

[Inici](#) > Neix ?AINA?, el projecte del Govern per garantir el català en l'era digital

---

## Neix ?AINA?, el projecte del Govern per garantir el català en l'era digital

El projecte està impulsat pel Departament de Polítiques Digitals amb la col·laboració del BSC.



**El projecte AINA generarà els recursos digitals i lingüístics necessaris per facilitar el desenvolupament d'assistents de veu, traductors automàtics o agents conversacionals en català**

**L'objectiu últim és que la ciutadania pugui participar en català en el món digital al mateix nivell que els parlants d'una llengua global, com ara l'anglès, i evitar així l'extinció digital de la llengua**

**El projecte té un pressupost de 13,5M€, que es finançaran amb els fons NextGenerationEU, i arrenca amb una aportació inicial de 250.000€ del Departament de Polítiques Digitals**

**El primer recurs generat és el 'corpus' del català per entrenar els algorismes d'Intel·ligència Artificial (IA), el més gran creat fins al moment, amb 1.770 milions de metadades associades a paraules**

## **El proper pas serà generar els models de la llengua, models de la parla i models per a la traducció utilitzant xarxes neuronals multicapa**

Dotar el català de recursos digitals i lingüístics perquè esdevingui una llengua competitiva al món digital i assegurar-ne així la seva supervivència futura és l'objectiu del projecte AINA, que el conseller de Polítiques Digitals i Administració Pública, **Jordi Puigneró**, ha presentat avui en roda de premsa, acompanyat per la directora general de Societat Digital, **Joana Barbany**, el director associat del Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC), **Josep Maria Martorell**, i la investigadora i colíder de la unitat de Minería de Dades del BSC, **Marta Villegas**, responsable del projecte.

Impulsat pel Departament de Polítiques Digitals, amb la col·laboració del BSC, el projecte AINA generarà corpus i models informàtics de la llengua catalana perquè les empreses que creen aplicacions basades en intel·ligència artificial (IA), com ara assistents de veu, traductors automàtics, agents conversacionals, etc., puguin fer-ho fàcilment en català.

### **Pressupost i abast del projecte**

El projecte AINA té un pressupost global de 13,5 milions d'euros per al període 2020-2024 i és un dels projectes prioritzats pel Departament de Polítiques Digitals per ser finançat amb els fons europeus NextGenerationEU. De moment, el projecte arrenca amb una aportació inicial de 250.000€ que el Departament de Polítiques Digitals ha assignat al BSC per ampliar els corpus de la llengua catalana i així obtenir models lingüístics que abastin les diferents variants i registres.

El BSC ja disposa d'un primer corpus textual del català, consistent en 1.770 milions de paraules, reunides en 95 milions de frases. Aquest corpus, el més gran que s'ha fet mai de la llengua catalana, s'ha obtingut a base de descarregar textos de diferents fonts digitals (planes web, arxius, etc), netejar-los i esborrar duplicitats.

La Generalitat de Catalunya ha proveït tota la informació de les seves pàgines web i del DOGC, cosa que ha suposat el 33% de tots els continguts descarregats, i han estat necessàries 2.000 hores de processadors del superordinador MareNostrum per revisar les dades obtingudes, desduplicar-les i eliminar tot allò que no fossin pròpiament oracions del català.

Aquest primer corpus es va fer amb el finançament del Plan de Impulso de las Tecnologías del Lenguaje, de la Vicepresidència d'Afers Econòmics i Transformació Digital del Govern espanyol.

Ara, amb l'impuls del Departament de Polítiques Digitals, es crearan nous corpus per incorporar les diferents variants dialectals del català, diferents registres lingüístics (col.loquial, literari, administratiu, etc.) i arxius de veu i imatge. La Corporació Catalana de Mitjans Audiovisuals aportarà tot el seu repositori documental.

Amb tota aquesta informació, el següent pas serà entrenar xarxes neuronals multicapa perquè "aprenguin el català" i generin models de la llengua, models de la parla i models per a la traducció. Aquests models també són molt costosos de fer perquè necessiten gran capacitat de càlcul (el que s'està construint en base al primer corpus textual farà servir 9.000 hores de GPU), i seran les bases sobre les quals es podran desenvolupar aplicacions basades en IA, com assistents de veu, predictors i correctors lingüístics, *xatbots*, aplicacions de resum automàtic, cerques intel·ligents, aplicacions per a l'anàlisi de sentiments o motors de traducció i subtitulació automàtica, entre d'altres.

Tots els models que crearà el BSC estaran a disposició de totes aquelles empreses o entitats que les vulguin fer servir, ja que es publicaran en obert i amb llicències permissives.

## **El món digital, una oportunitat i un repte per al català**

Això ha de permetre que el català faci un salt qualitatiu i quantitatiu en l'ecosistema digital. De fet, el món digital és avui una oportunitat i un repte per a la llengua catalana. Actualment, les tecnologies de veu i les aplicacions i interfícies de veu per a l'accés al món digital són estratègiques per al desenvolupament ple de la llengua en tots els sectors. La interacció entre les persones i la tecnologia ha entrat en una nova fase on cada vegada menys es fa a través de dispositius com el teclat, el ratolí o les pantalles tàctils, per donar pas a una nova forma d'interacció més natural a través de la veu i la parla. I aquest fet atorga especial rellevància a la llengua, que esdevé un dels principals vehicles d'interacció.

Aquesta nova interacció ha de ser possible fer-la, també, en català. En aquest sentit, el Govern té el ferm propòsit de garantir que la ciutadania pugui parlar i interactuar en català en el món digital al mateix nivell que els parlants d'altres llengües com l'anglès o el castellà, llengües que, ara per ara, tenen garantida la seva supervivència digital perquè darrere han tingut Estats que han invertit per dotar de recursos suficients pel que fa a les tècniques d'aprenentatge i xarxes neuronals en Intel·ligència Artificial.

Un estudi realitzat l'any 2011 per la xarxa europea d'excel·lència META-NET, realitzat per més de 200 experts en Tecnologies del Llenguatge, adverteix que més de 20 llengües europees, entre elles el català, s'enfronten a l'extinció digital si no reben més suport tecnològic en quatre àrees: la traducció automàtica, la interacció amb la veu, l'anàlisi textual i la disponibilitat de recursos lingüístics.

## **La IA i les Tecnologies del Llenguatge**

Les Tecnologies del Llenguatge són aquelles que ja utilitzem en el nostre dia a dia quan corregim de forma automàtica un text al correu electrònic, utilitzem un cercador web a Internet, traduïm de forma automàtica una pàgina web, donem ordres de veu al telèfon mòbil, interactuem amb assistents virtuals o seguim les indicacions del navegador GPS, entre d'altres. I són les tecnologies que permetran dialogar amb ordinadors, aparells domèstics i fins i tot amb el nostre vehicle de manera natural.

Les noves tecnologies d'Intel·ligència Artificial i les Tecnologies del Llenguatge es basen en l'aplicació d'algorismes sobre grans conjunts de dades de qualitat, però els conjunts de dades sobre els quals s'entrenen els algorismes són específics per a cada llengua.

En aquest sentit, les grans multinacionals com Google, Apple i Microsoft utilitzen els recursos digitals de l'anglès creats per l'Agència de Projectes de Recerca Avançada del Departament de Defensa dels EUA (DARPA), que han estat la base lingüística de la IA a nivell mundial, ja que generar aquests mateixos recursos per part d'una empresa resultaria molt costós.

## **Un projecte estratègic**

El projecte AINA s'emmarca en l'estratègia digital del Govern, a través de dues iniciatives liderades pel Departament de Polítiques Digitals: el Consell de Direcció interdepartamental per a la promoció del català a Internet i en les tecnologies digitals avançades, aprovat el desembre del 2018, i l'Estratègia d'Intel·ligència Artificial de Catalunya (Catalonia.AI), aprovat el febrer del 2020.

El primer compte amb la participació de la Direcció General de Societat Digital, la Direcció General de Política Lingüística, la Direcció General de Mitjans de Comunicació, l'Agència de Ciberseguretat de Catalunya, la Direcció General d'Atenció Ciutadana i la Fundació puntCAT, i té entre els seus objectius impulsar la presència del català en els assistents de veu. Per la seva banda, un dels eixos prioritaris de l'estratègia Catalonia.AI està vinculat a l'ús normalitzat de la llengua catalana en les interfícies com a element clau en el desplegament de la IA, ja que la llengua és l'element de comunicació de base per accedir,

utilitzar i interactuar amb aquestes tecnologies.

### **AINA, un nom on conflueixen llengua i tecnologia**

El projecte ha estat batejat amb el nom d'AINA en homenatge a la filòloga menorquina **Aina Moll**, figura central de la promoció i la normalització del català i primera Directora General de Política Lingüística de la Generalitat de Catalunya del 1980 al 1988. Ella va ser l'artífex del llançament, el 1982, de la primera campanya institucional de sensibilització sobre l'ús de la llengua '**El català, cosa de tots**', que, amb el popular personatge de la **Norma** al capdavant, tenia com a objectiu conscienciar la societat sobre la situació sociolingüística del català. Un any després, s'aprovava la primera llei de normalització lingüística.

AINA conté una referència a la tecnologia (AI: Intel·ligència Artificial) que farà possible la seva normalització en l'àmbit digital.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---

**Source URL (retrieved on 12 ago 2024 - 07:30):** <https://www.bsc.es/ca/noticies/noticies-del-bsc/neix-%E2%80%98aina%E2%80%99-el-projecte-del-govern-garantir-el-catal%C3%A0-en-lera-digital>