



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat d'Informàtica de Barcelona



Aplicació de 'Large Language Models' per a l'Anàlisi i Consulta de Documents Multimodals

Entrega 4: Document final

Autor: Albert Reina i Buxó
Director: Marc Alier Forment
Tutor GEP: Fernando Barrabes Naval
Data: 27 de febrer de 2025

Resum

Aquest projecte és lliura com a Treball de Final de Grau en el Grau d'Enginyeria Informàtica de la Facultat d'Informàtica de Barcelona.

L'objectiu principal del projecte és fer ús de “Large Language Models” per a extreure la informació d'imatges, taules i textos en documents amb la finalitat de calcular “embeddings” multimodals per a fer consultes a bases de dades semàntiques.

Resumen

Este proyecto se entrega como Trabajo de Fin de Grado en el Grado de Ingeniería Informática de la Facultad de Informática de Barcelona.

El objetivo principal del proyecto es usar “Large Language Models” para extraer la información de imágenes, tablas y textos en documentos, con la finalidad de calcular “embeddings” multimodales para hacer consultas en bases de datos semánticas.

Abstract

This project is submitted as a Bachelor's Thesis in the Computer Engineering program at the Faculty of Computer Science of Barcelona.

The main objective of the project is to use Large Language Models to extract information from images, tables, and texts in documents in order to compute multimodal embeddings for querying semantic databases.

Índex de continguts

1. Introducció i contextualització.....	5
1.1 Context.....	5
1.2 Identificació del problema.....	5
1.3 Actors implicats.....	6
1.4 Definicions.....	6
2. Justificació.....	8
2.1 Solucions prèvies.....	8
2.2 Justificació.....	9
3. Abast.....	10
3.1 Objectius.....	10
3.2 Impediments i riscos.....	10
4. Metodologia.....	11
4.1 Gestió del projecte.....	11
4.2 Control de versions.....	12
4.3 Validació i proves.....	13
5. Planificació temporal.....	13
5.1 Descripció de les tasques.....	13
5.1.1 Gestió del projecte.....	14
5.1.2 Desenvolupament.....	14
5.1.3 Documentació.....	16
5.1.4 Preparació defensa del TFG.....	16
5.2 Recursos.....	16
5.2.1 Recursos humans.....	16
5.2.2 Recursos materials.....	16
5.3 Estimacions.....	17
5.4 Diagrama de Gantt.....	18
5.5 Gestió de riscos.....	20
6. Gestió econòmica.....	21
6.1 Costos de personal per activitat.....	21
6.2 Costos genèrics.....	22
6.2.1 Hardware.....	22
6.2.2 Software.....	23
6.2.3 Electricitat.....	23
6.2.4 Internet.....	23
6.3 Contingències.....	24
6.4 Imprevistos.....	25
6.5 Total.....	25
6.6 Control de gestió.....	26
7. Informe de sostenibilitat.....	27
7.1 Autoavaluació.....	27
7.2 Dimensió econòmica.....	27
7.3 Dimensió ambiental.....	28
7.4 Dimensió social.....	28

8. Referències.....	29
----------------------------	-----------

1. Introducció i contextualització

El projecte *Aplicació de 'Large Language Models' per a l'Anàlisi i Consulta de Documents Multimodals* és un treball de Fi de Grau de modalitat A (centre) i pertany als estudis de Grau en Enginyeria Informàtica de la Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya, dins de l'especialitat de l'Enginyeria del Software. Aquest projecte està dirigit per Marc Aliet Forment, del Departament d'Enginyeria de Serveis i Sistemes de la Informació(ESSI).

1.1 Context

En els últims anys, els avenços de la intel·ligència artificial han fomentat el desenvolupament de models Large Language Models que han obtingut una capacitat, cada cop més gran, per a entendre i generar textos amb un grau de coherència i precisió elevadíssims. Però aquests models també tenen altres aplicacions que estan fora de l'àmbit del llenguatge natural. Un d'aquests àmbits és l'extracció d'informació en diversos formats, com per exemple, imatges, taules, textos. També es poden emprar per a fer anàlisis de dades en grans conjunts de dades o conjunts de dades amb formats difícils de processar per a les persones.

Actualment, la gestió i recuperació d'informació afrontant el repte de gestionar grans quantitats de dades en diferents tipus de formats com els que hem comentat anteriorment, text, imatges i taules. Tradicionalment, els sistemes de recuperació d'informació es basen en la cerca textual, això no és suficient per a processar els diferents tipus de dades que tenim en l'actualitat i és en aquest context que sorgeix la necessitat de millorar els models actuals i millorar l'eficiència en aquesta recuperació d'informació multimodal amb l'objectiu de millorar les consultes sobre les bases de dades semàntiques.

1.2 Identificació del problema

Donats els fets exposats en l'apartat anterior, el creixement en la mida de les dades a tractar pels sistemes actuals ens ha portat a un augment del nombre de documents que combinen diferents tipus de formats de dades. Tot i això, els sistemes actuals no són capaços de processar aquests diferents formats, ja que, estan dissenyats per a fer cerques textuais. Això redueix la seva capacitat de comprendre i processar les dades de manera semàntica.

El problema principal recau en el fet que la majoria de les bases de dades semàntiques utilitzen estratègies de cerca de coincidència exacta en paraules clau i, per tant, es dificulta la interpretació quan la informació es presenta en formats de complexitat més elevada.

Un dels altres problemes de l'enfocament tradicional és la integració de diferents tipus de dades en un mateix entorn de consulta. Això representa un problema actualment, ja que no hi ha metodologies estandarditzades per a fer les extraccions de dades de la informació multimodal de manera eficient. Per tant, sense una bona extracció i representació de les dades multimodals, les consultes a les bases de dades semàntiques perden precisió, afectant l'eficiència i la qualitat dels resultats obtinguts per aquestes consultes.

Com a resultat, aquest projecte intentarà generar un pipeline, amb l'ajut d'un model o la conjunció de varis, que permeti reduir al màxim aquests problemes exposats.

1.3 Actors implicats

Desenvolupador: El projecte està realitzat per un desenvolupador, autor d'aquest treball. El desenvolupador és l'encarregat de dissenyar, desenvolupar i documentar el model.

Director del treball de fi de grau: El projecte és dirigit per en Marc Alier Forment, professor del Departament d'Enginyeria de Serveis i Sistemes de la Informació (ESSI). Encarregat de guiar i supervisar el correcte desenvolupament del projecte.

Especialistes en intel·ligència artificial: Donat que el projecte es desenvolupa en el marc de les intel·ligències artificials, en concret dins de l'àmbit dels Large Language Models, els especialistes en intel·ligència artificial desenvolupen i document les diferents tècniques i models i, per tant, són part interessada en el projecte.

Empreses que analitzen grans quantitats d'informació: Qualsevol empresa que tingui la necessitat d'analitzar, gestionar grans quantitats d'informació en documents és part interessada en implementar el sistema del projecte.

Desenvolupadors de Codi Obert: Tots els desenvolupadors de les comunitats Codi obert interessants en reutilitzar part del codi implementat en el projecte amb l'objectiu de millorar o crear nous sistemes basats en el del projecte.

Usuaris finals: Aquesta categoria inclou tots aquells professionals en l'àmbit de la tecnologia i que fan ús de la intel·ligència artificial per a realitzar gestió documental, investigació. Anàlisis de dades i altres aplicacions.

1.4 Definicions

IA: Sigles d'intel·ligència artificial.

Model: Estructura entrenada amb dades per a executar tasques específiques.

LLM: Sigles de Large Language Model.

Large Language Model: Model d'intel·ligència artificial basat en xarxes neuronals, entrenat per a entendre i generar llenguatge natural.

Multimodalitat: Capacitat d'un sistema d'IA per a processar i combinar diferents tipus de dades.

Embeddings: Representacions numèriques d'un espai vectorial amb significats semàntics com paraules, imatges, etc.

Pipeline: Seqüència de passos d'un flux de processament de dades que transforma analitza i modela la informació.

Tasques NLP: Conjunt de processos dins del processament de llenguatge natural.

Base de dades semàntica: Sistema d'emmagatzematge optimitzat per a organitzar la informació basant-se en el significat o el context de les dades en lloc de coincidències exactes.

Repositori de codi: Espai on s'emmagatzema i es gestionen les versions del codi font d'un projecte.

Branca: Línia de desenvolupament dins d'un repositori.

2. Justificació

2.1 Solucions prèvies

El camp de la intel·ligència artificial existeix des de fa més de cinquanta anys i ha anat evolucionant a passos agerantats des del seu inici. Per això, prèviament a fer aquest treball he recollit algunes solucions, tant de codi obert com propietàries, que utilitzen els Large Language Models per a extreure informació de documents i generar-ne embeddings multimodals.

La primera solució és el model Luminous de la companyia alemanya Aleph Alpha. Luminous destaca per la seva capacitat de processar tant text com imatge, de manera multimodal. Això els permet integrar diferents tipus de dades en una única representació i facilita la recuperació de la informació i les anàlisis semàntiques de les dades generades pel model.

Una altra solució de codi obert és Spark NLP. És una biblioteca de processament de llenguatge natural que ofereix pipelines i models preentrenats per a resoldre tasques NLP. Spark NLP consta d'una extensió anomenada Spark OCR que permet extreure text d'imatges i documents escanejats i d'aquesta manera, generar els embeddings multimodals a partir de text i imatges.

Per una altra banda, tenim Amazon Titan. Titan forma part de la suite de models que ofereix Amazon i conté models multimodals que poden processar text i imatges per igual. Tots aquests models estan dissenyats per a integrar-se en aplicacions amb requisits de comprensió i generació de continguts i que reben les dades per diferents canals.

Una de les solucions més importants actualment és Gemini. Aquest model Large Language Model està dissenyat per Google i compte amb una gran capacitat per a processar tota mena de dades, per exemple text, imatges, àudio i vídeo entre d'altres. Aquesta multimodalitat converteix a Gemini en una de les solucions més robustes del mercat, pel que a multimodalitat es refereix.

Finalment, OpenAI ofereix diferents serveis que permeten la generació d'embeddings multimodals. Tot i que el seu producte estrella, ChatGPT 4 compta amb la capacitat de processar tant text com imatges no genera embeddings multimodals de manera directa i, per tant, no seria una solució a tenir en compte. Però sí que hi ha altres serveis que generen aquests embeddings i poden ser interessants en el context del nostre projecte.

Per exemple, l'API de embeddings d'OpenAI, tot i que actualment només compta amb la generació d'embeddings de text es preveu que pugui generar embeddings d'imatges i taules en un futur.

OpenAI també ofereix un model anomenat CLIP (Contrastive Language-Image Pretraining) que té per objectiu generar representacions multimodals imatge-text que permetin fer cerques semàntiques de manera eficient.

2.2 Justificació

Un cop exposades les solucions que hi ha al mercat actualment i tenint en compte el que aporten i el que no, podem afirmar que, no existeix una solució que, de manera eficient, processa la informació recollida en documents en diferents modalitats i en concret text, imatges i taules i posteriorment genera embeddings multimodals que permeten fer consultes sobre les bases de dades semàntiques on s'emmagatzemen aquests embeddings.

Això obre la porta a la creació d'un nou model o conjunció de diversos models en un nou sistema que ens permet fer aquest processament multimodal dels documents i la generació d'embeddings multimodals de text, imatge i taules. Aquest sistema permetria unificar el processament de dades i la cerca d'aquestes, millorant l'eficiència del procés.

A més, aquesta solució permet simplificar el procés d'integració amb altres eines o sistemes que requereixin diferents models per a fer el processament i generació d'embeddings de les dades que tracten. Unificant tot el processament en un mateix sistema i eliminant possibles incompatibilitats entre altres sistemes.

Per tant, la realització d'aquest projecte implica una millora en les actuals solucions que s'ofereixen en el mercat, tant de codi obert com propietàries.

3. Abast

Aquesta secció està dedicada a l'especificació dels objectius i subobjectius que es pretenen assolir amb aquest projecte, els diferents impediments i riscos que poden sorgir durant tot el procés de desenvolupament del projecte.

3.1 Objectius

El principal objectiu del projecte és desenvolupar un sistema que implementi un o més models Large Language Model que permeti extreure la informació de tipus text, imatge o taules de documents i la processi per a obtenir embeddings multimodals (únicament en les modalitats definides abans text, imatge i taules). Aquests embeddings han de ser compatibles amb la base de dades semàntica escollida.

Per tal d'aprofundir una mica més en aquest objectiu ens cal definir alguns subobjectius que ens permetin especificar una mica més.

- Preparar un entorn de desenvolupament amb les tecnologies escollides.
- Implementar un sistema que permeti processar les dades dels documents a tractar.
- Crear un sistema que implementi un o més models LLM (Large Language Model).
- Dotar aquest sistema de la capacitat de rebre les dades en el format correcte (documents).
- Implementar el codi que permet generar els embeddings multimodals.
- Assegurar que els embeddings generats són compatibles amb les bases de dades utilitzades.
- Definir una interfície per facilitar l'ús de l'eina.
- Documentar tot el procés.

3.2 Impediments i riscos

En tots els projectes apareixen obstacles i impediments que dificulten les tasques pendents de realitzar per a assolir els objectius marcats a l'inici. Per tant, és important estar preparats per a afrontar aquests obstacles de la millor manera i reduir l'impacte en el projecte.

Seguidament, s'exposen alguns dels possibles impediments i riscos que es poden trobar durant el desenvolupament del projecte.

Limitacions de recursos i temps: Els recursos limitats i la manca de temps degut als terminis ajustats del projecte poden resultar en una limitació de capacitat per a assolir tots els objectius marcats en el projecte de manera satisfactòria.

Desafiaments d'implementació: Els problemes de seguretat, eficiència o comptabilitat en les tecnologies emprades per al desenvolupament poden representar un obstacle en el correcte desenvolupament del projecte.

Inexperiència en les tecnologies: La falta d'experiència en talguna de les tecnologies emprades per l'equip de desenvolupament poden resultar en retards significatius en les tasques derivades del projecte.

Requeriments computacionals: Alguns dels models emprats per a l'extracció d'informació i generació d'embeddings són molt demandants de la maquinaria, això podria alentir el procés de proves de les diferents versions.

4. Metodologia

Amb l'objectiu de completar un projecte d'aquestes dimensions és necessari definir una metodologia de treball per tal de tenir un control de les tasques i els recursos i, d'aquesta manera, guiar el procés de desenvolupament mantenint el control i la qualitat esperada.

Tot seguit es defineix aquesta metodologia de gestió.

4.1 Gestió del projecte

La metodologia escollida per a fer la gestió del projecte és la metodologia **Kanban**, que és una gestió visual basada en columnes i targetes. Aquest mètode permet fer un seguiment molt clar del progrés del projecte i veure quines tasques queden pendents per a completar els objectius. Per tal de fer ús d'aquesta metodologia usarem l'eina **Trello**, on les diferents columnes representen els possibles estats de les tasques i les targetes les mateixes tasques. D'aquesta manera facilitem la priorització de les tasques més importants o urgents i aquelles que poden bloquejar altres tasques. A més, **Trello** ens permet definir en cada targeta (tasca) tot un seguit d'informació addicional i camps que són molt útils per a fer una gestió eficient d'aquestes.

Els possibles estats de les tasques són els següents:

Pendent: Tasques no començades.

En Procés: Tasques ja començades.

Bloquejat: Tasques que no es poden realitzar per un motiu.

Proves: Tasques pendents de ser provades o en proves.

Validació: Tasques pendents de validar abans de donar-les per acabades.

Finalitzada: Tasques acabades.

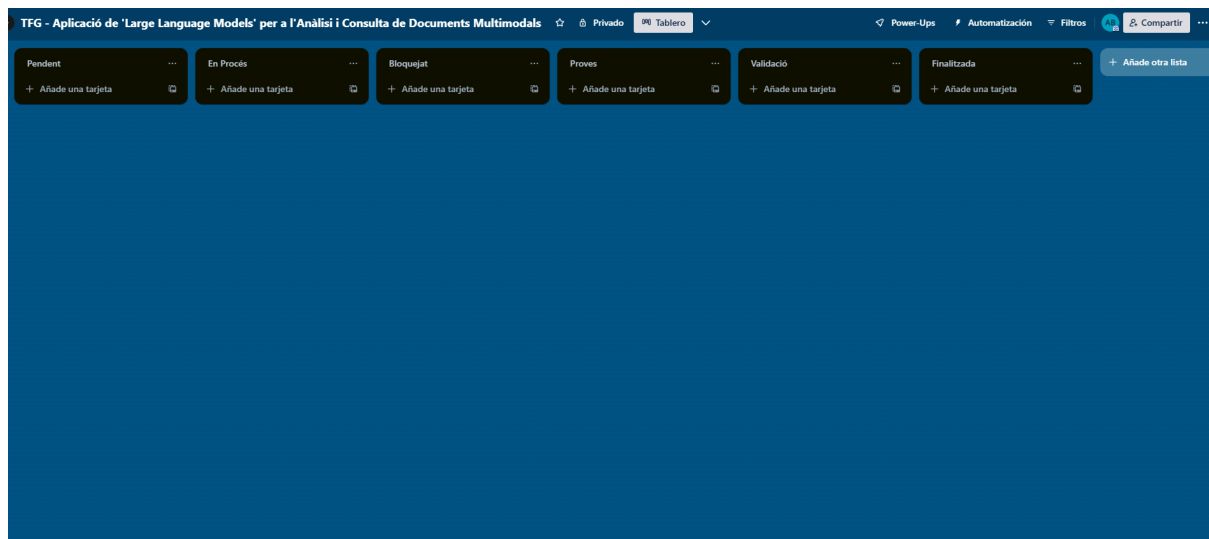


Figura 4.1: Tauler del Trello del projecte. Font: elaboració pròpia

4.2 Control de versions

Per a dur un control de versions del codi font del sistema s'utilitzarà *Git*. *Git* és una eina de control de versions de codi obert que permet fer un seguiment dels canvis en el codi font durant tot el projecte. Permet, addicionalment, treballar de forma col·laborativa amb altres desenvolupadors mitjançant l'ús de repositoris i branques.

Hi ha diferents plataformes que permeten usar *Git*, la més important és *Github*, que serà la que farem servir en aquest projecte. *Github* és un gestor de repositoris *Git*.

Dins del sistema de *Git* existeixen diferents maneres de treballar. En aquest projecte usarem *Gitflow*, que és un model de ramificacions que permet treballar de manera clara i estructurada. La utilització d'aquesta metodologia implica la definició d'una seria de branques i regles per a la correcta gestió. Les branques definides per *Gitflow* que s'implementaran en el projecte són les següents:

Main: És la branca principal del projecte i la que conté el codi estable del projecte. Els diferents commits que rebí aquesta branca representen les versions de codi que han estat provades i estan llestes per a implementar en producció.

Develop: És la branca principal del desenvolupament del projecte. Els diferents commits que rebí la branca seran les funcionalitats acabades i pendents de provar en el conjunt del sistema.

Tasca: Per a cada tasca de desenvolupament es crearà una nova branca amb el nom de la tasca a realitzar en ella. Rebrà els commits relacionats amb la funcionalitat de la tasca i és on es faran les primeres proves. Aquesta branca es crearà des de la branca develop sempre.

Fix: Les branques de fix s'utilitzaran per a solucionar els problemes crítics que puguin aparèixer a la branca Main i, per tant, es crearan a partir d'aquesta i sempre retornaran a

main. És important que aquestes branques es tornin a ajuntar amb Main i no amb altres branques que no continguin les versions estables del projecte.

4.3 Validació i proves

També és important, per tal de mantenir la qualitat durant tot el procés de desenvolupament, fer proves del codi i validar que les tasques fetes compleixen amb els criteris que s'esperen. Per tant, es definiran un conjunt de proves dins l'entorn de *Github*, utilitzant els sistemes de *Github Actions* per a assegurar que el codi del repositori funciona correctament.

Adicionalment i per a fer la validació de les tasques, durant el desenvolupament del projecte es mantindrà una comunicació regular entre el desenvolupador i autor del projecte i el director d'aquest, el Marc Alier Forment, amb l'objectiu de garantir el compliment dels objectius definits anteriorment amb la qualitat esperada. Per a dur a terme aquestes validacions es convocaran reunions periòdiques per a discutir els avenços del projecte i les següents passes d'aquest.

5. Planificació temporal

La planificació de les diferents tasques a desenvolupar durant el projecte és una part crucial per a la bona realització del projecte i ens permet assolir els objectius proposats de manera satisfactòria i dintre dels terminis acordats prèviament. És amb aquest motiu que en els següents apartats s'exposen les tasques que porten a la completa del projecte. També queden definides les corresponents estimacions de temps, dependències i recursos necessaris per a completar aquestes tasques. Totes les tasques queden reflectides en el diagrama de Gantt del projecte que ens permet tenir una visió general del projecte i l'organització d'aquest.

Aquest projecte es va iniciar, dins del marc del Treball de Final de Grau, el 19 de febrer de 2025 amb el curs de GEP (Gestió de Projectes) i té prevista la seva finalització durant l'última setmana de juny de 2025 amb la defensa del projecte davant del tribunal (aproximadament entre el dia 25 de juny i el dia 1 de juliol). Per tant, entre la data d'inici i la de finalització hi ha un total de noranta dies laborables (incloent-hi les dates d'inici i final).

El Treball de Final de Grau consta de 18 crèdits ECTS i cada crèdit representa un total de 30 hores de càrrega de treball. Doncs, la realització completa del projecte hauria d'ocupar unes 540 hores totals. D'aquesta manera, la càrrega de treball total distribuïda en els noranta dies laborables que tenim des de l'inici al final del treball és d'unues 6 hores diàries aproximadament (treballant únicament en dies laborables, sense comptar caps de setmana).

5.1 Descripció de les tasques

Per a facilitar la seva lectura, s'han classificat les tasques del projecte en quatre grups, en funció de la feina a fer en cadascuna d'elles. Els grups són: Gestió del projecte, Desenvolupament, Documentació del projecte i Preparació de la defensa del TFG. Cada

grup conté les tasques d'una manera detallada que faciliten estimar d'una forma més precisa la càrrega de treball i obtenir una millor planificació del projecte.

5.1.1 Gestió del projecte

En aquest primer grup, podem observar totes aquelles tasques de recerca, documentació i planificació del projecte que ens permeten fer un seguiment efectiu d'aquest.

- **GP1 - Contextualització i Abast:** Definició de l'abast del projecte en el context del seu estudi i definició de la metodologia de treball a seguir.
Càrrega de treball: 15 h
Dependències: N/A
- **GP2 - Planificació temporal:** Definició de les diferents fases del projecte, especificació de tasques i recursos.
Càrrega de treball: 10 h
Dependències: GP1
- **GP3 - Pressupost i sostenibilitat:** Definició del pla econòmic del projecte i elaboració de l'informe de sostenibilitat.
Càrrega de treball: 10 h
Dependències: GP2
- **GP4 - Integració del document final:** Agrupació dels documents GP1, GP2 i GP3 tenint en compte les observacions del tutor.
Càrrega de treball: 15 h
Dependències: GP3
- **GP5 - Reunions:** Trobades i comunicacions amb el director del TFG durant tot el projecte per a controlar el correcte desenvolupament d'aquest.
Càrrega de treball: 20 h
Dependències: -

5.1.2 Desenvolupament

El grup de tasques de desenvolupament és el més extens i agrupa totes les tasques relacionades amb el disseny, implementació i testatge de la solució del projecte. Per això s'han dividit altre cop en subgrups dins del grup de desenvolupament. Els subgrups són: Incepció, Implementació, Proves.

Incepció

- **IN1 - Definició arquitectura del sistema:** Especificació de la integració dels LLMs amb la base de dades semàntica. Disseny de l'estructura del procés de processament de dades.
Càrrega de treball: 20 h
Dependències: IN3
- **IN2 - Selecció de les tecnologies i eines:** Elecció del model o models. Elecció de les eines d'extracció de dades.
Càrrega de treball: 15 h
Dependències: IN3
- **IN3 - Preparació d'entorns:** Configuració de repositoris i entorn virtual de treball. Definició de les processos de CI/CD.
Càrrega de treball: 15 h

Dependències: -

- **IN4 - Definició dels formats d'entrada i sortida:** Definició de com es rebran els documents i especificació de l'estructura dels embeddings generats per a ser compatibles amb la base de dades.

Càrrega de treball: 10 h

Dependències: IN3

Implementació

- **IM1 - Implementació del processament de dades:** Desenvolupament del mòdul que carrega els documents. Implementació del model que extreu text i taules d'imatges. Normalitzador de dades extretes en un mateix format (JSON).
Càrrega de treball: 50 h
Dependències: IN
- **IM2 - Desenvolupament del generador d'embeddings multimodals:**
Implementació dels models de processament d'embeddings per a cada tipus (text, imatges i taules). Implementació del fusionador d'embeddings de diferents modalitats per a adaptar-los a la base de dades.
Càrrega de treball: 70 h
Dependències: IM1
- **IM3 - Integració amb la base de dades semàntica:** Connexió dels embeddings generats amb la base de dades. Implementació de consultes semàntiques per a recuperar la informació i optimitzar emmagatzematge i recuperació dels embeddings generats.
Càrrega de treball: 50 h
Dependències: IM2
- **IM4 - Creació de la interfície:** Implementació d'una interfície que permeti rebre documents, processar-los i retornar els embeddings.
Càrrega de treball: 30 h
Dependències: IM3

Proves

- **P1 - Proves unitàries:** Verificar el bon funcionament de cada mòdul de l'aplicació.
Càrrega de treball: 25 h
Dependències: -
- **P2 - Proves d'integració:** Validar que els embeddings de text, imatges i taules es generen correctament i són compatibles amb la base de dades.
Càrrega de treball: 25 h
Dependències: P1
- **P3 - Avaluació de precisió del model:** Mesurar la qualitat dels embeddings amb diferents mètriques que permetin garantir un nivell de qualitat esperat.
Càrrega de treball: 20 h
Dependències: P2
- **P4 - Optimitzacions de rendiment:** Reduir els temps de processament ajustant els valors del model i la base de dades.
Càrrega de treball: 20 h
Dependències: P3

5.1.3 Documentació

La documentació del projecte agrupa totes aquelles tasques relacionades directament amb la redacció de la memòria del projecte i es desenvolupen de manera paral·lela totes les tasques del projecte.

- **D1 - Anotacions i seguiment del procés:** Documentació, notes i apunts de tot el procés de desenvolupament, per tal de fer un seguiment del projecte.
Càrrega de treball: 30 h
Dependències: -
- **D2 - Redacció document de la memòria:** Redacció dels diferents apartats de la memòria final del TFG.
Càrrega de treball: 50 h
Dependències: D1

5.14 Preparació defensa del TFG

Un cop finalitzades totes les tasques de gestió, desenvolupament i documentació s'iniciaran les tasques de preparació de la defensa del projecte.

- **PD1 - Redacció exposició:** Elaboració d'un guió amb els punts més importants a destacar del projecte per a la defensa d'aquest.
Càrrega de treball: 25 h
Dependències: -
- **PD2 - Preparació del material visual:** Elaboració del material visual de suport per a la defensa del TFG.
Càrrega de treball: 15 h
Dependències: PD1

5.2 Recursos

Les tasques definides en l'apartat anterior requereixen l'ús de diferents recursos humans i materials per tal de ser completades amb èxit.

5.2.1 Recursos humans

El desenvolupament de la totalitat de les tasques serà dut a terme per l'autor d'aquest projecte únicament, que assumirà tots els rols del projecte segons la feina a fer en cada moment. Addicionalment, l'autor del projecte comptarà amb el suport del director del projecte Marc Alier Forment i el tutor del mòdul de Gestió de Projectes Fernando Barrabes Naval per a guiar el correcte desenvolupament del projecte.

5.2.2 Recursos materials

El projecte es realitzarà a la residència de l'autor, comptant amb un ordinador amb connexió a internet, webcam i micròfon. També es disposarà de diferent programari, que pot variar segons la tasca a realitzar.

Les eines d'ofimàtica utilitzades en el projecte són *Google Docs* [1], per a redactar la memòria i *Google Spreadsheets* [2] per a l'elaboració del pressupost. La presentació final del projecte s'elaborarà amb *PowerPoint* [3] de Microsoft.

Durant el projecte també es donarà ús a altres eines com *GanttProject* [4] per a generar el diagrama de Gantt, *Gmail* [5] i *Meet* [6] per a mantenir una comunicació directa amb el director del projecte i el tutor de GEP. També, per a fer les diferents entregues *Atenea* [7] i *el Racó de la FIB* [8].

L'editor de codi utilitzat principalment serà *Visual Studio Code* [9], el control de versions del codi es farà a *Github* [10] i la gestió del projecte a *Trello* [11]. Addicionalment, per a tenir un control de l'entorn de desenvolupament també es podrà fer servir *Docker* [12].

5.3 Estimacions

A la taula 5.1 s'observa tot el conjunt de tasques que formen el projecte i que permeten assolir els objectius marcats de forma satisfactòria. Les tasques estan ordenades segons el grup i subgrup al qual pertanyen. En cada una s'indica el codi de la tasca, el nom d'aquesta, la càrrega de treball en hores, les dependències de la mateixa i els recursos específics que es requereixen per a dur-la a terme.

Codi	Tasca	Càrrega de treball (en hores)	Dependències	Recursos
Gestió del projecte		Total: 70		
GP1	Contextualització i Abast	15	-	1,5,7
GP2	Planificació temporal	10	GP1	1,4,5,7
GP3	Pressupost i sostenibilitat	10	GP2	1,2,5,7
GP4	Integració del document final	15	GP3	1,2,4,5,7
GP5	Reunions	20	-	5,6,11
Desenvolupament		Total: 350		
IN	Incepció	Total: 60		
IN1	Definició arquitectura del sistema	20	IN3	1
IN2	Selecció de les tecnologies i eines	15	IN3	1

IN3	Preparació d'entorns	15	-	9,10,12
IN4	Definició dels formats d'entrada i sortida	10	IN3	1
IM	Implementació	Total: 200		
IM1	Implementació del processament de dades	50	IN	9,10,12
IM2	Desenvolupament del generador d'embeddings multimodals	70	IM1	9,10,12
IM3	Integració amb la base de dades semàntica	50	IM2	9,10,12
IM4	Creació de la interfície	30	IM3	9,10,12
P	Proves	Total: 90		
P1	Proves unitàries	25	IM	9,10,12
P2	Proves d'integració	25	P1	9,10,12
P3	Avaluació de precisió del model	20	P2	9,10,12
P4	Optimitzacions de rendiment	20	P3	9,10,12
Documentació del projecte		Total: 80		
D1	Anotacions i seguiment del procés	30	-	1,11
D2	Redacció document de la memòria	50	D1	1
Preparació de la defensa del TFG		Total: 40		
PD1	Redacció exposició	25	-	1
PD2	Preparació del material visual	15	PD1	3
		Total: 540		

Taula 5.1 Resum tasques del projecte. Font: elaboració pròpia

5.4 Diagrama de Gantt

A continuació, a la figura 5.1, podem veure el diagrama de Gantt amb totes les tasques ordenades per dependències i en l'ordre de realització.

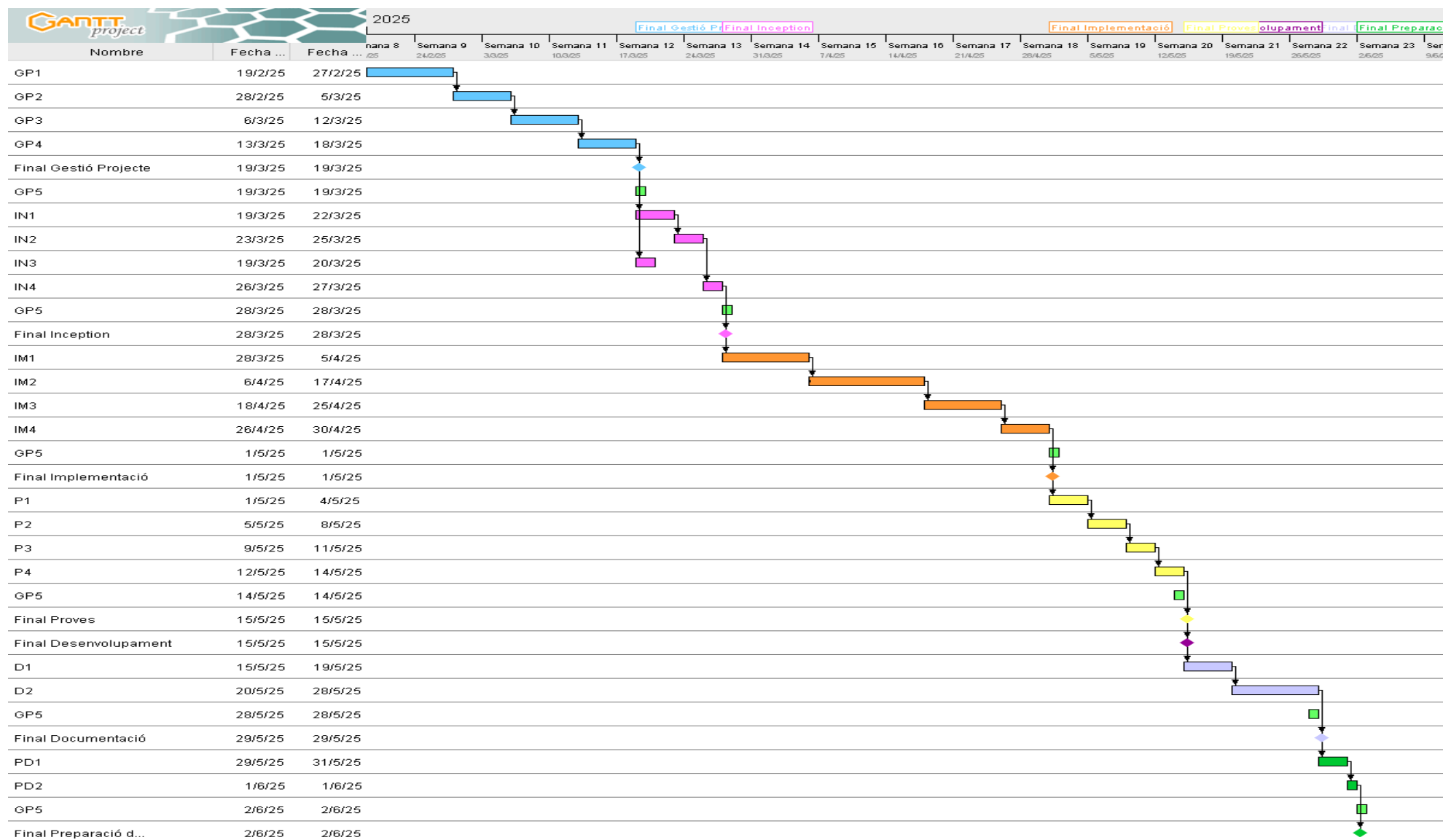


Figura 5.1 Diagrama de Gantt. Font: elaboració pròpia

5.5 Gestió de riscos

Amb l'objectiu de tenir una bona planificació temporal, és necessari tenir en compte els possibles riscos i pensar els possibles escenaris alternatius en cada cas per tal de resoldre els inconvenients derivats d'aquests. A la taula 5.2 podem observar el pla alternatiu per cada un dels riscos exposats en el punt 3.2 d'aquest document.

Risc	Impacte	Probabilitat	Pla alternatiu
Limitacions de recursos i temps	Alt	Mitjana	Priorització de les tasques crítiques i reajustaments del pla de treball per a garantir l'entrega en els terminis previstos.
Desafiaments d'implementació	Mitjà	Mitjana	Sobreestimació de la càrrega de treball de les tasques que requereixen l'ús de tecnologies on el desenvolupador no és expert.
Inexperiència en les tecnologies	Baix	Alta	Sobreestimació de la càrrega de treball de les tasques que requereixen l'ús de tecnologies on el desenvolupador no és expert.
Requeriments computacionals	Alt	Baixa	Preparació de models alternatius per a la substitució dels models més demandants.

Taula 5.2 Impacte i probabilitat de riscos. Font: elaboració pròpia

6. Gestió econòmica

La següent passa, després de definir la planificació de tasques del projecte, és definir una planificació econòmica per tal d'estimar el total dels costos que suposen la realització del projecte. Per això, en els següents apartats es defineixen i s'estimen els costos de personal per activitat, els costos genèrics i les contingències que formaran el pressupost total del projecte.

6.1 Costos de personal per activitat

Per tal de dur a terme una estimació correcta dels costos de personal per activitat (CPA), cal definir quins rols que portaran a terme cada una de les tasques del projecte definides en l'apartat de Planificació temporal. Amb això, el preu per hora fixat i el total de les hores treballades, obtenim el CPA total del projecte.

A la taula 6.1 es mostren els 5 rols que participen en el projecte, el sou brut per hora (en €) de cada un i el sou per hora de cada un dels rols amb l'increment del cost de la Seguretat Social. El cost afegit s'extreu multiplicant el sou brut per 1.3. Totes les dades representades en la taula 6.1 han estat extretes del portal *Glassdoor* que ofereix informació relacionada amb les posicions de treball de qualsevol àmbit i ubicació.

Rol	Sou brut (€/h)	Sou cost Seguretat Social (€/h)
Cap de projecte [CP]	22,12	28,76
Analista Software [AS]	16,35	21,26
Desenvolupador [D]	15,38	20
Dissenyador UX/UI [DU]	13,94	18,12
Tester [T]	12,02	15,63

Taula 6.1 Sous rols projecte en €/hora . Font: elaboració pròpia

Tots els rols definits en la taula anterior seran representats per l'autor del projecte.

Un cop definit el preu per hora dels diferents rols del projecte, ens cal multiplicar el preu per hora de cada rol pel nombre d'hores dedicades a cada tasca. La suma de tots aquests valors representarà el CPA total del projecte. A la taula 6.2 podem veure el desglossament d'hores de cada rol.

Tasca	Hores	Desglossament hores					Cost (€)
		CP	AS	D	DU	T	
GP1	15	15	0	0	0	0	431,4
GP2	10	10	0	0	0	0	287,6
GP3	10	10	0	0	0	0	287,6
GP4	15	15	0	0	0	0	431,4
GP5	20	20	0	0	0	0	575,2
IN1	20	0	20	0	0	0	425,2
IN2	15	0	15	0	0	0	318,9
IN3	15	0	0	10	0	5	278,15
IN4	10	0	10	0	0	0	212,6
IM1	50	0	10	40	0	0	1012,6
IM2	70	0	20	50	0	0	1425,2
IM3	50	0	15	35	0	0	1018,9
IM4	30	0	5	0	25	0	559,3
P1	25	0	2,5	2,5	0	20	415,75
P2	25	0	5	2,5	0	17,5	429,825
P3	20	0	2,5	2,5	0	15	337,6
P4	20	0	2,5	5	2,5	10	354,75
D1	30	30	0	0	0	0	862,8
D2	50	50	0	0	0	0	1438
PD1	25	25	0	0	0	0	719
PD2	15	15	0	0	0	0	431,4
TOTAL	540	190	107,5	147,5	27,5	67,5	12253,175

Taula 6.2 Costos per tasca i CPA total. Font: elaboració pròpia

6.2 Costos genèrics

Els costos genèrics (GC) inclouen tots aquells costos globals i que no van vinculats a cap tasca en específic. Un exemple de cost genèric seria el hardware, l'electricitat, l'espai, etc...

A continuació es descriuen els diferents costos genèrics que s'engloben en el projecte.

6.2.1 Hardware

En aquest projecte únicament usarem un recurs *hardware*, un ordinador portàtil Apple MacBook Air M1. Està estipulat que el període d'amortització pels equips *hardware* és de quatre anys. Per tant, tenint en compte que el MacBook Air M1 és del 2020 no computa pel càlcul de costos genèrics d'aquest projecte (vegeu Taula 6.3).

Hardware	Cost	Amortització
Apple MacBook Air (M1, 13 polzades, 2020)	1.609,00 €	0,00 €
Total		0,00 €

Taula 6.3 Costos totals Hardware. Font: elaboració pròpia

6.2.2 Software

Pel que fa al programari utilitzat en el projecte, es tenen en compte tots els programes esmentats anteriorment en l'apartat 5.2.2 recursos materials. Tots són d'ús gratuït o disposen d'una modalitat gratuïta i, per tant, no cal computar cap amortització d'aquests (vegeu taula 6.4).

Software	Cost
Google Docs	0,00 €
Google Spreadsheets	0,00 €
PowePoint	0,00 €
GanttProject	0,00 €
Gmail	0,00 €
Google Meet	0,00 €
Visual Studio Code	0,00 €
Github	0,00 €
Docker	0,00 €
Trello (versió gratuïta)	0,00 €
Total	0,00 €

Taula 6.4 Costos totals Software. Font: elaboració pròpia

6.2.3 Electricitat

Sabem que el MacBook Air M1 2020, segons les especificacions del fabricant consumeix uns 49,9 W/h, les hores de càrrega de treball estimades pel projecte són 540 hores i tenint en compte que el cost de l'electricitat aproximadament és de 0,1662 €/kWh, el cost total de l'electricitat pel projecte és:

$$(49,9 \text{ Wh}/1000) * 0,1662 \text{ €/kWh} * 540 \text{ h} = 4,48 \text{ €}$$

6.2.4 Internet

Un recurs imprescindible per a dur a terme el projecte és l'internet. Aquest té una quota mensual de 65€ aproximadament. Aquesta quota mensual equival a 0,0903 €/h. Per tant, sabent que el projecte té un total de 540 hores, el preu de l'internet pel projecte és:

$$540 \text{ h} * 0,0903 \text{ €/h} = 48,76 \text{ €}$$

Amb tot això, ja podem saber el cost total del projecte. A la taula 6.5 podem veure, de manera resumida el cost del projecte desglossat en els quatre conceptes esmentats anteriorment.

Concepte	Cost
Hardware	0,00 €
Software	0,00 €
Internet	48,76 €
Electricitat	4,48 €
Total	53,24 €

Taula 6.5 Costos totals costos genèrics. Font: elaboració pròpia

6.3 Contingències

Les contingències són un sobrecost afegit al CPA i al CG per tal de cobrir els obstacles i impediments no previstos durant el projecte. En el cas del sector de l'enginyeria del *software* i el nivell de risc que pren el projecte (moderada) els percentatges estan entre un 10% i un 20%. Per això el percentatge del projecte es fixa en un 15%.

A la taula 6.6 podem observar el cost total del projecte i l'import de les contingències.

Concepte	Cost
CPA	12.253,17 €
CG	53,24 €
Total	12.306,41 €
Contingències	1.845,96 €

Taula 6.6 Cost total projecte i import contingències. Font: elaboració pròpia

6.4 Imprevistos

A continuació, a la taula 6.7, podem veure el cost que representen en hores els diferents imprevistos i el percentatge de probabilitat que es doni un dels imprevistos. Per tant, el cost final dels imprevistos estaria format pel cost estimat de l'imprevist (hores dedicades multiplicades pel preu per hora del rol que ha de resoldre'l) multiplicat pel percentatge de què finalment passi l'imprevist.

Imprevist	Hores i Rol	Rol	Cost estimat	Probabilitat	Cost
Limitacions de recursos i temps	7	Cap de Projecte [CP]	201,32 €	50%	100,66 €
Desafiaments d'implementació	20	Desenvolupador [D]	400,00 €	50%	200,00 €
Inexperiència en les tecnologies	15	Desenvolupador [D]	300,00 €	75%	225,00 €
Requeriments computacionals	10	Analista Software [AS]	212,60 €	25%	53,15 €
Total	578,81 €				

Taula 6.7 Cost imprevistos. Font: elaboració pròpia

6.5 Total

El cost total del pressupost està format per la suma de costos dels apartats anteriors, que són: el cost de personal per activitat (CPA), els costos genèrics (CG), les contingències i els imprevistos. La taula 6.8 mostra els imports de cada un dels costos esmentats i el total del pressupost del projecte.

Concepte	Cost
Costos de personal per activitat	12.253,17 €
Costos genèrics	53,24 €
Contingències	1.845,96 €
Imprevistos	578,81 €
Total	14.731,18 €

Taula 6.8 Cost total del projecte. Font: elaboració pròpia

6.6 Control de gestió

Les estimacions anteriorment descrites poden no ser del tot exactes i poden variar tant en el cost econòmic com el cost en hores de temps. Per a aquest motiu, cal establir mecanismes de control que ens permetin avaluar les desviacions entre el pressupost calculat anteriorment i el cost real del projecte.

Per tal de fer aquesta comparació realitzarem un càlcul de la desviació del cost de cada tasca. Un cop una tasca hagi estat completada satisfactòriament, calcularem les hores reals que ha suposat la realització d'aquesta per a cada rol i les compararem amb les hores estimades al diagrama de Gantt. La diferència de les hores multiplicada pel preu per hora representarà la desviació sobre el pressupost. Per tant, quan la mètrica sigui més propera a 0 voldrà dir que l'estimació ha sigut correcta.

$$\text{Desviació del Cost per Tasca} = \sum_{i=1}^n (\text{Cost estimat hores} - \text{Cost real hores}) \text{Preu hora}$$

On:

- n és el nombre total de rols que desenvolupen la tasca.
- *Cost estimat hores* és el nombre d'hores estimades del rol per la tasca.
- *Cost real hores* és el nombre d'hores reals del rol per la tasca.
- *Preu hora* és el cost que suposa una hora de treball del rol.

El sumatori de les desviacions de costos de totes les tasques ens permetrà saber la desviació global respecte del pressupost del projecte:

$$\text{Desviació de Costos Global} = \sum_{i=1}^n (\text{Desviació del Cost per Tasca})$$

On:

- n és el nombre total de tasques del projecte.
- *Desviació del Cost per Tasca* és la mètrica de desviació del cost de la tasca.

Igual que en la mètrica anterior, si obtenim un valor negatiu, el pressupost general del projecte no s'haurà complert i, per tant, haurem de fer ús de la partida de contingències per cobrir la sobre despesa. També cal aclarir que el control de gestió aquí descrit se centra únicament les desviacions del CPA, ja que els costos genèrics, són invariables i, com a resultat, podem considerar que no hi haurà desviació en aquest aspecte.

7. Informe de sostenibilitat

7.1 Autoavaluació

Després de completar l'enquesta proposada a la rúbrica sobre les competències de sostenibilitat i he pogut reflexionar sobre els meus coneixements sobre el tema.

En relació amb la dimensió ambiental, crec que és l'àmbit en què millor em puc desenvolupar, donat que durant totes les assignatures del grau hem pogut tractar els aspectes de sostenibilitat en relació amb el medi ambient. A través dels diferents tallers, iniciatives i aprenentatges dins les assignatures he pogut adquirir una base de coneixement prou correcta per a analitzar, des del prisma de la sostenibilitat ambiental, aquest i altres projectes.

Un altre aspecte tractat en l'enquesta és la dimensió social del projecte. Aquest aspecte no està tan treballat durant el grau, però gràcies a les pràctiques curriculars que he realitzat durant el grau he rebut una formació mínima per a què, ajudat dels companys i el meu criteri, fer anàlisis prou consistents dels projectes en què formo part.

Finalment, la tercera dimensió és l'econòmica. Segurament és la dimensió en què menys experiència tinc i, per tant, en la que més he d'aprofundir per tal de fer una bona anàlisi de sostenibilitat d'un projecte. Durant el grau s'han donat poques pinzellades sobre aquesta dimensió i la formació que he rebut ha sigut principalment autodidacta o en les pràctiques d'empresa.

Doncs, tot i que considero que tinc un nivell prou alt per a fer una anàlisi de sostenibilitat d'un projecte, hi ha molt de marge per a millorar en l'àmbit de la sostenibilitat i més tenint en compte que la sostenibilitat és un aspecte fonamental i crític dels projectes en l'àmbit de l'enginyeria.

7.2 Dimensió econòmica

En relació amb el projecte que es tracta en aquest document, s'ha definit un pressupost ampli que pretén abastar objectius com la investigació de tecnologies, la implementació de models de LLM o la integració de models amb bases de dades semàntiques. Tot això implica un cost en els diferents rols que han de desenvolupar les tasques relacionades amb aquests objectius.

En l'estat de l'art del projecte, es parla de les diferents solucions ja implementades i en el mercat d'altres competidors i ens permet veure la varietat de productes amb els diferents costos de cada un d'ells.

A més la solució proposada, busca millorar les solucions ja implementades augmentant en un grau més el nivell de modalitat de les dades. També, en basar la solució en models i tecnologies de codi obert s'espera reduir els costos de desenvolupament d'aquesta.

7.3 Dimensió ambiental

En aquest aspecte, el projecte s'ha centrat, principalment, en l'impacte ambiental derivat del consum en el desenvolupament. En l'apartat de consum energètic i de materials.

Tot i que, en l'estat de l'art es planteja l'optimització del sistema per a aconseguir una eficiència el més gran possible i, per tant, reduir el consum del sistema, no és l'objectiu principal del projecte.

7.4 Dimensió social

Finalment, l'aspecte social del projecte és un dels àmbits més importants del projecte. S'espera que el projecte representi una millora respecte de les solucions actuals per als usuaris actuals i que a més a més serveixi de base per a altres projectes de codi obert similars.

8. Referències

Li, X., Zhang, Y., & Wang, J. (2023). Multimodal Large Language Models: A Survey. arXiv. <https://arxiv.org/abs/2311.13165>

Chen, L., Huang, T., & Zhao, M. (2024). Leveraging Large Language Models for Multimodal Search. arXiv. <https://arxiv.org/abs/2404.15790>

Ghosh, A., & Patel, S. (2021). Efficient Multi-Modal Embeddings from Structured Data. arXiv. <https://arxiv.org/abs/2110.02577>

Rodríguez, P., & Martínez, L. (2024). Semantic-Aware Representation of Multi-Modal Data for Data Ingress: A Literature Review. arXiv. <https://arxiv.org/abs/2407.12438>

Management Solutions. (2023). El auge de los large language models. Management Solutions. <https://www.managementsolutions.com/sites/default/files/minisite/static/72b0015f-39c9-4a52-ba63-872c115bfbd0/llm/pdf/auge-de-los-llm-03.pdf>

Google Cloud. (2024). Obtén incorporaciones multimodales. Google Cloud. <https://cloud.google.com/vertex-ai/generative-ai/docs/embeddings/get-multimodal-embeddings?hl=es-419>

García, J., & López, M. (2023). Bases de datos semánticas: Evolución y aplicaciones. Revista de Software y Bases de Datos, 15(2), 45-67. <https://revistas.unla.edu.ar/software/article/view/1278/1113>

Pérez, A., & Ramírez, C. (2021). Sistemas de extracción de información: Procesos y técnicas aplicadas. Universidad de Córdoba. <https://www.uco.es/investiga/grupos/labindoc/images/files/neoinstrumenta/articulos/neoins2013n3.pdf>

Fernández, L., & Torres, R. (2020). Extracción y representación de conocimiento a partir de corpus textuales. ResearchGate. Recuperat de https://www.researchgate.net/publication/265293565_Extraccion_y_representacion_de_conocimiento_a_partir_de_corpus

Zhang, Y., & Liu, H. (2024). Semantic-Aware Representation of Multi-Modal Data for Data Ingress: A Literature Review. arXiv preprint. <https://arxiv.org/abs/2407.12438>

Kim, D., & Lee, J. (2023). Multimodal Neural Databases: An Approach to Multi-Modal Data Querying at Scale. arXiv preprint. <https://arxiv.org/abs/2305.01447>

Wikipedia. (2024). Aleph Alpha. Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Aleph_Alpha

Wikipedia. (2024). Spark NLP. Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Spark_NLP

Wikipedia. (2024). Gemini (modelo de lenguaje). Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Gemini_%28modelo_de_lenguaje%29

Amazon Web Services (AWS). (2024). Amazon Titan Models. AWS Bedrock. <https://aws.amazon.com/es/bedrock/amazon-models/titan/>

OpenAI. (2021). CLIP: Connecting text and images. OpenAI. <https://openai.com/index/clip/>

OpenAI. (2021). CLIP (Contrastive Language-Image Pre-Training). GitHub. <https://github.com/openai/CLIP>

Atlassian. (s.f.). Flujo de trabajo de Gitflow. Atlassian. <https://www.atlassian.com/es/git/tutorials/comparing-workflows/gitflow-workflow>

Glassdoor. (s.f.). Salarios y opiniones sobre empresas. Glassdoor. <https://www.glassdoor.com>

Apple Inc. (2020). MacBook Air (M1, 2020) - Especificaciones técnicas. <https://support.apple.com/es-es/111883>

Morales Fernández, J. (2025, 10 de marzo). Precio kWh España a 10 Marzo 2025: Evolución y Comparativa. Tarifaluzhora.es. <https://tarifaluzhora.es/info/precio-kwh>