

Generació de Retrats Robots i Identificació de Persones



CIDAI-POC-2026-05

Drets reservats. Aquest treball està disponible sota la llicència Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International ([CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)).

Segons els termes d'aquesta llicència, podeu copiar, redistribuir i adaptar l'obra amb fins no comercials, sempre que l'obra sigui citada adequadament, tal com s'indica a continuació.

En qualsevol ús d'aquest treball, no s'ha de suggerir que el CIDAI doni suport a cap organització, producte o servei específic. No es permet l'ús del logotip CIDAI.

Si adapteu l'obra, heu de llicenciar-la amb la mateixa llicència Creative Commons o equivalent.

Si creeu una traducció d'aquest treball, heu d'afegir la següent exempció de responsabilitat juntament amb la cita suggerida: "Aquesta traducció no la va crear el Centre of Innovation for Data tech and Artificial Intelligence (CIDAI). CIDAI no es fa responsable del contingut ni de l'exactitud d'aquesta traducció. L'edició original en català serà l'edició autèntica i vinculant".

Qualsevol mediació relacionada amb disputes derivades de la llicència es durà a terme d'acord amb les normes de mediació de la [World Intellectual Property Organization](https://www.wipo.int/mediation/).

Cita suggerida. CIDAI-POC-2026-05 Generació de Retrats Robots i Identificació de Persones per Luis Lebron Casas, Fundació i2cat, CIDAI, 2026.

Llicència: [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1.	ABSTRACT	4
2.	IMPLEMENTACIÓ DE LA SOLUCIÓ	6
3.	RESULTATS	12
4.	CONCLUSIONS I POSSIBLES MILLORES	17
5.	AGENTS PARTICIPANTS	18
6.	BIBLIOGRAFIA	19



1. Abstract

1.1. Context i Problemàtica

El projecte de Retrats Robot IA sorgeix d'una necessitat crítica i creixent del cos de Mossos d'Esquadra, atès que actualment disposen d'una unitat centralitzada i altament especialitzada en la creació de retrats robot amb un personal molt reduït. Aquest factor provoca que la tècnica només s'utilitzi en casos molt específics i de suma gravetat, generant un coll d'ampolla operatiu. Aquest context evidencia la necessitat d'incorporar eines més precises, intuïtives i adaptables en la investigació policial diària. Addicionalment, els mètodes tradicionals requereixen llargues i feixugues sessions d'entrevista que sotmeten les víctimes i els testimonis a un alt desgast psicològic, allargant els processos d'investigació i provocant un risc elevat de victimització repetida.

1.2. Objectiu

L'objectiu principal és desenvolupar un pilot complet basat en Intel·ligència Artificial generativa d'imatges que es pugui integrar de forma fluida i segura amb el flux de treball habitual de la policia científica. Aquest sistema ofereix capacitats avançades dividides en mòduls:

- **Mòdul PORTRAITS:** Centrat en la generació assistida d'esbossos de retrats robots basats en descripcions detallades. Incorpora eines d'iteració i refinament continu (Text-a-Imatge, Imatge-a-Imatge i funcionalitats d'In-Painting) per permetre ajustaments precisos (com ara modificar únicament la forma del nas o els ulls) de manera gairebé instantània.
- **Mòdul IMAGIA:** Orientat a la generació d'imatges a partir de fotografies per a la identificació de persones. Aquest mòdul permet, a partir de poques fotografies d'entrenament, reimaginar un subjecte aplicant alteracions d'edat (envelliment o rejuveniment), condicions físiques, i inclou l'aplicació de tècniques de millora de la qualitat de la imatge així com la generació de retrats tridimensionals.

Assistent de veu multi-idioma: Integrat de manera transversal a l'eina per facilitar la interacció natural, reduint la barrera d'entrada tecnològica per als investigadors policials.

1.3. Impacte

L'adopció d'aquesta tecnologia permetrà estendre i “democratitzar” el procés de creació de retrats a un nombre molt més elevat de casos, convertint-se en una eina directament escalable a diferents comissaries de tot el territori català, sense la limitació o la dependència de la disponibilitat de personal exclusivament especialitzat. Des d'una perspectiva social i operativa, s'espera reduir dràsticament la victimització repetida gràcies a la rapidesa de les iteracions visuals, millorar la precisió en la identificació primerenca de sospitosos i facilitar enormement la càrrega de treball i difusió d'alertes per part dels professionals de la policia científica.

1.4. Solució

La solució tecnològica es basa en una arquitectura altament modular i resilient, dissenyada per a complir amb els estrictes requeriments de seguretat. Amb components de *Frontend* desenvolupats en Angular i *Backend* construït en Python, l'arquitectura es desplega íntegrament al núvol d'AWS. El sistema aprofita entorns serverless com ECS Fargate per a les API lleugeres i instàncies EC2 amb capacitat de GPU dedicada per al processament intensiu dels models d'IA. Mitjançant SQS, S3 i RDS, l'aplicació processa entrades en múltiples formats (text, imatge i veu), transformant-les en retrats robots amb filtres específics d'estil “foto de carnet” o “carbonet”, o bé aplicant complexes tècniques de millora i modelatge 3D mitjançant diversos models d'IA de propòsit específic.



2. Implementació de la solució

2.1. Fluxos de treball

El sistema empra una arquitectura asíncrona i desacoblada basada en cues de missatgeria (Amazon SQS) per gestionar de manera eficient les diferents càrregues de treball, tant d'inferència com d'entrenament. Aquesta estructura garanteix que les peticions intensives no bloquegin la interfície de l'investigador policial, permetent un escalat autònom i robust.

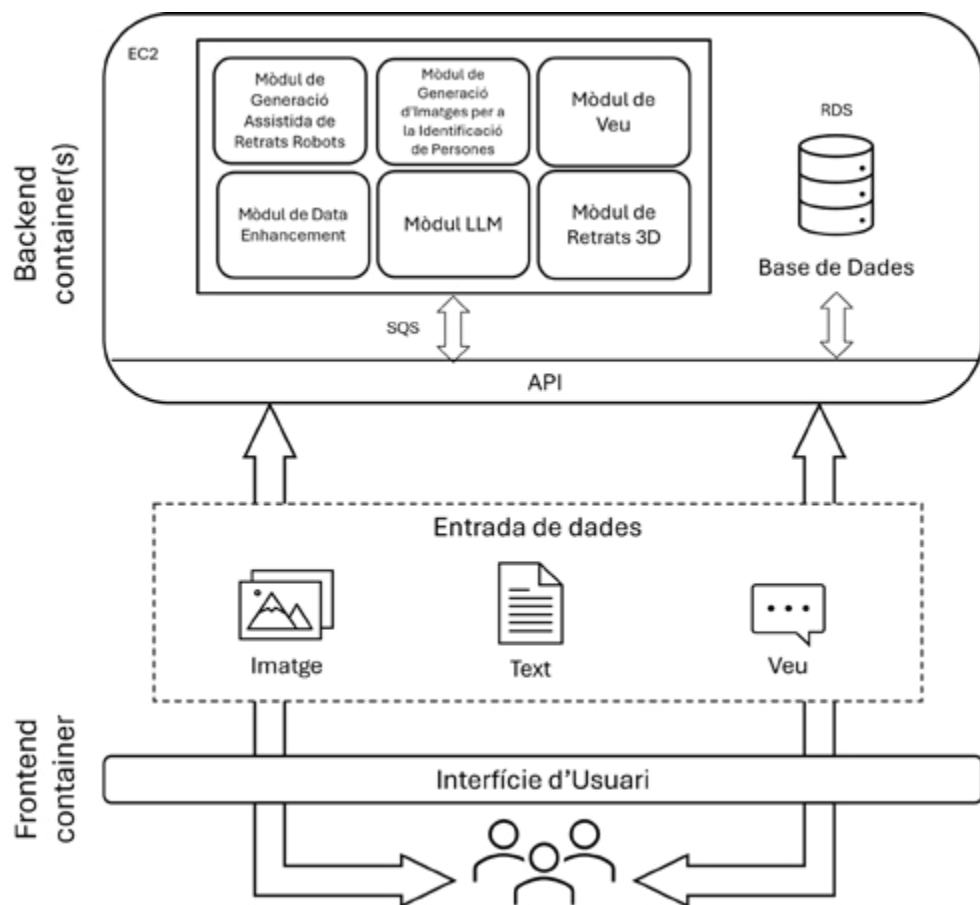
L'estructura de processament general segueix un cicle estandaritzat operatiu per a la plataforma, la qual es pot observar en la Figura 1:

Recepció i emmagatzematge: El *Frontend* envia la petició (imatge, àudio o text) a l'API pública. Si la petició inclou arxius pesants (com la gravació de veu o fotografies originals per a entrenament), l'API els desa de forma temporal i segura al servei d'emmagatzematge d'objectes Amazon S3.

- **Encuament:** L'API registra la petició a la base de dades i l'encua a la cua SQS corresponent segons el tipus de tasca (inferència estàndard, entrenament, o processament de models), adjuntant la referència del fitxer d'S3 si escau.
- **Processament (Consumidor):** Els contenidors Docker, actuant com a consumidors i desplegats als clústers d'ECS amb GPU, llegeixen el missatge de la cua de manera seqüencial. Descarreguen els arxius necessaris d'S3 i executen les operacions pertinents sobre els models d'IA de forma dedicada.

- **Notificació i neteja:** Un cop finalitzat el procés, el consumidor esborra l'arxiu temporal d'S3 i notifica a l'API que la tasca ha conclòs satisfactòriament. Finalment, l'API envia el resultat generat al *Frontend* en temps real a través de connexions WebSocket.

Figura 1: Esquema general de l'arquitectura.



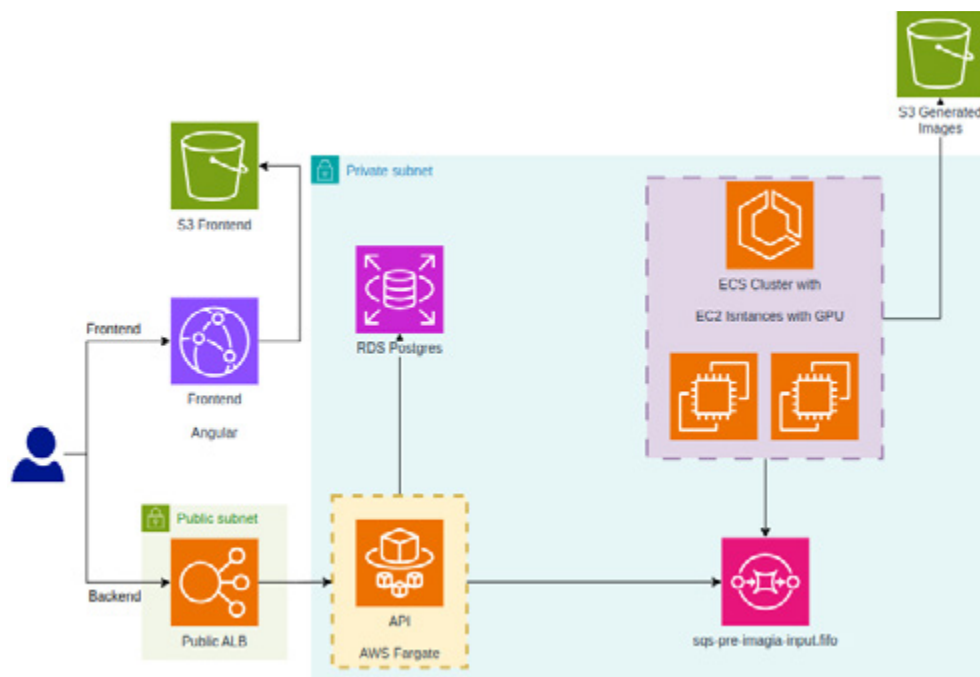
Aquest cicle tecnològic s'adapta i es divideix de manera especialitzada en funció del mòdul en ús:

Flux de treball per a PORTRAITS

Aquest mòdul està dissenyat per maximitzar la velocitat de resposta i facilitar la iteració contínua durant la creació d'esbossos. L'arquitectura especialitzada es pot veure en la Figura 2. En detall hi ha dos mòduls específics per a la inferència d'aquest flux de PORTRAITS:

- **Inferència d'Imatges amb Stable Diffusion XL [1]:** Les peticions de generació de nous retrats o de modificació de detalls (TXT2IMG, IMG2IMG, INPAINT) s'encuen a la cua d'inferència. El *Stable Diffusion inference cluster* recull aquestes sol·licituds i retorna els resultats generats a la interfície generalment en menys d'un minut (11 a 44 segons, depenent del nombre d'imatges).
- **Processament de Veu:** En cas que l'investigador utilitzi l'assistent de veu, la gravació d'àudio s'encua al clúster de models d'i2CAT. Allà, els models AINA o Whisper processen la transcripció gairebé en temps real, retornant el text perquè l'usuari el revisi abans de llançar la inferència de la imatge.

Figura 2: Esquema del mòdul de PORTRAITS.

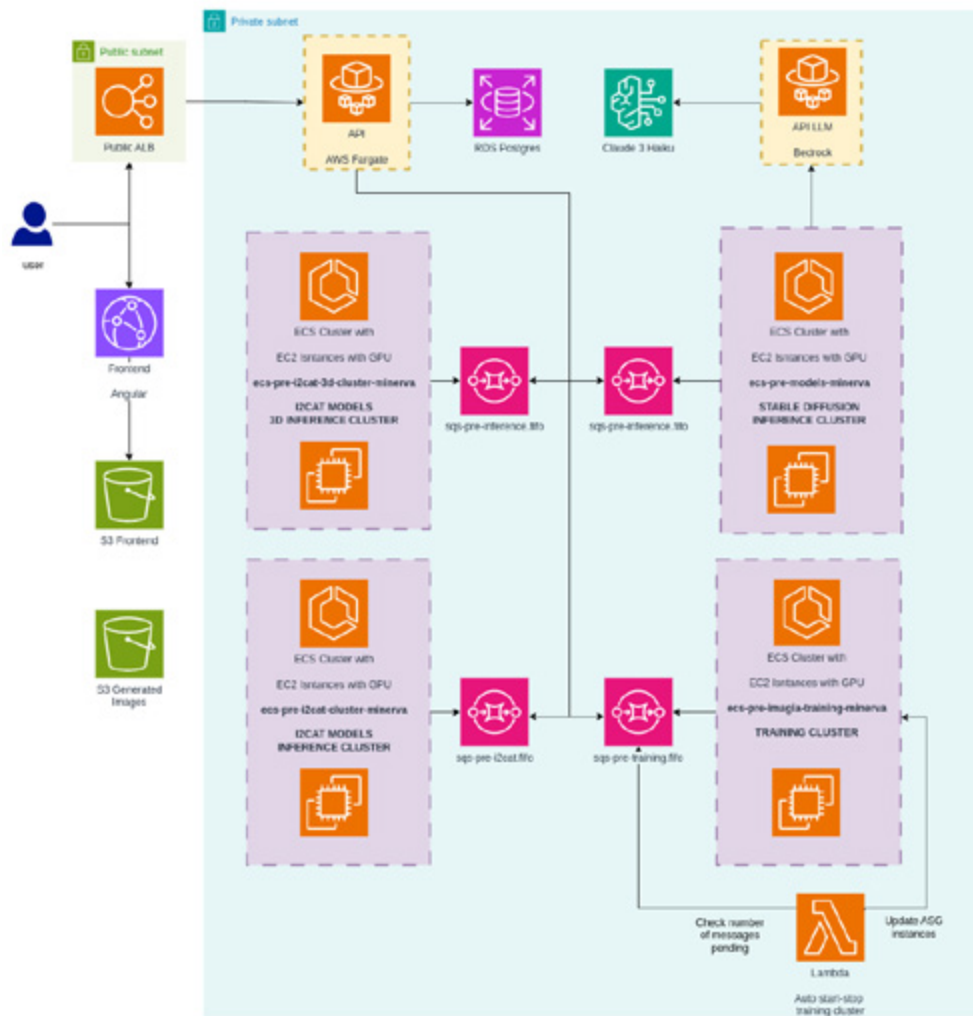


Flux de treball per a IMAGIA

Aquest mòdul gestiona processos de còmput molt més intensius, dedicats a la generació de noves imatges, restauració i modelatge tridimensional. La seva arquitectura específica està reflectida en la Figura 3. Consta dels següents mòduls addicionals:

- **Entrenament de Models (LoRA[4]):** El processament de fotografies d'entrenament per crear un model propi de la persona és la tasca més exigent (triga de 4 a 8 hores). Per optimitzar l'eficiència energètica i els costos, aquestes peticions van a una cua específica controlada per funcions AWS Lambda. Aquestes funcions encenen automàticament les instàncies EC2 amb GPU del *Training Cluster* només quan reben sol·licituds pendents, i les apaguen de seguida que l'entrenament finalitza.
- **Inferència amb Model Personalitzat:** Una vegada el model LoRA[4] està llest, les noves descripcions per reimaginar l'individu s'envien a la cua d'inferència estàndard. El clúster d'inferència d'Stable Diffusion XL [1] carrega llavors els pesos del LoRA[4] acabat d'entrenar per a generar la imatge alterada.
- **Data Enhancement i Retrats 3D:** Les peticions per a millorar fotografies de baixa qualitat o generar vídeos 360° es deriven cap al *I2CAT models 3D/inference cluster*. Aquest clúster especialitzat utilitza els models DifFace[2], Swimir[3] o Panohead[3] per processar l'arxiu i retornar el resultat d'alta resolució o el model tridimensional a l'investigador.

Figura 3: Esquema del mòdul de IMAGIA.



2.2. Modificacions de les imatges originals i generades

El sistema ofereix un conjunt d'eines avançades dissenyades específicament per al tractament i la manipulació detallada tant de les fotografies originals aportades com a prova forense (mòdul IMAGIA), com dels esbossos i retrats sintètics generats per la intel·ligència artificial (mòdul PORTRAITS).

Tractament d'Imatges Originals

Quan la investigació requereix processar fotografies reals d'un individu per entrenar el model d'IA, sovint es parteix de material imperfecte. El sistema permet optimitzar-lo mitjançant les següents eines:

- **Data Enhancement:** Si les fotografies aportades presenten problemes de resolució, enfocament, soroll digital o condicions d'il·luminació deficientes (fetes amb CCTV o dispositius antics), el sistema ofereix l'aplicació de tècniques de millora prèvies a l'entrenament. L'investigador pot escollir entre dues metodologies segons les necessitats del cas:
- **Swimir[3]:** Un model d'arquitectura Transformer que proporciona una millora subtil, reduint el soroll i augmentant la resolució mentre actua de forma molt conservadora amb la imatge original per evitar alteracions fisonòmiques innecessàries.

- **DiffFace[2]:** Un model de difusió dissenyat exclusivament per a la reconstrucció i escalat de cares humanes, capaç de lliurar un rostre de molt alta definició, idoni per a fotografies severament degradades (assumint sempre el risc inherent a possibles “al·lucinacions” de detall generades pel propi model).
- **Generador de Prompts (Captioning Automàtic):** Per tal que l'entrenament del model (LoRA[4]) sigui eficaç, les imatges requereixen descripcions que les acompanyin. El sistema integra una funcionalitat que analitza cada fotografia real pujada i en genera automàticament una descripció textual (caption). Això permet documentar automàticament trets físics, colors, indumentària i expressions, proporcionant a la IA les metadades necessàries per entendre perfectament el rostre de la persona abans de l'entrenament.

Modificació d'Imatges Generades

Per a la iteració interactiva i l'ajustament metòdic dels retrats robots sintètics (o esbossos), el sistema disposa de tres eines principals d'inferència:

- **TXT2IMG (Text-a-Imatge):** És l'eina de creació base. Permet generar imatges completament noves, similars a les d'un retrat robot, exclusivament a partir d'una descripció textual detallada (o dictada per veu). El sistema utilitza una lògica oculta per assegurar-se que els resultats mostrin una posició frontal i puguin adoptar de manera selectiva filtres visuals (ex. estil “carbonet” o “realista”). L'investigador disposa de

control sobre la Precisió (fidelitat forçada al text) i els Passos de procés (nombre de vegades que es repeteix el procés de generació del model).

- **IMG2IMG (Imatge-a-Imatge):** Funcionalitat dissenyada per aplicar canvis globals sobre una base visual existent. Permet introduir una imatge de referència (ja sigui generada prèviament, dibuixada a mà alçada o extreta d'altres eines policials) i, acompanyant-la d'un nou prompt descriptiu, generar-ne una evolució. A més dels paràmetres habituals, inclou el control d'Intensitat del canvi, que determina el pes que se li donarà a la nova instrucció escrita en detriment de la imatge base utilitzada, regulant així com d'allunyada o semblant serà la variació resultant.
- **INPAINT (In-painting):** L'eina de retoc focalitzat i d'alta precisió per excel·lència. Aquesta opció permet seleccionar, mitjançant una eina d'enquadrament, una àrea hiper-específica de la imatge (per exemple, només les celles, el nas o els llavis) i indicar-hi canvis descriptius textuals, com “celles més espesses”. El model aplicarà el canvi demanat exclusivament a la zona marcada, preservant intacta i coherent tota la resta de la fisonomia del rostre generat anteriorment, permetent polir el retrat fins al mínim detall amb el testimoni.

2.3. Model d'IA

La solució implementa una complexa bateria de models d'intel·ligència artificial, cadascun altament especialitzat per a executar tasques determinades dins de la plataforma:

- **Generació d'Imatges (Stable Diffusion XL):** S'utilitza el model de difusió latent Stable Diffusion XL [1] com a motor principal per a la creació base de tots els retrats, donada la seva gran capacitat de resolució i adherència al text.
- **LoRA[4]:** Tècnica essencial per a l'entrenament dinàmic i específic. En comptes de reentrenar completament el model base, cosa que requeriria recursos computacionals inassumibles, el LoRA[4] permet afegir i entrenar únicament un nombre reduït de nous pesos. D'aquesta manera, el sistema aconsegueix aprendre les faccions exactes d'una persona concreta utilitzant només entre 5 i 20 fotografies d'entrenament, en un temps de processament que oscil·la entre les 4 i les 8 hores.
- **Generació 3D (Panohead[3]):** Donat que sovint només es disposa d'una única imatge frontal 2D d'un subjecte, s'utilitza Panohead, un model basat en Diffusion Models combinat amb StyleGAN2 (usant la tècnica de tri-plane). Aquest model permet projectar i extreure un cap genèric 3D iterativament fins a coincidir amb el perfil de la fotografia frontal, aconseguint generar vistes en un eix complet de 360°.
- **Data Enhancement:**
 - **Swimir[3]:** Basat en arquitectures Transformer. S'utilitza per a aplicacions conservadores de denoising i super-resolució, on la prioritat absoluta és no alterar la veracitat de la prova gràfica original.
 - **DifFace[2]:** Basat en models de difusió, està exclusivament focalitzat en la restauració profunda i l'augment de qualitat de cares humanes fortament degradades.
- **Veu (AINA[6] + Whisper[7]):** La interacció vocal fa servir un model híbrid dissenyat per suportar ordres policials mitjançant una transcripció altament robusta i multi-idioma (optimitzat especialment per a l'ús de català, castellà i anglès).



3. Resultats

A continuació es detallen els resultats obtinguts amb el demostrador.

3.1. Mètriques de rendiment

L'avaluació exhaustiva dels models s'ha quantificat tenint en compte els temps de resposta en inferència/entrenament i aplicant mètriques de qualitat visual rigoroses per als models de Data Enhancement.

En el procés de generació iterativa d'imatges amb Stable Diffusion XL [1], s'ha determinat que el sistema requereix un paràmetre òptim d'entre 25 i 40 "passos" per lliurar la millor relació qualitat-fidelitat. Aquest ajustament resulta en temps de resposta extraordinàriament àgils per a l'entorn de l'investigador i el testimoni, mantenint un flux de treball interactiu:

- Generació d'1 imatge: ~11 segons.
- Generació d'un panell de 4 imatges: ~44 segons.

Entrenament complet d'un model LoRA[4] personalitzat: Entre 4 i 6 hores (temps operativament viable per a deixar processos actius en horari nocturn i disposar dels resultats l'endemà).

Pel que fa a la qualitat perceptual i al rendiment tècnic dels models de Data Enhancement, validats localment utilitzant el dataset CelebA[8] sobre infraestructures d'alt rendiment amb targetes GPU A6000, els resultats han evidenciat el següent:

Taula 1: Resultats utilitzan la base de dades de CelebA[8] per a calcular el rendiment dels models de Data Enhancement.

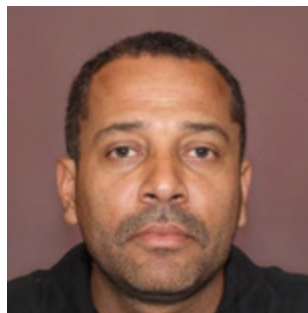
Model	# Params	VRAM	CLIP-IQA↑	SSIM↑	Temps (per imatge)
SwinIR	16 M	8GB	19.02	0.71	4 seg
DiffFace	176 M	8GB	56.41	0.67	14 seg

3.2. Exemples qualitatius

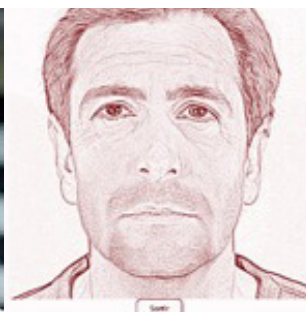
En aquest apartat s'exposen diversos exemples finals de les aplicacions desenvolupades. La Figura 4 mostra els resultats obtinguts amb el mòdul PORTRAITS. El procés ha consistit a partir d'una descripció inicial i realitzar diverses iteracions fins a seleccionar una imatge base satisfactòria; posteriorment, s'hi han aplicat ajustos detallats per assolir el resultat desitjat. Els resultats posen de manifest una dependència directa entre la interpretació humana de la descripció i l'execució del model d'IA. Un cas paradigmàtic és l'ús de l'expressió «pell morena»: mentre que en l'entorn espanyol s'associa sovint a trets nord-africans, els models entrenats majoritàriament en anglès vinculen aquests tons amb perfils sud-americans. Aquesta discrepància cultural es pot mitigar mitjançant l'ús de descripcions més específiques o instruint l'usuari sobre aquests matisos interpretatius.

Figura 4: Exemples de resultats del mòdul de PORTRAITS.

A l'esquerra trobem la imatge original, després la imatge resultat i al final la descripció inicial.



Sexe: home
Grup ètnic: sudamericana
Edat: 45
Pell: morena
Contorn de la cara: rectangular
Cabell: arrissat, curt, negre
Front: ample
Celles: fines, poc poblades
Ulls: prominents
Nas: curt, base ampla
Orelles: mitjanes
Boca: llavis gruixuts, comissures descendents
Mentó: quadrat
Pèl facial: barba molt curta
Línies facials: arrugues frontals, ulleres

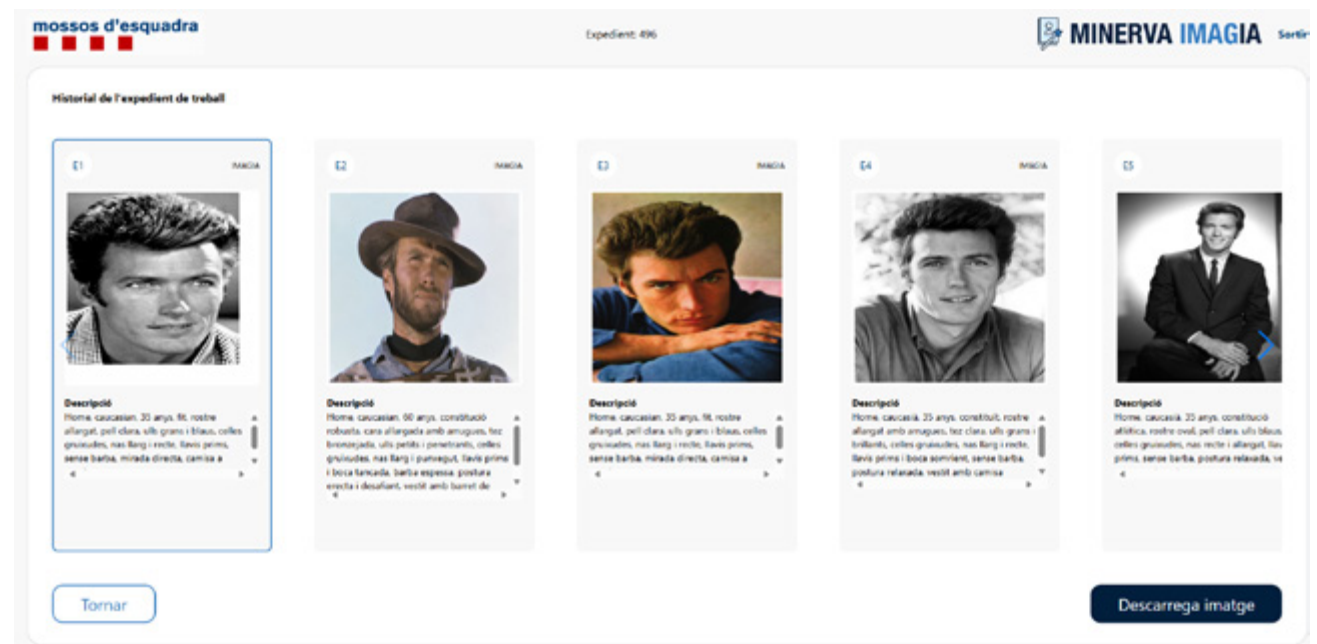


Sexe: Home
Grup ètnic: caucàsic
Edat: 49 anys
Pell: morena
Contorn de la cara: ovalat
Cabell: curt i canós
Front: front estret i baix
Celles: arquejades mig poblades i llargues
Ulls: amb parpelles superiors caigudes
Nas: nas mida mitjana amb la punta rodona
Orelles: enganxades
Boca: llavi superior prim i llavi inferior no tan prim
Mentó: forma aguda
Pèl facial: sense barba
Línies facials: arrugues nasobucals i arrugues infraorbitals

A la Figura 4 es mostra una prova feta per generar imatges de l'actor Clint Eastwood amb uns 80 anys a partir d'imatges de quan era jove, utilitzant el mòdul IMAGIA. Com que és una persona coneguda, disposem de moltes imatges de qualitat. En els resultats es poden veure petits biaixos, com ara que sempre manté les postures de les imatges de prova o que el model es decanta sempre per mostrar pèrdua de pèl. Els models utilitzats poden caure en aquest tipus de problemàtica, en què tenen tendència a enfocar-se massa en certes característiques genèriques i no focalitzar en les característiques diferenciadores de cada persona.

Figura 4: Exemple del mòdul IMAGIA.

En aquest exemple es demana generar imatges envellides del actor Clint Eastwood ara a partir d'imatges de ell en la època dels 60s que es poden veure en la primera fila amb les seves descripcions autogenerades.



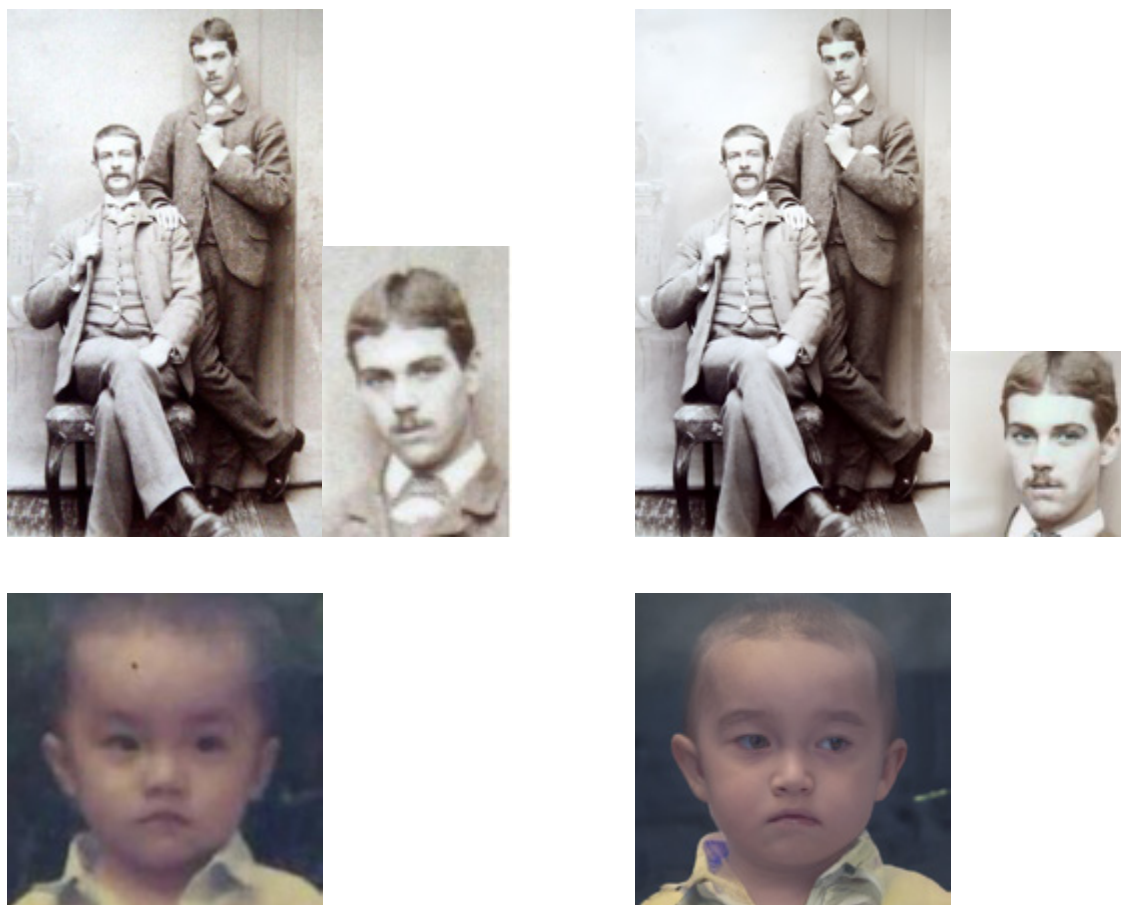
3.3. Limitacions del models

Des d'una perspectiva operativa i qualitativa, durant la fase de proves de camp amb la unitat científica, s'han visualitzat, documentat i advertit les següents desviacions (limitacions estructurals i errors d'algoritme) en el comportament dels models, les quals s'han de tenir molt presents en un entorn d'investigació penal:

DiffFace[2] (Risc d'Al·lucinacions): A causa de la seva naturalesa estocàstica i basada en models de difusió, si la imatge d'entrada està excessivament danyada, el model presenta el risc de patir "al·lucinacions" visuals. Això significa que podria inventar i introduir de manera artificial artefactes i detalls facials minúsculs que no estaven presents a la fotografia original, alterant potencialment elements identificatius clau. En el exemple de la Figura 6 es pot veure com el model modifica el color de una fotografia en tons sepia.

Figure 6: Exemple de al·lucinacions.

En el primer es veu com en una imatge en tons sepia afegeix ulls blaus.
En la segona es veu com converteix un nen asiàtic en un caucàsic.



Panohead[3] (3D Angles cecs): En la tasca de generar vistes de 360° partint exclusivament d'una única imatge frontal plana, el model té una forta tendència a generar detalls incongruents a les zones "cegues" (angles o parts del rostre i cap que romanien completament amagats a la fotografia original, com per exemple la forma posterior de les orelles o la disposició exacta del pentinat al clatell). Un exemple es pot veure a la Figura 7 on el model afegeix un color de cabell que no estava en la imatge original.

Stable Diffusion XL (LoRA[4] i perfils sensibles): S'ha observat que si a l'eina se li apliquen més de 60 passos de generació, la imatge resultant perd l'estètica funcional d'esbós i la fidelitat cau cap a una "sobre-definició" innecessària. A més, s'alerta que els entrenaments basats en imatges de menors d'edat (infants de menys de 12 anys aproximadament) generen resultats poc fiables i no garantits, a causa de l'absència de prou dades estructurals infantils en els conjunts d'entrenament del model base universal.

Figure 7

Exemple de que per falta d'informació el model no és capaç de reconstruir el darrere del cap. En aquest afegeix un color rosat a la part posterior.



4. Conclusions i possibles millores

La implementació amb èxit d'aquesta prova de concepte ha demostrat que és possible integrar models d'IA generativa multimodals d'una manera pràctica i intuïtiva per a l'entorn policial. L'ús estratègic de l'arquitectura d'AWS permet garantir una alta disponibilitat i l'escalabilitat desitjada per estendre l'eina a múltiples comissaries sense col·lapsar la infraestructura. El mòdul IMAGIA demostra una capacitat d'adaptació fidel treballant amb entrenaments LoRA[4] sobre espais latents, malgrat que la consistència i qualitat del resultat final depèn estretament de subministrar fotografies inicials ben enquadrades, il·luminades i pròximes al format "foto de carnet". L'aplicació redueix els temps de confecció dràsticament, afavorint el confort psicològic dels testimonis. Actualment la arquitectura es modula amb una certa dependència a AWS però la tecnologia és fàcilment substituïble per versions millors dels models o modificacions. S'ha treballat conjuntament amb la unitat de Mossos per garantir un nivell de seguretat adient. Malgrat això en el cas que es volgués desplegar en un entorn real, fora d'aquesta prova de concepte, es tindria que realitzar un seguit de tasques de documentació, control i modificacions del software per al correcte compliment del AI ACT.

Possibles millores:

- **Automatització de preprocessament d'imatges:** Integrar eines d'IA automàtiques a l'aplicació que permetin detectar la geometria facial, retallar l'entorn innecessari (*crop*) i fer *zoom* de manera intel·ligent. Això garantiria que la cara de la persona quedi correctament centrada i estandarditzada per a optimitzar de manera directa l'entrenament dels models LoRA[4], eliminant per complet la necessitat que l'investigador hagi de preparar i quadrar les imatges manualment en programari fotogràfic abans de pujar-les a la plataforma.
- **Afinament de l'Encoder per a entorns adversos:** Aprofundir en tècniques avançades d'entrenament específic sobre l'encoder del model Stable Diffusion XL [1] per aconseguir disminuir els artefactes quan la fotografia base ha estat extreta de càmeres de seguretat (CCTV). Aquestes imatges forenses reals solen contenir variacions d'il·luminació dràstiques, distorsions de lents, soroll d'alta ISO o angles zenitals no òptims que els models actuals encara lluiten per processar amb precisió impecable.
- **Millora de la fidelitat de les imatges i les descripcions en ètnies minoritàries i gènere:** Quan el model presenta un alt grau de llibertat, s'observen biaixos derivats dels conjunts de dades d'entrenament. Això dificulta la inclusió de persones pertanyents a ètnies minoritàries o amb característiques molt específiques.



5. Agents participants

Les següents són les entitats i agents essencials participants en el disseny i la implementació d'aquesta prova de concepte d'innovació tecnològica:

- **Mossos d'Esquadra (Unitat Central de Fotografia i Audiovisuals):** Responsables de la definició de la necessitat inicial, l'aportació del coneixement tàctic del cas d'ús i de l'establiment dels estrictes requeriments operatius. Així mateix, han actuat com a garants de la validació en entorn real, realitzant les proves de les funcionalitats i auditant l'exactitud i la fidelitat policial dels esbossos i les variacions gràfiques obtingudes.
- **IThinkUPC:** Encarregats exclusius del desenvolupament tecnològic de l'aplicació (plataforma PORTRAITS i IMAGIA). La seva intervenció cobreix el disseny complet de la interfície i de l'experiència d'usuari, la programació de la lògica de negoci i orquestració, així com la integració total de solucions, el desplegament en entorn cloud i l'administració de la infraestructura d'alt rendiment a AWS.
- **Fundació i2CAT:** Centre de recerca i innovació tecnològica que ha exercit com a proveïdor científic de base i administrador expert del subministrament directe de components d'intel·ligència artificial. Entre les seves contribucions destaquen el proveïment i integració de l'assistent de veu híbrid, la implementació arquitectònica dels models de projecció 3D i l'ajustament rigorós dels diferents algorismes d'inferència per a la resolució visual i el Data Enhancement.



6. Bibliografia

- Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., ... & Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.
- Yue, Z., & Loy, C. C. (2024). Difface: Blind face restoration with diffused error contraction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 9991-10004.
- An, S., Xu, H., Shi, Y., Song, G., Ogras, U. Y., & Luo, L. (2023). Panohead: Geometry-aware 3d full-head synthesis in 360deg. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 20950-20959).
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2022). Lora: Low-rank adaptation of large language models. Iclr, 1(2), 3.
- Projecte Aina. (n.d.). Inici - Aina kit. Barcelona Supercomputing Center. Retrieved March 31, 2026, from <https://langtech-bsc.gitbook.io/aina-kit>
- Radford, Alec, et al. "Robust speech recognition via large-scale weak supervision." International conference on machine learning. PMLR, 2023.
- Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).

Socis



Membres

