

El Govern invertirà 12M€ fins al 2026 per culminar el projecte AINA i garantir el futur de la llengua catalana al món digital

written by David Folch | 21 de desembre de 2022

El Govern invertirà un total de 12 milions d'euros durant els propers quatre anys al projecte AINA per garantir-ne el finançament fins a la seva culminació, l'any 2026, i assolir així el seu objectiu últim: ensenyar català a les màquines de manera que la ciutadania pugui relacionar-se amb elles i participar en el món digital en català al mateix nivell que els parlants d'una llengua global i garantir, així, el futur digital de la llengua catalana.

Per fer-ho, el Departament d'Empresa i Treball destinarà 3M€ anuals al projecte durant els propers quatre anys mitjançant una subvenció directa al Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC-CNS), l'entitat encarregada d'executar les diferents actuacions que preveu el full de ruta de l'AINA fins al 2026, any de finalització del projecte.

L'anunci l'han fet el dilluns el conseller d'Empresa i Treball, **Roger Torrent i Ramió**, i el director associat del BSC-CNS, **Josep Maria Martorell**, després de mantenir una reunió de treball en la qual també hi han participat la secretària de Polítiques Digitals, **Gina Tost i Faus**; el secretari de Política Lingüística, **Francesc Xavier Vila Moreno**; el director del BSC-CNS, **Mateo Valero**; i la responsable del projecte AINA i colíder de la Unitat de Minería de Textos del BSC, **Marta Villegas**.

El conseller d'Empresa i Treball, Roger Torrent i Ramió, ha afirmat que “volem que el català sigui llengua d'ús normalitzat en el món digital. En aquest sentit, el projecte AINA és una infraestructura de país que necessitava estabilitat pressupostària fins a la seva finalització per garantir que podrà culminar-se amb èxit i retenir el talent que el fa possible. Per això el Govern hi aposta fermament i en garanteix el finançament fins al 2026”. Torrent també ha afegit que “un cop culminat el projecte, podrem oferir les peces necessàries per a que qualsevol empresa o organització pugui crear les seves pròpies aplicacions i incloure el català de forma normalitzada en els seus serveis digitals”.

Així mateix, el titular d'Empresa i Treball ha remarcat que aquests 12 milions d'euros són “una inversió important a mig termini, que garanteix que el català tingui una presència normalitzada al món digital, facilita la digitalització en català al teixit empresarial, i reforça el sector de la Intel·ligència Artificial i les Tecnologies del Llenguatge al nostre país.”

Per la seva part, el director associat del BSC-CNS, Josep Maria Martorell, ha assegurat: “Disposar d'un finançament plurianual fins a l'any 2026 és una notícia excepcional que ens permet donar continuïtat a l'equip que treballa en el projecte AINA i seguir invertint en desenvolupar nova tecnologia. Tancar aquesta nova etapa d'AINA garantirà que mai més cap empresa gran o petita pugui tenir una excusa per no incorporar el català en els seus serveis textuais, visuals o d'àudio”.

Fites assolides i properes passes fins a la culminació del projecte

AINA és un projecte basat en tecnologies de dades i intel·ligència artificial (AI) impulsat pel Govern i el BSC, i obert a col·laboracions, amb l'objectiu últim d'aconseguir que la tecnologia entengui i parli la nostra llengua, és a dir,

ensenyar català a les màquines, de manera que la ciutadania pugui relacionar-se amb elles i participar plenament en el món digital en català.

Per assolir aquest objectiu, el projecte AINA està desenvolupant la infraestructura necessària perquè la inclusió del català a les aplicacions d'IA sigui prou atractiva i viable, tant per a les grans companyies tecnològiques com per a la indústria local, de manera que qualsevol empresa o organització pugui utilitzar els recursos generats per l'AINA, com ara els corpus (conjunts massius de dades) i els models de la llengua catalana, per desenvolupar les seves solucions o serveis específics (traductors, assistents personals, sintetitzadors de veu, classificadors de textos, etc.) en català.

A dia d'avui, el projecte AINA ha creat ja el "corpus de text" més gran que s'ha fet mai de la llengua catalana. Aquest corpus s'ha obtingut i continua creixent a base de descarregar textos de diferents fonts digitals en català (planes web, arxius, etc.) i processar-los per poder ser utilitzats com a dades d'entrenament per les xarxes neuronals que utilitzen els models de la llengua.

AINA ha començat a construir també un gran "corpus de veu" del català, que es nodreix principalment de les dades obtingudes a través de la iniciativa ["La nostra llengua és la teva veu"](#), impulsada pel Govern a principis del 2022 i consistent en una crida a la participació altruista de la ciutadania de parla catalana perquè doni la seva veu i validi l'aportada per altres persones a través de la plataforma de [Common Voice de Mozilla](#).

Concretament, les tasques del projecte AINA executades aquest 2022 són:

· **Desenvolupament de serveis lingüístics bàsics i transversals** (d'anonimització, de classificació de documents i

d'identificació d'entitats i conceptes clau, per exemple) que serveixen com a mòduls bàsics sobre els quals desenvolupar aplicacions i solucions per a l'usuari final

- **Compilació i preparació de dades massives i de qualitat** per entrenar models genèrics de llengua i models per a tasques específiques.

- **Publicació de nous models de la llengua**, generals i adaptats a una tasca concreta (per exemple, sistema de pregunta/resposta), llestos per servir de base per crear noves aplicacions.

- **Entrenament de models de reconeixement i síntesi de la parla** de qualitat per al català, que poden ser incorporats a aplicacions de veu més comuns del mercat.

- **Entrenament de motors de traducció automàtica de qualitat**, tant genèrics com adaptats a àmbits concrets,.

- **Implementació de prototips d'impacte** que permeten posar en producció els recursos generats i visibilitzar el seu ús i impacte

- **Difusió i promoció** de l'adopció de la tecnologia.

Entre els primers prototips desenvolupats al 2022, destaquen les noves veus sintètiques entrenades per AINA mitjançant el seu ús en un assistent virtual de l'empresa Bookline; una eina de transcripció automàtica (oTranscribe+) que permet i facilita l'edició garantint la privadesa de les dades; i un xatbot de veu que respon preguntes sobre el projecte AINA i que pot servir de base per crear altres experiències conversacionals en català.

Aquest 2023, es continuarà treballant en aquestes línies per ampliar els corpus de text i de veu, i els models de llengua entrenats a partir d'aquests corpus.

Al final del projecte AINA, el 2026, es disposarà de totes les

peces necessàries perquè qualsevol empresa o organització les pugui combinar per crear les seves solucions o serveis garantint que aquests entendran i parlaran correctament el català en qualsevol de les seves variants.