
This is the **published version** of the bachelor thesis:

Vila Simón, Pol. *Sistema d'Assistència Intel·ligent basat en RAG per a la Gestió d'Alertes Sensibles en Entorns Educatius*. Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/334399>

under the terms of the  license.

Sistema d'Assistència Intel·ligent basat en RAG per a la Gestió d'Alertes Sensibles en Entorns Educatius

Pol Vila Simon

30 de juny de 2026

Resum— Aquest treball presenta el disseny, el desenvolupament i l'avaluació d'un sistema d'assistència intel·ligent basat en l'arquitectura *Retrieval-Augmented Generation* (RAG) per donar suport a la gestió d'alertes educatives sensibles en entorns escolars. El projecte neix de la dificultat dels equips docents i responsables de convivència per interpretar la normativa i els protocols d'actuació davant situacions de risc que afecten menors. Per donar resposta a aquesta necessitat, es desenvolupa un sistema integrat a la plataforma b-resol, capaç d'analitzar alertes i consultes en llenguatge natural i recuperar informació rellevant d'una base documental orientada al domini educatiu. L'enfocament adoptat busca reduir les limitacions dels grans models de llenguatge (LLM), millorant la traçabilitat, la contextualització i la fiabilitat de les respostes. Finalment, el sistema s'avalua des d'una perspectiva tècnica, funcional i qualitativa amb professionals del sector.

Paraules clau— Intel·ligència artificial generativa, Retrieval-Augmented Generation (RAG), embeddings, bases de dades vectorials, protecció del menor, entorns educatius.

Abstract— This work presents the design, development, and evaluation of an intelligent assistance system based on a *Retrieval-Augmented Generation* (RAG) architecture to support the management of sensitive educational alerts in school environments. The project arises from the difficulty faced by teachers and coexistence coordinators in interpreting regulations and action protocols in situations involving risks affecting minors. To address this need, a system integrated into the b-resol platform is developed, capable of analyzing alerts and queries in natural language and retrieving relevant information from a documentary base oriented to the educational domain. The proposed approach aims to reduce the limitations of large language models (LLMs) by improving the traceability, contextualization, and reliability of the generated responses. Finally, the system is evaluated from technical, functional, and qualitative perspectives with professionals in the field.

Keywords— Generative Artificial Intelligence, Retrieval-Augmented Generation (RAG), embeddings, vector databases, child protection, educational environments.



1 INTRODUCCIÓ - CONTEXT DEL TREBALL

ACTUALMENT, el sistema educatiu disposa cada vegada més de canals digitals per facilitar la comunicació i la detecció precoç de situacions de risc que afecten els menors, com ara casos d'assetjament es-

colar, violència, discriminació, trastorns alimentaris o altres conflictes entre infants, amb la finalitat de protegir-los i garantir-ne el benestar. En aquest context, l'empresa Bcn-resol desenvolupa eines com b-resol, una plataforma digital que permet als menors comunicar incidències de manera immediata, segura i accessible a través d'una aplicació mòbil. Aquestes alertes són rebudes pel centre educatiu de l'alumne mitjançant una plataforma web que en facilita la gestió, el registre de les actuacions realitzades i la generació d'informes i estadístiques.

Tot i l'eficiència de b-resol en la recepció d'alertes, la seva gestió posterior continua representant un repte important.

- E-mail de contacte: polvila25@gmail.com
- Menció: Computació
- Treball tutoritzat per: Daniel Ponsa Mussarra (CVC)
- Curs 2025/26

La problemàtica principal es deu al fet que els equips docents i responsables de convivència no sempre disposen dels coneixements jurídics i procedimentals necessaris per interpretar correctament el marc normatiu vigent, format per la LOPIVI (Ley Orgánica de Protección Integral a la Infancia y la Adolescencia frente a la Violencia), lleis i decrets, els protocols del Departament d'Educació i la normativa interna de cada centre [1] [2]. Aquesta situació dificulta determinar el procés d'actuació adequat, incrementa la càrrega del centre i fa que la gestió de les incidències sigui menys proactiva, menys eficient i més dependent de la interpretació individual.

Com a conseqüència, aquesta situació incrementa el risc d'errors procedimentals i retards en l'activació de mesures de protecció, fet que pot mantenir el menor més temps desprotegit en una situació de risc i generar inseguretat dins del centre.

Davant aquest escenari, l'ús directe de grans models de llenguatge (LLM), de l'anglès *Large Language Models*, no constitueix una eina de suport prou fiable per a la gestió d'aquestes alertes. En tractar-se de models generalistes, poden generar respostes convincentes però incorrectes o insuficientment fonamentades, i no poden conèixer de manera fiable la normativa interna del centre ni els protocols específics de cada territori. Per aquest motiu, en un àmbit tan sensible, on les decisions afecten la protecció del menor, és necessari disposar de respostes contextualitzades, traçables i justificades a partir de fonts normatives verificables, un requisit que els models lingüístics generals no poden garantir per si sols.

Per donar resposta a aquestes limitacions, aquest projecte proposa l'ús d'una arquitectura Retrieval-Augmented Generation (RAG) [3][4]. Aquest enfocament combina la capacitat generativa dels LLM amb un sistema de recuperació d'informació sobre una base documental externa, controlada i validada. Així, quan un docent formula una consulta, el sistema recupera primer els fragments més rellevants de la normativa i dels protocols disponibles i genera la resposta a partir d'aquest context. D'aquesta manera, es redueixen errors, es millora la contextualització i s'afavoreix la traçabilitat de les respostes, alhora que es disminueix el risc d'al·lucinacions [5]. A més, aquest plantejament permet incorporar documentació específica del centre, un fet especialment rellevant per adaptar les respostes al context real d'aplicació.

En conclusió, l'arquitectura RAG es planteja com una solució més segura, precisa i adaptada a les necessitats dels centres educatius, sempre sota supervisió humana.

2 OBJECTIUS

L'objectiu general d'aquest projecte és desenvolupar un sistema de suport que ajudi els docents i els responsables de convivència dels centres educatius en la gestió d'alertes sensibles, guiant-los en la consulta, interpretació i aplicació del marc normatiu vigent a Catalunya i dels protocols específics aplicables a cada cas. S'implementa a través d'un sistema amb xat conversacional dissenyat per integrar-se dins l'ecosistema de b-resol.

Per assolir aquest objectiu general, s'han definit els següents objectius específics:

- **Processar i estructurar la documentació:** Recopilar, digitalitzar i classificar la normativa i els procediments rellevants per construir una base de coneixement coherent, actualitzable i optimitzada per al seu tractament semàntic.
- **Dissenyar i implementar l'arquitectura RAG:** Definir la base de dades vectorial i el model d'*embeddings* òptim dins l'ecosistema de b-resol. Aquest repte, també inclou integrar els processos de fragmentació documental, indexació i generació amb el LLM, garantint respostes contextualitzades, traçables i amb un risc mínim d'al·lucinacions.
- **Construir la interfície d'interacció:** Implementar un entorn conversacional intuïtiu i transparent que permeti als responsables de convivència i als equips docents introduir, en llenguatge natural, tant consultes com alertes relacionades amb situacions de risc o conflictes detectats al centre educatiu. L'objectiu és facilitar la interpretació de cada cas i visualitzar de manera clara tant les respostes generades com les fonts normatives i procedimentals emprades.
- **Validar i avaluar el sistema:** Analitzar el sistema final des d'una doble perspectiva. D'una banda, una avaluació tècnica per mesurar la qualitat i traçabilitat de les respostes; i de l'altra, una validació qualitativa per part de professionals del sector educatiu sobre la seva utilitat i adequació en la gestió d'alertes reals.

3 ESTAT DE L'ART

En els darrers anys, els grans models de llenguatge (LLM) s'han convertit en una de les tecnologies més destacades en l'àmbit del processament del llenguatge natural. La seva capacitat per comprendre consultes i generar respostes coherents ha impulsat el seu ús en tasques d'assistència, síntesi i consulta d'informació. A més, aquestes eines han popularitzat una nova manera d'accedir al coneixement, ja que permeten interactuar amb la informació de forma més natural i obtenir respostes contextualitzades sense necessitat de formular consultes complexes [6]. Tot i això, quan s'apliquen en entorns sensibles i regulats, com el de la gestió d'alertes en centres educatius, és necessari analitzar amb detall tant el seu potencial com les seves limitacions.

3.1 Limitacions dels LLM

En l'àmbit educatiu, diferents treballs destaquen el potencial dels LLM com a eina de suport per a estudiants, docents i institucions, especialment quan cal accedir de manera àgil a grans volums d'informació [6].

Tot i això, aquests models presenten limitacions importants quan s'utilitzen en entorns on la fiabilitat és crítica. Una de les principals són les al·lucinacions [5], és a dir, la generació de respostes aparentment certes però incorrectes o no fonamentades, així com la manca de traçabilitat sobre l'origen de la informació generada. En àmbits normatius o procedimentals, aquest problema és especialment greu, ja que una resposta pot semblar adequada i, tot i així, no correspondre amb la normativa real aplicable.

Una altra limitació clau és que els LLM no poden conèixer de manera fiable la documentació específica de cada organització ni la normativa pròpia de cada comunitat autònoma, com la de Catalunya, si aquesta no s'incorpora explícitament al sistema. Per tant, tot i el seu potencial, el seu ús directe no és suficient en un entorn on les respostes han de ser exactes, contextualitzades i traçables.

3.2 Arquitectures RAG

Per superar aquestes limitacions, les arquitectures RAG constitueixen una solució eficaç per combinar la capacitat generativa dels LLM amb informació específica de domini. Un sistema RAG complementa el coneixement intern del model amb documentació externa recuperable, de manera que la resposta es construeix a partir de fragments documentals rellevants [3].

Aquest enfocament es basa en la fragmentació dels documents, la seva representació semàntica mitjançant *embeddings* i la recuperació dels fragments més rellevants a través d'una base de dades vectorial [4]. Així, la generació es fonamenta tant en el coneixement entrenat del model com en una base documental específica, controlada i actualitzable.

En aquest treball, aquesta arquitectura és especialment adequada perquè permet operar sobre una base documental formada per normativa legal, protocols i circuits d'actuació. Això permet generar respostes més fonamentades, contextualitzades i traçables, reduint així el risc d'al·lucinacions i millorant la fiabilitat del sistema en un entorn especialment sensible.

En la mateixa línia, existeixen treballs relacionats amb el plantejament d'aquest projecte que avaluen arquitectures RAG per a preguntes i respostes sobre documents normatius o de procediments institucionals [7]. Aquests estudis mostren que la recuperació basada en fragments documentals millora la rellevància i la fidelitat de les respostes respecte d'un model sense recuperació. Això és especialment útil per a aquest treball, ja que reforça la idea que, quan la resposta ha d'estar recolzada en documentació normativa concreta, l'arquitectura RAG ofereix una base més adequada que l'ús directe d'un LLM.

En conseqüència, les arquitectures RAG constitueixen una alternativa especialment adequada per a la gestió d'alertes en un entorn amb menors, on la precisió, la justificació de la resposta i la supervisió humana són elements essencials.

4 METODOLOGIA I PLANIFICACIÓ

El desenvolupament d'aquest projecte s'ha dut a terme adoptant una metodologia àgil, de caràcter iteratiu i incremental. Aquest enfocament ha permès adaptar el sistema a mesura que s'aprofundia en la complexitat normativa i tecnològica, afavorint la integració progressiva dels components principals del sistema, tant del backend associat al motor RAG com de la interfície d'usuari.

El seguiment del projecte s'ha realitzat mitjançant reunions periòdiques amb el tutor acadèmic i reunions amb el tutor de Bcn-resol aproximadament cada dues o tres setmanes, amb la finalitat de revisar l'evolució del treball, validar decisions i recollir valoracions sobre l'adequació del prototip al context real d'ús.

Per a la gestió del desenvolupament s'ha utilitzat Git [8] com a sistema de control de versions, fet que ha permès mantenir un historial ordenat dels canvis i assegurar la traçabilitat entre les implementacions realitzades.

A nivell metodològic general, el cicle de vida s'estructura en tres grans etapes: (1) Identificació del problema, (2) Definició d'objectius, (3) Implementació i avaluació dels mètodes.

El projecte ha seguit aquestes etapes, agrupant i integrant les fases d'execució planificades de la manera següent:

(1) Identificació del problema i (2) Definició d'objectius
Aquestes etapes estableixen les bases del projecte i s'han realitzat mitjançant les dues primeres fases de treball:

- **Fase de Planificació i Especificació:** Definició de l'abast del projecte, establiment dels requisits funcionals i no funcionals, i distribució de la càrrega de treball (hores i paquets) en coordinació amb Bcn-resol.
- **Fase d'Exploració i Fonamentació:** Recerca bibliogràfica per construir el marc teòric legal i estudi de l'estat de l'art de les tecnologies d'IA generativa.

(3) Desenvolupament iteratiu i avaluació del prototip

Aquesta etapa central agrupa el nucli tècnic del projecte i la seva validació. El desenvolupament s'ha basat en un prototip funcional construït de manera iterativa i incremental mitjançant quatre iteracions principals. Cada iteració ha incorporat millores funcionals, proves internes i revisions posteriors amb el tutor acadèmic i Bcn-resol. Paral·lelament, la interfície desenvolupada amb Streamlit s'ha utilitzat com a entorn de prototipatge i validació contínua, permetent simular el comportament del sistema, visualitzar les respostes i verificar la traçabilitat documental.

- **Iteració I – Disseny inicial i prototip mínim:** Definició de l'estructura general del sistema i de la interacció entre els components principals, prioritzant la seguretat i la privacitat des del disseny. Construcció d'un primer prototip basat en una arquitectura RAG bàsica amb un únic document per validar el flux inicial de consulta, recuperació i generació.
- **Iteració II – Base documental i indexació vectorial:** Ampliació de la base documental, definició de l'estratègia de fragmentació semàntica i enriquiment dels fragments amb metadades. Generació dels *embeddings* i configuració de Qdrant com a base de dades vectorial.
- **Iteració III – Disseny Advanced RAG:** Incorporació dels components principals del flux RAG avançat: anàlisi prèvia de la consulta, avaluació de completitud, classificació de la intenció documental, enrutament, recuperació amb filtres i *reranking*.
- **Iteració IV – Integració funcional i validació final:** Millora de la generació de respostes mitjançant un *prompt* dinàmic, integració de la traçabilitat documental a la interfície i incorporació de funcionalitats específiques com el xat de protocol intern del centre i la generació automàtica de l'informe b-resol. Finalment, es van realitzar les proves completes del sistema amb casos representatius i la recollida del *feedback* final dels professionals.

- **Avaluació contínua:** L'avaluació no s'ha plantejat com una fase única al final del projecte, sinó com un procés continu. Al final de cada iteració s'han realitzat proves internes i revisions amb Bcn-resol, que han funcionat com a *sprint reviews*. Aquestes revisions han permès ajustar progressivament el sistema segons el feedback rebut.

5 ARQUITECTURA I IMPLEMENTACIÓ DEL SISTEMA

5.1 Visió general del sistema

El sistema desenvolupat es basa en una arquitectura RAG orientada a la gestió d'alertes educatives sensibles en el context de b-resol.

A nivell funcional, el sistema rep una alerta o consulta introduïda pel docent, i processa la informació per guiar-lo seguint tres fases principals:

- **Anàlisi de la consulta:** En aquesta primera fase, el sistema interpreta l'entrada de l'usuari mitjançant dos processos analítics executats en paral·lel. D'una banda, un mòdul de classificació analitza el contingut per identificar la naturalesa del cas, la categoria de risc i el nivell d'urgència. De manera simultània, un segon mòdul avalua la completitud i suficiència de la informació aportada. Aquesta doble anàlisi permet determinar si la consulta disposa de prou context per activar la fase de recuperació documental o si cal prioritzar una fase prèvia d'indagació.
- **Recuperació documental:** Si la consulta requereix suport documental, s'activa la fase de recuperació. El sistema transforma la consulta enriquida en una representació semàntica vectorial i executa una cerca sobre la base documental indexada mitjançant similitud semàntica. Aquest procés es complementa amb un mecanisme de prefiltratge basat en metadades i una fase posterior de *reranking*, que permet seleccionar els fragments candidats que presenten una major rellevància contextual abans de construir el context final que serà utilitzat pel model generatiu.
- **Generació:** Un cop seleccionats els fragments documentals més rellevants, el sistema construeix un context estructurat que s'envia al model generatiu. Aquest context inclou tant el contingut textual com les metadades necessàries per citar correctament les fonts utilitzades. A partir d'aquest context i d'un *prompt* generat pel propi sistema (les regles i instruccions de comportament que se li donen al LLM) adaptat al tipus de cas, el model genera una resposta final en català, amb un to institucional i empàtic, indicant les fonts documentals utilitzades.

Finalment, el sistema ha estat dissenyat per operar de manera integrada, adaptant-se a l'ecosistema de la plataforma b-resol. En lloc d'emetre accions aïllades, el sistema coordina les seves recomanacions amb les funcionalitats de l'aplicació (com ara el xat segur per a la indagació de dades o el panell de gestió de casos), garantint que les propostes d'actuació siguin totalment coherents amb la realitat operativa de b-resol.

5.2 Estructura modular

Per tal de garantir la robustesa i escalabilitat del sistema, la solució s'ha dissenyat seguint un enfocament desacoblat que s'estructura en quatre blocs fonamentals: la preparació de la base documental i indexació vectorial, l'anàlisi i preres recuperació de la consulta, la recuperació i reordenació semàntica de fragments, i finalment, la generació condicionada de la resposta. Cada component opera com una capa independent dins de l'arquitectura global, comunicant-se mitjançant estructures de dades ben definides.

Aquesta modularitat és clau per al manteniment del projecte, ja que permet l'actualització o substitució independent de qualsevol dels mòduls. A més, aquesta flexibilitat afavoreix futures ampliacions funcionals, com ara la incorporació de noves funcionalitats, protocols o l'adaptació de la base de coneixement a les necessitats de cada centre educatiu.

5.3 Preparació Documental i Base de Dades Vectorial

La qualitat d'un sistema RAG depèn en bona mesura de la fase prèvia d'ingestió, que comprèn la neteja, la segmentació i l'emmagatzematge dels fragments. En el sistema desenvolupat, aquesta fase ha permès consolidar una base documental formada per 926 fragments (*chunks*), extrets dels documents de la base documental i incorporats al projecte.

El primer pas ha consistit a processar els documents normatius originals, extreient-ne el text dels arxius PDF mitjançant la llibreria PyMuPDF [9], fet que ha permès obtenir un text net, eliminant capçaleres i peus de pàgina.

Per tal d'optimitzar la precisió de la futura recuperació, no s'ha aplicat un *chunking* uniforme basat únicament en una mida fixa de text. En el seu lloc, s'ha optat per una estratègia de fragmentació adaptativa o *semantic chunking* (vegeu Apèndix A.1), que respecta l'estructura lògica de cada tipus de document:

- **Circuits d'actuació:** Aquests diagrames visuals s'han fragmentat convertint cada node del diagrama de flux en un *chunk* independent, assegurant la captura individual de cada fase.
- **Protocols oficials i guies:** La divisió respecta la jerarquia de capçaleres del document original per tal d'aïllar seccions funcionals de forma íntegra.
- **Lleis i decrets:** S'ha implementat una segmentació basada en expressions regulars (*Regex*) [10] que aïlla estrictament cada article legal i els seus apartats numerats, evitant així la barreja de normatives i preservant l'exactitud de la capa jurídica.
- **Normativa pròpia del centre:** A diferència dels documents anteriors, els reglaments interns del centre no es fragmenten ni s'indexen vectorialment a la base de dades. S'ingesten íntegrament en temps real dins la finestra de context del model.

Per a l'emmagatzematge i la posterior cerca d'aquests fragments, s'ha optat per utilitzar una base de dades vectorial en lloc d'una base relacional tradicional (vegeu Apèndix

A.2). Aquesta elecció permet recuperar informació segons la similitud semàntica mitjançant *embeddings* [11], oferint la flexibilitat necessària per atendre consultes on el vocabulari de l'usuari no coincideix literalment amb el text normatiu. S'ha seleccionat Qdrant [12] com a gestor vectorial per la seva eficiència i per presentar una arquitectura *open source*, escalable i lleugera.

Un punt clau d'aquesta integració és l'ús avançat de metadades. Amb la finalitat de millorar la precisió i la traçabilitat del sistema, cada fragment segmentat es complementa amb informació clau, com el document d'origen, la pàgina, el títol, el tipus documental i la categoria de risc, que s'insereix directament al *payload* de Qdrant. Aquesta estructura permet aplicar filtres sobre les metadades (*payload filtering*) [13], restringint l'espai de cerca semàntica i millorant la rellevància dels fragments recuperats.

Finalment, es generen els *embeddings* de tots els fragments i s'indexen a Qdrant com a punts dins d'una col·lecció vectorial, garantint la traçabilitat completa de cada fragment recuperat.

5.4 Arquitectura Advanced RAG

El funcionament intern del sistema de suport dissenyat per a b-resol s'estructura al voltant d'un flux de treball seqüencial que optimitza la gestió de les alertes rebudes.

La Fig. 1 presenta l'esquema general d'aquesta arquitectura Advanced RAG, mentre que la seva versió detallada es pot consultar a l'Apèndix A.3. L'arquitectura es divideix conceptualment en tres grans fases: la prerecuperació, la recuperació i la generació i sortida.

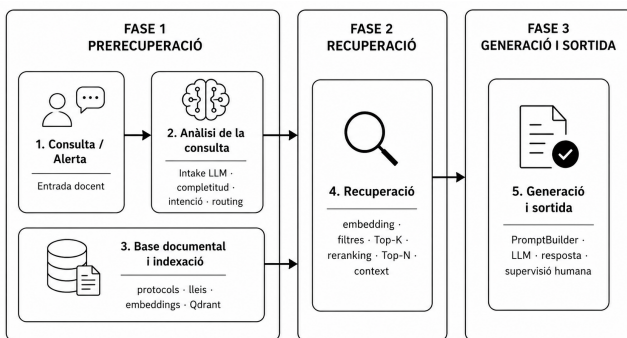


Fig. 1: Esquema Arquitectura Advanced RAG.

5.4.1 I. Anàlisi de la consulta i prerecuperació

La primera fase del flux d'execució correspon a l'anàlisi de l'entrada textual abans d'activar la recuperació documental. Aquesta entrada, introduïda pel docent, pot correspondre tant a una consulta com a una observació o a una alerta d'un alumne rebuda a través de la plataforma b-resol. Aquesta etapa es defineix com una fase de prerecuperació. Considerant que, en el context de b-resol, les alertes poden arribar incompletes, ambigües o anònimes, aquesta fase actua com un filtre semàntic, procedimental i de seguretat que interpreta, estructura i enruta la informació.

En primer lloc, la consulta és analitzada mitjançant un LLM configurat amb baixa temperatura per afavorir respostes més deterministes i reduir la variabilitat [5]. El model

extreu de la consulta una estructura de dades amb la categoria de risc, els indicadors detectats, el rol de l'informant i l'existència d'urgències. La finalitat és estructurar la informació de l'alerta o consulta per facilitar-ne la gestió posterior.

A continuació, un algoritme avalua la completitud de la informació. Aquesta etapa funciona com un filtre lògic abans del *retrieval*: penalitza les consultes massa breus o ambigües per determinar de forma segura si és possible activar directament la cerca RAG o si, per contra, cal prioritzar un procés d'indagació preventiva mitjançant preguntes de suport a través del xat. Si la consulta no conté dades mínimes, el sistema bloqueja l'activació del RAG i orienta la recollida d'informació addicional. Si la informació és parcial, el sistema permet avançar cap a la fase de recuperació, però conserva els elements pendents de confirmar. En alertes anònimes, aquesta lògica evita forçar identificacions directes i prioritza preguntes segures i no invasives.

Seguidament, s'aplica una heurística de classificació per determinar la intenció documental de la consulta (d'aplicació pràctica, legal, mixta o fora de domini). Això actua com un mecanisme de control que delimita l'espai de recuperació documental i permet diferenciar el que realment necessita el docent, reduint el risc de generar respostes desalineades amb les necessitats reals del cas plantejat.

Amb totes aquestes variables, el sistema executa un enrutament determinista (*routing*) basat en regles predefinides. Aquesta capa de decisió defineix el recorregut dins del *pipeline*, establint si s'autoritza la cerca RAG, el tipus de resposta a construir i les restriccions aplicables.

Finalment, quan l'enrutament autoritza la fase de recuperació, el sistema prepara la consulta mitjançant la generació d'una cerca enriquida. Per fer-ho, reescriu de manera automàtica la pregunta de l'usuari integrant-hi el context dels missatges anteriors, resolent així les referències conversacionals per no perdre el fil del xat. Per tant, el resultat d'aquesta fase és una consulta estructurada, independent i preparada per executar una recuperació altament precisa, controlada i coherent amb el cas analitzat.

5.4.2 II. Recuperació documental

La segona fase (*retrieval*) s'activa quan l'enrutament previ determina que la consulta requereix suport de la base documental. A partir de la consulta enriquida, s'utilitza un model d'*embeddings* que projecta la cerca en un espai vectorial de 384 dimensions [14]. Això permet comparar semànticament la consulta amb els fragments indexats, encara que l'usuari utilitzi un vocabulari diferent del dels documents oficials.

Abans d'executar la cerca, el sistema aplica un prefiltratge flexible a Qdrant sobre el *payload* dels fragments, és a dir, sobre les metadades estructurades associades a cada fragment [13]. Mitjançant un mecanisme de priorització no restrictiva, la recuperació dona preferència als documents alineats amb el context detectat a la Fase anterior, sense dependre únicament de la similitud vectorial.

La recuperació inicial es fa amb una estratègia Top-*K*, on Qdrant retorna un conjunt ampli de candidats ($K = 15$ fragments) avaluant-ne la proximitat mitjançant la mètrica de similitud del cosinus. Com que aquesta selecció ràpida no sempre és prou precisa, els fragments passen a una fa-

se de *reranking*. Aquí, un model *Cross-Encoder* reavalua exhaustivament la rellevància bidireccional de cada parell consulta-fragment [15], reduint el conjunt inicial a un sub-conjunt final Top- N (on $N = 5$). Aquests fragments de major valor conceptual conformaran el context definitiu per al model generatiu.

5.4.3 III. Generació de la resposta

La tercera i última fase de l'arquitectura RAG és la generació de la resposta. Amb els fragments definitius i les seves respectives metadades, es construeix un context documental estructurat. D'aquesta manera, el model generatiu pot construir una resposta fonamentada en fonts concretes i mantenir la traçabilitat de la informació utilitzada.

A continuació, el sistema construeix el *prompt* dinàmic que adapta les instruccions segons el tipus de resposta definit en les fases anteriors. El sistema està dissenyat per incorporar directrius específiques segons la categoria de risc, com ara prioritzar la protecció immediata en casos greus, evitar preguntes revictimitzants o mantenir un to proper al menor.

Finalment, el *prompt* complet s'envia al model de llenguatge, que genera una resposta en català formal, clara i orientada al suport de l'actuació docent. En aquesta arquitectura, el LLM s'utilitza com un component de generació condicionat pel context documental recuperat [3], per les regles del sistema i per les restriccions de privacitat establertes. Per aquest motiu, la resposta final ha de ser verificable i coherent amb els protocols i la normativa recuperats.

Alhora, el sistema adopta un enfocament sota supervisió humana [17]: la resposta generada actua com una recomanació de suport i no com una decisió automàtica. Per tant, la validació final continua depenent del docent o del professional responsable, reforçant la seva capacitat d'anàlisi i actuació sense substituir-ne el criteri ni la responsabilitat.

5.5 Models i serveis externs integrats

El sistema desenvolupat utilitza serveis externs d'intel·ligència artificial per a tres funcions diferenciades: la generació d'embeddings, la reordenació dels fragments i la generació de respostes en llenguatge natural. Aquests serveis operen sota una arquitectura modular dissenyada per permetre el canvi de models o proveïdors futurs sense alterar el funcionament general del sistema.

En primer lloc, per representar semànticament les consultes i els fragments documentals, s'executa localment el model *paraphrase-multilingual-MiniLM-L12-v2* [14]. Aquest transforma el text en vectors de 384 dimensions i s'ha seleccionat per la seva baixa latència, alta eficiència computacional i l'òptima compatibilitat multilingüe, fet especialment important ja que la base documental es troba en català i castellà.

Per afinar la cerca inicial de Qdrant, s'utilitza un model local de reordenació *Cross-Encoder* [15]. A diferència de l'enfocament vectorial estàndard, aquest model avalua conjuntament cada parell consulta-fragment, assignant una puntuació de rellevància bidireccional per assegurar que només se seleccionen els fragments més útils per construir el context documental final.

La classificació, el preprocés i la generació de llenguatge es deleguen a l'API de Google Gemini [16]. Les tasques de preprocés, extracció semàntica i classificació d'intencions s'assignen a *Gemini 2.5 Flash-Lite* per minimitzar la latència i el consum de *tokens*. En canvi, la generació final de la resposta s'executa a través de *Gemini 1.5 Flash*, garantint una redacció robusta i una alta eficiència.

5.6 Frontend i experiència d'usuari

Per facilitar la interacció dels equips directius i docents amb l'arquitectura RAG, s'ha desenvolupat una interfície web mitjançant Streamlit [18] (Fig.2), escollit per la seva agilitat en la integració de models d'IA i entorns interactius. Això ha permès construir un *frontend* adaptat al context escolar i a l'ecosistema de b-resol, oferint una experiència guiada que afavoreix una gestió més estructurada de les alertes i dels protocols, facilitant una actuació més precoç davant situacions de risc.

A nivell d'arquitectura, la interfície es basa en una gestió d'estats aïllada [19], que preserva el context de cada cas i evita interferències entre converses. Per aquest motiu, l'entorn s'organitza en dos espais de xat independents per gestionar funcionalitats diferenciades:

- **Assistent RAG Principal:** És l'entorn central on el docent introdueix les alertes en llenguatge natural i rep una resposta llegible i ordenada. Per facilitar una actuació més àgil al centre, s'habilita un botó dinàmic que permet descarregar directament en PDF el circuit d'actuació oficial del risc detectat. A més, es pot generar automàticament un informe estructurat de la incidència segons els criteris de b-resol.
- **Anàlisi del Protocol de Centre:** És un espai secundari que permet al docent carregar en format PDF el protocol exclusiu del seu centre (NOFC). El sistema n'extreu el text i l'injecta directament al model de llenguatge. Això permet resoldre dubtes ràpids sobre la normativa interna sense necessitat d'haver-la indexat prèviament a la base de dades vectorial.

Per configurar el comportament de l'assistent principal, la interfície incorpora un panell lateral dissenyat específicament per replicar la naturalesa de les alertes de b-resol. En aquest panell, el docent ha de classificar l'origen de l'alerta indicant si és «Anònima» o «Identificada». En aquest darrer cas, cal afegir-hi les dades associades (edat, curs i rol). Aquesta selecció condiciona completament l'actuació de l'assistent. Les variables associades s'utilitzen únicament de forma temporal durant la sessió per afinar la resposta. D'aquesta manera, es compleix estrictament el principi de minimització de dades personals i es protegeix la privacitat del menor.

Finalment, el *frontend* incorpora un component sòlid de transparència i traçabilitat. Mitjançant espais desplegable, l'usuari pot consultar informació interna del procés, com la classificació de la consulta, la categoria de risc estimada, el tipus de resposta seleccionada i els fragments documentals recuperats. Per a cada fragment es mostren dades imprescindibles, com el document d'origen i la pàgina. Això permet verificar d'on prové la informació utilitzada i reforça la seguretat del docent en la seva decisió. A més, s'ha integrat

un formulari perquè el docent pugui avaluar el sistema després de cada interacció. Aquest mecanisme permet obtenir mètriques aplicades sobre la utilitat de l'eina.

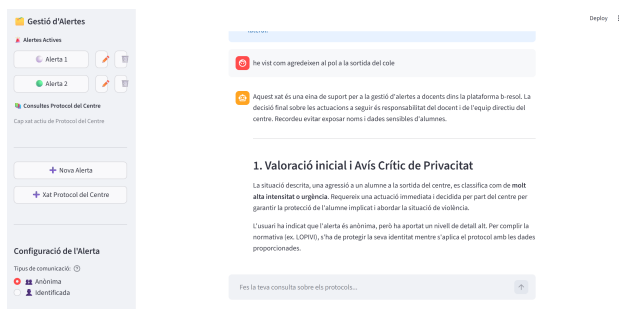


Fig. 2: Interfície d'usuari de l'Assistent d'Alertes de Convivència.

5.7 Privacitat, protecció de dades i supervisió humana

Atès que el sistema tracta alertes educatives sensibles, la privacitat s'ha integrat com a criteri de disseny fonamental. L'arquitectura aplica el principi de minimització de dades [20]. Així, en les alertes identificades, les metadades (com el curs o el sexe) s'utilitzen exclusivament de forma temporal durant la sessió, sense emmagatzemar-se a la base documental. Per contra, si l'alerta és anònima, s'evita qualsevol procediment invasiu que pugui forçar la identificació de l'informant.

La protecció de les dades es reforça mitjançant el desenvolupament d'un Mòdul d'Anonimització. Aquest component intercepta la consulta inicial del docent i substitueix automàticament les entitats personals (com els noms propis) per marcadors genèrics abans d'enviar la informació als models externs. Un cop el sistema genera la resposta, s'aplica el procés invers de desanonimització abans de mostrar-la per pantalla. Això garanteix el compliment estricte de la normativa, ja que les API externes no processen dades sensibles de menors.

Finalment, com que el LLM genera la sortida basant-se estrictament en els fragments recuperats, es garanteix la traçabilitat documental i es redueix el risc d'al·lucinacions. Tot el disseny respon a un enfocament *human-in-the-loop* [22]: el sistema sempre indica que la resposta s'ha d'entendre com una recomanació de suport i mai com una decisió automàtica. La presa de decisions i l'aplicació formal dels protocols recauen exclusivament sobre l'equip educatiu, mantenint la supervisió humana com a pilar indispensable.

6 RESULTATS

La validació d'aquest primer sistema s'ha plantejat des d'una perspectiva funcional, tècnica i qualitativa. L'objectiu no ha estat únicament comprovar que el sistema genera una resposta, sinó analitzar si les diferents fases del sistema actuen de manera coherent amb la naturalesa de cada consulta. Per aquest motiu, s'han definit casos de prova representatius del context de *b-resol*.

6.1 Casos de prova funcionals

S'han executat diversos casos de prova representatius, definits i revisats conjuntament amb els professionals de Bcn-resol, per validar el comportament del sistema en diferents escenaris. Les proves han permès verificar la detecció de consultes incompletes i la correcta selecció de la capa documental. També s'ha validat la gestió d'alertes anònimes mitjançant vies d'indagació segures que preserven la privacitat de l'informant, així com l'activació d'actuacions prudentes davant situacions potencialment urgents.

En conjunt, aquests resultats confirmen la coherència funcional dels principals components de l'arquitectura.

6.2 Validació qualitativa amb professionals de b-resol

A més de les proves funcionals, el sistema ha estat revisat de manera continuada amb professionals de Bcn-resol i educatius al llarg del desenvolupament del projecte. Aquesta revisió s'ha dut a terme principalment en reunions periòdiques de seguiment, on es presentaven els avenços implementats, es contrastaven decisions de disseny i es recollia *feedback* sobre la utilitat, la claredat i l'adequació del sistema.

Aquesta validació no s'ha plantejat com una avaluació estadística definitiva, sinó com un procés qualitatiu i iteratiu orientat a ajustar progressivament el comportament del sistema segons les necessitats reals del context de *b-resol*. En aquest procés s'han pogut valorar aspectes clau com la formulació de les respostes, el tractament de les alertes anònimes i la necessitat de mantenir sempre la supervisió humana.

El *feedback* rebut ha reforçat la importància de generar respostes prudentes, ben estructurades, traçables i vinculades a documentació oficial. Aquest procés de revisió continuada ha estat clau per adaptar el sistema a l'ecosistema de *b-resol* i consolidar-ne la utilitat pràctica.

6.3 Proves quantitatives

S'ha dissenyat una avaluació quantitativa per mesurar objectivament el rendiment de les diferents fases del *pipeline*.

Per dur-la a terme de manera segura i sense comprometre dades personals, es va construir un conjunt de consultes simulades representatives del context de *b-resol*. Aquestes consultes es van etiquetar manualment amb la categoria de risc i els fragments documentals esperats, servint com a base de referència (*ground truth*).

L'avaluació es va estructurar en dues dimensions:

- **Avaluació tècnica:** Es mesura la capacitat del sistema per analitzar correctament les consultes durant la fase de prè-recuperació, validant criteris essencials per al correcte funcionament del sistema, com la classificació de la categoria de risc, el nivell d'urgència i la puntuació de completitud. Per fer-ho, s'utilitzen mètriques com l'*Accuracy* per avaluar la classificació de la categoria de risc i del nivell d'urgència. Aquesta validació és especialment rellevant, ja que condiciona directament la qualitat de la recuperació documental posterior i, en conseqüència, el rendiment global del sistema.

- **Avaluació aplicada:** La qualitat de les respostes generades s'avalua mitjançant un mòdul integrat a la interfície d'usuari, que permet als professionals puntuar aspectes com la claredat, la traçabilitat i la utilitat de cada resposta, així com aportar suggeriments de millora i canvis.

Això permet generar mètriques de satisfacció en temps real, basades en l'ús real de l'eina en producció.

Aquestes proves permeten justificar quantitativament l'impacte de les decisions tecnològiques implementades.

6.3.1 Resultats de l'Avaluació Prerecuperació

A partir de l'execució de l'avaluació sobre el conjunt de dades de validació, format per 26 casos pràctics elaborats a partir d'enquestes als professionals de b-resol, s'han obtingut les següents mètriques de rendiment per al mòdul inicial.

La Figura 3 mostra de manera global els resultats obtinguts en les tres dimensions principals avaluades: la classificació de la categoria de risc, la detecció del nivell d'urgència i l'avaluació de la completitud de la informació.

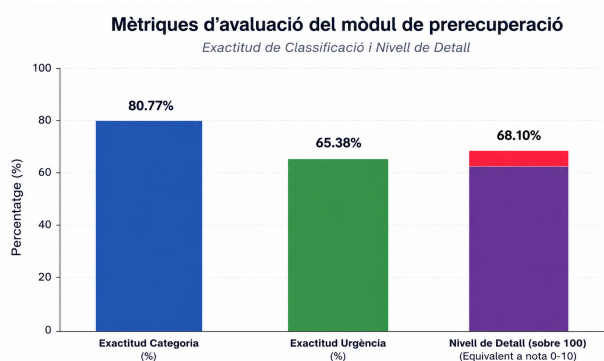


Fig. 3: Mètriques d'avaluació del mòdul de prerecuperació

A partir d'aquests resultats, es poden destacar les següents observacions:

- **Exactitud de Categoria (80.77%):** El sistema classifica correctament la categoria de risc principal en la majoria dels casos, mostrant una elevada capacitat de detecció inicial i orientant de manera efectiva la recuperació documental posterior.
- **Exactitud d'Urgència (65.38%):** La classificació del nivell d'urgència presenta una precisió inferior, però acceptable, a causa del component subjectiu inherent a aquesta variable i de la variabilitat existent entre professionals.
- **Avaluació del Nivell de Detall:** La puntuació mitjana de completitud assignada pel sistema ha estat de **6.58 sobre 10**, mentre que la valoració mitjana dels docents ha estat de **6.81 sobre 10**. Aquesta diferència de només **0.23 punts** mostra una elevada alineació entre el criteri algorítmic i el judici humà a l'hora de determinar la completitud de l'alerta.

Aquestes mètriques permeten validar quantitativament la solidesa del mòdul de prerecuperació i justificar la seva utilitat com a base per a les fases posteriors de recuperació documental i generació de respostes.

7 DISCUSSIÓ DELS RESULTATS

Els resultats obtinguts mostren que el sistema desenvolupat compleix els objectius plantejats i valida la viabilitat d'una arquitectura RAG aplicada a la gestió d'alertes educatives sensibles dins l'entorn de b-resol. La solució permet processar consultes en llenguatge natural, analitzar-ne el context, recuperar fragments documentals rellevants i generar respostes orientatives fonamentades en documentació oficial, configurant així un flux complet d'assistència documental supervisable.

Els resultats quantitatius reforcen la solidesa del mòdul de prerecuperació. L'exactitud en la classificació de categoria (80.77%) mostra una elevada capacitat per orientar correctament la recuperació documental en la majoria de casos. La classificació del nivell d'urgència ($\approx 70\%$), tot i ser inferior, manté un rendiment acceptable, especialment tenint en compte la naturalesa subjectiva d'aquest criteri. L'anàlisi detallada de la matriu de confusió (vegeu Apèndix A.5) mostra que els errors no es distribueixen de manera aleatòria, sinó que es concentren principalment en casos d'urgència mitjana que el sistema tendeix a elevar a alta. Aquest comportament conservador és especialment rellevant en un context sensible com la protecció de menors, ja que prioritza la prudència davant del dubte i evita infravaloracions crítiques en casos d'alta urgència. A més, la diferència mínima de 0.23 punts entre la puntuació de completitud generada pel sistema (6.58) i la valoració mitjana dels professionals (6.81) evidencia una alta alineació entre el criteri algorítmic i el judici humà.

Des del punt de vista funcional, els casos de prova mostren que els diferents components del sistema treballen de manera coordinada i adapten el seu comportament segons el tipus de consulta. En alertes incompletes o anònimes, la fase de prerecuperació evita activar protocols sense prou informació i prioritza la recollida de dades mínimes, contribuint a la minimització de dades personals i a la protecció de la identitat de la persona informant.

Des del punt de vista tècnic, la fiabilitat del sistema es fonamenta en la combinació de recuperació semàntica, filtres de metadades, reranking, restriccions de *prompt* i supervisió humana. La separació entre la capa pràctica i la legal ha demostrat ser especialment rellevant per adaptar la recuperació documental a la intenció de l'usuari i evitar barrejar procediments amb justificacions normatives.

Durant el desenvolupament s'han identificat diversos reptes, especialment en la preparació de la base documental i en la fragmentació semàntica dels protocols, circuits i normatives. Igualment, la recuperació de fragments fora de context ha requerit mecanismes específics com el *payload filtering*, el *Cross-Encoder Reranker* i la memòria conversacional per incrementar l'estabilitat i la precisió del sistema.

La validació qualitativa amb professionals de Bcn-resol ha estat clau per ajustar el to, l'estructura i la utilitat de les respostes. Aquest procés ha confirmat que el valor del sistema no rau únicament en generar text, sinó en oferir una orientació prudent, documentada i verificable.

Finalment, els resultats indiquen que la solució és tècnicament viable dins del marc d'un TFG d'Enginyeria Informàtica. El projecte integra tecnologies com Python, Qdrant, models d'embeddings, heurístiques de preprocesament, reranking, LLMs i Streamlit per construir una eina

aplicable, segura i amb valor real per a la comunitat educativa.

8 CONCLUSIONS I TREBALL FUTUR

Aquest Treball de Final de Grau ha permès desenvolupar un primer sistema funcional orientat a donar suport als docents i responsables de convivència en la gestió inicial d'alertes educatives sensibles dins la plataforma b-resol. La solució proposada parteix de la necessitat de disposar d'una eina capaç de generar respostes en llenguatge natural, però fonamentades en documentació oficial, traçables i adaptades al context específic de b-resol.

A partir dels resultats obtinguts, es pot afirmar que l'ús d'una arquitectura RAG és adequat per a aquest àmbit, ja que permet combinar la capacitat dels LLM amb una base documental controlada. D'aquesta manera, el sistema no depèn únicament del coneixement intern del LLM, sinó que recupera fragments rellevants de protocols, circuits i normativa abans de generar una resposta orientativa. Aquest enfocament és especialment rellevant en un context sensible com la protecció de menors, on les respostes han de ser prudentes, verificables i fonamentades.

El treball també posa de manifest que la qualitat del sistema no depèn només del model d'intel·ligència artificial utilitzat, sinó del conjunt de decisions d'enginyeria adoptades. Aspectes com l'estratègia de fragmentació, la qualitat de les metadades, els filtres aplicats a Qdrant, el *reranking* i la construcció del context documental tenen un impacte directe en la qualitat final de la resposta. Per això, el sistema s'ha dissenyat com una arquitectura completa de recuperació, control i generació, i no simplement com un xat connectat a un model de llenguatge.

Un dels punts clau és que el sistema incorpora criteris de privacitat, traçabilitat i supervisió humana des del disseny. En aquest sentit, la resposta generada s'ha d'entendre com una orientació documental de suport, però no com una decisió automàtica, ja que la interpretació final del cas i l'activació dels protocols continuen corresponent als professionals responsables.

Més enllà dels resultats estrictament tècnics, cal destacar l'impacte real que aquest sistema pot tenir un cop integrat en un entorn escolar. El projecte tracta de donar suport en situacions que afecten usuaris extremadament vulnerables, menors d'edat, on el temps de reacció i l'exactitud procedimental són crítics. En aquest sentit, la tecnologia desenvolupada no actua només com un motor de cerca de lleis, sinó que permet transformar la gestió de les alertes, passant del model actual purament reactiu a un model altament proactiu. En dotar els docents i responsables educatius d'eines d'assessorament estratègic i guies d'indagació immediates, s'accelera la protecció de l'alumne, es redueix el marge d'error humà davant l'estrès i es posiciona la Intel·ligència Artificial com una eina clau de prevenció i suport per a la protecció integral dels menors.

Com a treball futur, caldria ampliar el corpus documental, comparar diferents models d'*embeddings* i optimitzar les fases de recuperació i reordenació de resultats. Per dur a terme l'avaluació quantitativa amb consultes etiquetades, es proposa utilitzar entorns d'avaluació automatitzada basats en LLM com a jutges (com ara RAGAS o TruLens), la qual cosa permetria escalar les proves mètriques segons

l'estat de l'art actual. També seria necessari estudiar una integració més directa amb la plataforma b-resol i les seves funcionalitats, tenint en compte aspectes de seguretat, control d'accés i gestió de dades sensibles. En conjunt, el projecte mostra que els sistemes RAG poden ser una alternativa viable per donar suport documental en contextos educatius sensibles, sempre que es mantingui un enfocament prudent i amb supervisió humana.

ÚS DE LA IA GENERATIVA

En el desenvolupament d'aquest Treball de Final de Grau, l'ús de la Intel·ligència Artificial (IA) generativa s'ha limitat estrictament a funcions d'assistència tècnica i optimització. En cap cas aquestes eines han substituït el pensament crític, l'anàlisi de resultats ni l'autoria intel·lectual de l'estudiant.

Concretament, s'han definit els següents usos:

- **Assistent de programació i depuració:** S'han emprat eines d'IA per donar suport en la resolució d'errors de codi, la generació d'scripts rutinaris (com per exemple, funcions de neteja i preprocés de text) i l'optimització de consultes estructurades a les bases de dades vectorials del sistema RAG.
- **Suport en la redacció i estructuració:** S'han utilitzat els LLM com a eina de suport lingüístic per revisar la correcció gramatical, millorar la fluïdesa del text acadèmic en català i facilitar l'estructuració inicial dels diferents apartats de la memòria escrita.
- **Síntesi i exploració de l'estat de l'art:** L'IA ha servit com a eina d'exploració per resumir documentació tècnica extensa sobre frameworks RAG, extreure conceptes clau d'articles científics i ajuda en la interpretació de normatives legals complexes. Tota la informació obtinguda ha estat sotmesa a una validació posterior contrastant-la amb les fonts originals.

AGRAÏMENTS

Vull agrair al meu tutor acadèmic, Daniel Ponsa Mussarra, l'orientació, la disponibilitat i el seguiment al llarg del desenvolupament d'aquest treball. La seva guia ha estat clau per donar forma al projecte i per millorar-ne tant l'enfocament com el resultat final. També m'agradaria agrair a l'Albert Gil la seva ajuda en la part tècnica del projecte.

Aquest TFG representa, per a mi, el tancament d'una etapa de molts anys d'esforç. Per això, vull expressar un agraïment molt especial a la meua família, pel suport incondicional, per la confiança dipositada en mi i per haver-me ajudat a arribar fins aquí.

Igualment, vull agrair als meus amics el suport i la motivació al llarg d'aquest camí, i especialment als companys de carrera, amb qui he compartit aprenentatges, reptes i molts moments que han fet aquest recorregut més enriquidor.

Finalment, vull fer un agraïment especial a tot l'equip de Bcn-Resol i, en particular, al Luis Ribó, el meu tutor d'empresa, per haver-me donat l'oportunitat de desenvolupar aquest projecte, per la confiança i el suport constant.

REFERÈNCIES

- [1] LOPIVI. Llei Orgànica de Protecció Integral a la Infància i l'Adolescència enfront de la Violència.
- [2] Protocol d'actuació de la Generalitat de Catalunya en casos de violència a l'entorn escolar.
- [3] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., i Kiela, D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS, 2020.
- [4] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., i Wang, H. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv, 2023.
- [5] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., i Liu, T. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. arXiv, 2023.
- [6] Wang, S., Xu, T., Li, H., Zhang, C., Liang, J., Tang, J., Yu, P. S., i Wen, Q. Large Language Models for Education: A Survey and Outlook. arXiv, 2024.
- [7] Maharjan, A., i Yadav, U. Chunking, Retrieval, and Re-ranking: An Empirical Evaluation of RAG Architectures for Policy Document Question Answering. arXiv, 2026.
- [8] Git. Git Documentation. Disponible a: <https://git-scm.com/doc> [Consulta: 11 abril 2026].
- [9] PyMuPDF. *PyMuPDF Documentation*. Disponible a: <https://pymupdf.readthedocs.io/> [Consulta: 29 juny 2026].
- [10] Python Software Foundation. *re – Regular expression operations*. Python Documentation. Disponible a: <https://docs.python.org/3/library/re.html> [Consulta: 29 juny 2026].
- [11] Reimers, N., i Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019.
- [12] Qdrant. Qdrant Documentation. Disponible a: <https://qdrant.tech/documentation> [Consulta: 20 abril 2026].
- [13] Qdrant. Filtering - Qdrant Documentation. Disponible a: <https://qdrant.tech/documentation/concepts/filtering/> [Consulta: 24 abril 2026].
- [14] Hugging Face. sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2. Disponible a: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> [Consulta: 2 maig 2026].
- [15] SentenceTransformers. CrossEncoder - Sentence Transformers Documentation. Disponible a: https://www.sbert.net/docs/package_reference/cross_encoder/model.html [Consulta: 2 maig 2026].
- [16] Google AI for Developers. Gemini API Documentation: Models. Disponible a: <https://ai.google.dev/gemini-api/docs/models> [Consulta: 8 maig 2026].
- [17] European Commission, High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. European Commission, 2019.
- [18] Streamlit. Streamlit Documentation. Disponible a: <https://docs.streamlit.io/> [Consulta: 2 abril 2026].
- [19] Streamlit. Session State - Streamlit Documentation. Disponible a: https://docs.streamlit.io/develop/api-reference/caching-and-state/st.session_state [Consulta: 2 abril 2026].
- [20] Parlament Europeu i Consell de la Unió Europea. Reglament (UE) 2016/679, Reglament General de Protecció de Dades. Diari Oficial de la Unió Europea, 2016.
- [21] Explosion AI. *EntityRecognizer – spaCy API Documentation*. Disponible a: <https://spacy.io/api/entityrecognizer> [Consulta: 29 juny 2026].
- [22] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST, 2023.

APÈNDIXS

A.1 Tipologia documental i estratègia de fragmentació

Aquest apèndix mostra de manera visual la diversitat documental utilitzada en la base de coneixement i el flux de preparació aplicat abans de la indexació vectorial. La naturalesa dels documents (circuit, protocols i normativa legal) requereix estratègies de fragmentació específiques per preservar la coherència semàntica dels fragments i optimitzar la recuperació posterior.

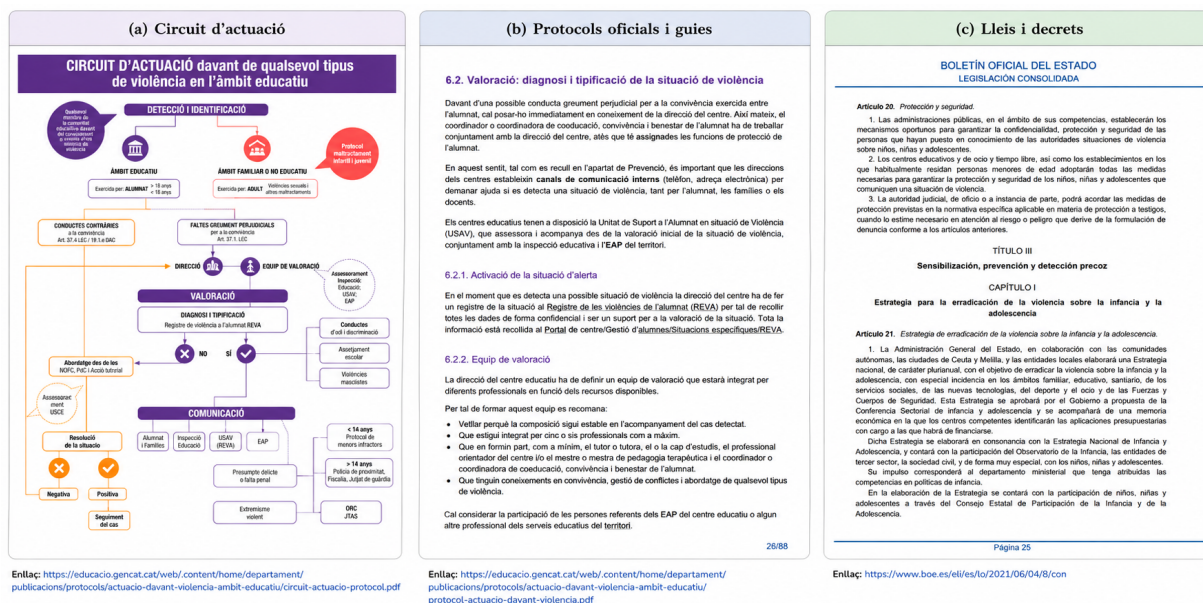


Fig. 4: Tipologies documentals utilitzades com a base de coneixement del sistema.

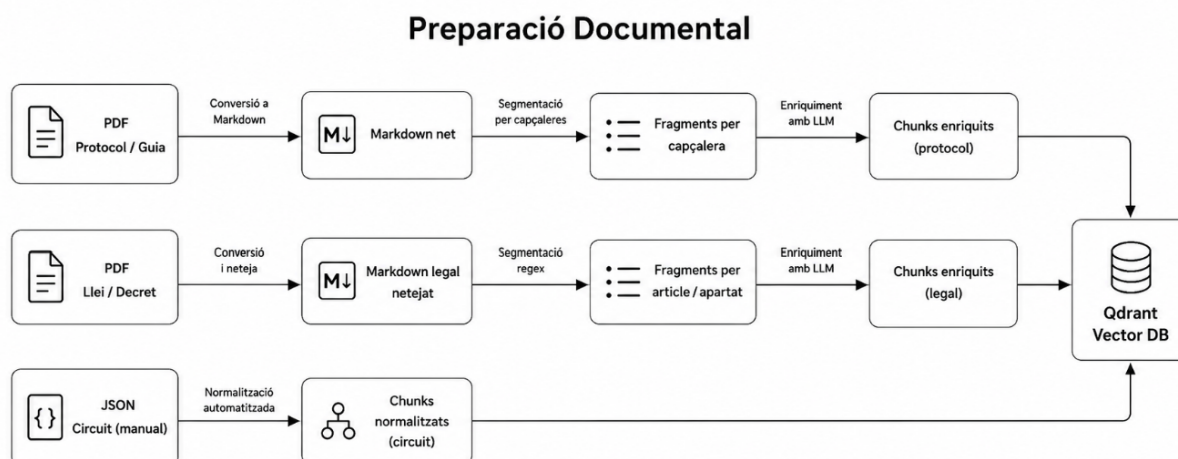


Fig. 5: Flux de preparació documental, fragmentació semàntica i indexació vectorial.

A.2 Distribució dels fragments de la Base de Dades Vectorial

Aquesta figura mostra la distribució semàntica dels fragments indexats a la base de dades vectorial, agrupats segons les diferents categories de risc, permetent visualitzar la separació i coherència temàtica del coneixement emmagatzemat.

Distribució Base de Dades Vectorial

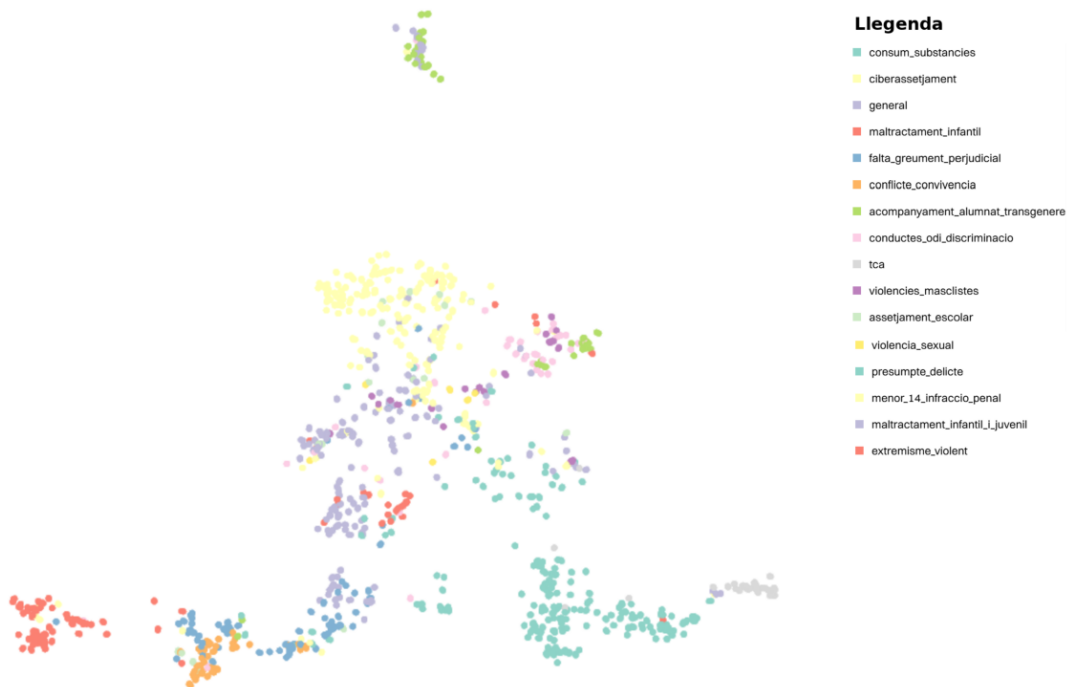


Fig. 6: Distribució dels fragments de la Base de Dades Vectorial.

A.3 Esquema detallat de l'arquitectura Advanced RAG

Aquest apèndix mostra l'estructura completa de l'arquitectura *Advanced RAG* desenvolupada per al projecte, detallant el flux funcional des de la recepció inicial de l'alerta fins a la generació final de la resposta. L'esquema permet visualitzar la separació modular entre les fases de prerecuperació, recuperació semàntica i generació assistida, així com la interacció entre els diferents components interns del sistema.

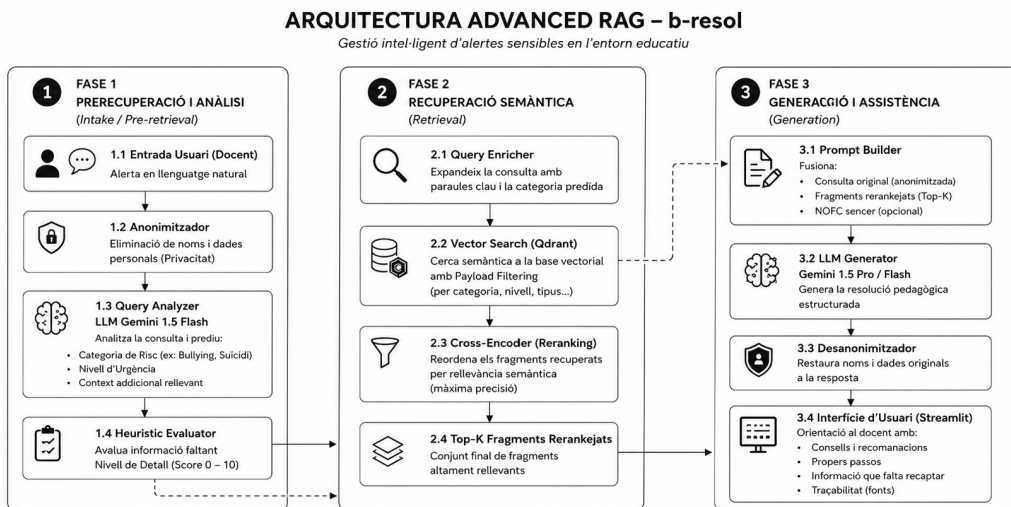


Fig. 7: Esquema detallat de l'arquitectura Advanced RAG integrada a l'entorn de b-resol.

A.4 Vídeo del funcionament complet del sistema

Aquest vídeo mostra de manera detallada el funcionament complet del sistema comentat durant l'article.

Enllaç: https://drive.google.com/drive/folders/1gmFM0Zfy-Lb4peHE0n7eJYymMqqfF_uV?usp=sharing

A.5 Matriu de confusió de la classificació d'urgència

Aquest apèndix mostra la matriu de confusió corresponent a l'avaluació del mòdul de prerecuperació en la tasca de classificació del nivell d'urgència.

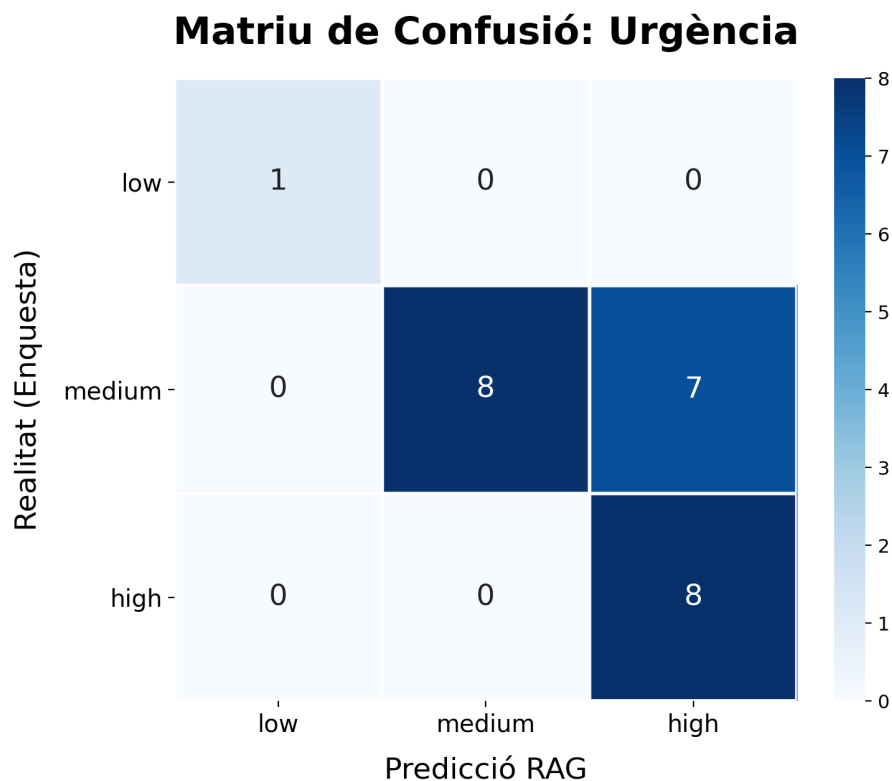


Fig. 8: Matriu de confusió del model en la classificació del nivell d'urgència.