies de la IA abans de posar-les en el mercat, en un entorn controlat de proves (*sandbox*), així com d'obligacions per a la indústria en matèria de transparència dels sistemes, perquè s'obrin a la cooperació amb verificadors de fets independents.

Tenint en compte el que hem explicat fins ara en aquesta primera part, fem una anàlisi molt breu de les diferents iniciatives (institucionals, normatives i d'altra índole) orientades a articular un control democràtic i legítim del desenvolupament i de l'ús de la IA(G) a Catalunya.

Les mesures que s'adopten a Catalunya per reforçar la sobirania digital nacional i el control democràtic del desenvolupament i l'ús de la IA

En primer lloc, cal tenir en compte que Catalunya va aprovar l'Estratègia d'Intel·ligència Artificial de Catalunya - Catalonia.AI l'any 2020. Aquest document està alineat amb l'Estratègia Europea, adoptada l'abril de 2018. L'estratègia catalana inclou, d'una banda, una primera revisió de les capacitats acadèmiques i industrials existents en el nostre territori. L'estratègia fou actualitzada el 2024, incorporant-hi les dades de l'informe «AI a Catalunya». L'estratègia, d'altra banda, estableix les prioritats i les línies d'actuació per promoure la recerca i la innovació, per crear i retenir talent especialitzat, així com per garantir les infraestructures necessàries i garantir l'accés a les dades, fomentant l'adopció de la IA en diversos sectors, i posant especial èmfasi en desenvolupar una IA ètica que sigui conforme a la normativa aplicable i a les nostres pràctiques socials i culturals.

Hi ha dos elements bàsics per al desenvolupament de la IAG: en primer lloc, la supercomputació, i, en segon lloc, grans conjunts de dades de qualitat.

Certament, la supercomputació és un dels pilars fonamentals per al desenvolupament de la IA, i ha tingut un paper fonamental en el desplegament de l'estratègia catalana en la matèria. El nou superordinador europeu MareNostrum 5, inaugurat el desembre de 2023 al Barcelona Supercomputing Center - National Supercomputing Center (BSC-CNS), l'actualització del qual es va aprovar el maig de 2024, té com una de les seves missions crítiques la creació de models lingüístics –en totes les llengües oficials de l'Estat– oberts i transparents, que evitin biaixos i que millorin la qualitat de les solucions d'IA que es posin a disposició del mercat o de la ciutadania.

A més, i com hem apuntat, el desenvolupament de la IA requereix també grans conjunts de dades d'alta qualitat, però aquests conjunts de dades són específics per a cada idioma. Aquesta gran quantitat de dades necessàries per formar models de llenguatge extens ha ampliat la bretxa entre els idiomes –els àmbits lingüístics– amb accés a aquestes dades i els que no hi tenen accés. Per abordar aquesta problemàtica i garantir que el català segueixi sent rellevant en les aplicacions d'IA i de processament del llenguatge natural (PNL), la Generalitat de Catalunya, en col·laboració amb el BSC-CNS i amb el suport financer del Projecte Estratègic Espanyol de Recuperació i Transformació Econòmica (PERTE) sobre la nova economia lingüística, ha posat en marxa el projecte Aina. Per consultar més informació sobre aquest projecte, vegeu el punt 1.3.

L'esforç català per donar suport a un ecosistema d'IA coordinat dins de l'estratègia Catalonia.AI està cristal·litzant en iniciatives com:

— **El Centre d'Innovació en Tecnologies de Dades i IA** (CIDAI), creat seguint el model de Digital Innovation Hubs establert per la Comissió Europea i configurat com un centre en xarxa al servei de les empreses i institucions. Està coordinat per Eurecat, Centre Tecnològic de Catalunya, i treballa en contacte constant amb, entre d'altres, la Unitat d'Intel·ligència Artificial Aplicada, que desenvolupa solucions innovadores (algorismes, mètodes, plataformes) basades en la combinació de tecnologies d'IA i de gestió del coneixement, especialment orientades al sector industrial, energètic i de la sostenibilitat. El CIDAI promou la transferència de coneixement i la realització de projectes conjunts entre entitats generadores de coneixement (universitats, centres de recerca i innovació), empreses proveïdores de tecnologia i serveis, i empreses i institucions usuàries demandants de solucions innovadores en IA aplicada. Entre els seus projectes d'alt impacte consten els següents:
  — Eines intel·ligents per al resum i generació de contingut audiovisual en notícies.
  — Eines per al sector agropecuari.
  — Eines per a la micromobilitat sostenible.
  — Eines per a la gestió de dades de transaccions bancàries.
  — Eines per a l'explotació de dades d'opinions.
  — Un model de clusterització (*clustering*) de visitants.

El **processament del llenguatge natural** (**PLN; en anglès, *natural language processing***, **NLP**) és una branca de la IA que es dedica a l'estudi i el desenvolupament de tècniques per entendre, interpretar i generar llenguatge humà de manera que sigui útil. El PLN és utilitzat, per exemple, en traducció o en la generació automàtica de contingut.

La clusterització de dades és una manera d'agrupar un conjunt de dades en grups o clústers que comparteixin característiques semblants. Mitjançant aquest mètode és possible descobrir patrons i estructures inherents a les dades, identificar relacions entre variables i segmentar en aquest cas els visitants basant-se en el seu comportament o altres atributs compartits.

— **L'Aliança de Recerca en Intel·ligència Artificial a Catalunya** (AIRA), que inclou la recent posada en marxa de la Unitat ELLIS de Barcelona, integrada a la xarxa paneuropea d'excel·lència en recerca i innovació. AIRA té com a objectiu promoure l'excel·lència en la recerca científica, la gestió del talent i l'acceleració del desenvolupament de solucions basades en la IA i també pretén esdevenir un referent al servei de l'excel·lència en la investigació científica, i d'impuls d'institucions acadèmiques i socials de l'ecosistema d'IA català, per demostrar els beneficis de la IA i accelerar la investigació científica i la generació de talent.
— **L'Aliança Catalunya Digital** (DCA-AI), que integra empreses, centres de recerca, administracions públiques i altres entitats que desenvolupen, integren, implementen o ofereixen solucions tecnològiques en l'àmbit de la IA. Entre els seus serveis hi ha la creació de més oportunitats de negoci per a les empreses de l'àmbit tecnològic, més participació i visibilitat en fires i congressos, més sinergies i accés a espais de treball amb membres de les diferents comunitats, connexió amb l'ecosistema d'innovació, accés a coneixement, tecnologia i infraestructures i també (no menys rellevant) accés a fonts de finançament.

En aquest context, l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC) estudia les conseqüències ètiques, socials i jurídiques de la implantació de la IA a Catalunya. La seva tasca també dona el suport necessari per poder complir amb la normativa aplicable a cada projecte que empri un sistema d'IA i, concretament, del Reglament d'intel·ligència artificial de la UE, que es va aprovar el 2024 (RIA). En particular, la metodologia de l'OEIAC està dissenyada per oferir un marc d'avaluació ètica a través de set principis bàsics:

1. Transparència i explicabilitat.
2. Justícia i equitat.
3. Seguretat i no-maleficència.
4. Responsabilitat i retiment de comptes.
5. Privadesa.
6. Autonomia.
7. Sostenibilitat.

A tall d'exemple, l'OEIAC ha desenvolupat una proposta d'autoavaluació organitzativa sobre l'ús ètic de dades i sistemes d'IA denominada model PIO (principis, indicadors i observables), que es basa en els principis ètics fonamentals establerts al RIA. El que és realment inte-

ressant d'aquesta iniciativa és el catàleg d'exemples que ajuden a adonar-se dels riscos legals i ètics que pot generar l'ús de la IA i a reconèixer-los.

En l'àmbit de l'Administració pública, a fi de bastir una governança adequada de l'ús de la IA en aquest sector, i per tal de desenvolupar la seva estratègia d'IA, la Generalitat de Catalunya ha creat el Comitè d'Ètica de les Dades per fomentar la reflexió ètica sobre el desplegament del model de dades i el seu ús per part de l'Administració pública catalana (Acord de Govern GOV/6/2021). Aquest comitè és un òrgan col·legiat de caràcter consultiu i transversal al servei de l'Administració de la Generalitat de Catalunya i del seu sector públic, que té els objectius següents:

— Garantir un espai de reflexió ètica en el desplegament del model de govern de les dades.
— Generar coneixement i orientacions entorn l'ús de les dades.
— Desplegar usos avançats de les dades i generar bones pràctiques i actitud per orientar el desenvolupament de l'Administració digital.

El Comitè d'Ètica de les Dades forma part de la Xarxa de Comitès d'Ètica de Catalunya.

En aquest sentit, també cal esmentar l'aprovació de l'Acord de Govern GOV/158/2023, pel qual s'aprova el model de governança de dades de l'Administració de la Generalitat de Catalunya i el seu sector públic.

Igualment, la Generalitat de Catalunya també ha aprovat la creació del seu registre d'algorismes (Acord de Govern GOV/45/2024), el model del qual es troba actualment en desenvolupament. Aquest registre forma part del model català d'Administració digital, tal com estableix la Llei 29/2010, d'ús de la tecnologia en el sector públic català.

Aquest mateix Acord de Govern (GOV/45/2024) ha creat la Comissió d'Intel·ligència Artificial. Entre les seves funcions principals s'hi troben les següents:

— Proposar al Govern les accions de promoció i supervisió de la IA en els serveis públics.
— Supervisar les metodologies d'avaluació dels drets i llibertats dels ciutadans en la prestació de serveis públics.
— Supervisar el compliment i l'adequació de l'aplicació dels sistemes d'IA amb el model de governança de dades.
— Posar en coneixement del Comitè d'Ètica de Dades les qüestions ètiques sobre l'ús de la IA a l'Administració pública; promoure mecanismes de bones pràctiques per aplicar la IA que tinguin en compte principis i valors ètics, i finalment.
— Proposar mesures de vigilància o supervisió humana, per prevenir o reduir al mínim els riscos derivats de l'aplicació de la IA, especialment en relació amb els riscos derivats de la protecció de dades personals.

A més, la societat civil –especialment activa a Catalunya– està molt implicada en la garantia d'espais de participació cívica en la regulació i governança de la IA. A tall d'exemple, diverses entitats catalanes participen en «AI Ciudadana», una coalició de disset organitzacions que treballen per la defensa dels drets humans en el context de les tecnologies digitals: Civio, Political Watch, AlgoRace, Observatorio de Trabajo, Algoritmo y Sociedad, Institut de Drets Humans de Catalunya, La Fede CAT, etc.

Aquestes organitzacions advoquen perquè es prenguin mesures per garantir que la implementació de la nova legislació europea sobre IA (el RIA) respecti els drets humans mitjançant quatre accions clau:

1. Prohibir el reconeixement biomètric i d'emocions a l'espai públic.
2. Crear un registre més complet d'algorismes amb informes públics obligatoris.
3. Garantir que els sistemes d'IA estiguin dissenyats per prevenir la discriminació.
4. Implicar la societat civil en la governança i les avaluacions d'impacte de la IA, així com en la configuració de polítiques públiques relacionades amb la IA.

Ja en l'àmbit municipal, cal destacar que l'Ajuntament de Barcelona ha aprovat la seva «Definició de metodologies de treball i protocols per a la implementació de sistemes algorítmics», també conegut com a «Protocol de Barcelona», que té com a objectiu la creació d'un protocol intern per a la implantació de sistemes algorítmics que fou pioner. Aquest protocol es vol aplicable a qualsevol sistema algorítmic impulsat per l'Ajuntament, i es pretén garantir que aquests sistemes s'utilitzen de manera proporcionada, supervisada i d'acord amb les normes legals, ètiques i tècniques que resultin d'aplicació.

Aquest document defineix els mecanismes de salvaguarda dels drets associats a cada etapa de la licitació i implantació d'un sistema algorítmic per part de l'Ajuntament, i estableix els òrgans de govern i supervisió que vetllaran perquè l'ús de la IA s'ajusti als principis ètics i a les normes legals vigents.

Igualment, a través de la xarxa Eurocities, les ciutats de Barcelona, Rotterdam (Països Baixos), Eindhoven (Països Baixos), Mannheim (Alemanya), Bolonya (Itàlia), Brussel·les (Bèlgica) i Sofia (Bulgària) han treballat conjuntament per desenvolupar un model comú de registre d'algorismes que garanteix l'ús adequat de les dades.

Els canvis i les actualitzacions de les polítiques nacionals d'IA després de la interrupció de la IAG

Fa anys que la Unió Europea treballa per establir una regulació pròpia de la IA que posi al centre de la tecnologia les persones i la garantia dels seus drets. Amb aquesta finalitat, el 25 d'abril de 2018 la Comissió Europea va establir una estratègia sobre IA que abordava els aspectes socioeconòmics, juntament amb un augment de la inversió en recerca, innovació i capacitat en matèria d'IA a tota la UE.

De la mateixa manera, es va aprovar un pla coordinat amb els estats membres per harmonitzar estratègies. Posteriorment, el 2019 la Comissió va nomenar un grup d'experts d'alt nivell que, l'abril de 2019, va publicar unes directrius per a una IA fiable. Posteriorment, la Comissió va aprovar i va publicar una comunicació mitjançant la qual s'adoptaven els set principis especificats en les directrius d'aquest grup d'experts. Aquests principis, que han d'informar de qualsevol normativa en matèria d'IA de la Unió, són els següents: acció i supervisió humanes; solidesa tècnica i seguretat; gestió de la privacitat i de les dades; transparència; diversitat, no-discriminació i equitat; benestar social i mediambiental, i rendició de comptes.

La UE ha considerat necessari elaborar una normativa comuna sobre la IA, bàsicament per dos motius: d'una banda, perquè la divergència que es podria generar si es permetessin normatives i aproximacions diferents en els estats membres posaria en risc el mercat únic; de l'altra, perquè es vol que aquesta regulació reflecteixi valors compartits en l'àmbit europeu, com la dignitat humana, la protecció de la privacitat i la sostenibilitat ambiental. Una de les característiques destacades de la regulació europea de la IA és la seva orientació antropocèntrica i la reivindicació de la supervisió i el control humà dels sistemes d'IA.

Tots aquests esforços previs van arribar a bon port quan el Parlament Europeu va aprovar el RIA de la UE, una norma històrica, impulsada durant la presidència espanyola del Consell de la Unió Europea, el segon semestre del 2023 i amb l'ampli suport de l'Eurocambra. La norma vol situar Europa com a líder en innovació, alhora que es pretén garantir la protecció dels drets de la ciutadania.

A més, el RIA té en compte, com a element clau de la regulació, els riscos sistèmics que poden generar els models d'IA de propòsit general, incloent-hi els grans models d'IAG, com el ChatGPT. Com s'ha mencionat, els models de propòsit general poden utilitzar-se per a una varietat de tasques i s'estan convertint en la base de molts sistemes d'IA.

El RIA defineix *risc sistèmic* com «un risc específic de les capacitats de gran impacte dels models d'IA d'ús general, que tenen unes repercussions considerables al mercat de la Unió a causa del seu abast o dels efectes negatius reals o raonablement previsibles en la salut pública, la seguretat, la seguretat pública, els drets fonamentals o la societat en el seu conjunt, que es pot propagar a gran escala al llarg de tota la cadena de valor».[1]

La Llei d'IA estableix els deures per als proveïdors de la IAG, però n'imposa encara més als proveïdors d'aquelles IAG que poden presentar els riscos sistèmics.

Un cop aprovat el RIA, a l'Estat espanyol, el Consell de Ministres va aprovar el 15 de maig l'Estratègia d'IA 2024, que dona continuïtat i reforça l'Estratègia d'IA (ENIA) pu-

1 Val a dir que no tots els sistemes d'IA d'ús general presenten un risc sistèmic. Només el tenen aquells sistemes que compleixen un dels requisits establerts a l'article 51 del RIA:
«(a) té capacitats de gran impacte avaluades sobre la base d'eines i metodologies tècniques adequades, [...]», o
«(b) sobre la base d'una decisió de la Comissió, adoptada d'ofici o arran d'una alerta qualificada del grup d'experts científics, té capacitats o un impacte equivalent als establerts a la lletra a), tenint en compte els criteris establerts a l'annex XIII».
Aquest annex XIII descriu els «Criteris per a la classificació dels models d'IA d'ús general amb risc sistèmic a què fa referència l'article 51», com per exemple, el nombre de paràmetres del model; la qualitat o la mida del conjunt de dades, per exemple mesurats mitjançant *tokens*; la quantitat de càlcul utilitzada per entrenar el model, mesurada en operacions de coma flotant; les modalitats d'entrada i sortida del model; els paràmetres de referència i les avaluacions de les capacitats del model; si les repercussions per al mercat interior són importants a causa del seu abast, cosa que es dona per descomptada quan s'hagi posat a disposició d'almenys 10.000 usuaris professionals registrats establerts a la Unió; el nombre d'usuaris finals registrats.

blicada el 2020, en compliment del Pla de Recuperació, Transformació i Resiliència (PRTR), i l'adapta als nous avenços tecnològics. L'impuls de la IA també està recollit a l'Agenda Espanya Digital 2026 com a element clau de caràcter transversal per transformar el model productiu i impulsar el creixement de l'economia espanyola. L'estratègia es desenvoluparà durant els anys 2024 i 2025 i està dotada amb 1.500 milions d'euros procedents fonamentalment del Pla de Recuperació, Transformació i Resiliència i de la seva addenda, que se sumen als 600 milions ja mobilitzats. És un pla dissenyat per desenvolupar i expandir l'ús de la IA de manera transparent i ètica, i està enfocat a facilitar-ne l'ús als sectors públic i privat.

Entre les iniciatives més destacades darrerament, es troba la publicació de l'informe «A Fourth Wave of Open Data? Exploring the Spectrum of Scenaris for Open Data and Generative AI» (*Una quarta onada de dades obertes? S'analitza la relació emergent entre les dades obertes i la IA generativa*), presentant diversos escenaris i recomanacions. Per exemple, aquest informe suggereix que la intersecció de dades obertes, específicament de dades obertes del govern o d'institucions de recerca, i la IA generativa pot millorar la qualitat de la producció d'IA generativa, i també pot ajudar a ampliar els casos d'ús d'IA generativa i democratitzar l'accés a les dades obertes. És més, podria ajudar a millorar la qualitat i la confiança en els resultats de la IA generativa. Els autors també subratllen que entendre com es podrien creuar exactament les dades obertes i la IA generativa i els requisits per fer que les dades obertes estiguin «a punt» per a aquestes interseccions segueixen sent àrees poc explorades, així com moltíssimes altres possibilitats per utilitzar les dades obertes.

Els autors elaboren cinc escenaris on les dades obertes i la IA generativa es podrien creuar i recomanen millorar la transparència i la documentació de les dades obertes, centrar-se en la qualitat i la integritat de les dades, promoure la interoperabilitat i els estàndards, millorar l'accessibilitat de les dades obertes i tenir en compte les consideracions ètiques.

No cal dir que la Generalitat participa activament en iniciatives de dades obertes en general i sobre dades obertes i IA en particular: per exemple, ja el 2020, es va publicar l'informe «Les dades obertes i la intel·ligència artificial, eines per a la igualtat de gènere».

El **model de llenguatge extens (MLE; en anglès, *large language model*, LLM)** és un tipus de sistema d'IA generatiu, que pot crear contingut nou en resposta a les ordres dels usuaris (*prompts*) basant-se en les seves dades d'entrenament. Estan entrenats amb grans quantitats de dades de diverses fonts. Són exemples de MLE els GPT (en anglès, General Pretrained Transformers), que constitueixen les bases de ChatGPT.
Atenció: la diferència entre GPT i ChatGPT és que ChatGPT és l'aplicació (amb interfície d'usuari) mentre que GPT és el «cervell» d'aquesta aplicació, que no ens és accessible sense la interfície d'usuari on podem col·locar els *prompts*.

ALIA, una infraestructura pública d'IA en castellà, català i les altres llengües cooficials, oberta i finançada al 100% amb recursos públics. És un projecte que està coordinat pel Barcelona Supercomputing Center - Centre Nacional de Supercomputació (BSC-CNS), i «es tracta de la primera infraestructura pública europea que, gràcies a les capacitats úniques de supercomputació del Barcelona Supercomputing Center, reforça la sobirania tecnològica d'Espanya i Europa en el desenvolupament d'una IA transparent, responsable i al servei de les persones».

Els plans en curs per a desenvolupar models de llenguatge extens (MLE) a Catalunya: l'àmbit (públic, privat o acadèmic) i els principals debats públics amb relació al desenvolupament de MLE nacionals

El desenvolupament de les tecnologies de la parla és vital per a les aplicacions d'IA, com els assistents virtuals i les eines de subtitulació multilingüe, que afavoreixen una comunicació efectiva entre llengües diverses. Mentre que les llengües tecnològicament avançades se'n beneficien significativament, les llengües amb recursos limitats, com el català, afronten desafiaments per obtenir els conjunts de dades necessaris per a models de reconeixement i síntesi de parla d'alta qualitat. Com a conseqüència, la vida digital en català ha hagut tradicionalment de passar a l'anglès o al castellà. Aquests desafiaments inclouen qüestions legals, tècniques i relacionades amb la diversitat.

Aina (https://projecteaina.cat/) és un projecte promogut pel Departament de Polítiques Digitals i Administració Pública, en col·laboració amb el Barcelona Supercomputing Center, a finals del 2020, amb l'objectiu de preservar i potenciar la competitivitat del català en l'àmbit digital mitjançant l'ús de la intel·ligència artificial i contribuir a la normalització lingüística del català en el món digital. El superordinador MareNostrum 5 habilita el processament de dades massives i l'execució de models lingüístics avançats i pioners en el sector.

El projecte té com a objectiu promoure i reforçar la presència de la llengua catalana en l'actual paradigma tecnològic, informàtic i audiovisual, així com crear i posar a disposició de la comunitat científica i la indústria els recursos necessaris perquè qualsevol persona pugui utilitzar el català en l'àmbit digital, i que permetin desenvolupar aplicacions com assistents de veu, traductors automàtics o agents de conversa en català. Aquest objectiu requereix la recopilació i processament de grans volums de dades de qualitat, així com dades anotades per entrenar i avaluar models per a tasques específiques. En aquest sentit, cal destacar dues característiques del projecte. En primer lloc, la seva orientació comunitària i col·laborativa. En el desenvolupament de recursos lingüístics, Aina utilitza dades textuals i de veu obtingudes gràcies a la col·laboració dels usuaris, entitats de la comunitat lingüística i d'altres sectors, i la iniciativa Aina Lab promou la creació d'una comunitat activa de desenvolupadors i desenvolupadores per impulsar la innovació tecnològica en català, utilitzant com a vehicle principal la plataforma Aina Challenge. D'altra banda, els models i conjunts de dades desenvolupats dins del projecte Aina formen l'Aina Kit, disponible per a totes les entitats, tant públiques com privades, que desitgin utilitzar-los, de manera oberta i amb llicències permissives.

La IA a les eleccions

El creixent impacte de la intel·ligència artificial (IA) en l'àmbit electoral és un fenomen que està transformant les estratègies de campanya i els processos democràtics en l'àmbit global. Per abordar aquesta qüestió, la secció d'aquest informe dedicada a l'ús de la IA a les eleccions compta amb tres apartats interconnectats que permeten entendre de quina manera aquesta tecnologia està influint en diferents dimensions del sistema electoral.

Per fer-ho, s'ha realitzat recerca documental, és a dir, s'ha consultat i citat articles de premsa, informes governamentals, documents legislatius i literatura científica que, a més, s'ha completat amb la realització d'una sèrie d'entrevistes i qüestionaris, algunes dutes a terme per via personal i d'altres via correu electrònic.[2]

En primer lloc, s'aborda l'ús de la IA en les campanyes electorals amb exemples concrets de Catalunya i d'altres països, en què es mostra com els partits polítics han començat a implementar la IA com un recurs per millorar la seva comunicació amb l'electorat.

En el segon apartat, s'exploren les mesures que s'estan adoptant per mitigar els riscos que aquesta tecnologia pot suposar per a la seguretat i la transparència del sistema democràtic.

2 Concretament, s'ha entrevistat partits polítics que han accedit a participar en l'estudi i que són: Candidatura d'Unitat Popular, En Comú Podem, Junts per Catalunya, Esquerra Republicana de Catalunya i el Partit Socialista de Catalunya. També s'ha entrevistat el Consell de l'Audiovisual de Catalunya, els Mossos d'Esquadra i l'Agència de Ciberseguretat de Catalunya.

Finalment, es posa l'accent en la importància de la competència digital de la ciutadania i en les iniciatives que promouen una millor alfabetització mediàtica per contrarestar els efectes negatius de la desinformació que se'n pugui derivar.

L'ús de la IA en les campanyes electorals catalanes i altres referències internacionals

La IA ha esdevingut un recurs amb un recorregut incert, especialment en matèria política i legal. L'emergència de legislació vigent en la regulació del seu ús és un primer pas determinant en l'establiment de protocols d'acció i resposta davant d'un fenomen de caràcter global. En aquest sentit, resulta especialment important l'ús responsable de la IA per part dels representants polítics, no només com a garants de la normativa vigent, sinó també com a impulsors conscients de la naturalesa de les seves implicacions. Per aquest motiu, en aquest apartat s'aborden els usos de la IA en les campanyes electorals catalanes, però també es recullen pràctiques realitzades en l'àmbit europeu i internacional.

L'ús actual de la IA en les campanyes electorals a Catalunya és limitat. La majoria de partits polítics catalans amb representació parlamentària admeten que han fet ús d'algunes eines d'IA, però de manera exploratòria.

Per exemple, hi ha casos paradigmàtics de l'ús de l'anomenada IA generativa per part d'alguns partits polítics que han implementat aquesta tecnologia per crear materials gràfics i audiovisuals de campanya que els permetin imaginar o visualitzar escenaris polítics hipotètics. Un dels primers usos que trobem és a les eleccions municipals del maig de 2023 quan En Comú Podem va presentar un espot amb imatges sintètiques amb la doble intenció de mostrar visions de les ciutats del futur i, alhora, posar de manifest la necessitat de regular la IA i de reivindicar el poder de decisió dels humans en aquests àmbits.

Figura 1. Part de l'espot electoral d'En Comú Podem generat amb imatges sintètiques.
Font: Compte de YouTube d'En Comú Podem.

A la precampanya de les eleccions espanyoles del juliol de 2023, Junts per Catalunya també va utilitzar la IA per produir un vídeo fals que recreava la imatge del president del Govern espanyol, Pedro Sánchez, el qual apareixia demanant disculpes per les seves polítiques, i que Junts per Catalunya va utilitzar per denunciar els incompliments i la manca d'autocrítica que, al seu parer, Sánchez mostra amb Catalunya.

Durant les eleccions al Parlament de Catalunya del maig de 2024, Ciutadans va crear el primer cartell de campanya electoral generat amb IA de la història de la política catalana i espanyola. El partit va renunciar a col·locar les imatges dels seus candidats en cartells electorals per recrear una imatge d'unitat entre el president del Govern espanyol, Pedro Sánchez, i el líder de Junts per Catalunya, Carles Puigdemont, amb la voluntat d'il·lustrar les afinitats polítiques que Ciutadans percep que tenen aquests dos polítics.

A banda d'aquests casos de creació de materials gràfics i audiovisuals, alguns partits polítics com el Partit dels Socialistes de Catalunya (PSC) i Junts per Catalunya (Junts) també han aplicat la IA en la distribució de continguts polítics mitjançant l'ús de bots de conversa. En campanyes electorals recents, el PSC ha utilitzat un canal de WhatsApp automatitzat per respondre les consultes de votants, que havia estat entrenat i supervisat amb informació proporcionada per persones del partit. Per la seva banda, Junts va posar en marxa un bot de conversa anomenat MireIA durant la campanya de les eleccions espanyoles del juliol de 2023 per donar a conèixer el seu programa electoral.

Partits com Candidatura d'Unitat Popular (CUP) també han aplicat eines d'intel·ligència artificial per subtitular automàticament els vídeos que difonen entre els votants per donar a conèixer el seu ideari polític. En aquest sentit, formacions com Esquerra Republicana de Catalunya (ERC) també admeten que estan considerant l'ús de la IA per a la subtitulació de materials audiovisuals, així com per agilitzar algunes tasques internes que poden ser mecàniques com, per exemple, tasques de traducció i transcripció de continguts.

Cal apuntar, que cap dels exemples anteriors va marcar les campanyes electorals i no va provocar cap controvèrsia política significativa. També és destacable el fet que cap partit va informar d'aquestes pràctiques la Junta Electoral. No obstant això, aquestes pràctiques han estat els primers passos en l'ús de la IA en campanyes polítiques a Catalunya i han plantejat preocupacions sobre la necessitat potencial de regular l'ús de la IA generativa per part dels partits polítics com a mesura per combatre la desinformació.

Figura 2. Cartell de Junts per Catalunya per promoure l'ús del bot de conversa durant la campanya. Font: Compte d'X de Junts per Catalunya.

En aquesta línia, la majoria dels partits polítics catalans expressen certes reserves sobre la implementació de la IA, especialment la de tipus generatiu, per les implicacions que pot tenir en la producció de materials sensibles que poden infringir qüestions com la propietat intel·lectual. Per aquest motiu, els partits polítics que han participat en les consultes realitzades en el marc d'aquest informe afirmen que encara no disposen d'un protocol per a la implementació d'eines d'IA, ja que consideren que el seu ús és de caràcter exploratori i ocasional. També expliciten que, en cas d'haver d'establir protocols, comptarien amb el suport dels serveis jurídics dels seus partits respectius i tindrien en compte la regulació aprovada a Europa en relació amb l'ús de la IA, l'anomenada AI Act.

A banda de l'aplicació de la IA per elaborar materials de campanya electoral, també cal tenir en compte que a Catalunya l'ús d'eines d'intel·ligència artificial també s'ha aplicat per assegurar més precisió en l'escrutini. La mateixa Generalitat de Catalunya va aplicar aquesta tipolo-

gia d'eines a les eleccions catalanes del maig de 2024 amb l'objectiu de millorar el recompte de vots.

Si fem un repàs internacional, detectem un ús de la IA desdibuixat i poc consolidat. A les darreres eleccions nord-americanes, els mateixos demòcrates van admetre haver fet ús de la IA en la captació de votants i en la identificació de contingut enganyós que pogués suposar una amenaça davant de la influència de Donald Trump a les xarxes socials. Paradoxalment, advertits dels perills de la IA generativa en la difusió de desinformació per part d'experts interns del partit, van fer-ne ús per a combatre-la, conscients del potencial transformador d'aquesta eina per a unes eleccions de transcendència mundial (Subramanian, 2024).

No obstant això, l'ús de la IA en campanyes electorals nord-americanes no és una novetat. Ja va ser durant les eleccions del novembre de 2012 que els estrategues de la campanya de Barack Obama van posar sobre la taula el potencial de la IA com un estalvi de temps i de recursos valuós, però potencialment problemàtic des d'un punt de vista ètic i legal (Subramanian, 2024).

Altres països, com, per exemple, l'Índia, han incorporat la IA per atendre la diversitat lingüística. En aquest sentit, les eleccions generals finalitzades al juny de 2024 han estat especialment rellevants en relació amb l'impacte de la IA. D'una banda, mentre que el candidat Modi posava en valor els usos ètics i democràtics de la IA utilitzats durant el seu mandat, com, per exemple, la traducció simultània i el doblatge de discursos polítics en temps real per a superar barreres idiomàtiques amb bona part de la ciutadania, la mateixa campanya electoral es veia especialment afectada pels usos problemàtics de la IA generativa i els hipertrucatges (*deepfakes*). Un dels exemples més mediàtics va ser la difusió de vídeos, creats a partir d'IA generativa, on apareixien les estrelles de Bollywood Ranveer Singh i Aamir Khan, amb milions de seguidors arreu del país, criticant obertament el primer ministre i difonent el seu suport ferm al principal partit de l'oposició al congrés (Landrin, 2024).

En la mateixa línia, durant la campanya de les eleccions generals del juliol de 2024 al Regne Unit s'han realitzat diverses estratègies per arribar a l'electorat, com, per exemple, la creació d'un avatar, en aquest cas, del candidat Steve Endacott, amb la intenció de mantenir una comunicació fluïda i personalitzada amb els seus potencials votants a través de la resolució de dubtes i el manteniment de converses (Grierson, 2024). Fins i tot, a països com Bielorússia l'ús d'avatars ha servit per burlar la censura política. La líder de l'oposició, Sviatlana Tsikhanóuskaia, va fer campanya des de l'exili a les eleccions del febrer de 2024 amb un candidat sintètic anomenat Yas Gaspadar aprofitant que la llei vigent impedeix fer campanya als líders opositors, però no a candidats creats per intel·ligència artificial (Stein, 2024).

Segons un informe independent realitzat per AI Forensics, una organització europea sense ànim de lucre, a les eleccions legislatives franceses del 2024 es van detectar fins a 51 imatges creades amb IA generativa i utilitzades majoritàriament per Agrupació Nacional (RN), Reconquête i Les Patriotes.

Aquesta font també apunta que els partits d'extrema dreta estarien utilitzant aquest tipus d'imatges per donar suport a missatges polítics directament vinculats a la primera volta de la campanya legislativa francesa, però també a les passades eleccions europees. Malgrat que es desconeix l'impacte real d'aquest tipus de pràctiques, adverteixen de les conseqüències vinculades a la desinformació i de les possibles negligències que suposa la preservació fervent de determinades narratives de marca (*storytellings*) elaborades amb el suport de continguts falsos (Schueler, 2024).

De fet, hi ha casos precedents que alerten del perill de no regular l'ús de la IA en períodes electorals. A Eslovàquia, per exemple, a les eleccions del 2023 van ser testimonis de com la IA es pot utilitzar per desacreditar un candidat. Dos dies abans dels comicis, es va publicar a Facebook un àudio en què se sentia com suposadament el líder del partit liberal Eslovàquia Progressista, Michal Simecka, conversava amb la periodista Monika Tódová, del diari *Denník N*, sobre com falsejar el procés electoral amb la compra de vots de minoria romaní o gitana. Els dos implicats van denunciar que la gravació d'àudio era falsa i agències com AFP també van afirmar que el material presentava signes d'haver estat manipulat mitjançant l'ús de la IA. La gravació, però, es va publicar 48 hores abans de l'obertura dels col·legis electorals, en una franja de temps en què se suposa que tant els mitjans de comunicació com els polítics s'han de mantenir en silenci per garantir l'anomenada jornada de reflexió, fet que també va complicar-ne el desmentiment.

El cas d'Eslovàquia és, per tant, un exemple de com la intel·ligència artificial pot sofisticar la difusió i propagació de desinformació, sobretot tenint en compte que la IA també pot permetre altres fenòmens, com el de les trucades robotitzades, és a dir, que una desinformació no només es difongui a través de canals de missatgeria instantània com WhatsApp o Telegram, sinó que també circuli a través de trucades telefòniques massives fetes de manera automàtica. De fet, en països com Estats Units és un fenomen que ja s'ha detectat, com per exemple quan una trucada robotitzada amb la veu de Joe Biden va apel·lar els demòcrates de l'estat de New Hampshire a no votar a les primàries (Stein, 2024).

A mode de conclusió, a continuació es presenta una taula que sintetitza les idees principals d'aquest apartat.

Idees principals de l'ús de la IA en campanyes electorals

Els partits polítics catalans han començat a utilitzar la IA generativa per crear materials gràfics i audiovisuals que reforcin els seus missatges de campanya. Aquest fet, entre d'altres, ha obert el debat sobre el seu impacte real en els processos democràtics, com són les eleccions.

L'automatització de la interacció amb els votants mitjançant bots de conversa ha permès flexibilitzar l'accessibilitat i la comunicació directa amb l'electorat.

En el context internacional, la IA s'ha emprat tant per captar votants com per combatre la desinformació, amb resultats variables segons els casos.

Malgrat el seu potencial, la manca de regulació clara i els dilemes ètics associats a la IA generativa continuen sent reptes importants per a la seva implementació.

Mesures per mitigar els riscos potencials de la IA a les eleccions i a les campanyes electorals

Les mesures per mitigar els riscos potencials de la intel·ligència artificial en períodes electorals a Catalunya actualment es centralitzen i es focalitzen en la tasca que desenvolupa l'Agència de Ciberseguretat de Catalunya, creada l'any 2016 amb l'objectiu de vetllar i garantir la seguretat de les xarxes de comunicacions electròniques i els sistemes d'informació de l'àmbit de la Generalitat de Catalunya i el seu sector públic, en col·laboració amb els organismes judicials i policials.

L'Agència, formada per una seixantena de professionals en plantilla, compta des del 2017 amb un protocol de seguretat que apliquen durant els períodes d'eleccions. La seva tasca consisteix sobretot en la supervisió i el monitoratge de webs implicades en el procés electoral, que comença abans de la convocatòria electoral i s'intensifica durant les eleccions.

Abans de la convocatòria electoral, des de l'ens revisen tots els dominis i sistemes implicats per fer-ne una fotografia inicial i, a partir d'aquí, detectar les amenaces que hi pot haver en diferents àmbits, com, per exemple, l'existència de dominis fraudulents i d'altres ciberatacs. D'una banda, s'examinen els webs implicats directament en el procés electoral, des de la convocatòria de les eleccions fins a la finalització dels comicis. També s'analitzen els sistemes que formen part dels anomenats Serveis Crítics i que inclouen Atenció Ciutadana, Emergències, Interior i Educació, per garantir que els sistemes dels col·legis electorals funcionin correctament. També es vetlla pel funcionament dels Processos de Participació, que són tots els serveis de participació que han de funcionar durant els comicis. A més, es supervisen les anomenades Infraestructures Transversals, que serien els sistemes de correu i comunicacions de la Generalitat; la Imatge Institucional de l'Administració –es monitoritzen les xarxes de tots els departaments i comptes associats a la Generalitat–; i també els Serveis Informatius –es vetlla perquè no hi hagi cap atac informàtic als mitjans de comunicació públics de Catalunya ni tampoc a les pàgines web d'informació que s'hagin activat durant el procés electoral amb l'objectiu de garantir un accés lliure a la informació.

Durant el procés electoral, que comprèn el període transcorregut entre la convocatòria electoral i la jornada de votació, tota aquesta tasca de supervisió i monitoratge es centralitza des d'un equip de coordinació que depèn del Centre d'Operacions de Ciberseguretat (SOC). L'equip està format per vuit persones que exerceixen diferents responsabilitats dins l'Agència i també per una vintena de col·laboradors que té l'ens.

També cal destacar que, durant el procés electoral, el Centre de Telecomunicacions i Tecnologies de la Infor-

mació (CTTI) i l'Agència es reuneixen diàriament, i també mantenen trobades periòdiques amb l'empresa proveïdora a qui s'adjudica la provisió de sistemes tecnològics que s'utilitzaran durant el procés electoral. En algunes d'aquestes reunions també hi participen Mossos d'Esquadra, especialment perquè també es fa un seguiment de xarxes socials de grups criminals i de l'anomenat web fosc (*dark web*) amb l'objectiu de detectar amenaces potencials. Fins i tot, durant els dies previs als comicis es fan diferents simulacres d'intents de ciberatacs per comprovar que tots els tallafocs funcionen.

Un cop finalitzat el procés electoral, l'Agència fa un informe de tancament a mode de resum per recopilar quins actors han participat en el procés de supervisió, els actius que s'han monitoritzat, les amenaces detectades i els simulacres realitzats. L'informe també s'acompanya de gràfics amb les amenaces i els incidents identificats amb la voluntat que les lliçons apreses puguin ser aplicables en el futur.

Segons l'Agència de Ciberseguretat de Catalunya, en les últimes eleccions autonòmiques del maig de 2024 no es va materialitzar cap amenaça, tot i que hi va haver intents d'atacs de denegació de servei des de col·lectius prorussos. Tot i així, no van aconseguir anar més enllà. De fet, des de l'ens realitzen informes mensuals en què alerten de possibles perills i amenaces, així com estudis anuals prospectius.

En matèria de ciberseguretat, Estats Units és un dels referents internacionals més importants del món. L'Administració Biden-Harris va implementar diverses mesures clau en el marc d'una ordre executiva amb la intenció de mitigar els riscos de seguretat associats a la IA. Entre aquestes accions s'inclouen la publicació de directrius de seguretat i el reforçament de capacitats a les agències federals per provar i avaluar sistemes d'IA amb l'objectiu d'abordar els riscos de manera efectiva. L'Institut de Seguretat en IA (AISI, per les seves sigles en anglès) ha desenvolupat guies tècniques per ajudar els desenvolupadors a gestionar potencials usos indeguts dels models de doble ús, prioritzant la prevenció de danys a les persones o a la seguretat pública (White House Government, 2024).

Així mateix, l'Institut Nacional d'Estàndards i Tecnologia (NIST) ha presentat marcs definitius per gestionar riscos associats a la IA generativa i garantir-ne un desenvolupament segur. Per la seva banda, el Departament d'Energia (DOE) ha creat bancs de proves i eines d'avaluació de models per analitzar la seguretat de la IA, especialment pel que fa a infraestructures de seguretat nacional. Addicionalment, l'Administració ha emès una crida a combatre l'abús sexual basat en imatges, incloses aquelles generades de manera sintètica mitjançant IA (White House Government, 2024).

Actualment, Estats Units ja disposa de mesures que reflecteixen un enfocament integral per promoure el desenvolupament i l'ús responsable de la IA, amb un abordatge proactiu dels riscos potencials per a la (ciber)seguretat de la ciutadania i l'estat (White House Government, 2024).

Per tal de concloure aquest apartat, es presenta la taula següent amb els punts més rellevants:

Idees principals de les mesures per mitigar riscos de la IA

Catalunya disposa d'un protocol de ciberseguretat que inclou la supervisió de dominis electorals i la prevenció de ciberatacs durant tot el procés electoral.

L'Agència de Ciberseguretat coordina esforços amb altres institucions i empreses tecnològiques per reforçar les defenses i simular possibles atacs.

A Estats Units, l'Administració Biden-Harris va implementar directrius específiques per avaluar i gestionar riscos derivats de la IA, especialment en l'àmbit electoral.

Les mesures enfocades a mitigar els riscos de la IA busquen protegir el procés electoral i, paral·lelament, establir marcs de col·laboració internacional per afrontar desafiaments comuns.

Competència digital, inclusió i qualitat democràtica

Més enllà dels usos de la IA generativa en les campanyes electorals per part dels partits polítics, resulta interessant analitzar quines mesures s'adopten per reforçar la competència digital crítica i la resiliència de la població davant d'aquest fenomen, així com les estratègies per combatre la desinformació i mitigar l'impacte de la IA en la cultura democràtica en general. En aquest sentit, la lluita contra la desinformació i el reforçament de la integritat de la informació en l'esfera pública digital són dos pilars fonamentals.

Precisament, els usos de la IA en campanya i l'impacte de la desinformació, potencialment agreujada per la creació de continguts falsos, tal com hem vist a la secció anterior, es deu principalment a dos factors: a l'absència de legislació (i/o a l'actualització de la vigent) i a la manca d'alfabetització mediàtica de la ciutadania. Ambdues qüestions són complementàries, principalment perquè la ràpida i quasi imprevisible evolució de la IA dificulta la vigència d'una legislació actualitzada. En aquest sentit, i per mitigar els possibles efectes negatius, cal apostar per millorar l'alfabetització mediàtica, en mitjans, però també en informació (AMI). Tal com anticipava Marisol López Vicente, directora general d'Innovació i Cultura Digital de la Generalitat de Catalunya, en una entrevista ja a finals de l'any 2023, «la disrupció que ha provocat [la IA] és tan gran que genera cert vertigen, però indubtablement també ens ajuda molt. [...] En el pla cultural, allò creatiu i allò humà s'haurà de posar molt més en valor i, evidentment, hi haurà d'haver una regulació» (Administración Pública Digital, 2023).

El concepte tradicional d'alfabetització mediàtica ha donat pas a l'AMI, que actualitza i amplia la lectura crítica i la producció de mitjans per incloure els nous aprenentatges i reptes de les tecnologies digitals. Així doncs, l'AMI és una competència augmentada (Kačinová i Sádaba Chalezquer, 2022) que, segons la UNESCO (2018), s'entén com un compendi de coneixements i habilitats necessaris per a la vida laboral i quotidiana, i que permet a les persones cercar, avaluar i utilitzar de manera crítica la informació i els continguts, tant difosos en els mitjans tradicionals com en les plataformes socials.

Tal com apunta Graves (2016), l'AMI és una eina fonamental per dotar d'autonomia la ciutadania en matèria informativa i apoderar-la amb la intenció de llegir i consumir continguts críticament i detectar falsedats a les xarxes socials o en continguts generats amb IA.

En l'àmbit europeu, Espanya destaca com un dels països menys competitius en matèria d'AMI. Malgrat les iniciatives realitzades a través d'institucions governamentals, com l'Estratègia nacional per a la inclusió sociodigital de la Generalitat de Catalunya, el pes majoritari recau en l'àmbit privat, especialment en plataformes de verificació (*fact-checking*) (Cucarella i Fuster, 2022). La influència europea en iniciatives com l'Estratègia nacional impulsada des de la Generalitat es basa en aproximacions d'impacte i referència com les proposades per Mariana Mazzucato des de l'Institut per a la Innovació i Propòsit Públic, afiliat a la University College London (Regne Unit). Precisament, aquest tipus d'iniciatives busquen determinar un full de ruta davant l'impacte de la digitalització i la bretxa digital, és a dir, procurar una societat digital inclusiva, sostenible i democràtica que pugui beneficiar-se equitativament dels usos de la tecnologia i els processos de digitalització. Per fer-ho, aquesta estratègia es proposa impulsar projectes que treballin per erradicar la bretxa digital i que, alhora, impliquin la col·laboració del sistema d'innovació de la quàdruple hèlix (administracions, centres de recerca, societat civil i sector privat) i fomentin un model de governança sorgit del treball metodològic d'innovació orientada a missions. L'execució òptima d'aquest tipus d'acords de govern, en el cas de Catalunya, va acompanyada de la creació del Comitè de Direcció de l'Estratègia, que actua com a òrgan interdepartamental de lideratge, debat i proposta que, en el cas que ens ocupa, afavoreixi la consecució de la inclusió sociodigital a Catalunya.

Per la seva banda, altres països europeus on l'alfabetització mediàtica gaudeix de major suport estatal, com, per exemple, Portugal i França, les accions enfocades a millorar l'AMI es duen a terme principalment per entitats públiques i associacions periodístiques (EDMO, 2024), arriben a un públic més ampli i, en certa manera, afavoreixen un accés universal a aquest tipus de coneixements.

Tanmateix, més enllà d'iniciatives específiques, les plataformes de verificació que promouen l'AMI de manera més regular, especialment a Espanya, són Newtral, Maldita.es i Verificat. Totes elles tenen un equip dedicat a la promoció de l'alfabetització mediàtica mitjançant cursos de formació en escoles primàries i secundàries, i en l'àmbit universitari. També cal destacar iniciatives que van més enllà de l'àmbit educatiu formal, com, per exemple, el programa d'empoderament ciutadà liderat per Verificat i destinat a població sènior anomenat «Train the Trainers», i la iniciativa BuloBús, de Maldita.es, que, amb el suport de la Google News Initiative, és un minibús que recorre l'Espanya rural amb l'objectiu d'ajudar la ciutadania a combatre la desinformació, especialment la població sènior. A més, Maldita.es és una plataforma referent en la creació de material ludificat en la lluita contra la desinformació. Actualment, entre d'altres, compten amb NoMoreHaters, un joc en línia dissenyat per contrarestar els discursos d'odi a les xarxes (García Ortega, 2022).

A Catalunya, també cal dir que el Consell de l'Audiovisual (CAC) impulsa des del 2017 el programa EduCAC, una web de materials didàctics per a famílies i professorat elaborada en col·laboració amb el Departament d'Educació que té l'objectiu de fomentar l'educació mediàtica en els àmbits educatius formals i informals, i que se suma a la iniciativa Els Premis El CAC a l'Escola, que des de l'any 2002 reconeix accions d'educació mediàtica als centres educatius catalans. A banda d'aquests recursos, el CAC també ofereix la possibilitat de realitzar tallers a centres educatius. Cal dir, però, que els materials disponibles encara no inclouen temes relacionats amb l'ús de la IA, tot i que està previst que es vagin actualitzant en aquest àmbit al llarg del 2025.

Justament, amb el propòsit de començar a treballar en aquesta línia, i coincidint amb la Setmana Mundial de l'Educació Mediàtica de la UNESCO, el CAC va organitzar l'octubre de 2024 la I Jornada sobre Desinformació i Intel·ligència Artificial amb l'objectiu de reflexionar sobre l'ús de la IA en la difusió de falsedats i com a part també d'una de les línies d'acció de la Plataforma per l'Educació Mediàtica de Catalunya, creada pel CAC l'any 2019. Des de l'ens, però, reclamen al Parlament de Catalunya una llei més ambiciosa que també els doti de més competències i que això també els permeti ampliar les seves accions en la lluita contra la desinformació i la promoció de l'educació mediàtica.

En un context de democratització a l'accés a la intel·ligència artificial, diferents iniciatives –especialment de l'àmbit privat– han ampliat les activitats educatives realitzades en aquest àmbit, com, per exemple, les mateixes plataformes de verificació d'informació. Organitzacions com Verificat han adaptat els seus tallers al camp de la IA amb l'objectiu de difondre eines per identificar falsedats generades amb aquesta tecnologia. A Catalunya, també ho han fet plataformes com Learn to Check, una iniciativa sense ànim de lucre formada per diferents investigadors i professors universitaris experts en l'àmbit de la desinformació que es dediquen a realitzar accions de transferència de coneixement per conscienciar la població en l'àmbit de la desinformació i la verificació digital.

De fet, la IA també s'ha convertit en un recurs utilitzat per les mateixes plataformes de verificació de fets a l'hora de verificar continguts. Per exemple, hi ha algunes experiències exploratòries amb ChatGPT (Cuartielles *et al*., 2023), així com amb eines d'intel·ligència artificial que permeten la identificació de falsedats recurrents mitjançant *claim matching*, és a dir, identificar afirmacions que comparteixen un significat comú malgrat que s'expressin de maneres diferents (Larraz *et al*., 2023). Tot i així, cal dir que es tracta d'un ús de la IA assistit per professionals que en comproven les dades i resultats, un model anomenat *human-in-the-loop* ('humà en el procés') (La Barbera *et al*., 2022) per assegurar-ne la fiabilitat.

L'ús de la IA generativa mitjançant la implementació de bots de conversa, en aquest cas en xarxes socials com WhatsApp, és un recurs que també ha demostrat ser útil per contrarestar la desinformació (Palomo i Sedano-Amundarain, 2018) i que diverses plataformes de verificació de tot el món implementen en les seves rutines de treball (Flores-Vivar, 2020; 2019), de nou de manera assistida. En els darrers anys, han sorgit iniciatives col·laboratives com FactChat o el bot de conversa de la Covid-19, llançats en les eleccions presidencials d'Estats Units del 2020 i durant la pandèmia, respectivament. Ambdues iniciatives s'emmarquen en la Xarxa Internacional de Verificació de Fets, la International Fact-Checking Network (IFCN, per les seves sigles en anglès), que compta amb la implicació de més de 80 plataformes de verificació d'arreu del món (Grau, 2020).

Com a conclusió, s'inclou la taula següent amb els punts clau sobre competència digital i democràcia:

Idees principals sobre competència digital i democràcia

L'alfabetització mediàtica i digital és essencial per dotar la ciutadania d'eines per detectar desinformació i fer un ús responsable de la IA.

Iniciatives com EduCAC o els tallers de Learn to Check i Verificat, plataforma de verificació, estan liderant l'educació mediàtica a Catalunya, adaptant-se als reptes de la IA.

En l'àmbit europeu, programes educatius més consolidats, com els de Portugal o França, serveixen de model per a promoure l'alfabetització mediàtica i l'ús i el consum crític dels continguts generats amb IA.

La col·laboració entre administracions, societat civil i el sector privat és clau per fomentar una cultura democràtica socioinclusiva en els usos i impactes de la IA.

Referències bibliogràfiques

Acord GOV/70/2024, de 26 de març, pel qual s'aprova elaborar l'Estratègia nacional per a la inclusió sociodigital. *Diari Oficial de la Generalitat de Catalunya*. https://dogc.gencat.cat/ca/document-del-dogc/?documentId=982043

Administración Pública Digital (9 de novembre de 2023). Entrevista a Marisol López Vicente (directora general de Innovación y Cultura Digital de la Generalitat de Catalunya) «Nos gustaría que saliese un proyecto cada año con el objetivo de transformar la cultura "tradicional"». https://www.administracionpublicadigital.es/protagonistas-aapp/2023/11/nos- gustaria-que-saliese-un-proyecto-cada-ano-con-el-objetivo-de-transformar-la- cultura-tradicional-marisol-lopez-vicente-directora-general-de-innovacion-y- cultura-digital-de-la-generalitat-de-catalunya

Cuartielles, R.; Ramon-Vegas, X.; Pont-Sorribes, C. (2023). «Retraining fact- checkers: The emergence of ChatGPT in information verification». *Profesional de la Información, 32*(5). https://doi.org/10.3145/epi.2023.sep.15

Cucarella, L.; Fuster, P. (2022). *Alfabetización mediática: contexto actual, legislación, casos de éxito, herramientas y recursos, y percepción y propuestas de especialistas y profesores*. Laboratorio de Periodismo. Fundación Luca de Tena. https://laboratoriodeperiodismo.org/wp-content/uploads/2023/02/informe-alfabetizacion-mediatica.pdf

European Digital Media Observatory (EDMO) (2024). Mapping the Media Literacy Sector. https://edmo.eu/resources/repositories/mapping-the-media-literacy-sector/

Flores-Vivar, J. M. (2020). «Datos masivos, algoritmización y nuevos medios frente a desinformación y fake news. Bots para minimizar el impacto en las organizaciones». *Comunicación y hombre*, 16, p. 101-114. https://doi.org/10.32466/eufv-cyh.2020.16.601.101-114

Flores-Vivar, J. M. (2019). «Inteligencia artificial y periodismo: diluyendo el impacto de la desinformación y las noticias falsas a través de los bots». *Doxa Comunicación*, 29, p. 197-212. https://doi.org/10.31921/doxacom.n29a10

García Ortega, A. (22 de novembre de 2022). *¿Crees que puedes escapar de la desinformación? Llegan los* escapes rooms *de alfabetización mediática*. Blog del Máster en Innovación en Periodismo. https://mip.umh.es/blog/2022/11/22/escapar-de-la-desinformacion-escapes- rooms-de-alfabetizacion-mediatica/

Grau, M. (4 de maig de 2020). «New WhatsApp chatbot unleashes power of worldwide fact-checking organizations to fight COVID-19 misinformation on the platform». International Fact-Checking Network (IFCN), Poynter Institute. https://www.poynter.org/fact-checking/2020/poynters-international-fact- checking-network-launches-whatsapp-chatbot-to-fight-covid-19-misinformation-leveraging-database-of-more-than-4000-hoaxes/

Graves, L. (2016). *Deciding what's True. The rise of political fact-checking in American journalism*. Columbia University Press.

Grierson, J. (10 de juny de 2024). «Brighton general election candidate aims to be UK's first 'AI MP'». *The Guardian*. https://www.theguardian.com/politics/article/2024/jun/10/brighton-general-election-candidate-uk-first-ai-mp-artificial-intelligence

Kaĉinová, V; Sádaba Chalezquer, C. (2022). Competencia mediática como una "competencia aumentada". *Revista Latina de Comunicación Social,* 80, p. 21-38. https://www.doi.org/10.4185/RLCS-2022-1514

Palomo, B.; Sedano Amundarain, J. A. (2018). «WhatsApp como herramienta de verificación de fake news. El caso de B de Bulo». *Revista Latina de Comunicación Social*, 73, p. 1384-1397. https://doi.org/10.4185/RLCS-2018-1312

Landrin, S. (21 de maig de 2024). «India's general election is being impacted by deepfakes». *Le Monde*. https://www.lemonde.fr/en/pixels/article/2024/05/21/india-s-general-election-is-being-impacted-by-deepfakes_6672168_13.html?utm_source=chatgpt.com

Larraz, I.; Míguez, R.; Sallicati, F. (2023). «Semantic similarity models for automated fact-checking: ClaimCheck as a claim matching tool». *Profesional de la Información*, 32(3). https://doi.org/10.3145/epi.2023.may.21

Schueler, M.; Romano, S.; Stanusch, N.: Buse Çetin, R.; Tabti, S.; Faddoul, M.; Lilley, I. (2024). «Artificial Elections. Exposing the Use of Generative AI Imagery in the Political Campaigns of the 2024 French Elections». AI Forensics. https://aiforensics.org/uploads/Report_Artificial_Elections_81d14977e9.pdf

Stein, J. (13 de març de 2024). «Belarus Dissidents Turn to AI Deep Fakes». CEPA. https://cepa.org/article/belarus-dissidents-turn-to-ai-deep-fakes/

Subramanian, C. (6 de maig de 2024). «Nervous about falling behind the GOP, Democrats are wrestling with how to use AI». AP News. https://apnews.com/article/ai-biden-campaign-democrats-2024-election- 520f22de-269ba1eff24d1544ca38d569

UNESCO (2018). «Media and Information Literacy». *Executive Board*, 205 EX/34 Rev. https://unesdoc.unesco.org/ark:/48223/pf0000265509

White House Government (26 de juliol de 2024). «FACT SHEET: Biden-Harris Administration Announces New AI Actions and Receives Additional Major Voluntary Commitment on AI». https://www.whitehouse.gov/briefing-room/statements- releases/2024/07/26/fact-sheet-biden-harris-administration-announces-new-ai-actions- and-receives-additional-major-voluntary-commitment-on-ai/

La intel·ligència artificial en els parlaments i en el sector públic

Aquest informe sobre la intel·ligència artificial (IA) en els parlaments i en el sector públic està organitzat de la manera següent: a la primera part –La IA en el Parlament– analitzem els possibles usos de la IA generativa (IAG) al Parlament de Catalunya, tenint com a punt de referència altres experiències internacionals, i a la vegada, analitzem els suggeriments de diferents autors de referència, que, per una banda, coneixen bé l'abast del treball parlamentari i, per l'altra, entenen el funcionament de les eines de la IA i són ben conscients dels seus riscos i possibilitats. A la segona part –La IAG en el sector públic–, molt més breu que la primera, posem alguns exemples de l'ús d'IAG en el sector públic català i centrem la nostra exposició en aquelles bones pràctiques que cal compartir i difondre perquè han demostrat la seva eficiència, eficàcia i compliment de la normativa europea i serven el respecte degut als drets fonamentals i als principis ètics aplicables.

La IA en el Parlament

La realitat actual de l'ús de la IA en els parlaments: algunes dades

Alguns estudis acadèmics revelen que el 2020 un 6% dels parlamentaris ja utilitzaven alguna eina amb IA per redactar lleis i un 30% estaven sospesant si utilitzar aquesta nova tecnologia a escala mundial (Citino, 2024, citant *World E-Parliament Report*). No podem oblidar que la IA no ha arribat als diferents àmbits de la vida social, econòmica, educativa, administrativa o parlamentària d'un dia per l'altre, sinó que és una eina que ha estat present i disponible des de fa temps, amb propostes i solucions certament menys potents que les que tenim a disposició avui dia, però aquestes eines ja apuntaven un canvi de rumb notable, i no es pot afirmar, doncs, que l'escenari hagi mudat del tot dràsticament, sinó que el canvi ha estat veritablement progressiu.

Actualment, la IA també s'està fent servir per donar suport a activitats de caràcter auxiliar en el procés legislatiu de les

cambres parlamentàries (Griglio i Marchetti, 2022, 362 i seg.). Existeixen ja programes específics per a la generació d'esmenes als textos de les propostes normatives i per gestionar-les, així com també s'han implementat programes intel·ligents de suport que permeten gestionar les consultes públiques i la participació dels ciutadans en les diverses tasques parlamentàries en què aquesta és rellevant.

Cal tenir en compte, però, que, com Fitsilis *et al.* (2025: 6) han posat de manifest, hi ha nombroses possibilitats per incorporar sistemes basats en IA en les tasques parlamentàries, però encara hi ha un nombre relativament reduït de parlaments d'arreu del món que hagi començat a desenvolupar i a emprar aplicacions d'IA per millorar el desenvolupament de les seves funcions.

Principis orientadors de l'ús de la IAG en l'àmbit parlamentari

Els parlaments afronten diversos reptes pel que fa a la introducció i regulació de l'ús de la intel·ligència artificial (IA) en el seu àmbit de treball. Aquests desafiaments són tant tècnics com ètics, socials i polítics, i impliquen equilibrar la innovació tecnològica amb la protecció dels drets fonamentals i els valors democràtics. Certament, els parlaments d'arreu del món estan encarant reptes similars a l'hora de planificar l'adopció d'eines d'intel·ligència artificial (IA) amb l'objectiu d'enfortir els seus processos institucionals, com ara la transcripció automatitzada, les eines de suport legislatiu i la participació ciutadana impulsada per la IA (Fitsilis, F.; Von Lucke, J.; De Vrieze, F., 2025). Per aquest motiu, el desenvolupament de directrius sobre l'ús de la IA als parlaments esdevé un pas necessari per gestionar l'evolució tecnològica (*ibid.*).

Estem molt d'acord amb les paraules següents, que reflecteixen acuradament el punt en què ens trobem actualment en aquesta qüestió:

«Per a la majoria dels acadèmics i professionals implicats en el desenvolupament parlamentari, la discussió sobre la IA als parlaments podria semblar antinatural o prematura. No obstant això, avaluant l'impuls i el volum d'inversions en tecnologies d'IA, sembla segur suposar que aquest *ParlTech* és literalment *ante portas*. En conseqüència, les decisions estratègiques sobre la introducció d'aplicacions basades en IA al parlament i les seves limitacions s'han de prendre a curt termini. La producció científica sobre els parlaments hauria d'estar millor preparada per adaptar-se a qüestions relacionades, com ara l'estudi de l'ètica de la IA i l'extensió de les àrees tradicionals de la investigació parlamentària, com ara l'elaboració de lleis i la supervisió, cap als serveis millorats per la IA» (Fitsilis, 2021, 629).[1]

Més enllà d'això, a continuació assenyalem els principals reptes regulatoris a tenir en compte si el que es pretén és utilitzar la IA als parlaments en múltiples àmbits:

1. Velocitat de l'evolució tecnològica. La IA avança molt més ràpidament que els processos legislatius, la qual cosa dificulta l'adopció de marcs normatius actualitzats i eficaços (EPTA Report, 2023).
2. Marc normatiu basat en el risc. La UE ha desenvolupat el Reglament d'IA (RIA), que classifica els sistemes segons el seu risc (baix, alt o prohibit). Entendre aquesta norma i el seu enfocament és capital per determinar quin és l'abast real que pot tenir l'ús de la IA en qualsevol àmbit del sector públic, tot i que la norma no abordi directament aquesta qüestió.
3. Garantir la supervisió humana i el funcionament transparent dels sistemes. És essencial garantir que els sistemes d'IA siguin transparents, fiables i subjectes a supervisió humana per evitar abusos o errors, d'acord amb el RIA.

Per abordar la qüestió relativa a l'ús que la IAG pugui tenir als parlaments, la doctrina especialitzada ha utilitzat un enfocament principal i prescriptiu. És a dir, qualsevol projecte que pretengui implementar l'ús de sistemes d'IA a un parlament hauria de tenir en compte una sèrie de principis, que han resultat de l'anàlisi de diverses experiències reals de l'ús d'aquestes eines tecnològiques en l'àmbit parlamentari.

Fitsilis *et al.* (2014, 13), en el seu llibre sobre l'ús de la IA als parlaments, descriuen aquells principis l'aplicació dels quals resulta rellevant a l'espai de treball parlamentari. En aquest informe, necessàriament breu, proposem centrar-nos en alguns d'aquests principis i seleccions i n'expliquem l'aplicació en l'àmbit parlamentari. Aquests principis només són suggeriments i segurament n'hi ha molts d'altres que poden orientar-nos sobre com desplegar l'ús de la IAG als parlaments.[2]

1 Traducció pròpia dels autors.

2 Altres principis són la inclusió i diversitat, l'adaptació als avenços tecnològics, la cooperació interparlamentària, la participació pública i el compliment normatiu.

Principis rellevants per a la IA als parlaments	Finalitat del principi	Aplicació a l'espai de treball parlamentari
Responsabilitat i transparència	Garantir decisions i aplicacions d'IA comprensibles, traçables i justificables.	Entendre el funcionament de la IA en diferents processos (per exemple, quines dades utilitza) del treball parlamentari i exigir explicacions als proveïdors sobre el seu funcionament. Tenir una documentació clara i accessible sobre aquest funcionament, crear un canal de comunicació amb els proveïdors i exigir un compromís de transparència, traçabilitat i rendició de comptes.
Autonomia de qui pren la decisió	Mantenir l'autonomia de qui pren la decisió sense manipulacions.	Informar-se sobre les diferents maneres de manipulació algorítmica i *nudging* i preparar un protocol de preservació de l'autonomia decisòria a escala parlamentària: p. ex. la falsificació de corrents d'opinió (*astroturfing*), perfils falsos de xarxes socials, etc.
Ús ètic i responsable de la IA	Mantenir els estàndards ètics i evitar el mal ús o el biaix en les aplicacions d'IA.	Conèixer els estàndards i revisar-los, així com ser informat sobre els canvis i les actualitzacions d'aquests estàndards ètics.
Supervisió humana i explicabilitat	Mantenir el control humà sobre els sistemes d'IA, però també tenir la capacitat de proporcionar una explicació orientada a diferents públics (per exemple, operador legal, ciutadà, etc.).	Evitar sempre la delegació completa de les decisions a la màquina i ser conscients que el seu ús no deslliura de la responsabilitat personal i professional de les decisions preses. En cas de dubte i/o poca transparència, manca d'informació o qualsevol altra causa que no permeti entendre el procediment i/o resultat, cal atenir-se a utilitzar el sistema i immediatament informar l'equip tècnic i altres col·laboradors del problema detectat.
Mitigació de riscos i avaluació de l'impacte per als drets fonamentals (IDF)	Identificar i abordar els riscos potencials associats a la implementació de la IA i detectats per l'avaluació de l'IDF.	Contribuir a l'avaluació de l'impacte per als drets fonamentals: familiaritzant-se amb models existents i, quan calgui, contribuint a preparar aquesta avaluació.
Confiança pública	Construir i mantenir la confiança pública en les institucions parlamentàries que utilitzen eines i serveis d'IA.	Explicar en cada ocasió que s'usi la IA què es fa al parlament en general i el treball individual i informar-ne amb claredat i transparència.

Observant aquesta taula podem apreciar fàcilment tant els beneficis del compliment d'aquests principis com els riscos d'ignorar-los. De fet, la violació d'aquests principis segurament implicaria conseqüències negatives per a la fiabilitat dels parlaments i la seva legitimitat democràtica. És d'enorme importància que el Parlament de Catalunya tingui presents aquests principis i el que representen com a garantia dels valors democràtics que ha de preservar.

Prerequisits per a l'ús de la IA o la IAG en l'àmbit parlamentari

Hi ha una sèrie de qüestions o prerequisits que els parlaments haurien de tenir en compte en el moment d'implementar sistemes d'IA o d'IAG. Com ja vàrem advertir en el cas de l'Administració pública en un altre treball (Velasco, 2022), si el Parlament aposta per l'ús i el desenvolupament d'eines d'IA o de programari intel·ligent, una de les qüestions fonamentals que hauria de tenir present és la necessitat de disposar de conjunts de dades de qualitat que puguin alimentar els sistemes, que o bé adoptaran decisions directament o bé oferiran prediccions sobre les quals basar la presa de decisions, tant en procediments legislatius concrets com en els processos de disseny i execució de tasques pròpies de la cambra.

Qualitat de les dades

Si, com tot apunta, ens encaminem cap a un model de presa de decisions basat en l'ús intensiu de dades, no hi ha dubte que els parlaments han d'adoptar totes les mesures necessàries per garantir que les da-

des que gestionen siguin de qualitat, és a dir, que no estiguin esbiaixades ni contaminades. En cas contrari, si les dades són de baixa qualitat, les decisions que se'n derivin no assoliran els resultats esperats i poden vulnerar drets fonamentals –com ara el principi d'igualtat– si condueixen a mesures discriminatòries (Velasco Rico, 2022).

Un tema especialment rellevant quant l'ús de sistemes d'IA o d'IAG és la sobirania de les dades i de les infraestructures a emprar. S'ha de recordar aquest principi, el de sobirania digital, sobretot quan els serveis i productes relacionats amb la IA són externalitzats a tercers o proveïts per ells. La major part dels models i sistemes d'IA provenen dels EUA, i hauria de ser aquest el camí a emprendre per les cambres parlamentàries europees, com apuntarem més avall.

També caldria anar amb molt de compte quan s'utilitzen sistemes d'IA entrenats amb dades sintètiques, ja que poden afectar la integritat i la precisió de les dades històriques.[3]

Capacitació del personal i nous processos selectius

Com ja hem apuntat més amunt, en segon lloc, cal adoptar amb urgència mesures que garanteixin una formació adequada del personal al servei de les cambres parlamentàries i dels mateixos parlamentaris en relació amb les eines d'intel·ligència artificial que ja s'estiguin utilitzant o que es prevegin incorporar. El personal ha de comprendre les implicacions, avantatges i riscos que comporta el seu ús, tant per al conjunt de l'organització com per als drets dels ciutadans. D'altra banda, i quant als processos de selecció de personal dels parlaments, urgeix repensar-los, atès que, per la seva configuració actual i el contingut dels temaris, no són adequats per identificar els nous perfils –absolutament imprescindibles– que necessitarà el Parlament per desenvolupar, gestionar i controlar els usos i les eines d'IA i el programari intel·ligent (Velasco Rico, 2022).

Falsificació de corrents d'opinió (***astroturfing***): estratègia de manipulació de l'opinió pública que consisteix a simular el suport espontani a una idea o una persona, com ha passat en diverses eleccions en què s'han utilitzat les tàctiques d'aquesta eina per difondre desinformació, desacreditar oponents i manipular la percepció pública a favor de certs candidats o partits polítics, com ha passat a les eleccions dels Estats Units (2016) (Schoch *et al.* 2022).

Les dades sintètiques són dades que reflecteixen les dades del món real, però que han estat creades per IA, que les genera molt ràpidament. Les dades sintètiques no contenen dades personals, però les propietats estadístiques d'aquestes es preserven. És una solució molt vàlida en les situacions d'escassetat de dades per fer proves amb sistemes, per investigar sobre el seu funcionament o per entrenar-les sense haver de fer front als problemes que es generarien si això es realitzés amb dades personals i/o amb dades protegides per secret comercial o industrial, o per propietat intel·lectual.

3 D'altra banda, l'Autoritat Catalana de Protecció de Dades (APDCAT) promou l'ús de dades sintètiques per protegir la privadesa en l'àmbit de la recerca i el desenvolupament tecnològic. Més sobre aixo: https://apdcat.gencat.cat/ca/sala_de_premsa/notes_premsa/noticia/ Jornada-Datos-Sinteticos

En particular, Von Lucke i Sanders (2024)[4] observen que:

«La manca de transparència dels models de llenguatge extens (LLM), la manca de coneixements sobre com alimentar-los i millorar-los i el repte aclaparador per als usuaris de valorar correctament les respostes [...]. La discussió es va centrar també en la incertesa i en un interès seriós sobre com estan canviant els conjunts d'habilitats [de les persones que treballen als parlaments] i quines mesures formatives són necessàries per ensenyar les noves habilitats necessàries.»

Mesures per evitar la captura per part de les empreses proveïdores de solucions tecnològiques

En tercer lloc, el desenvolupament de sistemes l'està desenvolupant bàsicament el sector privat, però amb una forta col·laboració del sector públic, a través d'un diàleg que es manifesta de múltiples maneres organitzatives o procedimentals. Aquest diàleg s'hauria de traduir en relacions equilibrades en la contractació per part de les cambres parlamentàries de solucions d'IA desenvolupades per empreses privades. Tant la contractació pública com la col·laboració publicoprivada en aquest àmbit exigeixen que els parlaments disposin de professionals amb un nivell de formació i experiència equivalent, capaços d'entendre, verificar, controlar, auditar, modificar i, si cal, desenvolupar els productes que proporcioni el sector privat. Això permetria evitar situacions de dependència o captació per part del sector privat, com ja ha passat en diversos projectes d'administració electrònica. En aquest sentit, caldria establir garanties per a la contractació de tecnologies d'IA per part de les entitats del sector públic i incorporar requisits de transparència, justificació, no-discriminació i respecte pels deures en matèria de privacitat. Aquest mateix equilibri s'ha de preservar en totes les iniciatives en què el Parlament pugui col·laborar amb el sector privat en la recollida, gestió i explotació de dades que posteriorment alimentaran els sistemes d'IA (Velasco Rico, 2022).

A parer nostre, com diem, seria oportú i desitjable, abans d'utilitzar sistemes en propietat d'altres països (dels EUA o de la Xina), explorar el que ofereix la IA de codi obert o també si hi ha alguna alternativa desenvolupada a escala estatal o europea. Per exemple, el fet que Catalunya sigui la seu del Barcelona Supercomputing Center (BSC) podria ser una gran oportunitat en aquest sentit.

El projecte ALIA, desenvolupat pel BSC, és la primera infraestructura pública d'IA en castellà i en les altres llengües cooficials (https://alia.gob.es/).

Aquesta iniciativa s'emmarca en un projecte més ambiciós que s'anomena ILENIA (impuls de les llengües en IA), un altre projecte del BSC, que resta «dins del projecte estratègic per a la recuperació i transformació econòmica (PERTE) Nova Economia de la Llengua (NEL), que

4 Traducció pròpia dels autors.

pretén aprofitar el potencial del castellà i de la resta de llengües oficials com a factor de creixement econòmic i de competitivitat internacional en àrees com ara la intel·ligència artificial, la traducció, l'ensenyament, la producció i la divulgació cultural, la recerca i la ciència» (https://proyectoilenia.es/ca/sobre- nel/).

Mesures d'alineament ètic i normatiu

Tenint en compte els riscos potencials que l'ús d'eines d'IA pot representar per a l'interès general i per als drets i llibertats de les persones, convindria fomentar l'adopció de codis de conducta sectorials i professionals, així com per al sector públic, amb un pla d'impuls específic. Per exemple, es podria exigir l'adhesió a codis amb uns continguts mínims com a requisit per poder treballar o contractar amb les cambres parlamentàries per a la provisió de solucions d'IA, i també incloure'n el contingut en els programes formatius del personal al servei dels parlaments. Això hauria de garantir la igualtat entre els licitadors i assegurar, des del disseny mateix, el compliment dels principis de privacitat, transparència, etc. (Velasco Rico, 2022)

Mesures per trencar resistències al canvi provocat per la implementació de la tecnologia

També cal tenir present que, més enllà de l'adopció de cartes, codis ètics i altres instruments de dret tou (*soft law*), també caldria promoure un canvi cultural dins els mateixos parlaments –a partir de la formació, però no únicament a través d'ella– que permeti corregir certes inèrcies tecnòfobes presents en determinades instàncies i superar les resistències al canvi que podrien manifestar-se en molts àmbits del dia a dia de les cambres parlamentàries.

Tasques en què es pot aplicar la IA en l'àmbit parlamentari

Així mateix, la doctrina ha expressat la seva preocupació per la manca del debat i d'indicacions clares sobre els usos de la IAG en els parlaments arreu del món (Von Lucke i Sanders, 2024).

Els experts han analitzat les principals aplicacions parlamentàries de la IA i de la IAG i les agrupen al voltant de les diverses tasques o funcions que es poden desenvolupar en l'àmbit parlamentari. Aquesta categorització es basa en prescripcions d'experts i en les dades empíriques recollides de l'anàlisi de les experiències desenvolupades en tres òrgans parlamentaris: el Parlament hel·lènic, l'Honorable Cambra de Diputats de la Nació Argentina i el Parlament Federal del Canadà.

Aquestes tasques són set (Fitsilis *et al*., 2024):

1. Tasques amb relació als parlamentaris:
   a. Subtitulació en temps real dels discursos dels diputats al parlament
   b. Sistemes de votació fiables al ple i les comissions
   c. Generació de continguts per a discursos i preguntes escrites
   d. Suport en la recuperació d'informació
2. Tasca legislativa:
   a. Examen de propostes legislatives per a les interaccions amb altres regulacions
   b. Recomanacions sobre legislació basades en llacunes, problemes i altres lleis rellevants identificats
   c. Esborranys de text per a un posterior processament
   d. Millor regulació i implementació de polítiques digitals
3. Tasca de control parlamentari i diplomàcia parlamentària:
   a. Anàlisi dels mitjans de comunicació sobre les activitats del parlament
   b. Anàlisi de dades de les xarxes socials sobre les activitats del parlament
   c. Detecció de manipulació de l'entorn d'informació
   d. Mesures per reduir el biaix/discriminació en les propostes d'eliminació basades en IA
4. Foment de l'educació cívica i cultura nacional:
   a. Funcions de cerca intel·ligents al lloc web del parlament
   b. Transparència mitjançant dades obertes (enllaçades)
   c. Visualització d'arguments i debats
   d. Facilitar l'aportació del públic als tràmits parlamentaris
5. Administració parlamentària, edificis del parlament, servei de conducció i policia:
   a. Ajudants virtuals per a persones amb discapacitats
   b. Programari de ciberseguretat
   c. Serveis de traducció
6. Mesa parlamentària, direccions parlamentàries i eleccions:
   a. Detecció de contingut fals generat per IA destinat a manipular el procés democràtic

   b. Automatització de processos
   c. Gestió de projectes

7. Serveis de recerca/científics:
   a. Cerca intel·ligent de documents
   b. Gestió avançada del coneixement
   c. Verificació de fets

Riscos concrets de l'ús de la IA en l'àmbit parlamentari i mesures de mitigació

Riscos de seguretat i l'opció de proveir-se de sistemes propis

Una vegada assenyalades les tasques en què es pot aplicar la IA o la IAG en l'àmbit parlamentari, s'ha d'emfasitzar que sorgeixen preguntes i dubtes sobre possibles noves dependències a les quals es poden veure sotmeses les persones que treballen als parlaments (i no pensem només en els parlamentaris, sinó en un ampli ventall de personal, d'experts, consellers, etc.).

Com ja s'ha dit, la IAG té capacitat de crear automàticament textos o esborranys de text sobre la base d'altres textos o esborranys, dades i altres tipus de documents. Segurament els actuals models lingüístics aniran perfeccionant i millorant aquests processos. No cal oblidar, però, que, a diferència de molts altres sectors, els parlaments tracten informació molt sensible i, de vegades, fins i tot informació reservada i, per tant, constitueixen una institució molt rellevant pel que fa a la seguretat nacional. En conseqüència, pot ser molt oportú que els parlaments, tenint en compte els riscos i la importància del seu treball, optin per proveir-se de sistemes d'IA per a les solucions protegides pròpies a fi d'evitar influències externes que podrien influir en el treball dels parlamentaris individualment considerats, però també dels grups parlamentaris i disrompre els processos deliberatius.

En el cas que una cambra parlamentària decidís proveir-se de sistemes propis, aquestes solucions haurien de venir acompanyades no només d'una guia d'ús de l'eina (per exemple, hauria d'incloure com s'han de fer els *prompts* i quins *prompts* o temes queden prohibits o restringits), sinó que també haurien d'incloure bones pràctiques i casos d'ús concret que altres ens públics o parlaments podrien voler compartir.

Així mateix, abans de posar en marxa qualsevol eina d'IAG, s'haurien de desenvolupar algunes sessions obertes a tot el personal de la cambra, o bé oferir-les segmentades per especialitats, per analitzar de forma crítica els resultats (esborranys de documents, anàlisis de dades, etc.) que s'hagin generat pel nou sistema d'IAG. Aquestes sessions haurien de servir d'entrenament per afinar les competències de les persones a l'hora de detectar fallades, problemes, errors o al·lucinacions presents en els textos generats sense supervisió humana.

Fer front a riscos futurs i incerts, l'establiment d'un diàleg permanent amb experts i el *human-in-the-loop*

La tecnologia basada en IA i en IAG està en efervescència i en plena evolució, de tal manera que els riscos que genera actualment poden ser diferents dels riscos que pugui generar en el futur. En aquest sentit, i per poder fer front als nous riscos que eventualment puguin sorgir, una opció seria establir relacions de confiança amb els experts en la matèria, a través de procediments o de solucions organitzatives, per habilitar un canal de comunicació i aprenentatge continu sobre els riscos de la IA que segurament aniran desenvolupant-se, canviant i modificant-se amb l'avenç de la tecnologia. Els experts subratllen dos àmbits en què això es produirà amb més intensitat:

(1) la identificació i el tractament dels aspectes ètics del funcionament dels sistemes basats en IA en un parlament, i
(2) les limitacions de l'ús de la IA al parlament i per desenvolupar tràmits parlamentaris (Von Lucke, Fitsilis, Gagnon, 2024).

Amb relació al primer dels àmbits identificats, cal poder individualitzar i tractar qüestions ètiques relacionades amb l'ús de sistemes d'IA. Per això, és necessari tenir coneixements sobre ètica i sobre els principals problemes que ha generat l'ús de la IA, com ara poca transparència en el procés de presa de decisions. En l'àmbit parlamentari, com en molts altres espais del sector públic, segurament no hauria de ser legal, i èticament no és correcte i no és acceptable adoptar una proposta generada pel sistema d'IA sense preguntar-nos com el sistema ha arribat a prendre-la. Aquesta reflexió és necessària sobretot si afecta directament les persones. És un deure moral per a qui adoptarà aquesta decisió saber sobre la

base de quines dades s'ha generat una determinada decisió i quin ha estat recorregut de raonament del sistema per arribar-hi.

Quant al segon dels àmbits esmentats, és important que les persones al servei del parlament que emprin sistemes d'IA entenguin que qualsevol ús de la IA és al capdavall la seva responsabilitat tant ètica com legal i que l'ús de la IA no els eximeix de la seva responsabilitat professional personal. Cal ser-ne conscient i entendre que la IA és una eina que no sempre genera un resultat fiable i que cal revisar tot el que aquesta fa, genera, recomana i soluciona. Això és la manera no només de supervisar la IA, sinó també d'evitar qualsevol tipus de problema en el futur. Aquí estem fent referència al conegut concepte de *human-in-the-loop*, o humà al bucle, que fa referència a sistemes en què les persones participen activament en algun dels passos de qualsevol tipus de procés, generalment en la seva supervisió, en la presa de decisions o en la valoració i l'ajust dels seus resultats, i s'asseguren així que aquests processos funcionen de manera eficient, precisa i ètica.

Riscos sobre la pèrdua del control humà de les decisions, el biaix d'automatització i la reserva d'humanitat

L'ús intensiu que s'està començant a desplegar davant dels nostres ulls, en temps real, de la IA i sobretot de la IAG genera un altre risc concret que no podem menystenir i no que és un altre que el biaix d'automatització, que soscava realment la supervisió que els humans podem realitzar respecte dels resultats que el sistema proporciona.

El biaix d'automatització és la tendència dels éssers humans a confiar excessivament en sistemes automatitzats o en les recomanacions d'aquests, fins i tot quan existeixen evidències –o fins i tot percepcions pròpies– que contradiuen aquestes recomanacions. Aquesta sobreconfiança pot portar a errors en la presa de decisions, especialment en contextos en què la supervisió humana és essencial.

Hi ha força literatura acadèmica que ha abordat aquesta qüestió. Fent un repàs ràpid d'alguns treballs especialment interessants respecte d'aquest fenomen, podem assenyalar-ne algunes conclusions rellevants. Així, en l'àmbit sanitari i mèdic, Goddard, Roudsari i Wyatt (2012) analitzen com el biaix d'automatització influeix en l'ús de sistemes de suport a la decisió clínica, i destaquen la necessitat de comprendre i mitigar aquest biaix per millorar la seguretat del pacient. Per la seva banda, Skitka, Mosier i Burdick, ja l'any 1999, analitzaven com la confiança excessiva en l'automatització pot afectar negativament la presa de decisions, especialment en entorns crítics. Específicament en l'àmbit públic, Alon-Barkat i Busuioc (2022) examinen com el biaix d'automatització pot influir en les decisions del sector públic i subratllen la importància de la supervisió humana en l'ús de sistemes d'IA.

Certament, la possibilitat d'acceptar, en qualsevol àmbit, però especialment en l'àmbit parlamentari, sense més, que els éssers humans, massivament, per un biaix cognitiu, concloguin que els resultats generats per la IA són sempre millors que els poden obtenir directament els humans i que no cal revisar-los críticament i atentament seria un error molt greu. La gravetat rau a com es podria alterar l'equilibri de la presa de decisions parlamentàries, equilibri plasmat en els reglaments de les cambres i que tan difícil és i ha estat d'aconseguir i que encara esdevindrà més difícil d'ajustar i de mantenir amb el pas del temps i amb l'adveniment de l'ús generalitzat d'aquestes tecnologies.

Un risc que pot passar desapercebut, però que en l'àmbit parlamentari pot esdevenir crític, és que l'ús de sistemes d'IAG pugui tenir un impacte directe en la capacitat dels presidents de les cambres i en la capacitat de les meses per exercir la seva discreció a l'hora d'interpretar les normes que regulen els procediments parlamentaris, cosa que és, òbviament, la part fonamental del paper del president i del rol de les meses. Si es confia massa en les solucions proporcionades pels sistemes esmentats, sense comprovar-les, i sobretot, per exemple, en la capacitat de la IA de categoritzar les esmenes que es puguin presentar a les propostes normatives i avaluar-ne l'admissibilitat, això pot portar a greus problemes i produir desconfiança social, sobretot si aquests sistemes s'utilitzen en la tramitació de normes sobre assumptes controvertits, d'alta importància socioeconòmica i/o política o que afectin els drets fonamentals o les estructures democràtiques.

Tal com ha explicat la doctrina italiana, la naturalesa complementària del reglament parlamentari amb les disposicions constitucionals és tal que permet un cert marge d'apreciació en la seva interpretació (Rivosecchi, 2004). És a dir, en les democràcies liberals s'atorguen

amplis poders als parlaments, i cada parlament articula aquests poders a través del seu reglament, que és la norma clau per articular la vida parlamentària i el joc institucional de la democràcia dels grups polítics. Per tant, és essencial que es prenguin mesures de precaució per superar el risc que genera l'ús de la IA en el sentit explicat més amunt, per tal que el poder de la presidència de l'assemblea parlamentària no es dilueixi ni es perdi (Citino, 2024). En definitiva, si introduïm la IA sense una reflexió crítica prèvia ens arrisquem, per exemple, a debilitar –potser fins a fer-los irreconeixibles– els poders dels actors humans que governen les institucions, en aquest cas, els parlaments, i debilitarem, en conseqüència, els seus fonaments democràtics.

En aquest punt, resulta clau recórrer al concepte *reserva d'humanitat*, que entre nosaltres ha desenvolupat Ponce Solé (2022), i que es refereix a la necessitat de mantenir la intervenció humana en determinades decisions, especialment aquelles que impliquen valors ètics i empatia, aspectes que les màquines no poden reproduir. L'autor argumenta que certes decisions no haurien de ser delegades completament a sistemes automatitzats, ja que la manca d'empatia de les màquines podria conduir a resultats inacceptables des d'un punt de vista humà. Així –sosté l'autor– la supervisió humana és essencial en l'ús de la IA per garantir que les decisions automatitzades respectin els drets fonamentals, la legalitat i els principis ètics. Ponce destaca que la reserva de humanitat és clau perquè les màquines no tenen empatia i, per tant, no poden gestionar adequadament certes decisions que requereixen comprensió humana.

Ponce Solé es basa, en part, en els posicionaments de Larson (2022), que ha posat de manifest i ha recordat de forma incansable la incapacitat dels sistemes d'IA de fer inferències. La inferència fa referència al procés lògic d'arribar a una conclusió a partir del que se sap i s'observa. Els humans estem fent inferències constantment. Larson recorda que hi ha dos tipus d'inferència: inductiva i deductiva. No obstant això, la majoria d'inferències que fem es basen a observar un fet i formular hipòtesis sobre les seves causes possibles: què ha passat? per què ha passat això? Aquest tipus d'inferència s'anomena raonament abductiu, un terme poc conegut encunyat per un matemàtic i filòsof del segle xix. L'abducció implica qüestionar constantment per què passa alguna cosa i formular hipòtesis sobre el món. En opinió de Larson, això és el que fem els humans, que és diferent de l'anàlisi estadística, i actualment no és quelcom que puguin fer els ordinadors.[5]

Aquestes aportacions subratllen la importància de preservar un paper actiu per als humans en processos decisoris, especialment en contextos en què les consideracions ètiques i empàtiques són fonamentals, com en l'àmbit parlamentari.

L'experiència del Parlament Europeu, com a punt de partida, i l'ús de la IA al Parlament Federal del Brasil

El Parlament Europeu segurament ha estat la institució representativa més informada fins ara sobre qüestions relacionades amb la IA: no és casualitat que la primera norma jurídica que la regula amb un abast important a escala mundial (el RIA) hagi estat aprovada a la UE. Així, el Parlament Europeu, que ha disposat de molta informació, capacitats i recursos, ha adoptat diversos posicionaments rellevants (abans de l'aprovació del RIA també) i ja utilitza activament solucions d'IA de la seva unitat d'arxius.

El Parlament Europeu ha implementat diverses eines que utilitzen la IA per millorar l'eficiència i l'eficàcia dels seus processos. Entre les solucions adoptades s'inclouen els bots de conversa (*chatbots*), que automatitzen processos en diferents àrees i ajuden a oferir respostes ràpides i efectives a les preguntes formulades pels usuaris. A més, el Parlament Europeu utilitza un sistema automatitzat basat en la IAG capaç de resumir textos i un editor per fer breus resums, i amb això es facilita la comprensió de documents complexos (Chamber of Deputies' Supervisory Committee on Documentation Activities of Italy, 2024).

A parer nostre, seria oportú seguir molt de prop el camí que ha recorregut el Parlament Europeu pel que fa a l'ús

5 El terme *raonament abductiu* va ser encunyat pel filòsof i lògic nord-americà Charles Sanders Peirce al segle XIX. Peirce va introduir aquest concepte per descriure un tipus d'inferència no deductiva que es distingeix de la inducció i la deducció, i que es basa en la formació d'hipòtesis explicatives per a fenòmens observats. Segons Peirce, l'abducció és el procés de generar noves idees o explicacions, i és fonamental en el procés de descobriment científic. Font: Stanford Encyclopedia of Philosophy (darrera consulta, 06.04.2025).

de la IAG i seguir les seves passes, ja que així podem estar segurs que el marge d'error en implementar solucions a la cambra catalana es podria reduir molt. Òbviament, això no vol dir que el Parlament català no disposi de la llibertat d'explorar les possibilitats que li pot oferir tant la IA com la IAG, amb independència del que hagin fet altres cambres parlamentàries. Amb tot, considerem que aquesta exploració caldria fer-la en col·laboració amb altres ens públics i/o privats i també involucrant sempre la societat civil.

Una altra experiència significativa és la que s'ha desenvolupat al Parlament Federal del Brasil. Allà s'ha usat el sistema Ulysses, que és una gamma completa i integrada de mòduls impulsats per IA, cadascun d'ells dissenyat per millorar un aspecte particular del flux de treball legislatiu. El paquet està compost de:

1. Agregació de continguts i organització temàtica a la pàgina web parlamentària;
2. Automatització de la distribució de sol·licituds del conseller legislatiu;
3. Anàlisi semàntica en les peticions de redacció legislativa;
4. Anàlisi i gestió de les esmenes legislatives;
5. Transformar el manteniment de registres legislatius amb la transcripció basada en IA;
6. Anàlisi semàntica dels comentaris del públic en les propostes legislatives;
7. Integració de l'autenticació biomètrica amb Infoleg per a les deliberacions parlamentàries;
8. Aplicació del reconeixement facial per a la comunicació i fotos de parlamentaris.

Què hauria de tenir en compte (criteris, indicadors, estàndards internacionals, etc.) una guia d'ús de la IA en l'àmbit parlamentari?

Les directrius de la Westminster Foundation for Democracy

La Westminster Foundation for Democracy (WFD) ha presentat unes directrius per a la integració responsable de la IA en els parlaments. Aquestes directrius aborden aspectes com les implicacions ètiques, la transparència, la rendició de comptes i la integració de la IA en l'entorn parlamentari. A més, s'hi fan consideracions sobre la intel·ligència artificial general i l'autonomia humana, la privacitat i la seguretat, així com la governança i supervisió dels sistemes d'IA.

Aquestes directrius tenen com a objectiu assegurar que la implementació de la IA en els parlaments es faci de manera que reforci la democràcia i no la distorsioni. És fonamental que els parlaments adoptin aquestes tecnologies de manera responsable per mantenir la confiança pública i garantir que els processos legislatius siguin conformes a la normativa que els regula, a l'ètica i que servin el respecte degut al principi de transparència.

Per a cada directriu, es respon a un conjunt de preguntes clau: Per què importa la directriu? Hi ha exemples coneguts? I, com es pot implementar això? Cada directriu acaba amb uns suggeriments sobre com utilitzar les noves eines i com aquests suggeriments es poden adaptar als diversos i distints –en abast i ambició– projectes d'implementació de la IA en els parlaments actuals.

Com ja hem exposat, les directrius posen l'accent en principis ètics, com ara la responsabilitat, la transparència i l'equitat. Les directrius que ofereix la WFD subratllen la importància de respectar la dignitat humana, la privadesa i la diversitat cultural, alhora que s'aborden els biaixos en les dades i els algorismes. S'hi destaca la promoció de l'autonomia humana en la presa de decisions i, així mateix, es reconeix l'impacte potencial de la IAG en aquest àmbit. En l'informe que estem analitzant, les consideracions sobre les mesures per salvaguardar la privadesa i la seguretat dels treballs parlamentaris són crucials, i s'advoca per implementar mesures sòlides per preservar les dades personals i prevenir els ciberatacs.

Les directrius descriuen com la governança i la supervisió eficaces són elements clau per alinear l'ús de la IA amb els valors democràtics i garantir-ne la transparència. El disseny i el funcionament del sistema d'IA haurien de prioritzar la interoperabilitat, la transparència, la fiabilitat i la seguretat, juntament amb el control humà dels sistemes d'IA. En el document que analitzem també es fa èmfasi en la creació de capacitats i en la formació amb l'objectiu de dotar els parlamentaris i el personal al servei de les cambres de les habilitats i els coneixements necessaris per a un ús responsable de la IA.

Recomanacions extretes de l'observació científica plasmada en treballs acadèmics

A més de les directrius publicades per la WFD, hi ha treballs acadèmics especialment rellevants per al nostre objecte d'estudi, com és el de Fitsilis, von Lucke i De Vrieze (2024). La seva obra ens ajuda a identificar i a avaluar de forma crítica algunes decisions que s'han pres en altres cambres sobre com utilitzar i extreure el màxim profit de la IA en l'àmbit de les activitats parlamentàries i el funcionament de les cambres, tot alineant aquestes decisions amb la garantia de l'estat de dret.

Les directrius que proposen els autors citats a la seva obra s'organitzen en sis seccions:

1. Principis ètics (que ja hem presentat breument en apartats anteriors)
2. IA generativa i l'autonomia humana
3. Privadesa i seguretat
4. Governança, gestió de les dades i supervisió
5. Disseny del sistema i seu funcionament
6. Desenvolupament de capacitats i formació

No entrarem ara en detall a la qüestió relativa als principis ètics (1) que han de guiar l'ús de la IA o de la IAG als parlaments, ja que aquest tema ha estat tractat àmpliament més amunt. Però les directrius contingudes als punts del 2 al 6 mereixen una atenció especial i més detallada per veure'n l'aplicabilitat en el context del Parlament català.

Pel que fa a la IA generativa i a l'autonomia humana (directriu 2), cal tenir en compte, com s'ha dit més amunt, que s'han de prendre molt seriosament els riscos que la IA (generativa però no només) crea per a l'autonomia humana. Com es permeti que la IA interactuï amb l'autonomia humana dependrà molt de les decisions que s'hagin pres sobre el disseny dels sistemes, però també dels àmbits d'aplicació dels sistemes. Si el que es pretén és garantir l'autonomia humana s'han d'establir quadres clars de responsabilitats eticojurídiques per a totes les parts implicades en el disseny dels sistemes.

En l'àmbit parlamentari no es tractaria, doncs, de substituir els éssers humans, sinó d'ajudar-los i assistir-los en tasques intel·lectualment complexes i que demanen esforços intel·lectuals (per tant, no estem parlant de l'automatització de tasques avorrides i repetitives que moltes vegades són les que defineixen altres àmbits de l'aplicació d'IA). Assistir l'ésser humà en l'àmbit parlamentari –no substituir-lo– és també una forma de garantir que els principis democràtics no estan en risc i que els éssers humans segueixen prenent totes les decisions en els assumptes polítics.

Això requereix:

— La determinació de quin ha de ser el paper concret de la IA al Parlament: és a dir, hauria de quedar clar que la IA és una eina de consulta que pot generar idees i/o nous enfocaments, que pot suggerir aspectes que potser les persones no han pensat, però sempre serà una eina i no substituirà l'ésser humà en la presa de decisions.
— Una formació específica per als parlamentaris que emfasitzi que la IA es pot equivocar, pot generar resultats (continguts, recomanacions, etc.) esbiaixats, pot al·lucinar (inventar-se respostes), i que cal parar molta atenció en la revisió del que el sistema proposa i genera.
— Cal establir clarament l'obligació que els humans supervisin els resultats que produeixin els sistemes d'IAG.
— Cal prioritzar que el disseny dels models i dels sistemes d'IA estigui centrat en l'ésser humà, per exemple, tenir en compte que la IA no té empatia, com ja hem apuntat més amunt, tenint present que molta de l'activitat parlamentària té repercussions socials sobre els grups de població vulnerables.
— L'establiment de mecanismes de supervisió humana sòlida, constant i professional, la qual cosa significa que mai es pot abaixar la guàrdia i que la responsabilitat de la presa de decisions és de l'humà i no delegada a la IA.
— Cal establir mecanismes que afavoreixin la transparència respecte de l'ús d'aquests sistemes d'IA o d'IAG. Això és, cal que se sàpiga i es publiciti quan s'ha d'utilitzar o s'ha utilitzat un sistema d'IA, en quina fase del procés parlamentari, amb quina justificació o per a quina necessitat, quin en va ser el resultat, qui el va revisar, etc.
— Cal establir canals de col·laboració amb experts en ètica de la IA, com ja hem apuntat, i amb institucions acadèmiques i organitzacions de la societat civil. Les cambres parlamentàries han d'estar informades sobre els avenços tecnològics en IA que poden afec-

tar l'autonomia humana i la presa de decisions en els processos parlamentaris.
— L'aprenentatge i la formació al llarg de la vida dels parlamentaris sobre la IA també té importància i va més enllà del seu ús al Parlament com a institució, atès que el Parlament és també el lloc que prendrà decisions sobre l'ús d'IA a l'Administració pública i al sector privat, i això suposa i implica la necessitat d'entendre, conèixer i estar al corrent del desenvolupament tecnològic sobre el qual després es discuteix i s'aprova la normativa que afecta tota la societat. Per això, el Parlament hauria de plantejar-se la creació d'un grup de treball específic dedicat a mantenir al dia els membres del Parlament, però també l'Administració parlamentària, sobre tot allò que es refereix al desenvolupament de la IA, per mitjà de l'organització de reunions amb els experts, etc. Aquest paper podria ser assumit per CAPCIT.

Quant a la privadesa i la seguretat (directriu 3), cal tenir en compte que els sistemes d'IA o d'IAG que s'empren en l'àmbit parlamentari sovint processen dades sensibles, com ara informació personal o dades de seguretat nacional o de seguretat pública. Sense les mesures que garanteixin un nivell de privadesa i de seguretat adequat, hi ha un risc cert de violacions de seguretat dels sistemes que comprometin les dades, de robatoris d'identitat i altres conseqüències perjudicials, i això sense tenir en compte la possible manipulació de les dades del sistema i els problemes plantejats per eventuals ingerències externes que pretenguin manipular els processos parlamentaris. Una prevenció defectuosa d'aquests riscos pot soscavar la confiança pública en el procés parlamentari i, de retruc, erosionar el sistema democràtic.

En particular, cal tenir en compte alguns aspectes clau per a la ciberseguretat:

— Tant la ciberseguretat com la privadesa han de ser dos aspectes que s'abordin «per disseny». Amb tot, tampoc no podem oblidar que sovint la part més dèbil de qualsevol sistema de ciberseguretat són les persones (a aquest tema es dedica l'enginyeria social, que descriu el conjunt de tècniques que usen els cibercriminals per enganyar les persones perquè els enviïn dades confidencials, dades personals, infectin els seus ordinadors amb un virus maliciós o obrin enllaços a llocs infectats, que després segrestin el sistema per exigir els diners a canvi), i és per això que s'ha de tenir ben present que la (ciber)seguretat sense una participació activa humana no és factible.
— La privadesa i seguretat dins del context de la IA parlamentària són crucials no només per protegir la privadesa de les persones individuals, sinó també per garantir la seguretat de les comunicacions a través de la intranet i per protegir la mateixa institució de possibles danys i amenaces de ciberseguretat.
— El Parlament pot garantir la seguretat dels sistemes d'IA exigint proves rigoroses del compliment de les normes i dels estàndards de ciberseguretat i amb la utilització del sistemes de xifratge. Cal establir un monitoratge continu, avaluacions de vulnerabilitat i protocols de resposta a les possibles situacions crítiques i adoptar les mesures (bones pràctiques) per prevenir qualsevol tipus de risc en aquest sentit.
— També és molt important fomentar una cultura de conscienciació sobre la seguretat dins l'espai de treball parlamentari, en què les persones estiguin alerta i siguin proactives a l'hora d'identificar i denunciar problemes de seguretat de manera fàcil i intuïtiva. Com a punt de partida podria servir un document recent de la CommonWealth Parlamentary Association (2025), en què el concepte de seguretat parlamentària es desglossa en molts enfocaments, com el de seguretat personal, física, dels empleats, ciberseguretat i seguretat tècnica.
— També seria molt important disposar d'una unitat de ciberseguretat dedicada a monitorar i millorar contínuament la seguretat dels sistemes d'IA: aquesta unitat també podria encarregar-se de fomentar la cultura de seguretat i prevenció i tenir al corrent tots els altres equips de treball i unitats tècniques dels desenvolupaments en ciberseguretat.
— Igualment, seria beneficiós col·laborar amb experts i agències governamentals de ciberseguretat per obtenir orientació i suport sobre com protegir eficaçment els sistemes d'IA parlamentaris: la xarxa existent de les institucions dedicades a la ciberseguretat pot ser una font important de recursos i de suport en situacions de crisi.
— S'ha de tenir en compte que molts parlamentaris volen treballar amb IA generadora de text. Per a aquest servei, caldria utilitzar bots de conversa interns i restriccions d'accés per garantir que les dades confidencials no es revelin involuntàriament a tercers no autoritzats.

— Així mateix, caldria demanar i invertir en auditories externes dels sistemes d'IA, la qual cosa podria aportar una garantia addicional al compliment de la normativa reguladora del seu ús en l'àmbit parlamentari.

No és mai sobrer posar en relleu les qüestions relatives a la ciberseguretat perquè les dades són clares i molt preocupants: només el 2022, l'Agència de Ciberseguretat de Catalunya va gestionar més de 4.424 milions de ciberatacs. No ens podem fer una idea de quant augmentarà aquesta xifra quan el Parlament català oficialment s'introdueixi algun sistema basat en la IA(G), però el que és segur és que augmentaran els ciberatacs. En aquest sentit, també és determinant que les decisions sobre ciberseguretat romanguin en mans humanes i no delegades a la IA i/o que resultin completament automatitzades. Òbviament la detecció de ciberatacs i moltes altres tasques poden i han de ser gestionades per la IA, però les decisions crítiques han de ser dels professionals que seran cridats a rendir els comptes, tractar aspectes ètics i, en general, respondre per l'eficiència del sistema que garanteixi la seguretat i la ininterrupció del treball parlamentari.

En relació amb la governança, la gestió de les dades i la supervisió dels sistemes (directriu 4), s'ha de recordar que caldria que els sistemes d'IA es desenvolupessin i s'utilitzessin d'una forma que s'alineï amb els valors democràtics i la supervisió parlamentària dels resultats generats per la IA(G) en garanteix la legitimitat, mentre que la IA(G) pot augmentar la confiança social en les institucions democràtiques i fer avançar el bé i l'interès comuns.

Qualsevol ús de la IA al Parlament s'hauria de basar en situar aquest ús dins una més àmplia estratègia digital de la institució. És, per tant, necessària una supervisió (ètica, tècnica, legal, etc.) que garanteixi que aquesta estratègia i l'ús pràctic de la IA(G) estiguin alineades i que permeti al Parlament com a institució tenir un paper més actiu en el desenvolupament de la IA(G). L'alineació de la IA amb el conjunt l'estratègia digital de la cambra és important perquè la IA contribueix eficaçment als objectius parlamentaris, alhora que s'arrenglera amb els esforços generals de transformació digital de la institució i en millora l'eficiència, la transparència i la responsabilitat.

Un altre aspecte molt rellevant que s'ha de tenir en compte quan es vol implementar un projecte d'IA en qualsevol organització, i, per tant, també ho és en l'àmbit parlamentari, és el relatiu als protocols per a la governança i la gestió de les dades que nodreixen els sistemes d'IA per assegurar-se que els resultats que proporcionen són acurats, complets i segurs. Aquests protocols permeten garantir els alts nivells de transparència i rendició de comptes, que són imprescindibles per a l'ús de la IA als parlaments.

Si bé avui dia existeixen protocols per a la governança i la gestió de dades en diferents sectors, encara no en tenim cap d'específic per al sector parlamentari i creiem que aquest buit s'hauria d'omplir com més aviat millor. És més, aquests protocols en l'àmbit parlamentari han de ser molt més exhaustius i rígids que per a alguns altres sectors. Aquests protocols han d'incloure els controls de qualitat de les dades, encriptació, auditoria, controls d'accés a les dades, compliment de la normativa de protecció de dades, etc. Per poder acomplir aquest objectiu òbviament és necessària l'estreta col·laboració amb experts en protecció de dades i amb l'Agència Catalana de Protecció de Dades.

Una altra proposta que volem deixar aquí apuntada és la creació d'un comitè ètic que supervisi l'ús dels sistemes d'IA al Parlament (o estendre a aquest àmbit les competències d'algun òrgan o comitè existent). Aquest comitè podria fer controls regulars sobre els sistemes, respondre preguntes i dubtes dels parlamentaris quant a l'ús legal i ètic de la IA al Parlament, crear coneixement i detectar i difondre bones pràctiques d'ús. Així mateix, també podria informar el públic sobre l'ús que es fa de la IA al Parlament i contribuir a generar i a mantenir una atmosfera de confiança, seguretat i credibilitat.

Potser, gràcies a aquesta iniciativa, es podria pensar a posar en marxa una xarxa a la UE de comitès ètics dels parlaments dedicats a la supervisió de l'ús de la IA en aquest àmbit. Aquesta xarxa podria posar en comú les problemàtiques detectades, compartir solucions i crear protocols comuns per afrontar els riscos que hem estat comentant.

Quant al disseny del sistema i seu funcionament (directriu 5), la manera de dissenyar el sistema d'IA defineix no només les seves funcionalitats, sinó també els riscos que genera, la manera de treballar dels actors involucrats, les dinàmiques per interactuar amb el sistema i molts altres aspectes. Per tant, s'ha de tenir molt en compte que

tots aquests aspectes es programen i que, en definitiva, com que serà un sistema, no és més que una decisió humana que cal negociar, decidir i consensuar amb tots els actors rellevants de l'àmbit parlamentari.

En aquest punt s'ha de destacar que en l'àmbit parlamentari pren una particular rellevància l'explicabilitat o la interpretabilitat del sistema del sistema d'IA(G). Un sistema d'IA és interpretable quan som capaços d'entendre com i per què pren les decisions que pren. Dit d'una altra manera senzilla: imaginem que la IA ens recomana una pel·lícula o aprova un préstec bancari. El fet que sigui interpretable vol dir que podem saber clarament quins criteris ha utilitzat per arribar a aquesta decisió, quines variables ha tingut en compte (per exemple, gènere de pel·lícula, historial de visionaments, nivell d'ingressos, etc.) i com ha influït cada variable en la decisió final. Això contrasta amb sistemes més complexos, com les xarxes neuronals profundes, que sovint són com una caixa negra: sabem què hi entra i què en surt, però no exactament com es decideix el resultat.

Per què és important que els sistemes d'IA emprats al Parlament siguin interpretables?

— Perquè permet identificar possibles errors o biaixos.
— Perquè ajuda les persones a confiar més en les decisions preses per la IA.
— Perquè facilita el compliment de normes ètiques i legals.

En definitiva, un sistema interpretable és un sistema transparent, del qual entenem les raons darrere de les seves decisions.

Aquesta explicabilitat no només és fonamental per als parlamentaris i per a les altres persones involucrades en el treball i el funcionament del Parlament, sinó també per al conjunt de la societat, que ha de poder comptar amb i confiar en aquests sistemes i en les seves capacitats per ajudar els humans en els processos de presa de decisions d'impacte social. El Parlament pot dotar-se de sistemes d'IA transparents i exigir als desenvolupadors o proveïdors que utilitzin models interpretables, i també demanar-los que proporcionin documentació comprensible sobre el sistema i sobre el seu ús, així com que estableixin mecanismes de supervisió per garantir la rendició de comptes. Hi ha moltes maneres d'incrementar el nivell de comprensió de les persones (tant de les persones que treballin al Parlament com de la societat) sobre com funciona la IA, sobre quines dades utilitza, i sobre com arriba a les conclusions i/o decisions, etc.

Els models d'IA interpretables ens permeten entendre el perquè d'un resultat que generen. A més, és important entendre'l perquè així podem entendre quan i com aquest resultat podria no ser correcte (per exemple, a causa de biaixos). Un model interpretable és el contrari d'un model de caixa negra, en què no podem entendre com el sistema ha arribat al resultat que ens proporciona.

En aquest sentit, també és important donar a conèixer de forma clara les limitacions de les explicacions dels algorismes d'IA per evitar expectatives massa elevades, així com malentesos i idees errònies sobre les capacitats reals dels sistemes.

Vist l'escrutini públic de l'activitat parlamentària i la rellevància de la seva funció, el Parlament es posiciona com un banc de proves sense precedents per fomentar i construir la confiança social en els sistemes d'IA(G) en general, però també, i sobretot, la confiança en el seu ús per les institucions democràtiques i per l'Administració pública. Per tant, s'ha de procedir amb extrema cura quant a l'ús de la IA (G) en aquest context, atès que no es poden reproduir els errors que en altres àmbits s'han detectat fins ara (violacions massives de drets humans per part de sistemes d'IA, generació de biaixos, manipulacions en diferents àmbits, escletxes de seguretat, etc.). El Parlament hauria d'establir uns estàndards molt alts i exigents perquè serveixi com a exemple i model per a altres institucions, empreses, entitats i la societat.

En relació amb el funcionament del sistema, també és fonamental atenir-se als estàndards d'interoperabilitat i compatibilitat entre diferents sistemes i plataformes digitals desplegats al Parlament. Aquesta interoperabilitat s'aconsegueix a través d'esquemes i processos de dades estandarditzats. Per exemple, emprant l'estàndard d'ús generalitzat en diferents parlaments a tot el món, des del Parlament Europeu fins al Parlament d'Uruguai, que s'anomena Akoma Ntoso, també emprat per la Generalitat de Catalunya.

Akoma Ntoso és un estàndard obert, desenvolupat específicament per representar documents legislatius, parlamentaris i judicials en format digital. El nom prové de la llengua akan de Ghana i significa «cors connectats» o «cors units».

Aquest estàndard té com a objectius principals:

— Facilitar la interoperabilitat entre diferents sistemes d'informació legislativa, parlamentària i judicial.
— Millorar l'accés i la transparència en la gestió de documents públics, com ara lleis, esmenes, debats parlamentaris i decisions judicials.
— Assegurar la preservació a llarg termini de la informació legal en format digital, independentment de les plataformes tecnològiques.

Akoma Ntoso utilitza un format XML especialitzat que permet etiquetar de forma estructurada i semàntica diferents elements dels documents jurídics, com articles, clàusules, autors, dates, versions i referències internes o externes. Això facilita enormement la cerca, indexació, visualització, reutilització i comparació automatitzada d'informació jurídica i parlamentària.

Aquest estàndard va ser desenvolupat originalment en el context de les Nacions Unides i ha estat adoptat per molts parlaments, governs i institucions d'arreu del món com a base per als seus sistemes de gestió documental legislativa i jurídica (Barabucci *et al.*, 2009).

El Parlament també hauria d'invertir en el desenvolupament de sistemes d'IA parlamentaris robusts i fiables que comptin amb capacitats de detecció i correcció d'errors, per exemple, a través del testeig, monitoratge, revisió posterior a la implementació. També hauria de comptar amb sistemes a prova de fallades (fail-safe). En tot això la col·laboració amb els equips tecnicoinformàtics és fonamental.

Una nota a part mereix l'avaluació del risc dels sistemes d'IA(G) sobre la base de l'actual normativa reguladora de la IA a la Unió Europea (RIA). Sobre la base de la seva classificació, els sistemes d'IA(G) que arribin a utilitzar-se als parlaments es qualificarien, en gran part, com a sistemes d'alt risc i, per tant, haurien de complir requisits específics: aquest compliment involucra tots els actors involucrats en el seu disseny i implementació i, en conseqüència, es necessiten alts nivells de col·laboració, confiança, seguretat i confidencialitat.

Finalment, pel que fa al desenvolupament de capacitats i la formació (directriu 6), la formació pot ajudar a desenvolupar coneixements i habilitats entre els parlamentaris i el personal al servei de la cambra. Això és primordial per garantir un ús efectiu, eficient, responsable i ètic i legalment raonable de la IA.

Aquesta formació hauria d'incloure, però no limitar-se a aspectes relatius a la comprensió del funcionament de la IA, de les seves aplicacions potencials i del seu impacte en la societat, així com les consideracions ètiques i legals que cal tenir en compte quan s'empra. També hauria de permetre difondre les bones pràctiques i aquelles experiències prèvies amb resultats no satisfactoris, que haurien de servir de lliçons per a tots els agents implicats. Sens dubte, no ens cansarem de repetir-ho, els programes de formació sobre la IA són molt necessaris tant dins com fora dels parlaments.

La IAG en el sector públic

Innovar amb IA generativa per millorar l'eficiència i la qualitat en la prestació dels serveis públics: casos d'ús a la Generalitat de Catalunya

En aquest darrer apartat de l'informe, volem donar compte d'alguns usos de la IAG en l'àmbit de l'Administració de la Generalitat de Catalunya i del seu sector públic. No és, òbviament, per raons d'espai i de pertinència, un catàleg exhaustiu el que aquí desenvoluparem, sinó que seran únicament uns breus apunts per situar el lector en la realitat de l'ús de la IAG en l'àmbit del sector públic català.

Per al cas concret de la IA generativa, la Generalitat de Catalunya va posar en marxa a finals de 2023, en cooperació amb Capgemini i Google Cloud, un primer xat d'IA generatiu experimental Català per generar respostes automàtiques a consultes, queixes i suggeriments que els ciutadans enviaven a l'Oficina de Gestió Empresarial. Donat el seu caràcter experimental, el personal de l'Oficina revisa les totes respostes generades.

La Generalitat també ha utilitzat la IA generativa, com a prova pilot, en el butlletí oficial per simplificar i resumir textos legals complexos en benefici del públic en general. El projecte es va posar en marxa en cooperació amb Inetum, una corporació privada. Segons els informes, la prova es va considerar reeixida i útil per promoure el

desenvolupament futur de la IAG en l'Administració pública catalana.

A més, cal recordar que l'Administració Oberta de Catalunya (AOC) el 2019 va llançar el seu primer bot de conversa utilitzant intel·ligència artificial. Aquest assistent virtual tenia l'avantatge d'estar operatiu les 24 hores del dia, amb un cost baix, però només era capaç de respondre a consultes simples. Actualment, l'AOC ha fet un salt qualitatiu en el servei suport als usuaris amb la implementació d'un bot de conversa d'última generació basat en intel·ligència artificial generativa. Segons la mateixa AOC, «en l'àmbit de l'atenció a la ciutadania, aquesta és una solució líder a l'Estat i equiparable a les iniciatives més avançades a nivell internacional».

Actualment el bot de conversa és operatiu en quatre serveis:

1. L'idCAT: certificat per la identificació digital. Ofereix suport per obtenir el certificat, per usar-lo i per renovar-lo.
2. L'idCAT Mòbil: ajuda a resoldre dubtes sobre la seva obtenció, ús i renovació.
3. L'e-Notum: per a notificacions i altres comunicacions entre administracions públiques catalanes. Resol dubtes i dona suport en la gestió de notificacions electròniques.
4. VALid: que és l'integrador d'identitats, com idCAT Mòbil, certificats digitals i Cl@ve. Assisteix la ciutadania en les consultes més freqüents durant el procés d'identificació a la pàgina de VALid.

Aquest sistema ofereix avantatges considerables respecte al model anterior. Si comparem la possibilitat d'entendre el llenguatge natural, el canvi és enorme. Això també implica, per exemple, una capacitat més gran de donar respostes més personalitzades i pertinents als casos individuals dels ciutadans.

Així mateix, l'AOC informa que, des del maig del 2024 quan es va llançar el xat, més d'11.000 persones al mes van utilitzar el sistema per a les seves consultes, la qual cosa demostra que la inversió en aquesta eina ha estat justificada i que el públic la utilitza. Òbviament, s'espera que el seu ús anirà augmentant sobretot quan el xat també pugui tractar casos molt més complexos.

Amb tot, les dades que ofereix l'AOC a aquest respecte són clares: el sistema encerta la resposta el 88% de casos. També ha generat un estalvi del 90% en l'atenció a una consulta d'un usuari. El cost mitjà de l'atenció humana és aproximadament de 15 €, mentre que atendre els dubtes dels usuaris a través del bot de conversa té un cost d'uns 1,5 €. Això significa un estalvi enorme de diners públics, que poden ser redirigits a tractar altres assumptes i serveis que necessiten més presència i assistència humanes.

Tot això no vol dir que el personal al servei de les administracions no estigui implicat en el funcionament correcte dels serveis perquè no s'ha obviat establir un sistema de monitoratge continu de les respostes proporcionades pel personal responsable de la decisió final. L'AOC explica que aquesta supervisió humana del sistema va ser particularment intensa durant el seu període de llançament, sobretot per compensar els biaixos de l'algorisme i reduir les respostes errònies i a fi de no perjudicar la confiança inicial que els ciutadans havien posat en aquest sistema d'IA.

Segons declara la mateixa AOC, la supervisió humana és molt rigorosa i es basa en:

1. Revisions setmanals: es revisen les converses per identificar respostes incorrectes proporcionades pel sistema. Per a cada error detectat, s'anota la resposta correcta i es reclassifiquen les frases dels usuaris amb una resposta errònia.
2. Revisions mensuals: aquestes revisions impliquen l'avaluació de 150 converses seleccionades aleatòriament, juntament amb les ja identificades com a errònies, per detectar patrons d'error (quins errors són més freqüents?) i corregir respostes inadequades i/o incompletes.
3. Gestió d'intents: quan els errors no es poden solucionar simplement reclassificant frases, s'ajusten els *prompts* dels intents existents o es creen nous intents que porten a les respostes correctes i exhaustives.

Els serveis que s'inclouran en el bot de conversa són els següents:

1. Representa: gestió de les representacions i apoderaments que facilita que una persona actuï en nom d'una altra.
2. e-FACT: punt general de factures electròniques de les administracions públiques de Catalunya.

Altres exemples d'ús d'IA en el sector públic de Catalunya

D'altra banda, més enllà de la IAG, el sector públic català utilitza algorismes predictius, per exemple, en el sector de l'execució penal i penitenciària per a l'avaluació i gestió del risc del conjunt de la població reclusa catalana (Riscanvi). L'avaluació del risc dels reclusos i la probabilitat de reincidència a l'hora d'avaluar la concessió de permisos de presó als reclusos es fa mitjançant un algorisme anomenat Protocol de valoració del risc (RisCanvi), que actualment utilitza la seva versió 3.0 i que s'actualitza cada tres anys. El seu ús, però, no està exempt de controvèrsia, com afirmen algunes auditories externes i també locals i més recents. En aquest sentit, i no només relacionat amb RisCanvi sinó també amb l'ús de la IA per part del Departament de Justícia de Catalunya en general, el Parlament ha instat el Govern mitjançant la promulgació de la Resolució 921/XIV sobre l'ús de la IA en el sistema penitenciari a:

a) Garantir que els algorismes d'IA hagin estat examinats i aprovats per l'Observatori d'Ètica en IA de Catalunya (OEIAC) abans de desplegar-los.
b) Vetllar perquè els algorismes i els codis font corresponents siguin públics, per tal de poder dur a terme tots els controls pertinents.
c) Garantir que s'estableixin mecanismes per revisar, cada dos anys, els nous programes aprovats que utilitzen intel·ligència artificial.
d) Vetllar perquè el Departament de Justícia revisi en un termini de dotze mesos els criteris del programa RisCanvi.

Referències bibliogràfiques

Alon-Barkat, S.; Busuioc, M. (2022). «Human-AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice. A: *Journal of Public Administration Research and Theory*. https://doi.org/10.1093/jopart/muac007

Barabucci, G. *et al.* (2010). «Multi-layer Markup and Ontological Structures in Akoma Ntoso». A: P. Casanovas *et al.* (Ed.): AICOL Workshops 2009, LNAI 6237, p. 133-149. Springer.

Chamber of Deputies' Supervisory Committee on Documentation Activities of Italy. 2024. Report on Using Artificial Intelligence to Support Parliamentary Work. https://comunicazione.camera.it/sites/comunicazione/files/notiz_prima_pag/allegati/Rappporto_IA_ENG_WEB_V2.pdf

Commonwealth Parlamentary Association (2025). Parlamentary Security. An Introductory Guide. https://www.agora-parl.org/sites/default/files/agora-documents/__Parliamentary%20Security_Introductory_Guide_FINAL%20%281%29.pdf

Citino, Y. M. (2024). «Leveraging automated technologies for law-making in Italy: Generative AI and constitutional challenges». A: *Parlamentary Affairs*, 20: 1-23.

Fitsilis, F., Von Lucke, J.; De Vrieze, F. (2025). «Inception, development and evolution of guidelines for AI in parliaments». A: *The Theory and Practice of Legislation*, 1-24. https://doi.org/10.1080/20508840.2025.2474791

Fitsilis, F.; Von Lucke, J.; De Vrieze F. (2024). *Guidelines for AI in Parliaments.* https://www.wfd.org/sites/default/files/2024-07/wfd-ai-guidelines-for-parliaments-2024-english.pdf

Fitsilis, F. (2023). «A Paradigm Shift for Parliaments». A: *International Journal of Parliamentary Studies*. https://brill.com/view/journals/parl/3/1/article-p1_001.xml

Fitsilis, F.; Micros, G. (2022). «Crowdsourcing the Digital Parliament - the Case of the Hellenic OCR Team». A: *Working Paper of the 15th Wroxton Workshop 30/31 July 2022*. https://wroxtonworkshop.org/wp-content/uploads/2022/07/2022-Fitsili s-Mikros-.pdf

Fitsilis, F. (2021). «Artificial Intelligence (AI) in parliaments – preliminary analysis of the Eduskunta experiment». A: *The Journal of Legislative Studies*, 27(4): 621-633.

Goddard, K.; Roudsari, A.; Wyatt, J.C. (2012). «Automation bias: a systematic review of frequency, effect mediators, and mitigators». A: *Journal of the American Medical Informatics Association*.

Griglio, E.; Marchetti, C. (2022). «La "specialità" delle sfide technologiche applicate al drafting parlamentare: dal quadro comparato all'esperienza del senato Italiano». A: *Osservatorio sulle fonti*, XV(3): 361-386.

Larson, E. J. (2022). *El mito de la Inteligencia Artificial: Por qué las máquinas no pueden pensar como nosotros lo hacemos*.

Ponce Solé, J. (2022). «Reserva de humanidad y supervisión humana de la Inteligencia artificial». A: *El Cronista del Estado Social y Democrático de Derecho*, número 100 (setembre-octubre de 2022), 58-67.

Rivosecchi, G. (2004). *Il Parlamento nei conflitti di attribuzione*. Milan: Cedam.

Schoch, D. *et al.* (2022). «Coordination patterns reveal online political astroturfing across the world». A: *Scientific Reports,* 12, 4.572. https://www.nature.com/articles/s41598-022-08404-9

Skitka, L.J; Mosier, K.L; Burdick, M. (1999). «Does automation bias decision-making?». A: *International Journal of Human-Computer Studies*. https://doi.org/10.1006/ijhc.1999.0252

Velasco Rico, C. I. (2022). «Tecnologías disruptivas en la Administración Pública: Inteligencia artificial y Blockchain». A: *La Administración Digital*, 227-256.

Von Lucke, J.; Sander, F. (2024). «A few Thoughts on the Use of ChatGPT, GPT 3.5, GPT-4 and LLMs in Parliaments: Reflecting on the results of experimenting with LLMs in the parliamentarian context». A: *Digital Government: Research and Practice.* https://dl.acm.org/doi/10.1145/3665333

Von Lucke, J.; Fitsilis, F.; Gagnon, S. (2024). «Using Artificial Intelligence in Parliament - Initial Results from the Canadian House of Commons». *Proceedings EGOV-CeDEM-ePart conference*, September 1-5, 2024, Ghent University and KU Leuven, Ghent/Leuven, Belgium. https://ceur-ws.org/Vol-3737/paper7.pdf

Intel·ligència artificial i regulació

La ràpida evolució de la intel·ligència artificial (IA), el seu abast global (tant en termes territorials com sectorials) o la complexitat de les cadenes de valor en què s'insereix, entre altres factors, han dificultat l'establiment d'un marc regulador propi. No obstant això, s'estan fent els primers passos a escala internacional. Destaquen els esforços de la Unió Europea amb un conjunt de marcs reguladors. La Llei d'intel·ligència artificial (AI Act) es complementa amb regulacions addicionals en relació amb dades –el Reglament general de protecció de dades (General Data Protection Regulation, GDPR); la Llei europea de governança de dades (European Data Governance Act), la Directiva de dades obertes (Open Data Directive) i la Llei de dades (Data Act)–, productes (proposta per a una directiva de responsabilitat de la IA, actualment aturada), mercats digitals –la Llei de serveis digitals (Digital Services Act) i la Llei de mercats digitals (Digital Markets Act)– i ciberseguretat –Llei de ciberresiliència (Cyber Resilience Act)–. Altres regulacions generals, com les vinculades amb la protecció dels consumidors, la competència en els mercats, la seguretat dels productes o la propietat intel·lectual, així com les pròpies dels sectors verticals (per exemple, de dispositius mèdics), també interactuen amb els sistemes que utilitzen IA. En els anys vinents, aquesta legislació es continuarà desenvolupant i desplegant, incloent-hi les mesures per al seguiment de la seva aplicació.

A escala més tècnica, una àrea de treball molt activa en els pròxims anys serà la plasmació dels requisits reguladors i els principis ètics, com la transparència, la responsabilitat i la justícia, en funcionalitats i dissenys tècnics que els desenvolupadors d'IA puguin implementar. Això implica activitats com la definició de mètriques i indicadors per avaluar el compliment d'aquests principis, el desenvolupament d'eines i metodologies per garantir que els sistemes de IA siguin ètics per disseny, la creació de registres d'auditoria, el desenvolupament de sistemes de detecció d'anomalies, de tècniques de privacitat per disseny o d'eines i metodologies que permetin als usuaris comprendre com els sistemes d'IA prenen decisions, així com identificar possibles biaixos o errors, per esmentar alguns exemples.

L'objectiu d'aquest capítol és, en primer lloc, descriure algunes consideracions en relació amb la regulació de la intel·ligència artificial, tenint en compte aspectes comuns a altres indústries (i, per tant, potencialment abordables amb aproximacions similars), aspectes tècnics específics de la intel·ligència artificial o aspectes derivats de la seva ràpida i massiva adopció. Posteriorment, es descriuran decisions reguladores rellevants a l'entorn internacional (desembre 2024).

La regulació de la intel·ligència artificial: reptes i perspectives

Regulació de la IA: consideracions generals

En les últimes dècades, el concepte d'intel·ligència artificial (IA) ha evolucionat significativament, passant de ser una noció exclusiva del vocabulari científic a referir-se a un ecosistema complex al voltant d'un sector industrial en creixement, capaç de penetrar en gairebé tots els àmbits de la vida quotidiana. En conseqüència, la indústria de la IA s'ha consolidat com un sector amb implicacions profundes en l'àmbit social, econòmic i polític.

La regulació pot ser un instrument actiu que promogui la innovació en una direcció que signifiqui veritable progrés, entenent «progrés» en tota la seva complexitat. Aquesta promoció de la innovació en un entorn democràtic ha de ser transparent i responsable, permetent la supervisió pública i la rendició de comptes, i garantint la seguretat i el benestar públic. Ha d'equilibrar el desenvolupament industrial amb el benestar social i planetari, de manera proporcional a l'impacte social, cultural, ambiental o geopolític de cada indústria. Aquest equilibri es persegueix mitjançant la definició, supervisió i revisió contínues de regles clares i adequades a les característiques de la indústria, de manera que promoguin el seu fi últim, que és fomentar els valors socials desitjats per la societat.

Les múltiples complexitats que implica la integració de la indústria de la IA en els nostres sistemes socials requereixen una complexificació intel·ligent equivalent per part dels sistemes de regulació i control (Innerarity, 2020), influïda per les propietats dels sistemes complexos (vegeu al quadre 1 algunes característiques que defineixen un sistema complex). La regulació de la IA dependrà no només de regles rígides i prescriptives, sinó també d'estratègies reguladores flexibles i adaptatives que reconeguin, per una banda, la diversitat de contextos i aplicacions de la IA i, per l'altra, la transversalitat dels seus impactes (vegeu el quadre 2 amb alguns exemples d'aproximacions regulatòries). (Coglianese & Crum, 2024.)

Quadre 1 - Sistemes complexos

Un sistema complex és un conjunt d'elements interconnectats que, a través de les seves interaccions, exhibeixen comportaments i propietats que no es poden deduir simplement de la suma de les seves parts individuals (a diferència d'un sistema exclusivament complicat). Exemples de sistemes complexes són el clima, el cervell humà, els mercats financers i, en general, els sistemes socials.

Els sistemes complexos es defineixen per propietats com:

– Interconnexió i interdependència: els elements del sistema s'influeixen mútuament, i generen efectes en cadena i retroalimentacions.
– Emergència: el sistema global exhibeix propietats no presents en els seus components individuals.
– Adaptabilitat i autoorganització: el sistema es modifica en resposta a canvis, mantenint l'estabilitat o evolucionant.
– No linealitat: petits canvis poden causar grans conseqüències, i dificultar la predicció del comportament.
– Dependència de la trajectòria: la història del sistema influeix en la seva evolució futura, i en fa difícil la reproducció.
– Diversitat: la varietat d'agents que interactuen augmenta la capacitat d'adaptació del sistema.

Quadre 2 - Exemples d'aproximacions regulatòries

En aquelles situacions en què l'aplicació sigui en un entorn on totes les variables són conegudes (o en aquelles parts d'un procés més ampli que són mesurables) i existeix una evidència científica sòlida, les aproximacions tradicionals estan basades en els estàndards de rendiment o la regulació de la divulgació de la informació. Exemples:

Estàndards de rendiment: establiment de normes i estàndards basats en els resultats o el rendiment dels productes i processos, en lloc de prescriure mètodes específics.

Exemple: límits per a la presència de patògens o contaminants en els aliments o límits al contingut de nutrients específics (sucres, greixos o sal, per exemple).

Regulació per divulgació d'informació: es basa en la idea que proporcionar informació clara i accessible als consumidors els permetrà prendre decisions informades i, per tant, regular indirectament el comportament de les empreses.

Exemple: l'obligació de proporcionar informació sobre el contingut de nutrients, com ara calories, greixos, sucres i sal, permet als consumidors comparar el valor nutricional dels aliments, o sistemes com Nutri-Score, per tal de facilitar la interpretació de la informació nutricional.

Aproximacions reguladores com la responsabilitat ***ex post*** i la regulació basada en la gestió, àmpliament presents en altres sectors, s'utilitzen en contextos específics per gestionar els riscos, fomentar que els actors siguin responsables i permetre una regulació més flexible i dinàmica, encara que pot resultar costosa i complexa d'implementar i, en el cas de la responsabilitat *ex post*, pot ser més lenta a l'hora d'actuar i generar confiança en els ciutadans, ja que els danys ja han estat causats. Exemples:

Responsabilitat *ex post*: se centra en la responsabilitat per les conseqüències negatives de l'ús d'un producte, en lloc de prevenir-les per endavant. És a dir, és l'obligació de respondre per les conseqüències d'una acció o decisió després que aquesta s'hagi produït.

Exemple: dret de la competència (per exemple, abús de posició dominant).

Regulació basada en la gestió: requereix que les organitzacions desenvolupin i implementin sistemes de gestió que garanteixin el compliment dels objectius reguladors i que demostrin la seva eficàcia a través d'auditories i altres mecanismes de control.

Exemple: sistemes de gestió de la seguretat i salut laboral, de la qualitat o de gestió ambiental.

Un dels majors reptes possiblement sigui la capacitat de les entitats reguladores per mantenir una vigilància constant i una acció activa. Això requereix recursos significatius, **proporcionals** a l'abast del seu impacte, incloent-hi capital financer, tecnològic i, especialment, humà, amb una diversitat d'expertesa essencial tant en els organismes reguladors com en la capacitat de col·laboració amb actors diversos. A més, cal establir mecanismes per integrar la vigilància activa i el *feedback* dels usuaris finals.

La narrativa es converteix en una eina clau per a la legitimació de la regulació, ja que té el poder de modelar el discurs públic i influir en les decisions polítiques. Les narratives poden orientar el desenvolupament tecnològic, i orientar la investigació i la innovació cap a determinades aplicacions i sectors. A més, les narratives predominants poden jugar un paper fonamental en la formació de la percepció pública sobre la IA. El risc inherent a les narratives dominants a l'entorn de la IA, sovint promogudes per la indústria, resideix en la seva capacitat per a prioritzar interessos particulars per sobre dels impactes socials, ambientals i culturals. Aquestes narratives solen centrar-se en aspectes com el progrés econòmic, la generació d'ocupació o el desenvolupament tecnològic, mentre minimitzen o invisibilitzen les conseqüències negatives associades a les seves activitats, com la normalització dels impactes ambientals (presentades com a costos inevitables del progrés), els impactes en les comunitats locals o la concentració del poder econòmic i polític. En promoure's com a essencials per al desenvolupament nacional o global, presentant-se com una activitat neutral, aquestes narratives poden afavorir monopolis i oligopolis i limitar la competència, així com la imposició de marcs tecnocràtics, mitjançant l'ús d'un llenguatge que emfatitza la complexitat tècnica de les seves operacions, posicionant-se com els únics actors capaços de gestionar aquests recursos. Això pot excloure altres actors, com les comunitats locals o experts independents, del procés de presa de decisions, o desplaçar el focus en les conseqüències a llarg termini (quadre 3).

En aquest sentit, la participació de la societat civil sembla una condició indispensable per a una regulació i supervisió de la IA que siguin efectives. Aquesta participació no es limita a les dimensions de repre-

Quadre 3 - Analogies i narratives

Les analogies ajuden a explicar allò desconegut comparant-ho amb un àmbit familiar que comparteix estructures comunes, i tenen el potencial de facilitar una comprensió més profunda de l'impacte de les TIC. Les analogies juguen un paper important en la governança de la tecnologia, ja que poden modelar la innovació, les polítiques i les perspectives acadèmiques a través de l'emmarcament dels problemes i la configuració de les percepcions sobre la seva viabilitat. Tot i això, les analogies poden ser tant útils com enganyoses, ja que els usuaris poden no identificar adequadament les connexions entre l'ús de les TIC i els seus impactes menys evidents, o bé poden arribar a conclusions errònies a partir d'aquestes comparacions. Per exemple, identificant les TIC amb una eina passiva, com un ganivet, que es pot utilitzar per al bé o per al mal (l'ús del qual, en qualsevol cas, està regulat en diferents sentits, com la impossibilitat de portar-lo a múltiples espais), pot ser una analogia limitada i que doni lloc a malinterpretacions o interpretacions interessades. Analogies més eficients amb altres entorns industrials amb implicacions individuals, socials, culturals i ambientals àmplies, com la indústria alimentària, es mostren més efectives. Aquestes analogies ajuden a visibilitzar la combinació de diferents enfocaments reguladors en entorns complexos, i proporcionen una millor comprensió dels riscos i les responsabilitats associades a les TIC (Ruiz-García & Hernández-Leo, 2025).

sentació i transparència de tot sistema democràtic, sinó que es manifesta com una font essencial de coneixement, experiències i perspectives, disponibles en l'àmplia gamma d'organitzacions, grups i individus que representen diferents interessos, valors i experiències. Aquesta diversitat és crucial per identificar i comprendre els impactes complexos i en constant evolució de la IA, i representa un dels grans reptes relacionats amb la governança de la IA en societats democràtiques com la nostra.

Particularitats de la indústria de la IA

La regulació de les indústries basades en el coneixement, com la farmacèutica o les indústries culturals, presenta desafiaments significatius en entorns democràtics a causa dels seus cicles d'innovació ràpida, la seva naturalesa tècnica complexa, la seva importància social i el seu abast global. Aquestes indústries es caracteritzen per **avenços ràpids** que sovint tensen els processos reguladors tradicionals. El **coneixement especialitzat** necessari per comprendre aquestes tecnologies afegeix una capa addicional de complexitat, ja que els reguladors poden no tenir l'experiència per avaluar els riscos i beneficis de manera efectiva. Però una confiança excessiva en l'experiència tècnica pot deixar de banda perspectives crucials de les ciències socials, les humanitats i la ciutadania, donant lloc a polítiques que no tenen en compte les complexitats de la societat. A més, **l'abast global** d'aquestes indústries significa que les regulacions d'un país poden no coincidir amb les d'altres i crear desafiaments per a l'aplicació i coordinació transfronterera. La naturalesa de la **propietat intel·lectual** d'aquestes indústries també complica la regulació, ja que les empreses poden ser reticents a compartir informació propietària essencial per avaluar el compliment de la regulació.

La indústria de la IA es caracteritza per un conjunt de tecnologies dinàmiques i adaptatives que evolucionen dins de sistemes sociotècnics més amplis, els quals es retroalimenten mútuament. Aquesta evolució constant, sovint imprevisible a curt termini, representa un desafiament significatiu per als marcs reguladors existents. La interconnexió de la indústria de la IA amb altres sistemes socials, econòmics i polítics accentua la complexitat de la seva regulació, ja que els canvis en un component poden generar efectes en cadena.

Aquesta complexitat es veu amplificada per **l'amplitud tecnològica** de la IA, que abasta una gran varietat de tecnologies amb diferents nivells de complexitat i risc. La interacció entre aquestes tecnologies i altres elements, incloses les institucions humanes, crea cadenes de valor altament complexes, fet que planteja el repte de decidir si la regulació ha de centrar-se en les tecnologies en si mateixes o en els processos que habiliten. Per exemple, la regulació europea de la IA (AI Act) exclou els sistemes basats en regles, tot i que aquests poden tenir un gran impacte social.

El model dominant d'IA actual es basa en l'extracció i explotació de grans quantitats de dades, juntament amb l'ús intensiu de grans quantitats de recursos computacionals. Aquesta dinàmica, combinada amb la gran concentració empresarial i les altes inversions realitzades sota les promeses generades en el sector, planteja **riscos de monopoli** amb implicacions econòmiques, socials i tecnològiques que van més enllà de l'àmbit purament tècnic.

La **multiplicitat d'usos** de la IA afegeix una capa addicional de complexitat. Una IA pot ser utilitzada per recomanar un restaurant o per determinar qui obté un lloc de treball, amb implicacions molt diferents en cada cas. Les seves aplicacions abasten una àmplia gamma de sectors, des de la salut fins a les finances i l'aplicació de la llei, cada una amb necessitats reguladores específiques i consideracions ètiques pròpies. En contextos ben definits, com la classificació de documents o l'optimització de cadenes logístiques, la IA pot assolir un alt grau d'objectivitat. Però quan s'aplica a sistemes complexos que impliquen éssers humans i sistemes socials, les decisions sobre què formalitzar estan subjectes a valors, judicis i interessos dels dissenyadors i usuaris de la tecnologia. La regulació ha de ser prou flexible per tenir en compte les particularitats de cada sector, ja que un enfocament únic podria passar per alt els riscos específics de cada àmbit. Així, la IA en el sector financer pot influir en el comportament del mercat, la qual cosa pot afectar, al seu torn, els algoritmes de negociació impulsats per IA, i crear mecanismes de retroalimentació que amplifiquen conseqüències no desitjades. Els reguladors han de dissenyar marcs de treball que permetin supervisar i ajustar aquests efectes en cadena.

La **diversitat d'escenaris** en què s'aplica la IA genera reptes addicionals, especialment en contextos sensibles

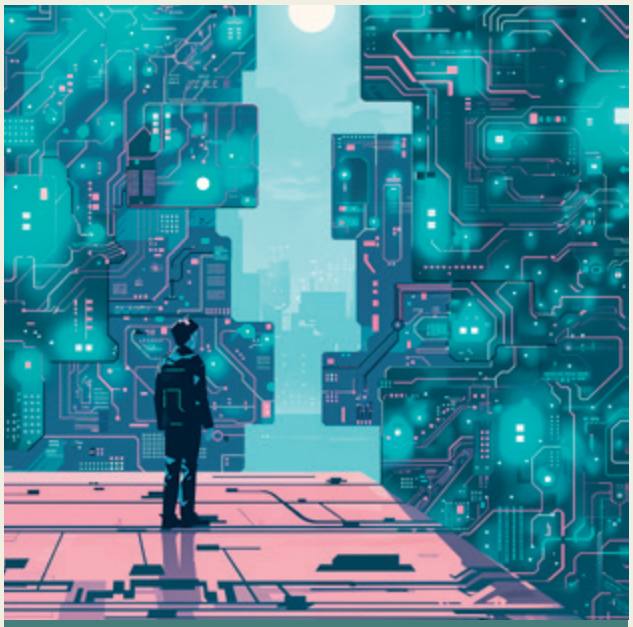
com la interacció amb menors o persones vulnerables. Els riscos associats a l'ús de la IA varien considerablement segons l'àrea d'ús, però també l'escenari específic d'aquest ús. En el cas de l'educació, per exemple, l'automatització de tasques rutinàries que no impliquin dades personals pot tenir un baix risc, mentre que l'ús d'IA en activitats que impliquin interacció amb menors o prediccions sobre comportaments futurs pot comportar riscos significatius. Això exigeix una regulació específica per a cada cas, tenint en compte els contextos i riscos particulars.

Un aspecte transversal a totes les aproximacions és abordar **l'opacitat** en la qual opera la indústria de la IA. L'opacitat en entorns tecnològics és gairebé una característica fundacional del desenvolupament informàtic, que va provocar la reacció dels moviments de codi obert en el passat. Aquesta opacitat, però, es va fent complexa amb l'anàlisi de dades massives i les aproximacions en intel·ligència artificial que se'n deriven. Si un agent (ciutadà o regulador, per exemple) actua en un context opac, no és capaç de jutjar factors com, per exemple, si les dades explotades suposen un risc posterior per a la seva privacitat, o si es faran servir en altres contextos diferents amb implicacions que no pot anticipar. Aquestes situacions impliquen el propi qüestionament del consentiment informat per a l'ús de les dades, ja que estar «informat» implicaria conèixer les implicacions d'aquest consentiment. Aquesta opacitat es manifesta de maneres semblants a altres indústries i, tradicionalment, s'aborda amb una combinació de més transparència en la interacció amb el producte (per exemple, informació alimentària) i controls quan les conseqüències poden ser greus (per exemple, seguretat alimentària). Aquest tipus d'opacitat tracta fonamentalment dels desequilibris de poder que es poden generar entre consumidors, productors i entitats reguladores. L'acció mateix de recollir dades, com a acte de poder amb possibles conseqüències, requereix, per tant, la promoció de la transparència en la seva recopilació i ús.

Una opacitat específica de la IA, especialment l'aprenentatge automàtic, és l'anomenada *de caixa negra*. La caixa negra descriu un procés en què la relació (estocàstica) entre l'entrada i la sortida no és clara, ni tan sols per als experts, la qual cosa implica la incapacitat de justificar la sortida del sistema. A més, a diferència d'altres camps científics i innovadors, el desenvolupament de la IA, quan aborda problemes relacionats amb el comportament humà, no es fonamenta en un coneixement «objectiu» de les característiques físiques o biològiques de l'entorn. En canvi, es basa en construccions socials, polítiques i econòmiques que determinen quins processos cal formalitzar i optimitzar, i que sovint impliquen una simplificació de processos complexos. Aquesta simplificació es deu tant a la dificultat de descriure completament aquests sistemes (per la incapacitat de traçar les operacions que donen lloc a un resultat concret) com als valors, judicis, motivacions o interessos de les organitzacions que els implementen (que poden orientar-se a diversos objectius entre el benefici comú i la maximització de guanys), les limitacions pràctiques en la recopilació i processament de dades o els prejudicis incrustats en les dades.

Els riscos associats a l'ús dels sistemes d'IA, per tant, no es limiten a característiques merament tècniques (com pot ser la seguretat o robustesa en altres àmbits, especialment lligats en el cas de la IA a la qualitat i representativitat de les dades amb les quals s'entrenen els models), sinó que inclouen riscos epistèmics, generats al llarg de diverses etapes del procés de disseny i ús de sistemes basats en IA. Aquestes etapes abasten la identificació del problema, i la selecció de quins aspectes s'han de modelar i quins s'ometen; la translació de conceptes a construccions operatives, que implica la traducció de conceptes abstractes a formats comprensibles per a una màquina, incloses les decisions relacionades amb les dades utilitzades (per exemple, legitimitat en relació amb la propietat intel·lectual, la seva adequació real al problema o la seva representativitat) i la lògica interna de les decisions; l'avaluació de resultats, inclosa la manera d'informar sobre la precisió, robustesa o fiabilitat de les respostes, les seves limitacions o les possibles conseqüències negatives (inclosa la seva consideració rigorosa durant tots els processos de desenvolupament i explotació).

Finalment, tot i que els algoritmes i les decisions que generen disposen d'una aura d'immaterialitat (reforçada pel desenvolupament de conceptes com «el núvol»), la seva adopció massiva durant els últims anys, i el seu esperat creixement en el futur, han fet aflorar en l'àmbit públic la dimensió material del seu disseny, producció i ús.



Quadre 4 - Cartografia de la IA

El llibre de Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (publicat en castellà per Edicions Ned), n'és un exemple paradigmàtic. Crawford aborda la suposada immaterialitat de la IA, i exposa els seus costos energètics, l'explotació laboral i la perpetuació de desigualtats socials. Aquesta obra ha influït en exposicions, com «Calculating Empires», que aprofundeixen en la relació entre la IA i les estructures de poder històriques.

Així mateix, el Taller Estampa, guardonat amb el Premi Ciutat de Barcelona de Cultura Digital 2024, ha desenvolupat la «Cartografia de la IA generativa». Aquest projecte és un exemple contemporani de com la cartografia de la IA pot adoptar formes innovadores i crítiques. Aquesta cartografia se centra en les IA generatives i en com aquestes repliquen i amplifiquen biaixos socials.

Quins elements abasta la indústria de la IA?

La «cartografia de la IA» és un camp emergent que busca visualitzar i analitzar les complexitats de la intel·ligència artificial, i revelar-ne les infraestructures, conjunts de dades, impactes socials i implicacions geopolítiques (quadre 4). Aquesta cartografia crítica va més enllà de la simple descripció tècnica, i es centra en les conseqüències materials i polítiques de la IA. Aquestes dimensions determinen la base de la seva governança, i es poden agrupar en:

Recursos naturals. El desenvolupament de la indústria de la IA i la seva infraestructura associada depenen críticament de l'extracció i l'ús de recursos naturals, que sovint es passen per alt. La fabricació de dispositius electrònics, com ara xips, sensors i servidors, requereix minerals rars i metalls preciosos, com ara liti, cobalt, terres rares i or. La fabricació de carcasses, cables i altres components requereix la producció de plàstics. La interconnexió global depèn de xarxes de cables submarins que travessen els oceans o xarxes satel·litàries. La seva fabricació i instal·lació requereixen grans quantitats de materials i energia. Els centres de dades consumeixen enormes quantitats d'energia per alimentar els servidors, els sistemes de refrigeració i altres equips, així com grans quantitats d'aigua per refrigerar els servidors. Tots aquests processos generen grans quantitats de residus electrònics, que contenen materials tòxics i contaminants. La seva reparació i el reciclatge són sovint inexistents, i el reciclatge de materials rars i complexos pot ser difícil i costós. Fins i tot la seva interacció amb l'escenari geopolític internacional és creixent, a causa de la necessitat de garantir el subministrament de minerals crítics (sovint, no abundants) requerits per a la fabricació d'infraestructures i dispositius (necessaris tant per a aquesta indústria com per a d'altres com les d'equipaments d'energia renovable, vehicles o armes).

Recursos humans. La indústria de la IA depèn d'un ampli espectre de recursos humans, que abasta des d'especialistes altament qualificats fins a treballadors poc qualificats que fan tasques essencials. Dintre dels especialistes, cal destacar que no només els perfils tècnics són necessaris per a la creació i el manteniment de sistemes d'IA, sinó també personal expert dels àmbits d'aplicació (medicina, educació, dret, etc.) que disposen del coneixement per determinar quins són els aspectes rellevants a considerar

per a la resolució dels problemes. A més, és crucial la participació d'experts amb coneixements transversals, que aborden aspectes com les consideracions legals i ètiques (per exemple, la privacitat dels usuaris o la discriminació algorítmica), la seguretat (per exemple, la protecció contra ciberatacs o la seguretat de les dades), la usabilitat (per exemple, el disseny d'interfícies intuïtives o l'experiència de l'usuari) o l'accessibilitat (per exemple, la garantia que els sistemes d'IA siguin utilitzables per persones amb discapacitat), per esmentar alguns d'aquests coneixements crítics. Els treballadors poc qualificats executen tasques clau com la recopilació i etiquetatge de dades per entrenar els models, així com de filtratge de resultats generats pels models o generats pels usuaris, les condicions precàries dels quals demanen una atenció creixent. Cal igualment destacar que el disseny de diferents tecnologies considerades sota el concepte de la IA es basa en les contribucions no remunerades, i sovint desconegudes, fruit de l'extracció de dades dels usuaris que interactuen amb els sistemes (per exemple, la informació compartida amb plataformes). Les dades extretes, classificades i avaluades a partir d'aquestes interaccions converteixen els usuaris en els proveïdors de les dades necessàries per a l'entrenament dels sistemes, amb poc control sobre el seu ús i possibles conseqüències posteriors (potencialment desconegudes fins i tot per a les entitats que realitzen aquest processament). Encara que d'una naturalesa diferent, l'entrenament dels models que es basa en productes culturals com àudio, vídeo i text s'aprofita críticament del treball creatiu de persones, pel qual no reben reconeixement ni compensació de manera general.

Impactes individuals i socials. La indústria de la IA no només transforma tecnologies, sinó que també té un impacte profund en les persones i la societat en general. Aquest impacte multidimensional exigeix una reflexió crítica i una acció col·lectiva per tal de maximitzar les seves possibilitats, i forma part indispensable de la comprensió dels fenòmens que interaccionen amb la indústria de la IA.

Cada interacció digital deixa una empremta. La recopilació massiva de dades, l'intercanvi i el processament per entrenar models pot proporcionar coneixements valuosos per a la presa de decisions en diversos sectors, però al mateix temps generen preocupacions sobre la privacitat i la seguretat de les dades personals, així com sobre la capacitat d'extreure patrons ocults i revelar informació sensible, amb implicacions per a la vigilància i el control.

La nostra creixent dependència de la IA també influeix en la manera en què pensem, actuem i ens relacionem entre nosaltres. Els algoritmes que curen la informació modelen la nostra visió del món, i sovint reforcen les creences existents o redueixen l'abast al qual estem exposats. Per exemple, les plataformes de xarxes socials utilitzen algoritmes per personalitzar el contingut que veiem, i creen un «filtre bombolla» que pot afavorir la informació amb la qual ja estem d'acord i limitar, així, la nostra exposició a idees contràries. L'aplicació de la IA pot facilitar la difusió de desinformació i la manipulació de l'opinió pública, amb implicacions per a la democràcia, com la pèrdua de confiança en les institucions i els mitjans de comunicació, la polarització de la societat o la manipulació de la participació ciutadana. L'accés desigual a la tecnologia i les habilitats digitals crea una escletxa digital que pot exacerbar les desigualtats socioeconòmiques i culturals, o crear-ne de noves.

La comoditat que ofereix la IA en diversos àmbits pot millorar la nostra capacitat de presa de decisions, però també pot canviar la nostra manera d'abordar la resolució de problemes, de vegades reduint el pensament crític a favor de solucions automatitzades, per exemple, en l'ús de sistemes d'IA en processos de presa de decisions socials, en què les persones responsables poden confiar excessivament en les recomanacions de l'algoritme, fins al punt de no qüestionar les seves decisions, tot i que la seva pròpia experiència suggereixi una opció diferent. La interacció amb sistemes d'IA pot tenir efectes psicològics i socials en els individus, com ara la dependència, l'aïllament o la desinformació, l'afectació en el benestar individual i la cohesió social i l'alteració de les dinàmiques socials, les relacions interpersonals i la formació de comunitats. La incorporació de la indústria de la IA en àrees d'alta rellevància social, com la justícia, l'educació o la sanitat, planteja qüestions de transparència, responsabilitat i equitat.

L'automatització impulsada per la indústria de la IA pot contribuir a un augment de l'eficiència i la productivitat, per mitjà de l'automatització de tasques repetitives o l'optimització de processos, però també provocar la desaparició de llocs de treball i la reestructuració de les relacions laborals. A mesura que les eines d'IA es converteixen en part integrant de les indústries i organitzacions, poden canviar les dinàmiques de poder, i afectar l'ocupació, les estructures socials i l'accés als recursos. Per exemple, en

Quadre 5 – Dinàmiques sectorials en l'evolució regulatòria: canvis en les capacitats, actors i estructures

Els impactes generats en diversos sectors o entorns poden provocar-hi canvis en les dinàmiques, que es manifesten de diverses maneres, com ara (Maas, 2021):

– Facilitar comportaments antics ja regulats amb noves eines no cobertes per regulacions prèvies, o provocar canvis absoluts o relatius en la capacitat d'actuar i obrir espais d'acció que estaven fora de l'abast de molts actors, la qual cosa comporta la necessitat de revisar les normes existents (incloses la capacitat i l'agilitat en la supervisió) per evitar buits o contradiccions regulatòries en marcs legals ja establerts. Per exemple, l'ús del reconeixement facial en seguretat, la consideració dels biaixos algorítmics en preses de decisions automatitzades, la manipulació informativa en línia, la manipulació de la propietat intel·lectual, la difamació i la protecció de la imatge o les suplantacions d'identitat.
– Alterar les posicions relatives dels actors, i modificar qui ocupa els papers dominants, tot i que les regles generals poden romandre pràcticament inalterades. Per exemple, en entorns rellevants com l'educació, la sanitat o la seguretat, poden generar dependències d'empreses privades en serveis essencials tradicionalment coberts amb infraestructures públiques, o, en el cas de països sencers, si no es disposa de capacitat o serveis propis.
– Modificar les dinàmiques estructurals en les societats internacionals i canviar la influència dels actors i els mitjans a través dels quals exerceixen la seva influència (per exemple, la transició de la força militar a la propaganda computacional) i, fins i tot, les normes i identitats dels actors, i alterar-ne, així, els objectius i els comportaments en el context global.

el sector de la fabricació, pot portar a la pèrdua de llocs de treball en determinades indústries mentre genera una concentració de poder en les empreses tecnològiques que desenvolupen aquestes eines. La concentració de dades i tecnologia en mans d'unes poques empreses (basades, principalment, en els Estats Units) pot generar asimetries de poder, limitar la competència i empobrir l'entorn industrial local si no es disposa de la capacitat de desenvolupar un ecosistema robust i innovador.

Referències bibliogràfiques

Coglianese, C.; Crum, C. R. (2024). «Regulating Multifunctionality» (SSRN Scholarly Paper 5059426). Social Science Research Network. https://doi.org/10.2139/ssrn.5059426

Innerarity, D. (2020). *Una teoría de la democracia compleja*. Galaxia Gutenberg. http://www.galaxiagutenberg.com/libros/una-teoria-de-la-democracia-compleja/

Maas, M. M. (2021). «Aligning AI Regulation to Sociotechnical Change» (SSRN Scholarly Paper ID 3871635). Social Science Research Network. https://doi.org/10.2139/ssrn.3871635

Ruiz-García, A.; Hernández-Leo, D. (2025). «Food for thought: an analogy for digital sovereignty in education». 19th International Conference of the Learning Sciences (ICLS) 2025-ICLS Proceedings. Learning, 2025 (Accepted).

Marc normatiu internacional i de la Unió Europea

Pel que fa al marc regulador actual, a escala internacional trobem diverses iniciatives tant per part de l'Organització de Cooperació i Desenvolupament Econòmics (OCDE) com per part del Consell d'Europa i de la Unió Europea per adreçar o regular la intel·ligència artificial (IA). A continuació es presentaran les característiques principals dels instruments més importants adoptats per part d'aquestes institucions internacionals.

OCDE

Els principis de la IA de l'OCDE, adoptats el 2019 i actualitzats el 2024, són la primera norma intergovernamental per regular la intel·ligència artificial. Aquests principis volen promoure una IA innovadora i fiable que respecti els drets humans i els valors democràtics.

Es componen de cinc principis i cinc recomanacions que tenen per objectiu orientar, de forma pràctica i flexible, els responsables polítics i els diferents actors que formen part de l'ecosistema en el disseny, desenvolupament, comercialització i implementació de sistemes d'IA.

Els principis d'IA de l'OCDE són: primer, assegurar el creixement inclusiu, el desenvolupament sostenible i la generació de benestar; segon, respectar els drets humans i valors democràtics, incloses l'equitat i la privadesa; tercer, assegurar la transparència i l'explicabilitat; quart, la solidesa i la seguretat dels sistemes d'IA, i, finalment, la responsabilitat pels danys que aquests sistemes puguin causar.

D'altra banda, les recomanacions que proposa l'OCDE en matèria d'IA s'articulen en, primer, invertir en recerca i desenvolupament en IA; segon, fomentar un ecosistema inclusiu que permeti el desenvolupament de sistemes d'IA; tercer, donar forma a un entorn polític i de govern interoperable i propici per a la creació i el desenvolupament de sistemes d'IA; quart, construir el capital humà i preparar-se per a la transformació del mercat laboral, i, finalment, la cooperació internacional per a una IA fiable.

Consell d'Europa

La preocupació i l'activitat normativa del Consell d'Europa amb relació a l'impacte de la intel·ligència artificial en les nostres societats i la urgència a assegurar que aquesta se centri en els humans i sigui respectuosa amb els drets fonamentals venen de lluny.

L'any 2010, el Comitè de Ministres va adoptar una recomanació sobre la protecció de les persones físiques pel que fa al tractament automàtic de dades personals en el context de l'elaboració de perfils en referència a les dades personals que es tracten mitjançant el càlcul, la comparació i la correlació estadística, que abordava la necessitat d'una salvaguarda addicional per protegir les dades personals i la privacitat.

D'altra banda, l'Assemblea Parlamentària del Consell d'Europa (PACE) va adoptar una Resolució sobre la vigilància massiva el 2015 que abordava les preocupacions generades per la recollida de grans quantitats de dades personals per part de les agències d'intel·ligència i l'ús de dades personals per a fins il·lícits per part d'actors estatals o no estatals. La PACE va abordar la necessitat d'una regulació legal adequada i la necessitat d'una supervisió efectiva.

Posteriorment, l'any 2017 la PACE va publicar una Recomanació sobre convergència tecnològica, IA i drets humans que destacava la necessitat d'enfortir la regulació existent i abordar la necessitat de governança d'internet que no depengués de grups privats o estats. Es va crear un Comitè d'Experts en Intermediaris d'Internet (MSI-NET) amb un mandat de dos anys (2016-2017) sota el Comitè Directiu dels Mitjans de Comunicació i la Societat de la Informació (CDMSI). Va elaborar un estudi titulat «Algorismes i drets humans» sobre «les dimensions dels drets humans de les tècniques automatitzades de tractament de dades i les possibles implicacions reguladores». Aquest és un dels primers documents del Consell d'Europa per abordar de manera integral les conseqüències de l'ús generalitzat dels algoritmes. El text exposa diversos principis i accions, entre els quals la transparència, la rendició de comptes, els marcs ètics i una millor avaluació de riscos, pel que fa als possibles impactes dels algoritmes sobre els drets humans.

L'any 2019 el Consell de Ministres va adoptar una Declaració sobre les capacitats manipuladores dels processos

algorítmics que advertia del risc d'utilitzar processos algorítmics per manipular el comportament social i polític dels individus i va subratllar la necessitat de marcs de protecció addicionals. Aquesta declaració inclou àrees clau d'acció com l'avaluació de l'impacte dels drets humans, les consultes públiques, la protecció de dades i la privacitat, i la supervisió independent.

El mateix any, el comissari de Drets Humans del Consell d'Europa va emetre una Recomanació, anomenada *Unboxing Artificial Intelligence: 10 steps to protect human rights*, que enumerava una sèrie de mesures que han de prendre les autoritats nacionals per evitar l'impacte negatiu dels sistemes d'IA en les persones, i maximitzar-ne també el potencial.

Aquesta activitat política i legislativa ha culminat amb el Conveni marc del Consell d'Europa sobre intel·ligència artificial, drets humans, democràcia i estat de dret, que representa el primer instrument de dret internacional imperatiu en matèria d'intel·ligència artificial. La Comissió Europea va signar el Conveni el 5 de setembre de 2024 en nom de la Unió Europea.

El Conveni marc té per objectiu assegurar que les activitats que tenen lloc al llarg del cicle de vida dels sistemes d'intel·ligència artificial respecten els drets humans, la democràcia i l'estat de dret mentre s'asseguren i es possibilita la innovació i el progrés tecnològic.

El Conveni marc va optar per la definició d'intel·ligència artificial proporcionada per l'OCDE segons la qual un sistema d'intel·ligència artificial és «un sistema basat en una màquina que, per a objectius explícits o implícits, dedueix, de l'entrada que rep, com generar sortides com prediccions, continguts, recomanacions o decisions que poden influir en entorns físics o virtuals».

La definició captura la naturalesa dinàmica i multidisciplinària dels sistemes d'intel·ligència artificial i té en compte les diferents perspectives en la governança de la intel·ligència artificial a escala global. Aquesta definició d'intel·ligència artificial és la que finalment s'ha adoptat al Reglament d'intel·ligència artificial recentment adoptat a la Unió Europea i que es presentarà a continuació.

El Conveni marc introdueix un marc legal que s'aplica en el transcurs del disseny, desenvolupament i aplicació de sistemes d'IA al llarg del seu cicle de vida i que cobreix els sistemes d'IA per part de les autoritats públiques –inclosos actors privats que actuen en nom seu– i actors privats. El Conveni marc ofereix als actors públics dues alternatives per complir els seus principis i obligacions a l'hora de regular el sector privat: poden optar per estar directament obligades per les disposicions del Conveni pertinent o, com a alternativa, prendre altres mesures per complir les disposicions del tractat respectant plenament les seves obligacions internacionals en matèria de drets humans, democràcia i estat de dret.

El Conveni marc no s'aplica a les qüestions de defensa nacional ni a les activitats de recerca i desenvolupament, excepte quan les proves dels sistemes d'IA poden interferir amb els drets humans, la democràcia o l'estat de dret.

Els principis inclosos al Conveni marc en relació amb el desenvolupament i l'ús de sistemes d'IA són: la dignitat humana i l'autonomia individual, la igualtat i la no-discriminació, el respecte a la privadesa i la protecció de dades personals, la transparència i la supervisió, la rendició de comptes i la responsabilitat, la fiabilitat i la innovació segura. El Conveni marc està centrat en l'ésser humà i adopta un enfocament basat en el risc per al disseny, desenvolupament i ús de sistemes d'IA que garanteix la prevenció dels usos nocius dels sistemes d'IA i promou l'ús d'aquesta tecnologia digital per al bé de la societat, inclòs mitjançant la permissió d'una innovació segura.

Pel que fa al marc de gestió de riscos i impactes, s'afirma que cada part ha d'adoptar mesures legislatives i altres per a la identificació, avaluació, prevenció i mitigació de riscos i impactes per garantir que els sistemes d'IA respectin els drets humans, el funcionament de la democràcia i l'estat de dret. El Conveni marc també exigeix que cada agent estableixi mecanismes de supervisió efectius per controlar i avaluar el compliment de les obligacions del seu ordenament jurídic intern.

El Conveni marc estableix un mecanisme de seguiment per determinar fins a quin punt s'estan implementant les seves disposicions. També serveix com a fòrum de cooperació internacional per facilitar la informació sobre diversos aspectes de la intel·ligència artificial.

A escala sectorial, la Comissió Europea per a l'Eficiència de la Justícia (CEPEJ) del Consell d'Europa ha pres iniciatives relatives a l'ús de sistemes d'intel·ligència artificial en l'àmbit judicial per la seva especial importància

en el respecte als drets fonamentals, inclús o sobretot quan s'utilitzen sistemes automatitzats o d'intel·ligència artificial.

El desembre de 2018 el CEPEJ va adoptar la Carta ètica de l'ús de la intel·ligència artificial als sistemes judicials i els seu entorn. Aquesta Carta ètica és el primer text europeu que estableix els principis ètics relacionats amb l'ús de la intel·ligència artificial (IA) en els sistemes judicials.

La Carta proporciona un marc de principis que poden guiar els responsables polítics, legisladors i professionals de la justícia quan s'enfronten al ràpid desenvolupament de la IA en els processos judicials nacionals.

El punt de vista del CEPEJ, tal com s'estableix en la Carta, és que l'aplicació de la intel·ligència artificial en l'àmbit de la justícia pot contribuir a millorar l'eficiència i la qualitat i ha d'aplicar-se de manera responsable i en compliment dels drets fonamentals garantits en particular pel Conveni europeu de drets humans (CEDH) i el Conveni del Consell d'Europa sobre la protecció de dades personals. Per al CEPEJ, és essencial assegurar que la IA continuï sent una eina al servei de l'interès general i que el seu ús respecti els drets individuals.

El CEPEJ ha identificat els següents principis bàsics a respectar en el camp de la IA i la justícia: primer, el principi de respecte dels drets fonamentals, i assegurar que el disseny i la implementació d'eines i serveis d'intel·ligència artificial siguin compatibles amb els drets fonamentals; segon, el principi de no-discriminació, i evitar específicament el desenvolupament o la intensificació de qualsevol discriminació entre individus o grups d'individus; tercer, el principi de qualitat i seguretat, pel que fa al tractament de decisions i dades judicials, utilitzant fonts certificades i dades intangibles amb models concebuts de manera multidisciplinària, en un entorn tecnològic segur; quart, el principi de transparència, imparcialitat i equitat, i fer accessibles i comprensibles els mètodes de tractament de dades, autoritzant auditories externes, i cinquè –el principi «sota control d'usuari»–, excloure un enfocament prescriptiu i garantir que els usuaris siguin actors informats i controlin les seves opcions.

Per al CEPEJ, el compliment d'aquests principis s'ha de garantir en el tractament de les decisions judicials i de les dades per algoritmes i en l'ús que se'n faci.

Unió Europea

La ràpida evolució de la tecnologia i el creixement exponencial de la intel·ligència artificial han fet que la Unió Europea s'hagi preocupat i ocupat de la digitalització del continent i de les normes jurídiques que es vol que assegurin un equilibri entre el desenvolupament industrial i el respecte dels drets fonamentals. Es vol trobar un equilibri entre el potencial econòmic generat per la digitalització i la intel·ligència artificial i la necessitat d'assegurar que aquestes es desenvolupin de forma segura i respectant els principis i drets fonamentals de la societat europea.

L'estratègia europea de digitalització i desenvolupament de la intel·ligència artificial es fonamenta en el desenvolupament i utilització de la intel·ligència artificial segura des del punt de vista tecnològic i centrada en els éssers humans.

L'any 2021 la Comissió Europea va presentar el seu paquet relatiu a la intel·ligència artificial. Aquest paquet inclou una comunicació relativa a com es vol que sigui el desenvolupament i la utilització de la intel·ligència artificial a Europa, la presentació d'un Pla coordinat entre els estats membres i les autoritats comunitàries en matèria d'intel·ligència artificial, i la creació d'un marc normatiu de la intel·ligència artificial i dels riscos que aquesta presenta.

A escala estratègica, aquests últims cinc anys la Unió Europea ha presentat les bases estratègiques per a com es veu el rol de la intel·ligència artificial a Europa i, sobretot, com es vol que aquesta tecnologia jugui un paper a les societats europees.

En l'àmbit del dret positiu, la norma més important que s'ha adoptat és el Reglament 2024/1689 conegut com el Reglament d'intel·ligència artificial. Aquest reglament és el resultat d'una proposta legislativa de la Comissió Europea de l'any 2021. D'alguna manera buscava el mateix efecte extraterritorial –l'efecte Brussel·les– que va tenir lloc posteriorment a l'adopció el Reglament de protecció de dades. En aquest sentit, el Reglament té efectes extraterritorials, donat que gran part de les seves previsions s'apliquen independentment de si els productors de sistemes d'intel·ligència artificial es troben dintre de l'àmbit de la Unió Europea o en països tercers.

Després de negociacions i diverses esmenes introduïdes a les diferents institucions que formen part del procedi-

ment legislatiu a escala europea –sobretot per la seva rellevància les introduïdes pel Parlament Europeu el passat mes de juny del 2023–, l'any 2024 es va adoptar el text final del Reglament d'intel·ligència artificial.

El Reglament d'intel·ligència artificial és un text ambiciós que té per objectiu proporcionar un marc normatiu per al desenvolupament, comercialització i utilització d'eines d'intel·ligència artificial en el mercat europeu.

S'aplica a qualsevol proveïdor o entitat responsable d'un sistema d'intel·ligència artificial si aquest està pensat per ser utilitzat dins la Unió Europea.

El tret definitori del Reglament europeu d'intel·ligència artificial és la seva perspectiva neutral des d'un punt de vista tecnològic i la seva estructura normativa sobre la base de la classificació dels riscos que presenten els diferents sistemes d'intel·ligència artificial.

El Reglament preveu quatre nivells de risc:

1. **Risc inacceptable.** Són sistemes que presenten uns riscos que s'entenen inacceptables i per tant són sistemes que estan prohibits en l'àmbit de la Unió Europea. Exemples d'aquests sistemes són sistemes que poden manipular les decisions dels individus, sistemes que poguessin crear puntuacions per a ciutadans o sistemes que poguessin explotar les vulnerabilitats dels individus sobre la base de les seves característiques personals.
2. **Risc elevat** (Annex III del Reglament). Aquests són sistemes que estaran subjectes a controls rigorosos. Els sistemes que s'entenen que presenten un risc elevat són aquells que poden afectar de forma directa els drets fonamentals dels ciutadans. Aquests sistemes es podrien utilitzar en sectors sensibles com als sistemes de salut, l'educació, l'Administració pública o els sistemes judicials.
3. **Risc limitat.** Aquests sistemes, no obstant el seu risc limitat, estan subjectes a obligacions de transparència.
4. **Mínim risc o risc zero**. Aquests sistemes no tenen cap tipus de restricció a l'aplicació del Reglament europeu.

De manera més específica val la pena destacar:

— La proposta de directiva de responsabilitat extracontractual de la intel·ligència artificial
— La proposta de directiva per adaptar la responsabilitat civil extracontractual a la intel·ligència artificial

Estats i regions dins de la Unió Europea

A escala europea, els estats membres estan adoptant mesures tant per assegurar el compliment del Reglament europeu d'intel·ligència artificial com per facilitar el desenvolupament d'aquests sistemes, crear mecanismes per supervisar-los i tenir mecanismes per eventualment sancionar els danys i incompliments que es poguessin derivar de la utilització de sistemes d'intel·ligència artificial.

Espanya

En l'àmbit espanyol, aquests últims anys l'estat ha adoptat normes i ha creat organismes per donar compliment al Reglament europeu d'intel·ligència artificial. L'any 2023 va ser un any especialment important normativament parlant, donat que es van adoptar tres iniciatives que val la pena destacar: en primer lloc, la creació de l'Agència Espanyola de Supervisió de la Intel·ligència Artificial; en segon lloc, l'adopció de la normativa per donar compliment a obligacions previstes al Reglament europeu d'intel·ligència artificial, i, finalment, l'adopció de normativa relativa a la possible aplicació de sistemes intel·ligents artificials en procediments judicials.

La creació de l'Agència Espanyola de Supervisió de la Intel·ligència Artificial (AESIA)

L'any 2023 es va aprovar la creació de l'Agència Espanyola de Supervisió de la Intel·ligència Artificial amb seu a la Corunya. Aquesta agència té per objecte donar compliment a l'obligació que preveu el Reglament europeu d'intel·ligència artificial amb relació a, primer, la designació d'una autoritat nacional de supervisió que supervisi i executi el que estableix el Reglament; segon, coordinar les activitats i obligacions dels estats membres; tercer, ser un punt de contacte amb la Comissió Europea, i, finalment, actuar com a representant de l'Estat a l'Oficina Europea d'Intel·ligència Artificial.

L'adopció de normes jurídiques per donar compliment a les obligacions previstes pel Reglament europeu d'intel·ligència artificial

El 2023 es va adoptar el Reial decret 817/2023, de 8 de novembre, que estableix un entorn controlat de proves per a l'assaig del compliment de la proposta de Reglament del Parlament Europeu i del Consell pel qual s'esta-

bleixen normes harmonitzades en matèria d'intel·ligència artificial.

Aquest Reial decret fa referència als sistemes qualificats del risc del Reglament europeu d'intel·ligència artificial de poder afectar la salut, seguretat o drets fonamentals de les persones.

Implementació de sistemes d'intel·ligència artificial en l'àmbit judicial

La justícia és un dels àmbits que requereixen digitalització, modernització i l'adopció de sistemes que puguin agilitzar determinades tasques processals que permetin als diferents professionals de l'Administració de justícia i del poder judicial dedicar el seu temps a tasques amb valor afegit i no subjectes a automatització.

Una part important de l'agenda del govern espanyol en matèria de justícia se centra en millorar la seva rapidesa i fiabilitat i reduir el col·lapse judicial que experimenten avui molts jutjats espanyols. Per tal d'assolir aquest objectiu, l'any 2022 es va presentar un projecte de llei de mesures d'eficiència digital del servei públic de justícia. A dia d'avui, però, aquesta llei no s'ha aprovat i, per tant, no està en vigor. No obstant això, l'any 2023 es va adoptar el Reial decret llei 6/2023, de 19 de desembre, pel qual s'aproven mesures urgents per a l'execució del Pla de recuperació, transformació i resiliència en matèria de servei públic de justícia, funció pública, règim local i mecenatge, conegut com el Reial decret d'eficiència processal, per crear les bases per enfocar l'Administració de justícia a generar dades que eventualment poguessin ser utilitzades per entrenar i/o implementar sistemes intel·ligència artificial en alguns procediments judicials.

El decret vol agilitzar els procediments judicials, reenfocar la visió de la justícia i proporcionar un marc jurídic que permeti gestionar les dades generades pel poder judicial. L'impacte de la tecnologia en l'àmbit de la justícia no es limita a la implementació de sistemes automatitzats o d'intel·ligència artificial en procediments judicials, sinó que també té a veure amb la transformació dels jutjats en sistemes generadors de dades que permetin conèixer empíricament la realitat de la justícia i a aprendre, a través de sistemes de *big data* i d'intel·ligència artificial, quins són els casos tipus, les accions judicials més freqüentment interposades o les respostes judicials a les demandes judicials.

El Reial decret llei 6/2023 no preveu explícitament la utilització de sistemes intel·ligència artificial però preveu les modificacions processals que eventualment s'haurien de dur a terme per tal de poder utilitzar les dades judicials per entrenar i implementar sistemes d'intel·ligència artificial.

França

França, per una república digital (*France, Loi pour une république numérique*) (https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746)

La llei francesa per una república digital, aprovada el 7 d'octubre de 2016, té un àmbit d'aplicació general i intersectorial que comprèn diferents sectors de la indústria digital, diferents però relacionats com són el dret a la privacitat, la protecció de dades i la protecció dels consumidors.

La llei imposa l'obligació a les plataformes digitals intermediàries a proporcionar «una [informació] justa, clara, i transparent». D'altra banda, la llei va posar a disposició de forma gratuïta i lliure totes les sentències judicials.

També, al seu article 2, aquesta llei proporciona el dret a accedir a les regles que defineixen i dissenyen els algoritmes que tracten dades personals i que eventualment puguin ser utilitzats per les administracions públiques. D'altra banda, la llei també atorga el dret a obtenir informació sobre l'aplicació i utilització d'algoritmes quan aquests duguin a terme un tractament de dades personals que puguin resultar en decisions que afectin els ciutadans de forma individual.

Llei relativa a l'organització i transformació del sistema sanitari (*Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé*) (https://www.legifrance.gouv.fr/loda/id/JORFTEXT000038821260/?isSuggest=true)

La llei relativa a l'organització i transformació del sistema sanitari es va aprovar el juliol de 2019 i té per objectiu incentivar el desenvolupament de sistemes d'intel·ligència artificial en l'àmbit de la salut. Aquesta llei té un àmbit d'aplicació que inclou tant administracions públiques, i per tant hospitals públics, com privats, així com persones jurídiques.

La llei se centra en el dret a la privacitat, la protecció de dades, el dret a la salut i el dret al consentiment informat tenint present, com a principi rector, el dret a la no-discriminació.

La llei presenta tres novetats fonamentals: en primer lloc, la llei preveu la integració completa de les dades clíniques de salut dels ciutadans en un Sistema Nacional de Dades Sanitàries (SNDS). Aquestes dades sanitàries comprenen les dades d'assegurances mèdiques, hospitalàries, causes mèdiques de defunció, dades relatives a una eventual discapacitat, compensacions per incapacitat i qualsevol dada d'entitats asseguradores de salut. Aquesta previsió amplia la base de dades de l'SNDS, bastant àmplia en aquests moments, donat que inclou dades de tractaments mèdics quan els tràmits són reemborsats per la seguretat social, dades relatives a la pèrdua d'autonomia, dades personals proporcionades per les enquestes de salut, dades recollides durant les revisions i controls mèdics obligatoris, dades recollides pels serveis de protecció maternoinfantil i dades de salut recollides durant les visites d'informació i prevenció de medicina del treball.

En segon lloc, hi ha hagut una ampliació dels objectius i finalitats de l'SNDS. Aquesta llei preveu que el sistema nacional de salut, a més de prevenir, tractar i gestionar la salut dels ciutadans, s'ocupi també de la recerca com estudis, avaluació i innovació en els àmbits de la salut. (Codi de salut pública, CSP, art. L. 1461-1, III.)

Les dades de salut de l'SNDS, no obstant això, no seran d'accés lliure, donat que la llei preveu que:

«Només es pot autoritzar l'accés a les dades personals de l'SNDS per al seu tractament, que o bé contribueixi a una d'aquestes finalitats i sigui d'interès públic o sigui necessari per a l'exercici de les tasques dels corresponents serveis de l'estat, establiments públics o organismes encarregats d'una missió de servei públic, en les condicions que s'estableixin per decret.»

Finalment, la llei preveu una simplificació del procediment d'accés a les dades de l'SNDS per tal d'afavorir incentivar el desenvolupament de sistemes d'intel·ligència artificial en l'àmbit de la salut. L'Institut Nacional de Dades Sanitàries ha estat substituït per una plataforma de salut –el Health Data Hub– que és una plataforma de naturalesa pública formada per representants de l'estat, de pacients i usuaris del sistema sanitari, de productors de dades de salut tant públics com privats i finalment per usuaris de dades de salut.

La missió del Health Data Hub és molt àmplia, donat que, a més de recollir, organitzar i posar a disposició les dades de l'SNDS i informar els pacients i promoure els seus drets, el Health Data Hub podrà actuar com a subcontractista d'un responsable del tractament (CSP, art. L. 1462-1, 6°).

Llei relativa a la bioètica (*LOI n° 2021-1017 du 2 août 2021 relative à la bioéthique*) (https://www.legifrance.gouv.fr/loda/id/LEGIARTI000043886059/2021-08-04/)

La Llei relativa a la bioètica, aprovada l'any 2021, s'aplica tant a autoritats públiques com a agents privats –tant individus com persones jurídiques– en l'àmbit del disseny i desenvolupament de sistemes d'intel·ligència artificial –especialment en sistemes que utilitzin l'aprenentatge automàtic (*machine learning)* o l'aprenentatge profund (*deep learning)*– en l'àmbit de la salut.

La llei té per objectiu garantir els drets de privacitat dels usuaris, així com garantir els drets a la informació i de transparència en l'àmbit sanitari. El punt de partida d'aquests drets es troba en el «tractament algorítmic de dades, l'aprenentatge del qual s'ha dut a terme a partir de dades massives» quan aquestes hagin estat utilitzades per un dispositiu mèdic per a la prevenció, diagnòstic o tractament de malalties. En aquests casos, la llei imposa l'obligació d'informar la persona interessada.

Aquesta informació s'haurà de proporcionar de manera que l'usuari la pugui entendre. D'altra banda, els pacients tindran dret a accedir a les dades que s'hagin utilitzat i tractat i als resultats que s'hagin pogut obtenir, així com a conèixer el disseny de l'algoritme que eventualment les ha tractat. Aquesta informació s'haurà de facilitar de manera que l'usuari la pugui entendre.

Portugal

A Portugal es va aprovar la Carta portuguesa de drets humans en l'era digital el mes de maig de 2021 (*Carta Portuguesa de Direitos Humanos na Era Digital - Lei n.º 27/2021, de 17 de maio*). (https://diariodarepublica.pt/dr/legislacao-consolidada/lei/2021-164870244?_ts=1709227594349)

Aquesta Carta té la vocació de garantir la protecció dels drets fonamentals a qualsevol àmbit en què s'utilitzin sistemes d'intel·ligència artificial. Així doncs, la Carta s'aplica en la utilització de la intel·ligència artificial per part d'administracions públiques, sistemes sanitaris, mercats digitals, protecció de dades i en qualsevol combinació d'estructures que impliquin diferents sectors.

La Carta consta de 21 articles que garanteixen drets i llibertats i proporcionen garanties per als ciutadans en entorns digitals, tals com, per exemple:

a. L'accés gratuït a internet. La llei preveu que tothom tingui dret a accedir gratuïtament a internet independentment de la seva ascendència, gènere, raça, llengua o religió.
b. La igualtat de gènere i habilitats digitals. L'Estat ha de promoure programes que assegurin la igualtat de gènere en entorns digitals així com de competències digitals per a tots els ciutadans.
c. La connectivitat. L'Estat ha de garantir una bona connectivitat a tot el territori de Portugal, independentment de la densitat de població que tingui el territori.
d. Tarifa social per a l'accés a internet. La llei preveu la possible creació d'una tarifa social per als clients econòmicament vulnerables.
e. La necessitat de combatre la desinformació i protegir els consumidors. La llei preveu l'adopció de mesures per combatre el contingut il·legal a les xarxes socials, així com protegir els consumidors que es consideren vulnerables en entorns digitals.
f. Drets de la personalitat i llibertats protegides per la Constitució portuguesa. La Carta protegeix drets constitucionals com són la privadesa, la llibertat d'expressió, el dret de reunió, el de manifestació i el desenvolupament d'habilitats digitals.

Itàlia

El 20 de maig de 2024, el govern italià va presentar al Parlament el projecte de llei que estableix disposicions i delega competències al govern en l'àmbit de la intel·ligència artificial, amb l'objectiu de coordinar i garantir l'aplicació de les disposicions del Reglament d'intel·ligència artificial de la Unió Europea, sense solapar-s'hi, establint «normes per a l'ús correcte, transparent i responsable, en una dimensió antropocèntrica, de la intel·ligència artificial, orientada a aprofitar les seves oportunitats» (article 1).

L'objectiu del projecte de llei és protegir els drets fonamentals, la democràcia, l'estat de dret i la sostenibilitat ambiental d'acord amb els possibles riscos i el grau d'impacte de la IA, així com fomentar la innovació per al benestar de tots els ciutadans. Per fer-ho, la proposta està formada per 26 articles dividits en 6 parts:

1. Normes subjacents als principis bàsics.
2. Normes sectorials específiques.
3. Governança, autoritats nacionals i accions de promoció.
4. Normes per a la protecció dels drets d'autor.
5. Sancions penals.
6. Provisions financeres.

La primera part (articles 1-6) preveu uns principis bàsics vinculats a les finalitats i àmbit d'aplicació de tot el projecte. Aquests principis interactuen de manera diferent segons els camps d'aplicació específics i de la fase del cicle de vida dels sistemes d'intel·ligència artificial considerada. Es tracta de transparència, proporcionalitat, seguretat, protecció de dades, confidencialitat, exactitud, no-discriminació, igualtat de gènere, sostenibilitat, fiabilitat, seguretat, qualitat, idoneïtat i transparència de les dades utilitzades, així com el respecte a l'autonomia humana i al poder de decisió.

Si el text finalment entrés en vigor, aquests principis s'aplicarien a la investigació, l'experimentació, el desenvolupament, l'adopció i l'aplicació de sistemes i models d'intel·ligència artificial.

Catalunya

A Catalunya el Govern de la Generalitat impulsa l'anomenada ***Estratègia d'intel·ligència artificial de Catalunya***, que desplega un programa d'actuacions específiques per enfortir l'ecosistema d'intel·ligència artificial que hi ha a Catalunya per tal de liderar la generació de coneixement, l'aplicació social i empresarial i la creació de solucions basades en intel·ligència artificial que fomentin el creixement econòmic i la millora de la vida de les persones.

El programa d'actuacions de l'Estratègia s'estructura entorn a diferents eixos que comprenen des de la indústria fins a la recerca, el talent, les infraestructures de dades, l'adopció de sistemes d'intel·ligència artificial i la promoció del desenvolupament d'una intel·ligència artificial ètica.

Finalment, l'informe ***La intel·ligència artificial a Catalunya***, presentat l'abril de 2024, identifica l'ecosistema de companyies, universitats, centres de recerca i innovació i institucions de suport del sector de la intel·ligència artificial a Catalunya.

Estats i regions de fora de la Unió Europea

Regne Unit

El 2023, el Regne Unit va presentar un projecte de llei de la intel·ligència artificial. El projecte, presentat per la Cambra dels Lords, consta de nou articles. L'article 1 encarrega al secretari d'estat la creació, per reglament, de l'Autoritat d'Intel·ligència Artificial. Les funcions de l'Autoritat d'Intel·ligència Artificial, descrites a l'article 1, apartat 2, inclourien:

1. Garantir que els reguladors rellevants tinguin en compte la IA amb un enfocament uniforme.
2. Coordinar una revisió de la legislació vigent per avaluar-ne la idoneïtat per abordar els reptes i les oportunitats de la intel·ligència artificial.
3. El seguiment i l'avaluació de l'eficàcia del marc normatiu global.
4. Donar suport a les iniciatives *sandbox* (article 3).
5. Acreditar auditors d'IA independents.
6. Educar les empreses i els ciutadans i promoure la participació ciutadana (art. 6).

L'Autoritat d'Intel·ligència Artificial també haurà de vigilar l'aplicació dels principis reguladors establerts per l'article 2 del projecte de llei i adreçats als reguladors i a les empreses.

En particular, la regulació de la intel·ligència artificial hauria de proporcionar: (a) seguretat i robustesa; (b) transparència i explicabilitat adequades; (c) equitat; (d) responsabilitat i governança; e) impugnabilitat (*contestability*) i reparació.

Les empreses que desenvolupin i distribueixin sistemes d'intel·ligència artificial hauran de garantir la transparència i el compliment de les lleis de protecció de dades i propietat intel·lectual, així com respectar els procediments definits pel secretari d'estat d'acord amb l'article 5 del projecte de llei.

Els sistemes d'intel·ligència artificial hauran de complir la legislació en matèria d'igualtat i ser inclusius des del seu disseny, per evitar la discriminació i satisfer les necessitats dels col·lectius vulnerables (persones grans, discapacitats, desfavorits econòmicament). També hauran de generar dades que siguin traçables, accessibles, interoperables i reutilitzables.

Finalment, és important assenyalar que actualment el govern del Regne Unit està en procés d'elaboració d'un projecte de llei per regular sistemes d'intel·ligència artificial de frontera –en terminologia del govern del Regne Unit–. Aquests models són models d'intel·ligència artificial general que tenen una capacitat fins i tot major que els models més avançats amb què comptem actualment. El govern del Regne Unit veu en la regulació de la intel·ligència artificial de frontera una oportunitat de liderar i de contribuir a la regulació de la intel·ligència artificial, que en aquests moments està relativament poc desenvolupada.

La intel·ligència artificial de frontera, de naturalesa eminentment internacional, presenta reptes jurídics específics, donat que les normatives nacionals no permeten regular la situació de manera comprensiva i general. No obstant això, el govern del Regne Unit entén que una bona llei del Regne Unit tindrà efectes extraterritorials: permetrà tenir un avantatge competitiu respecte d'altres sistemes jurídics i posicionar-se en un rol de lideratge en el disseny del model que finalment s'implementi en la regulació de la intel·ligència artificial de frontera.

Estats Units

Els Estats Units tenen una estructura federal que comporta una distribució de competències entre els estats i el govern federal. La regulació de la intel·ligència artificial, donat el seu impacte interestatal, forma part de l'agenda legislativa del govern federal nord-americà.

Avui els Estats Units no compten amb una norma equivalent el Reglament d'intel·ligència artificial com el que s'ha adoptat a Europa. A data d'avui, el govern federal dels Estats Units utilitza la normativa federal ja en vigor. A l'espera del desenvolupament d'un cos normatiu federal que reguli la intel·ligència artificial, els desenvolupadors, fabricants i els usuaris de sistemes intel·ligència artificial estan subjectes a l'entramat normatiu tant local com estatal vigent.

Això no significa una absència de dret positiu. En aquests moments hi ha més de cent propostes de llei sectorials pendents d'aprovació al Congrés dels Estats Units en matèria d'educació, propietat intel·lectual, robots, riscos biològics o seguretat nacional. Val la pena destacar que la majoria d'aquestes propostes normatives inclouen el desenvolupament de protocols de millors pràctiques que seran de compliment voluntari. La por a dificultar la innovació i posar en perill la competitivitat en el sector tecnològic fa que el legislador nord-americà sigui tremendament cautelós a l'hora d'introduir restriccions i requisits per a aquest tipus de sistemes i tecnologies en general.

Les propostes normatives presentades fins al moment, tant per part del govern federal com per part d'alguns dels estats dels Estats Units, tenen per objectiu que els sistemes intel·ligència artificial siguin segurs des del punt de vista de la protecció de dades, de la privacitat, dels drets fonamentals i dels nivells de risc. Aquesta seguretat es vol obtenir a través de la imposició d'obligacions per desenvolupar sistemes d'intel·ligència artificial responsables que minimitzin els danys que eventualment poguessin causar.

Iniciatives del govern federal nord-americà

L'any 2023 el govern nord-americà presidit per Joe Biden va adoptar una ordre executiva anomenada «Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence» amb l'objectiu d'aprofitar els beneficis i avantatges de la intel·ligència artificial i a la vegada controlar i mitigar els riscos que presenta. L'ordre executiva se centra en el rol de les agències federals i dels desenvolupadors de models fundacionals i els imposa estàndards federals i obligacions de compartir informació crítica i de seguretat amb el govern federal dels Estats Units.

L'ordre executiva se centra en les agències federals i els desenvolupadors de models fundacionals, imposa el desenvolupament d'estàndards federals i requereix que els desenvolupadors dels sistemes d'IA més potents comparteixin els resultats de les proves de seguretat i altra informació crítica amb el govern dels Estats Units.

Val la pena assenyalar que l'Administració del president Donald Trump preveu limitar o mitigar l'impacte d'aquesta ordre executiva.

D'altra banda, la proposta per al projecte de llei de drets en intel·ligència artificial (*Blueprint for an AI Bill of Rights*) aprovada l'any 2022 inclou cinc principis i recomana pràctiques per al disseny i la implementació de sistemes automatitzats.

El rol de les agències governamentals. La Federal Trade Commission

La Federal Trade Commission (FTC) ha adoptat iniciatives per regular la intel·ligència artificial. Per exemple, ha alertat els participants de mercats, sobretot digitals, amb relació a la utilització de sistemes d'intel·ligència artificial que poguessin provocar tractaments discriminatoris i ha imposat sancions a empreses per pràctiques deslleials i per haver causat danys als consumidors a través dels de sistemes d'intel·ligència artificial.

El rol dels estats en la regulació de la intel·ligència artificial

Els diferents estats d'Estats Units també han tingut iniciativa en matèria de regulació d'intel·ligència artificial. La majoria de les normes estatals adoptades tenen un àmbit d'aplicació del territori amb independència de si el sistema d'intel·ligència artificial ha estat desenvolupat dintre o fora de les seves fronteres. No és possible llistar-los tots aquí, però, a títol d'exemple:

Colorado

L'estat de Colorado va adoptar una normativa en matèria d'intel·ligència artificial amb la Llei d'intel·ligència artificial de Colorado (Colorado AI Act). Aquesta llei crea deures per als desenvolupadors i per als usuaris de sistemes d'intel·ligència artificial classificats d'alt risc amb independència de la seva mida. La llei entrarà en vigor al 2026.

Califòrnia

Califòrnia ha adoptat diferents normes en matèria d'intel·ligència artificial amb relació a la privacitat, transparència i responsabilitat, sobretot respecte al govern.

El model seguit a l'estat de Califòrnia és semblant a l'experiència relativa a protecció de dades: la Llei de protecció de la privacitat de Califòrnia (California Privacy Protection Act, CPPA), que també regula els sistemes de decisió automatitzada, es va inspirar fortament en el Reglament general de protecció de dades.

Xina

La regulació de la intel·ligència artificial a la Xina es troba en un estadi molt preliminar. L'any 2023 es van presentar les mesures provisionals en matèria de gestió dels serveis proporcionats amb intel·ligència artificial generativa, les Mesures d'intel·ligència artificial (AI Measures). Aquestes mesures provisionals van ser el resultat d'una feina conjunta entre l'Administració Xinesa del Ciberespai, la Comissió Nacional de Desenvolupament i Reforma, el Ministeri d'Educació, el Ministeri de Ciència i Tecnologia, el Ministeri d'Indústria i d'Informació Tecnològica, el Ministeri de Seguretat Pública i l'Administració de la Ràdio i Televisió Nacional.

Les Mesures d'intel·ligència artificial tenen un enfocament normatiu general per «promoure un desenvolupament saludable i una aplicació regulada de la intel·ligència artificial generativa, salvaguardar la seguretat nacional i els interessos públics socials i protegir els drets i interessos legals dels ciutadans, persones jurídiques i altres organitzacions». Alguns sectors com, per exemple, el sector de la salut, el sector automobilístic o el sector financer, compten amb la seva normativa específica relativa a la utilització de sistemes intel·ligents artificials en els seus processos de producció o de serveis. Les Mesures d'intel·ligència artificial no inclouen una categorització de risc dels diferents sistemes d'intel·ligència artificial, tot i que alguns sistemes i alguns usos estiguin subjectes a uns majors controls i requisits de seguretat, com serien, per exemple, sistemes d'intel·ligència artificial generativa que tinguin la capacitat de mobilitzar l'opinió pública.

Els usuaris de sistemes d'intel·ligència artificial generativa han de fer-ne un ús lícit i responsable per tal de protegir la informació dels usuaris i garantir els seus drets d'accés, còpia, rectificació i oblit. Els usuaris de sistemes d'intel·ligència artificial generativa han de prendre mesures per tal de garantir que els seus serveis són segurs i actuar per tal d'eliminar qualsevol contingut o conducta il·legal que pogués tenir lloc. Finalment, les Mesures d'intel·ligència artificial preveuen que els usuaris d'aquests sistemes cooperin amb les autoritats en el desenvolupament de les seves activitats de supervisió proporcionant l'assistència i la informació que siguin necessàries.

Els proveïdors de sistemes d'intel·ligència artificial generativa han de complir diferents requisits: respectar els valor essencials del socialisme –els sistemes no poden qüestionar el poder de l'estat o posar en perill la seguretat nacional–, han d'evitar que aquests sistemes generin discriminació per raó de nacionalitat, religió o gènere, entre d'altres, tant en la fase d'entrenament, com en la de generació o d'implementació. Cal també que els sistemes respectin els drets de propietat intel·lectual de tercers. Finalment, caldrà també que s'adoptin mesures per millorar la transparència, precisió i seguretat dels sistemes d'intel·ligència artificial generativa i dels serveis que es proporcionen amb la seva intervenció. En cas de vulneracions o incompliments de les Mesures d'intel·ligència artificial, les autoritats tenen competència i jurisdicció per imposar sancions econòmiques, així com sancions penals quan fossin aplicables.

Aquestes Mesures d'intel·ligència artificial s'apliquen a empreses que proporcionen sistemes o serveis d'intel·ligència artificial generativa als usuaris finals dins de les fronteres de la Xina, independentment de la procedència de l'empresa o del proveïdor que proporciona el servei.

És important assenyalar que hi ha diferents iniciatives legislatives i normes ja en vigor a la Xina que no regulen directament la intel·ligència artificial, però que afecten el seu desenvolupament i el seu ús.

El Consell Assessor del Parlament sobre Ciència i Tecnologia (CAPCIT) és un òrgan del Parlament creat el 2008 amb l'objectiu de coordinar la informació i l'assessorament en matèria científica i tecnològica que necessitin els diputats i els òrgans del Parlament. El CAPCIT és un òrgan de composició mixta, format pels diputats que en són membres i per representants de les principals institucions científiques i tecnològiques de Catalunya: l'Institut d'Estudis Catalans (IEC), la Fundació Catalana per a la Recerca i la Innovació (FCRi), la Reial Acadèmia de Ciències i Arts de Barcelona (RACAB), l'Associació Catalana d'Entitats de Recerca (ACER) i una representació de les universitats públiques i privades. Per a més informació: http://www.parlament.cat/capcit