

Intel·ligència artificial en el tractament d'anuncis oficials

1. INTRODUCCIÓ

En un món digitalitzat, l'Administració pública s'enfronta a reptes cada vegada més complexos. La creixent quantitat d'informació i la necessitat de millorar l'eficiència en la prestació de serveis requereixen solucions innovadores. Així mateix, la integració de tecnologies emergents en les administracions públiques europees ha demostrat ser un impuls fonamental per a la millora de l'eficiència, la transparència i la qualitat dels serveis públics.

En aquest context, el Servei del Butlletí Oficial de la Província de Barcelona i Documentació i Publicacions Oficials (SBOPiDiPO) de la Diputació de Barcelona (DB) ha adoptat diverses solucions innovadores basades en intel·ligència artificial (IA). D'una banda, un sistema per optimitzar la redacció dels resums dels anuncis publicats al Butlletí Oficial de la Província de Barcelona (BOPB). I, d'altra banda, una solució per a la classificació temàtica d'anuncis en butlletins i diaris oficials al Cercador d'Informació i Documentació Oficials (CIDO), coneguda com a CIDOBOT. Aquests projectes representen un avenç significatiu en la modernització dels processos administratius i en la millora de l'eficiència i l'eficàcia en el funcionament dels serveis públics.

Aquest document presenta una anàlisi de l'impacte i els avantatges d'aquestes solucions tecnològiques, abordant aspectes clau com la millora en la qualitat i eficiència dels serveis públics, l'augment de l'obertura governamental, l'orientació en l'usuari, i la seva escalabilitat i replicabilitat.

2. EL SERVEI DEL BUTLLETÍ OFICIAL DE LA PROVÍNCIA DE BARCELONA I DOCUMENTACIÓ I PUBLICACIONS OFICIALS DE LA DIPUTACIÓ DE BARCELONA

L'SBOPiDiPO té la missió de garantir el dret d'accés a la informació de les administracions públiques per contribuir a la transparència, mitjançant la tramitació, la publicació i la difusió de la informació oficial que afecta l'àmbit territorial català, tot contribuint així al compliment de les Lleis de transparència, accés a la informació i bon govern, 19/2013 i 19/2014, estatal i catalana, respectivament. Presta els seus serveis a partir de dos productes, el Butlletí Oficial de la Província de Barcelona i el Cercador d'Informació i Documentació Oficials.

El BOPB és el diari oficial on es publiquen, des del 3 de novembre de 1833, les disposicions de caràcter general i les ordenances, actes, edictes, acords, notificacions, anuncis i altres resolucions de les administracions públiques i de l'Administració de Justícia d'àmbit territorial provincial de Barcelona, quan així estigui previst en una

disposició legal o reglamentària. També s'hi poden publicar altres actes o anuncis que aquestes administracions o anunciants particulars ordenin inserir.

Des de l'1 de gener de 1998 es garanteix l'accés universal i gratuït a l'edició electrònica del butlletí a través del portal web del BOPB (<https://bop.diba.cat/>). El web centralitza tota la informació referent al butlletí: els serveis de consulta, la tramitació electrònica dels anuncis i la notificació de la publicació, mitjançant l'ús de certificat electrònic o altres mitjans segurs d'identificació i autenticació reconeguts a l'empara de la legislació vigent.

La Diputació de Barcelona presta amb mitjans propis, de forma directa, aquest servei públic, el qual es troba regulat al Reglament aprovat pel Ple corporatiu en la sessió de 29 d'octubre de 2020 i publicat en el BOPB del 25 de gener de 2021.

El CIDO és un servei de difusió de la informació de les administracions públiques per contribuir a la seva transparència, mitjançant la difusió de la informació oficial que afecta la ciutadania de Catalunya.

L'equip del CIDO elabora fitxes de normes i procediments del sector públic a través de l'agrupació dels anuncis oficials que les componen, i que les institucions publiquen de manera independent i molt dispersa. Així en garanteix la traçabilitat, i permet conèixer exactament en quina fase i en quin estat es troben aquestes normes i procediments.

Ahora, s'analitza documentalment el contingut dels anuncis per identificar la informació més rellevant i convertir-la en dades obertes estructurades, fàcilment recuperables a partir de les necessitats dels usuaris.

A més, des de l'any 2021 el CIDO ofereix tot el seu sistema d'informació en format obert a través d'una [API -interfície de programació d'aplicacions de dades obertes-](#) mitjançant una arquitectura de software per a sistemes estàndard.

El servei resultant és un conjunt de cercadors temàtics accessibles des del portal <https://cido.diba.cat>, que agrupa els següents conjunts de dades de tot Catalunya:

- Ocupació pública
- Ajuts, subvencions, beques i premis
- Contractació pública
- Normativa local
- Convenis col·lectius
- Legislació

El Cercador inclou un sistema d'alertes que permet rebre notificacions de les novetats i actualitzacions de les normes i/o procediments que l'usuari desitja seguir. Des d'*El meu CIDO* l'usuari pot gestionar múltiples alertes, de qualsevol dels cercadors.

3. INTEL·LIGÈNCIA ARTIFICIAL EN LA REDACCIÓ DELS RESUMS DELS ANUNCIS DEL BOPB

El BOPB gestiona una mitjana anual de 35.000 anuncis, i cadascun necessita un resum per ser publicat. Fins ara, aquest procés s'ha fet manualment, amb la participació tant dels anunciants com del personal de l'SBOPiDiPO.

Els anunciants han de proposar un resum inicial en el formulari d'inserció d'anuncis. Sovint, aquests resums no compleixen els estàndards necessaris i presenten errors o omissions, cosa que implica que l'equip de l'SBOPiDiPO hagi de llegir, corregir i resumir de nou cada anunci. Aquest procés no només consumeix molt de temps, sinó que també està subjecte a errors humans que poden comprometre la qualitat i la consistència dels resums publicats.

Per altra banda, algunes de les accions portades a terme per l'SBOPiDiPO, com la catalogació en clau d'Objectius de Desenvolupament Sostenible de l'Agenda 2030 o la classificació temàtica dels anuncis per incorporar el projecte ELI- Identificador Europeu de Legislació- per facilitar l'accés, l'intercanvi i la interconnexió de la informació jurídica publicada en els sistemes d'informació jurídica nacionals, europeus i mundials, per posar en marxa una xarxa d'informació legal, disponible com a conjunt de dades obertes i susceptibles de reutilització, requereixen una gran quantitat de recursos humans i materials. Per això, era necessari implementar una solució que millorés aquests processos i reduís la càrrega de treball del personal.

Per tot això, la redacció dels resums dels anuncis representa una inversió significativa de recursos humans, tant per part dels anunciants com del personal del BOPB, i destaca la necessitat d'una solució que pugui alleugerir la càrrega de treball per a ambdues parts. Amb l'augment constant del volum d'anuncis i la creixent complexitat dels continguts, es va identificar la necessitat d'una eina que assegurés la precisió i l'eficiència en la gestió dels anuncis. La modernització dels processos administratius s'ha convertit en una prioritat per a moltes institucions públiques, alineant-se amb les tendències globals de digitalització i innovació tecnològica.

Així doncs, recentment s'ha desenvolupat un algorisme basat en intel·ligència artificial que confecciona automàticament els resums dels anuncis per publicar en el BOPB. Aquesta intel·ligència artificial es basa en el que s'anomena *Machine Learning* (aprenentatge automàtic), un sistema que, a partir de l'anàlisi de dades, és capaç d'adquirir coneixements per resoldre problemes mitjançant exemples i, a més, d'aprendre i millorar amb l'experiència. Hi ha diversos tipus d'aprenentatge automàtic i, per les característiques de la informació que es tracta al BOPB, s'ha considerat utilitzar l'aprenentatge supervisat, una tècnica per determinar un resultat a partir d'uns dades prèvies d'entrenament.

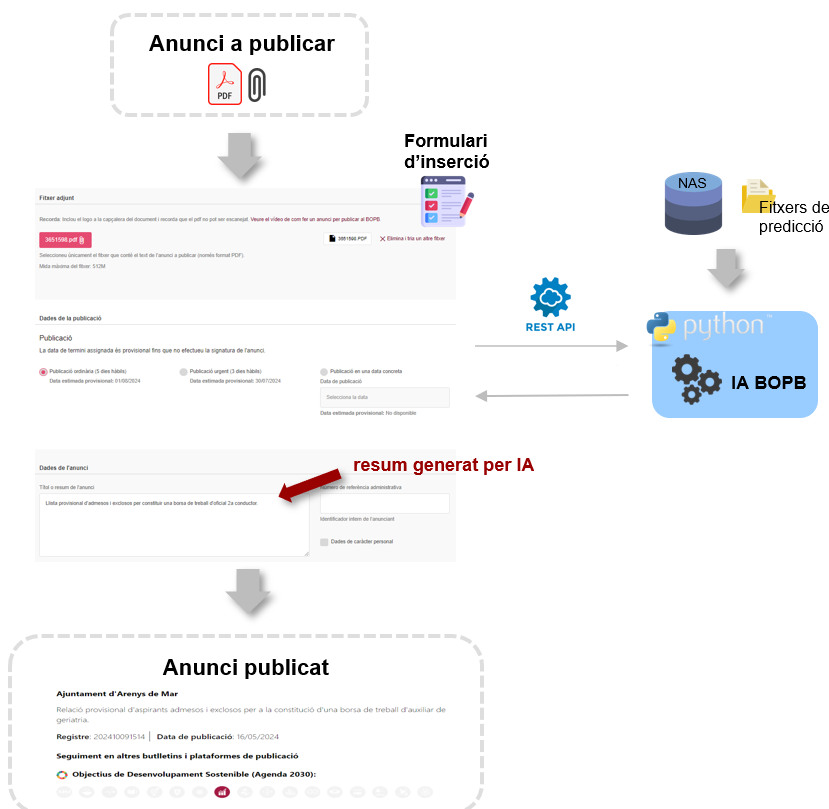
En casos més complexos, es requereix l'ús de xarxes neuronals, on la informació travessa la xarxa i, mitjançant un comportament anàleg al de les neurones del cervell, retorna un resultat; és el que s'anomena *Deep Learning* (aprenentatge profund). Una xarxa neuronal és un model matemàtic que utilitza fórmules concretes per calcular probabilitats que permeten assignar categories als registres d'entrada a la xarxa. Per aquest motiu, va ser necessari transformar les paraules dels anuncis en vectors numèrics compatibles amb l'entrada de les xarxes neuronals.

En el cas del BOPB, el tipus de xarxa neuronal emprada ha estat l'arquitectura *Transformer*, la qual és capaç d'analitzar el context d'un text i, per tant, extreure significat mitjançant el seguiment de relacions en dades seqüencials com les paraules d'aquesta frase.

Per aconseguir aquest model, el sistema ha entrenat amb uns 275.000 anuncis oficials publicats al BOPB entre els anys 2013 i 2022, amb un temps total d'entrenament d'unes 120 hores. D'aquesta manera, s'aconsegueix un model, format per 300 milions de paràmetres, capaç de generar resums dels anuncis de la mateixa forma que ho faria una persona.

Així, quan un anunciant li planteja al sistema un text d'un anunci oficial (ja sigui en català o castellà), automàticament processa el contingut de les dues primeres pàgines de l'anunci i retorna un resum amb un màxim de 100 paraules. A més, queda constància quan l'anunciant modifica manualment el resum proposat per la IA. Això permet analitzar el grau de precisió del sistema i valorar en quin moment és necessari un nou entrenament.

Imatge 1. Funcionament del procés d'automatització de la IA del BOPB



Font: elaboració pròpia

El sistema d'IA es troba en un servidor propi de la DIBA i, a més, és adaptable i escalable, capaç de gestionar un volum creixent d'anuncis sense comprometre la qualitat.

4. CIDOBOT: INTEL·LIGÈNCIA ARTIFICIAL PER A LA CLASSIFICACIÓ TEMÀTICA DELS ANUNCIS EN BUTLLETINS I DIARIS OFICIALS

L'any 2018, l'SBOPiDiPO va adaptar l'aplicatiu de gestió interna del CIDO amb la incorporació de noves funcionalitats que permeten millorar el rendiment en la catalogació d'anuncis. Entre aquestes funcionalitats, destaca un mòdul d'IA (pioner en la institució provincial), que realitza l'anàlisi documental previ de la informació pública que prové dels butlletins oficials per a la seva classificació per àmbits temàtics o procedimentals (oposicions, subvencions, contractació, normativa local i convenis col·lectius).

Aquest model, conegut com a CIDOBOT, es basa en un sistema de xarxes neuronals (*Deep Learning*), de la mateixa manera que la IA del BOPB, però en aquest cas es fa servir una arquitectura diferent. El sistema ha entrenat amb els resums dels anuncis publicats durant un any i mig en els butlletins oficials i la seva classificació per àmbit (realitzada per un equip humà de manera manual). A través de l'ajust dels paràmetres d'una funció que relaciona els resums dels anuncis amb el seu àmbit temàtic, el sistema aprèn d'aquests exemples i és capaç de donar una classificació per àmbit correcta.

Per aconseguir l'automatització del procés de classificació, en primer lloc, va ser necessària una fase d'entrenament. Es van tenir en compte les següents fonts d'informació: Butlletí Oficial de la Província de Barcelona (BOPB), Butlletí Oficial de la Província de Girona (BOPG), Butlletí Oficial de la Província de Lleida (BOPL), Butlletí Oficial de la Província de Tarragona (BOPT), Diari Oficial de la Generalitat de Catalunya (DOGC) i Butlletí Oficial de l'Estat (BOE). I es van considerar les següents categories o àmbits: contractació, convenis col·lectius, normativa local, oposicions, subvencions i no classificar.

Imatge 2. Funcionament del procés d'automatització del CIDOBOT: fase d'entrenament



Font: elaboració pròpia

El sistema de classificació utilitza una xarxa neuronal per a cadascun dels butlletins, fent servir com a dades d'entrenament tots els anuncis que es van classificar manualment per l'equip de persones del CIDO des de l'1 de gener de 2017 fins al 30 de juny de 2018. Es va decidir utilitzar una xarxa neuronal per a cada butlletí perquè al BOE els anuncis

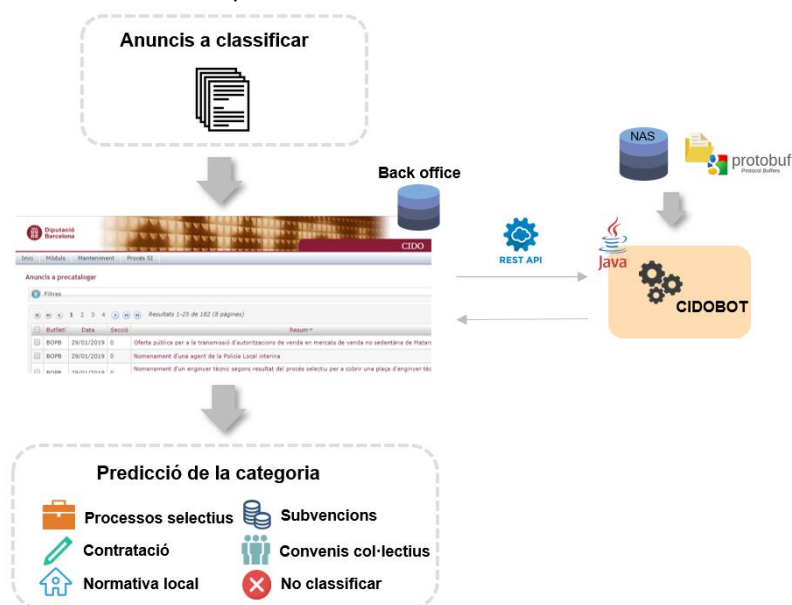
es publiquen en castellà, mentre que en els altres la majoria apareixen en català, i perquè cada butlletí publica els resums de forma que les paraules més repetides són significativament diferents, en funció de la informació que correspon al seu àmbit.

Com s'ha comentat anteriorment, una xarxa neural requereix transformar els textos d'entrada a la xarxa en vectors numèrics. En el cas de CIDOBOT, aquesta transformació numèrica es basa en construir un diccionari de les paraules més repetides per assignar així un valor numèric a cada paraula. Llavors, per a cada anunci es pot construir un vector numèric que representi quantes vegades apareix cada paraula del diccionari en el resum de l'anunci.

Per construir aquests vectors numèrics, primer és necessari un preprocessament del resum de cada anunci, que consisteix a eliminar els signes de puntuació i separar les paraules, descartant aquelles que tenen una longitud inferior a 4 caràcters. A partir d'aquí, es calcula quines són les 5.000 paraules més repetides de cada butlletí, de manera que cada paraula queda associada a un valor numèric que correspon a la posició que ocupa en el diccionari.

Una vegada que cada paraula té associat un valor numèric, es crea una matriu per a cada butlletí amb una fila per a cada anunci i 5.000 columnes, on cada entrada de la matriu representa el nombre de vegades que apareix cada paraula en l'anunci. Aquesta matriu es guarda en la seva carpeta corresponent per ser utilitzada posteriorment en la fase de predicció.

Imatge 3. Funcionament del procés d'automatització del CIDOBOT: fase de predicció



Font: elaboració pròpia

Així doncs, quan al sistema se li planteja el resum d'un nou anunci oficial, retorna una predicció de la classificació per àmbit (oposicions, subvencions, contractació, normativa local, convenis col·lectius o no classificar), de la mateixa manera que ho faria una persona; a més, informa del grau de confiança de la seva predicció.

Aquest grau de confiança serveix per, en el cas que no es consideri suficient, establir una comprovació humana. De la mateixa manera, es recullen els possibles errors en l'assignació de la classificació perquè el sistema sigui capaç de reaprendre en un nou entrenament futur. De fet, el sistema ha estat entrenat de nou amb posterioritat a la seva posada en explotació per aconseguir un major grau de precisió en la classificació retornada.

5. CONTRIBUTIÓ A LA QUALITAT I L'EFICIÈNCIA DELS SERVEIS PÚBLICS

Amb la implementació de la IA en el tractament d'anuncis oficials s'ha aconseguit una millora considerable tant en la qualitat dels serveis públics com en l'eficiència interna:

Augment de l'eficiència

La implementació de la IA fa que els processos siguin més ràpids i eficients, reduint el temps global de les tasques i millorant la resposta a les necessitats dels usuaris. Aquesta millora en l'eficiència també es tradueix en una reducció de costos operatius i un increment en la capacitat de gestionar grans volums de dades.

Millora de la precisió i la consistència

La IA ha demostrat una precisió sorprenent tant en la generació de resums com en la classificació temàtica. La seva capacitat per identificar temes específics i generar resums uniformes i precisos ha minimitzat els errors humans i garantit la integritat de la informació.

Optimització de recursos

En reduir la càrrega de treball manual, la IA ha permès una assignació més eficient dels recursos disponibles. Això permet redistribuir el personal per abordar altres àrees crítiques, més estratègiques i de més valor afegit, com la presa de decisions i la planificació de polítiques.

6. AUGMENT DE L'OBERTURA DEL GOVERN

L'SBOPiDiPO, des dels seus inicis, ha contribuït significativament a l'augment de l'obertura del govern. Automatitzant i millorant la precisió tant en la classificació d'anuncis amb CIDOBOT, com en la generació de resums dels anuncis del BOPB, es facilita un accés més ràpid i fiable a la informació pública. Aquesta millora en l'accés a la informació és un pilar fonamental per a la transparència governamental, ja que permet als ciutadans estar millor informats sobre les activitats i decisions de l'Administració pública.

Per la seva banda, el CIDO també ha millorat la traçabilitat dels procediments administratius, proporcionant un seguiment detallat i accessible dels processos des del seu inici fins a la seva finalització. Aquesta capacitat de seguiment és crucial per garantir

la rendició de comptes i perquè els ciutadans puguin verificar el compliment de les normatives i regulacions per part de les administracions públiques.

A més, la integració del CIDO amb altres sistemes d'informació pública, com els portals de transparència municipals del Consorci Administració Oberta de Catalunya i el portal Feina Activa del Servei d'Ocupació Pública de Catalunya, ha ampliat l'abast i la disponibilitat de la informació pública. Aquesta integració no només millora la transparència, sinó que també fomenta la col·laboració entre diferents entitats i nivells de govern, enfortint l'ecosistema de dades obertes.

Així doncs, la IA del BOPB i CIDOBOT no només han millorat l'eficiència interna, sinó que també han augmentat l'obertura del govern al públic:

Accés a la informació pública

La classificació automàtica dels anuncis i la generació de resums ha fet que la informació sigui accessible de forma més ràpida per als ciutadans. A més, el CIDO ha contribuït a la transparència governamental en millorar la traçabilitat dels procediments administratius.

Compartició de dades amb altres organitzacions

El CIDO ha col·laborat amb altres institucions i organitzacions per compartir les seves dades. Això ha ampliat encara més la difusió de la informació pública.

7. ORIENTACIÓ EN L'USUARI

El disseny i desenvolupament dels sistemes d'IA de l'SBOPiDiPO s'ha centrat en les necessitats dels usuaris finals, en aquest cas els empleats públics. Aquesta tasca totalment manual no aportava cap valor per al desenvolupament professional de les persones, tal com es va evidenciar per la major part del personal, de manera que es va decidir la seva automatització. Així, les persones de l'organització han pogut dedicar-se a altres tasques de més valor afegit i més gratificants a nivell personal, com el desenvolupament de polítiques estratègiques.

Les persones de l'organització valoren molt positivament aquests sistemes d'IA, motiu pel qual ja s'estan valorant i desenvolupant altres solucions per a l'optimització de processos administratius:

Classificació per Objectius de Desenvolupament Sostenible (ODS)

El sistema d'IA podria ser programat per classificar els anuncis segons els Objectius de Desenvolupament Sostenible (ODS). Aquesta capacitat ajudaria a alinear les publicacions amb les prioritats globals de sostenibilitat i facilitaria el seguiment i l'anàlisi de les contribucions dels anuncis a aquests objectius. A més, proporcionaria una eina valuosa per a la transparència i la rendició de comptes en relació amb els compromisos de desenvolupament sostenible.

Catalogació automàtica dels anuncis oficials

A banda de la preclassificació automàtica per part de CIDOBOT, la IA podria anar un pas més enllà i analitzar el contingut dels anuncis i extreure la informació més rellevant per als ciutadans. Aquesta informació passaria a ser estructurada dins del sistema del CIDO i, d'aquesta manera, contribuiria encara més a l'eficiència interna i la transparència de les administracions públiques.

Detecció de patrons i tendències

Amb la seva capacitat analítica avançada, l'algorisme d'IA podria identificar patrons i tendències en els anuncis publicats. Aquesta informació seria útil per a l'anàlisi estratègica i la presa de decisions, permetent a l'SBOPiDiPO anticipar necessitats futures i adaptar-se millor als canvis en les demandes de publicació. Això podria incloure l'anàlisi de freqüència de certs tipus d'anuncis, temàtiques emergents o variacions estacionals.

Assistència en la redacció i edició

A més de generar resums, l'algorisme podria ser utilitzat per assistir en la redacció i edició dels anuncis, oferint suggeriments de millora i correcció automàtica d'errors. Aquesta funcionalitat garantiria una major qualitat en els continguts publicats i reduiria la necessitat de revisions manuals, millorant l'eficiència i la consistència del servei.

Automatització de la gestió documental

El sistema d'IA podria ser ampliat per gestionar automàticament altres aspectes documentals, com la indexació, l'arxiu i la recuperació de documents. Aquesta funcionalitat optimitzaria la gestió de la informació, facilitant l'accés a documents històrics i d'arxiu, i assegurant una millor organització i preservació dels registres oficials.

8. ESCALABILITAT I REPLICABILITAT DE LA SOLUCIÓ

L'arquitectura de la IA de l'SBOPiDiPO ha estat dissenyada per ser altament escalable. Això significa que la solució pot gestionar un augment en el volum de dades sense comprometre el seu rendiment.

A més, la capacitat de la IA per integrar-se amb altres sistemes i plataformes d'informació pública facilita la seva expansió a nous àmbits i geografies. Està dissenyada per ser replicable en altres administracions públiques i sectors de l'administració. La solució tecnològica és flexible i, amb un ajust de les dades prèvies d'entrenament, pot ser adaptada a diferents contextos i requeriments específics.

Encara que la solució no es troba en codi obert, qualsevol Administració pública que ho desitgi pot beneficiar-se'n. Només cal posar-se en contacte amb l'SBOPiDiPO i es gestionarà l'acord de col·laboració corresponent.

En resum, l'arquitectura IA desenvolupada és una solució escalable i replicable:

Escalabilitat

Els sistemes d'IA permeten gestionar un volum creixent d'anuncis sense comprometre la qualitat i poden adaptar-se a futurs reptes i necessitats de l'organització.

Adaptabilitat a altres solucions i administracions públiques

La tecnologia subjacent utilitzada, basada en l'aprenentatge automàtic, pot aplicar-se a altres sistemes de classificació o generació de resums amb dades prèvies per a l'entrenament. A més, qualsevol Administració pública pot beneficiar-se d'aquesta solució, adaptant-la a les seves pròpies necessitats i processos.

Replicabilitat en diferents sectors

El sistema d'IA desenvolupat no es limita a l'àmbit de les publicacions oficials. El seu enfocament en la classificació temàtica i generació de resums podria aplicar-se en altres contextos, com l'organització de documents o la categorització de contingut en biblioteques.