NLP A LES AULES

Gerard Prats Gili

**Resum**—Aquest projecte busca analitzar l’estat d’art de leis eines de processament de llenguatge natural en l’actualitat, i analitzar el paper que han tingut les Tecnologies de la Informació i la Comunicació en el sector educatiu des d’un punt de vista tècnic i filosòfic. Després d’aquest anàlisi, es procedeix al disseny i al desenvolupament d’una Prova de Concepte que busqui implementar les tecnologies que s’han descobert amb la recerca en un cas d’ús concret de l’àmbit educatiu. Aquesta eina busca ajudar als alumnes amb dificultats d’aprenentatge, promovent la interactivitat i l’accessibilitat, i també busca aportar al professorat amb eines per poder entendre les necessitats individuals de cada alumna mitjançant les TIC. També es presenten propostes de millores i de funcions que es poden afegir un cop finalitzat el treball, per continuar desenvolupant la eina en un futur.

**Paraules clau**—tecnologies de la informació, TIC, GPT, processament de llenguatge natural, docència, interacció, chatbots, aprenentatge.

**Abstract**—This project seeks to analyze the State of the Art of Natural Language Processing (NLP) tools, and to examine the role that Information and Communication Technologies have had in the educational sector from a technical and philosophical perspective. After this analysis, we will proceed to design and develop a Proof of Concept for an application that implements the technologies discovered throughout the research to solve a specific problema. This tool seeks to assist students with learning difficulties, promoting interactivity and accessibility, and also aims to provide teachers with tools to understand the individual needs of each student through information gathered from the technologies employed. Proposals for improvements and funcions that can be added once the project is completed are also presented, so that the tool can continue to be developed in the future.

**Index Terms**—information technologies, ICTs, GPT, natural language processing, teaching, students, interaction, chatbots, learning.

—————————— ◆ ——————————

# 1 Introducció - Context del treball

E

L camp de la intel·ligència artificial (IA) ha experimentat un creixement sense precedent durant aquests últims anys, revolucionant la forma en què realitzem diverses tasques en la nostra vida quotidiana. Aquest impacte s’estén a molts àmbits, com ara l’educació, on les tecnologies emergents com el processament del llenguatge natural (NLP) i els models com Generative Pre-trained Transformer (GPT) estan proporcionant noves oportunitats per a l’ensenyament i l’aprenentatge. Aquest projecte pretén explorar les implicacions d’aquestes tecnologies i les seves aplicacions en l’àmbit educatiu, i abordar les consideracions ètiques que aquestes innovacions fan aparèixer.

Tot i que encara no sabem l’efecte complet que tindran les tecnologies emergents com GPT sobre la nostra societat, hi ha molts articles que han explorat les seves ramificacions. S’han plantejat debats sobre les possibles aplicacions de les eines d’IA en l’educació, els treballs i altres aspectes de la nostra vida quotidiana. Però també, simultàniament, els dilemes que apareixen amb aquesta tecnologia han accelerat les converses que busquen analitzar la IA des d’un punt de vista filosòfic i ètic per assegurar-nos que aquestes siguin beneficioses i no perjudicials per a nosaltres.

Per assolir l’objectiu d’aquest treball, presentem els següents objectius.

* Anàlisi de la intersecció entre la recerca en tecnologia i l’educació, amb un èmfasis en la IA i les Tecnologies de la Informació i la Comunicació (TIC).
* Introducció al concepte de NLP i GPT, explicant els algoritmes i les arquitectures utilitzades en l’actualitat.
* Fer ús del coneixement de l’estat d’art per dissenyar una eina que utilitzi les tecnologies emergents per ajudar a solucionar un problema concret en l’àmbit de l’educació.

————————————————

1. E-mail de contacte: gpratsgili@gmail.com
2. Menció realitzada: Computació
3. Treball tutoritzat per: Josep Lladós Canet (Ciències de la Computació)
4. Curs 2022/23

“Mes” de 2023, Escola d’Enginyeria (UAB)

# 2 Context i estat d’art

## 2.1 Les TIC en l’educació i les noves formes d’aprenentatge

Per entendre el rol de les Tecnologies de la Informació i la Comunicació (TIC) a l'educació, primer haurem de saber què són aquestes eines. Relativament, les TIC són eines molt recents que ens permeten tractar la informació i presentar-la de formes que abans no era possible. Consuelo Belloch Ortí, en el seu article *Las Tecnologías de la Información y Comunicación (T.I.C.)* [1], va exposar les diferents definicions que podem trobar per a aquest tipus de tecnologies, i hi podem intuir uns conceptes comuns:

* La **interactivitat**: La informació es crea o es mostra mitjançant la interacció d’un usuari amb el seu ordinador o amb una aplicació. A causa d’això, podem dir que les TIC són interactives i que s’adapten a l’ús que se’ls hi doni.
* La **interconnexió**: Les eines TIC, freqüentment, no estan aïllades. En molts casos (com pot ser, per exemple, les eines que ofereix Google) es poden connectar diferents tecnologies per tal d’oferir un nou servei que cobreixi necessitats diferents. Això, alhora, permet **innovació**.
* La **instantaneïtat** i la **immaterialitat**: Tot i que l’article separa aquestes dues qualitats, realment fan referència al mateix concepte. Les TIC fan que la informació sigui altament **disponible**, de forma ràpida i global.
* La **diversitat**: No hi ha un únic cas d’ús per aquestes tecnologies, i es poden integrar per millorar tota mena de processos en una gran varietat d’àmbits.

Per aquests motius i altres, les TIC han sigut una eina clau per l’evolució dels nostres treballs. La gran varietat d’usos que se’ls hi pot donar, la disponibilitat que tenen i la seva interactivitat que busca disminuir la complexitat de certes tasques es pot aplicar a molts sectors. Entre ells, l’educació ha sigut un dels camps afectats. Un article de la pàgina de Innovació en Formació Professional (IFP), *¿Cúales son las herramientas TIC en la educación?* [2], exposa les raons per les quals això passa: les TIC permeten facilitar les tasques tant a l’alumnat com al professorat.

* **Google Classroom**: Una plataforma de comunicació entre l’alumnat i el professorat que, a més, serveix per gestionar entregues i fer disponible material rellevant per les assignatures.
* **Kahoot**: Una plataforma que transforma l’educació en un joc, que fomenta la participació de l’alumne.
* **Office 365**: Un paquet d’eines que permet ajudar tant amb la creació de contingut (Microsoft Word, PowerPoint, Excel), la gestió de projectes (Microsoft Project), la comunicació (Outlook, Microsoft Teams) i l’accés de dades (OneDrive).

Podem veure que aquestes eines són una extensió del que hem vist al context històric de **l’annex A1**. Tal com hem vist amb l’evolució de l’educació, aquestes eines ajuden a presentar la informació de forma més immersiva i també la fan més accessible per a tothom. Per aquest motiu, si volem desenvolupar una nova eina que utilitzi les TIC per ajudar en el sector educatiu, hauria de ser: **interactiva**, **accessible** i, a més a més, **ajudar tant a l’alumnat com al professorat**.

## 2.2 Estat d’art dels models d’intel·ligència artificial actuals: NLP i GPT

**Natural Language Processing (NLP)** es refereix al processament de llenguatge natural realitzat per algoritmes d’intel·ligència artificial per tal d’analitzar texts escrits amb llenguatge natural (és a dir, el llenguatge que parlem en el nostre dia a dia). És una branca de la informàtica que té intersecció amb la **lingüística**, i involucra processos de **reconeixement de veu**, d’**anàlisi** sintàctica (estructural) i semàntica (el significat) de text i la **generació de text**.

John Hutchins (2014) [3] indica que té el seu origen al segle XX, en els anys 50, amb l’aparició d’un traductor automàtic. Aquest era molt limitat, i només podia traduir una quantitat limitada de frases, però va començar una època d’optimisme respecte a la **traducció automàtica**. Des del 1954, hem vist l’aparició de molts tipus de models d’intel·ligència artificial per tal d’intentar millorar el processament de llenguatge natural. Hem passat de models que implementaven diccionaris i normes de gramàtica **predefinides**; a **models probabilístics** (com bé podria ser un **Bag of Words**), que busquen patrons o variables a partir de les freqüències d’una paraula en un text; a **Recurrent Neural Networks (RNNs)**, que tenen en compte el significat general d’una paraula i la posició d’aquesta a una frase de forma seqüencial.

Alguns d’aquests models encara s’utilitzen avui en dia, però l’estat d’art quant a NLP des del 2017 són els **Transformers**. L’article *Attention Is All You Need* [4] de Google va proposar aquesta nova arquitectura basada en **capes d’atenció**. La Figura 1 mostra l’arquitectura dels transformar.

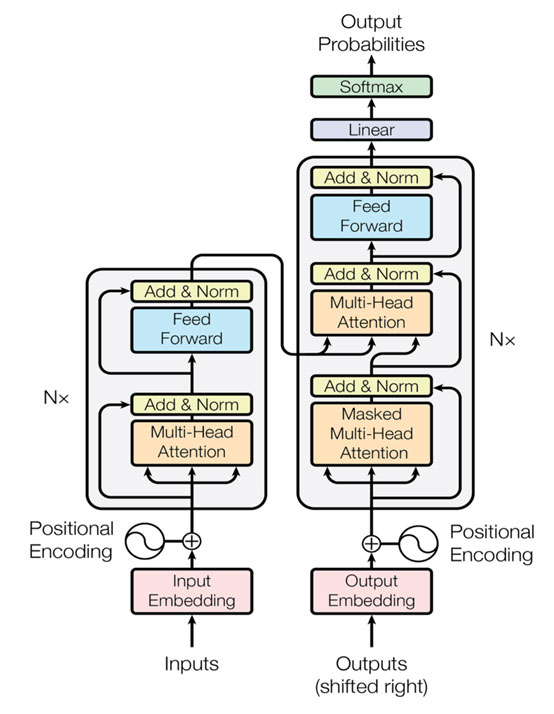


Figura 1. Arquitectura d’un Transformer.

Els Transformer estan basats en un mecanisme d’atenció. Mitjançant una fase d’entrenament i tenint en compte la posició de la paraula respecte als altres elements de la frase, es generen uns **embeddings** per unes paraules donades amb **context** en forma de **vectors**. Després, també s’afegeix una capa **d’atenció** que considera l’impacte que té una paraula donada sobre la resta de paraules d’un text donat.

Els beneficis d’aquesta arquitectura respecte als RNNs és la seva velocitat, a més del seu alt grau de paral·lelització a la fase d’entrenament. Els resultats que aporten a més, gràcies al fet que no són seqüencials i a la incorporació de context i atenció, són molt millors que els models anteriors de NLP. Des de l’aparició dels Transformers, hi ha hagut un creixement exponencial en l’evolució de les tecnologies de NLP. Entre els nous models que han aparegut a conseqüència dels Transformers, podem trobar el **Generative Pre-trained Transformer (GPT)**.

GPT es refereix a una sèrie de models per OpenAI que utilitzen Transformers per generar texts. Els models més anteriors (GPT-1 i GPT-2) es poden trobar alliberats a Hugging Face, mentre que els nous models (GPT-3.5 i GPT-4) només es poden accedir mitjançant APIs. Es caracteritzen per les grans quantitats de dades que s’utilitzen per entrenar-los, i per la gran quantitat de paràmetres que tenen els nous models.

A picture containing text, screenshot, diagram, plot

Description automatically generated

Figura 2. Resultats de GPT-3.5 (blau) i GPT-4 (verd) en exàmens, segons l’article d’OpenAI.

Gràcies a l’estructura de Transformers, GPT pot generar text que pot ser útil per moltes àrees d’estudi, tal com podem veure amb els resultats que pot obtenir a exàmens (il·lustrat per la Figura 2). Es pot utilitzar per fer xatbots avançats que generen text que sembla humà, es pot utilitzar per agilitzar tasques de programació i de resumir, i per assistir amb traduccions. Una de les seves debilitats, tal com s’exposa al paper *GPT-4 Technical Report* [5], és que té una tendència a al·lucinar dades.

Per aquest motiu, es fa ús de **bases de dades vectorials** per emmagatzemar dades. Ja que els models de Transformers utilitzen **embeddings** (representacions vectorials del text), podem transformar les nostres dades en vectors i emmagatzemar-les en una base de dades (Redis, Weaviate o Pinecone, per exemple). Això ens permet, després, buscar informació mitjançant la comparació de vectors. Per exemple, si emmagatzemem diferents articles en aquestes bases de dades, tindran un **vector** **associat que representa el seu contingut i context**. Això em permet buscar temes específics, i comparar la **representació vectorial** del concepte que volem buscar amb els vectors de la base dades per obtenir els resultats més propers (Figura 3).

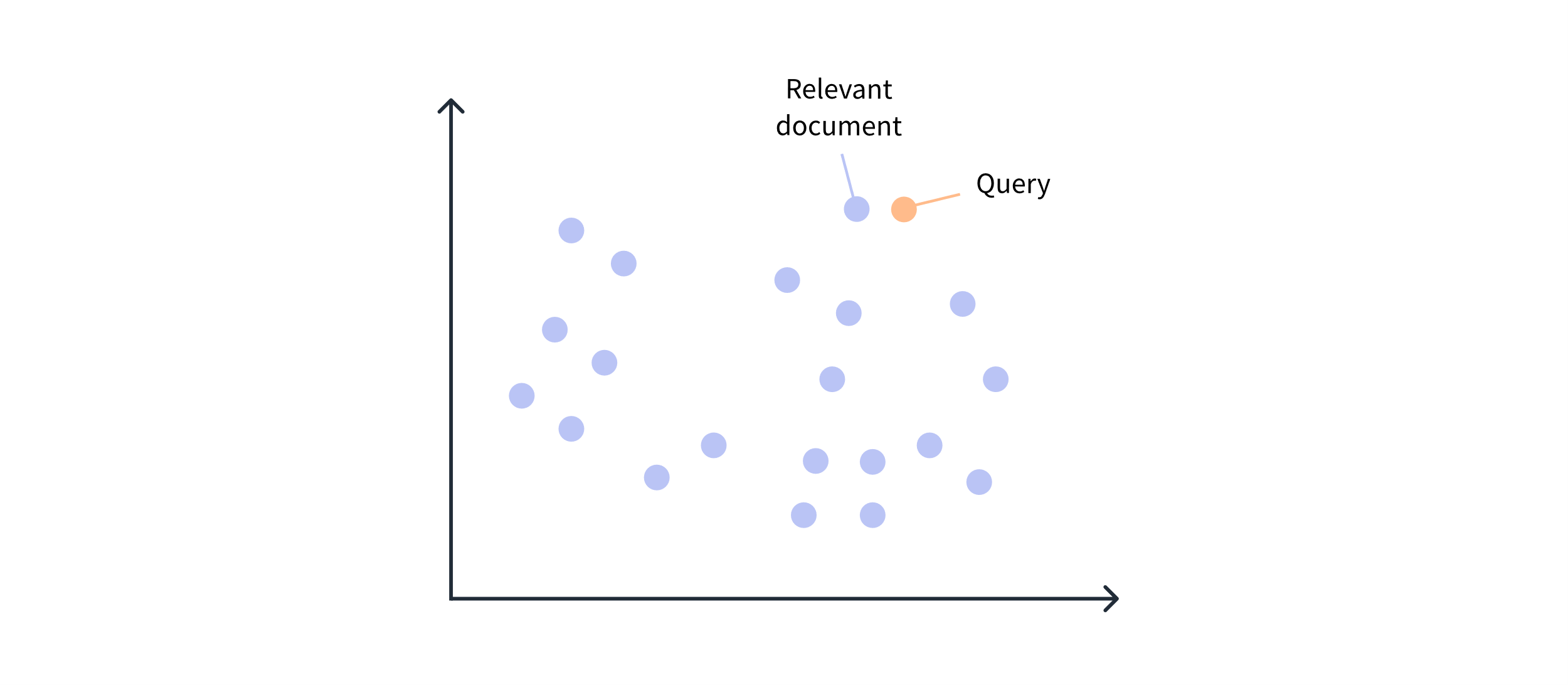


Figura 3. Cerca semàntica de documents. NLP Course: <https://huggingface.co/learn/nlp-course/chapter5/6?fw=tf>

Podem utilitzar la informació que tenim en les bases de dades vectorials afegir informació als nostres **prompts** (els inputs de GPT) i així afegir una capa de **context** per evitar aquesta al·lucinació de dades al restringir les respostes que aporta GPT a la informació que hi ha en el prompt. Entrarem en més detall sobre aquesta implementació en l’apartat de metodologia, ja que serà una eina que utilitzarem per realitzar la nostra aplicació.

## 2.3 Conclusions a partir de l’estat d’art

A partir del que hem vist analitzant el context actual de les TIC a l’educació i l’estat d’art de les eines de NLP, podem concloure el següent: la tecnologia, en el context de l’educació, s’ha de centrar a ajudar tant als estudiants com als professors per igual. S’ha de promoure la interactivitat i l’accessibilitat, per tal de permetre que més gent tingui accés a l’educació i per minimitzar les dificultats que es poden presentar. Actualment, una de les majors dificultats de l’educació és la del **tractament d’alumnes amb necessitats diferents**, ja que és necessari que hi hagi un grau de generalització per tal de dissenyar un pla d’estudi efectiu.

També, tal com entrarem en més profunditat en **l’annex A2**, sempre ha d’haver-hi un grau de monitoratge o de control per part d’un professor per evitar l’ús irresponsable de l’eina i per evitar que, en cas d’error, aquest perjudiqui l’alumnat de forma negativa. S’ha de promoure la col·laboració entre alumnat i professorat mitjançant l’eina, i dependre completament de la intel·ligència artificial no només seria inviable, sinó que també seria irresponsable.

Per això, durant l’apartat de **metodologia** d’aquest informe, veurem com podem aplicar el que hem après amb aquesta recerca per tal de dissenyar i crear una eina orientada a la gent amb TDAH. El que s’espera és que, a partir del disseny d’una eina que ajudi en un cas d’ús concret, es pugui utilitzar allò que s’ha presentat en aquest article pel desenvolupament d’eines futures.

# 3 Metodologia

## 3.1 Explicació de l’eina a realitzar

Tal com s’ha mencionat a les conclusions de l’apartat anterior, l’objectiu d’aquesta eina és la de proporcionar un suport pedagògic personalitzat a l’alumnat. Més concretament, s’ha tingut en compte les necessitats dels alumnes amb TDAH: proporcionar una manera d’aprendre que els permeti **participar activament** per tractar els problemes de concentració, ajudar-los a **dividir tasques grans en objectius més petits** i que els doni **motivació per acomplir tasques mitjançant la retroalimentació.** Alhora, també es busca ajudar al professorat a entendre les necessitats exactes de l’alumnat per tal de saber com dirigir les classes.

Per aconseguir això, desenvoluparem un sistema de documentació que evolucioni amb el mateix estudiant amb **GPT-4**. Es podrà **enviar documents a l’eina** per tal de permetre que l’alumnat pugui **realitzar preguntes** que els ajudin a entendre el contingut dels documents que ha enviat (una base de dades que acumuli el coneixement de l’alumne, amb arxius de text, àudio, vídeo...). Per tal d’evitar que la eina al·lucini dades falses, **només utilitzarà la informació que apareix en el mateix document**. Es pot utilitzar un sistema de xat, similar al que podem veure amb ChatGPT i Bing Search, per tal de parlar amb l’eina. L’objectiu és que no es puguin passar només arxius de text, sinó que també es podran passar arxius de vídeo, d’audio i altres tipus de documents per tal d’obtenir informació.

Per tal d’ajudar tant a professors com alumnes, hi hauria dos tipus d’usuaris: l’**alumne**, que pot utilitzar l’eina per estudiar o per organitzar-se, i el **docent**, que podrà veure els xats que té l’alumnat amb l’eina per tal de comprendre amb quines coses té més dificultats l’alumne. Un compte de docent també podrà enviar documents als seus alumnes mitjançant l’eina, per aportar materials d’estudi que poden ser rellevants o tasques a realitzar.

Hem de considerar una sèrie de riscos abans de començar a desenvolupar l’eina, a partir de les preocupacions que s’han expressat respecte a ChatGPT en el context de l’educació.

* S’ha d’evitar que l’eina aporti les respostes als alumnes directament, per evitar que l’alumne la pugui utilitzar **sense reflexionar** sobre el contingut del document. Això es pot realitzar mitjançant **prompt engineering**, la tècnica de modificar els inputs que es passen a l’API de GPT per donar indicacions de com haurien de ser les respostes. Si indiquem a GPT-4 que les respostes que ens aporta no han de respondre directament les preguntes de l’alumnat, sinó que haurien de fer preguntes per tal que raonin les respostes ells mateixos, podem evitar que s’utilitzi per a plagiar.
* Per assegurar-nos que no al·lucina les dades, a més de només utilitzar informació que aparegui en els documents dels alumnes, també hauria de proporcionar a l’alumne amb una indicació de la localització exacte de la informació que ha trobat (pàgina, paràgraf, etc.) per verificar-la.
* Els professors han de poder veure les conversacions que es tenen amb l’eina, i obtenir un resum de les preguntes que es fan. Això permet que el professor supervisi l’ús de l’eina i, a més a més, entengui la situació dels seus alumnes en tot moment. D’aquesta forma, es poden fer tutories personalitzades amb els alumnes a partir de les dificultats que veuen que han tingut i permet fer un seguiment més exacte.

Llavors, amb aquestes consideracions, tenim una idea exacta de com hauria de ser la nostra eina.

1. Requerim un sistema d’usuaris que ens permeti diferenciar entre professorat i alumnat.
2. Un cop hem entrat en el nostre compte, tindrem una pantalla on es podran enviar els documents que tenim per tal que l’eina els processi. En cas que el que envia un document sigui un professor, s’hauria de poder compartir aquest document amb els seus alumnes.
3. S’ha de tenir una interfície de xat a la qual fer preguntes sobre els nostres documents.
4. S’ha de poder tenir accés a tots els nostres documents i seleccionar sobre quins documents volem fer les preguntes. En cas que cap es seleccioni, es realitzarà una cerca general.
5. El professor ha de tenir accés als xats dels alumnes i preguntar quines han sigut les preguntes més freqüents que s’han realitzat.

En el següent apartat, veurem l’arquitectura de l’aplicació i una explicació sobre la implementació dels mòduls que realitzarem.

## 3.2 Arquitectura de l’aplicació i mòduls

Amb l’objectiu de realitzar una prova de concepte que presenti les funcionalitats bàsiques de l’eina al final d’aquest projecte, desenvoluparem l’aplicació **de forma incremental**. Definirem una sèrie de **mòduls independents** que després es puguin unir per tal de crear l’eina. Això, a més a més, és vital a causa de la possibilitat que apareguin nous models en un futur que aportin millors resultats als quals utilitzem en un principi; hi ha, per tant, la necessitat que sigui fàcil substituir-los canviant el mínim nombre d’elements possibles.

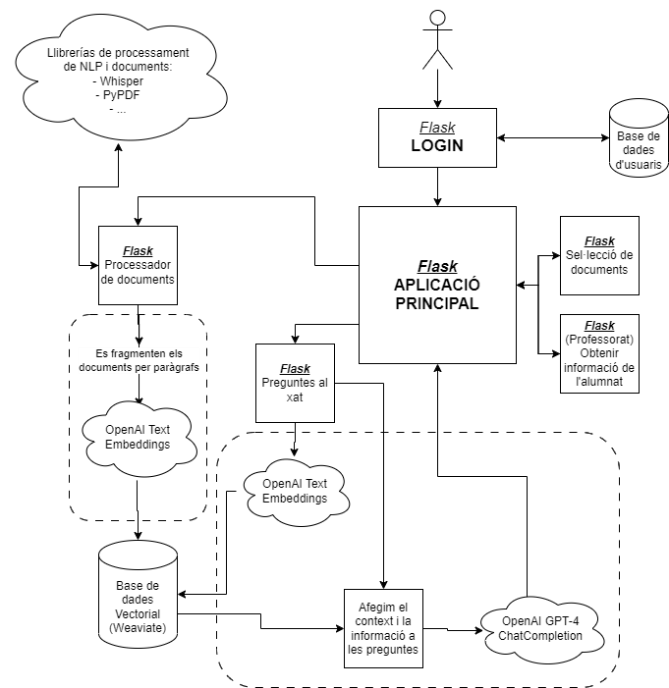


Figura 4. Arquitectura de l’aplicació.

A la Figura 4, podem veure l’arquitectura proposada per l’aplicació que realitzarem. En aquest apartat veurem com hem implementat els diferents mòduls que hi veiem.

## 3.2.1. Aplicació Web

L’aplicació s’ha realitzat amb **Flask**. És un framework de Python que permet crear aplicacions web i ens aporta tot allò que necessitem.

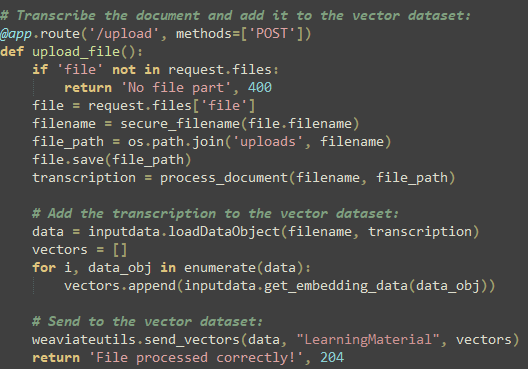


Figura 5. Exemple de codi de Flask.

Com veiem a la Figura 5, Flask ens permet assignar funcions a rutes per poder cridar-les com a API. També ens proporciona amb funcions per renderitzar contingut HTML, i així presentar una interfaç web a l’usuari amb la qual poder interactuar amb l’aplicació o per mostrar-li allò que sigui necessari.

El mòdul de **login**, **la interfície de l’aplicació** i les funcions que requereixen input de l’usuari es fan mitjançant Flask. Aquestes són les finestres que podem visualitzar gràcies a Flask.

* El **login** és una implementació bàsica amb correu, usuari i contrasenya (xifrada, per tal d’assegurar que el sistema és segur). També es considera fer que el login estigui lligat amb el compte de Google a la versió final de l’aplicació per tal de fer que el sistema sigui més segur. En el moment de crear un compte, es pregunta a l’usuari si es un alumne o un professor.
* **L’aplicació principal** permet a l’usuari utilitzar totes les funcions de l’aplicació. Té una àrea en la qual podem inserir els nostres documents (àudio, vídeo i qualsevol arxiu de text) per tal que el sistema ho pugui processar. També, a l’altre costat, té una finestra on es pot xatejar amb l’eina a partir dels documents que s’han enviat fins ara. S’ha escollit un estil similar al que es podria veure amb serveis de missatgeria com WhatsApp per aportar als usuaris amb un grau de familiaritat.

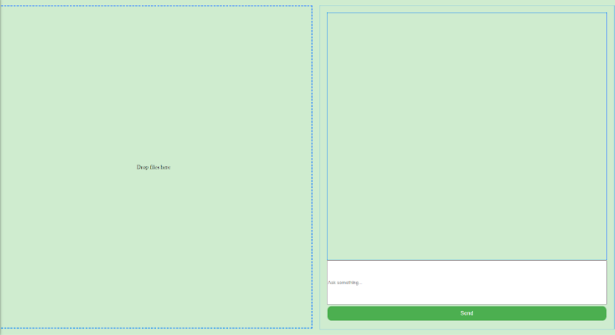


Figura 6. Pantalla principal de l’aplicació.



Figura 7. Exemple de xat amb l’eina.

* Una pantalla on **escollir els documents.** Cada usuari té accés als documents que ha enviat fins ara, i tindrà diferents opcions. Pot escollir si vol que aquest document s’utilitzi per fer preguntes i pot eliminar el document si veu que no el necessita.
* Una pantalla on **els professors puguin visualitzar informació sobre els seus alumnes**. Aquí poden visualitzar els xats que han tingut els seus alumnes, i utilitzar aquests logs com a documents per tal de fer preguntes sobre l’ús que l’alumne li ha donat a l’eina.

Aquestes són les funcions a les quals dona accés la nostra aplicació web creada amb Flask. Ara, descriurem els mòduls que duen a terme les funcions descrites anteriorment.

## 3.2.2. Processament de documents

El **processador de documents** utilitza diferents tecnologies segons el tipus del document d’entrada. Hem creat un mòdul anomenat *transcription* que s’encarrega d’extreure el text.

* Per arxius de text (txt), llegeix el contingut del document sense cap tracte especial. Les funcions base de Python per obrir arxius ens serveixen.
* Per arxius PDF i de Word, utilitzem les llibreries especialitzades en aquest tipus de documents, **PyPDF2** i **DOCX** respectivament. Aquestes llibreries tenen funcions per extreure el text del document.
* Per arxius d’àudio, utilitzem **Whisper AI**. Whisper és una eina d’intel·ligència artificial desenvolupada per OpenAI que permet fer transcripcions a partir d’arxius d’àudio. Tot i que és una eina de codi obert, els nostres tests unitaris ens han mostrat que és millor utilitzar l’API gràcies a la seva velocitat de resposta.
* Per vídeos, utilitzem la llibreria **moviepy**, que ens permet extreure l’àudio a partir d’un arxiu de vídeo. Després, podem utilitzar l’àudio que hem extret per fer el mateix processament que hem realitzat anteriorment amb Whisper AI.

Aquesta llibreria ens permet obtenir el text de tots els documents que necessitem, tot i que sempre es poden afegir nous mètodes de processament si volem afegir nous tipus de documents en un futur.

Un cop hem processat el text, necessitem tenir una forma d’emmagatzemar-los en la base de dades per fer cerques semàntiques. Per fer això, fem ús de les **bases de dades vectorials** que s’han explorat en l’apartat de l’estat d’art. Fem ús d’una base de dades vectorial de codi obert anomenada *Weaviate*, tot i que es put substituir per altres alternatives en un futur. Per inserir els documents a la base de dades, però, primer els hem de **vectoritzar**.

1. Per vectoritzar els documents, s’han de transformar en **embeddings**. Com estem utilitzant GPT-4, actualment, utilitzem el sistema d’embeddings que ofereix OpenAI com a part de la seva API. Com accepta un límit de 8.000 tokens per cada crida a l’API, hem de **fraccionar els documents**.
2. Utilitzem *tiktoken*, la llibreria que utilitza OpenAI per definir els encodings dels caràcters, per mesurar quants tokens necessitarem per representar un text d’entrada.
3. Utilitzem el *TextSplitter* de *LangChain* per dividir el text d’entrada en fragments de 400 tokens cadascun. Això ens permet tenir un parell de paràgrafs d’informació per cada fragment, i passar aproximadament 5 fragments d’informació a GPT-4 com a context.
4. Passem tots els fragments al model “text-embedding-ada-002” d’OpenAI, mitjançant l’API, per transformar-los en vectors que podem emmagatzemar a la nostra base de dades de Weaviate.

Havent realitzat això, hem processat els documents d’entrada independentment del tipus de document i estan preparats per realitzar cerques semàntiques.

## 3.2.3. Realització de consultes als documents

Per fer consultes als nostres documents, primer haurem d’obtenir informació sobre els documents rellevants a allò que estem preguntant. Això es pot saber mitjançant la **cerca semàntica** de la qual hem parlat en l’apartat de l’estat d’art. Com tenim tots els fragments dels documents en la nostra base de dades, vectoritzats, el que necessitarem és transformar la nostra pregunta en un vector que es pugui comparar amb els de la base de dades vectorial.

* Farem una sol·licitud a l’API d’OpenAI per obtenir el vector corresponent a la nostra pregunta. Com sabem que el nostre input ha de ser inferior a 8.000 tokens, no caldrà fraccionar la pregunta abans de passar-la al model d’embeddings de OpenAI.
* Utilitzem la funció *query* de la llibreria Python de Weaviate per obtenir els 5 vectors més propers al que hem obtingut. La manera en què es realitza aquesta comparació és amb la **similitud del cosinus**, que compara la projecció sobre l’eix del cosinus de l’angle que generen els dos vectors.
* Un cop hem obtingut els 5 vectors més propers, i els documents associats a aquests vectors, podem afegir aquests fragments al nostre prompt inicial, i afegir la pregunta que ha realitzat l’alumne per tal d’enviar-la al model de **ChatCompletion** d’OpenAI, que respon mitjançant el model de GPT que indiquem (en el nostre cas, com tenim accés a GPT-4, utilitzem aquest).

Seguint aquests passos, obtenim la resposta amb GPT-4. Per tal de tenir una conversació continua amb el xatbot, integrem cada resposta de l’usuari i del xatbot amb el prompt inicial. Hem de vigilar mai superar els 8.192 tokens d’entrada màxims de GPT-4, i per tant haurem de seleccionar les respostes més importants o les més recents per evitar superar el límit.

## 3.2.4. Sel·lecció de documents

Es tracta d’una interfície senzilla que permet visualitzar tots els documents que s’han enviat per l’usuari. La base de dades dels usuaris conté un identificador de tots els documents que han enviat, per poder-los identificar. A partir d’aquesta interfície, tenen tres opcions:

* **Visualitzar el document.**
* **Escollir el document per anàlisi**: Així, si volen obtenir resultats que només es refereixin als documents escollits, sabrem que aquest és un dels documents dels quals desitgem obtenir dades. Això es pot realitzar mitjançant la creació d’una segona taula per cada usuari que només tingui els vectors dels documents seleccionats. Si volen fer preguntes específiques, mirarà aquesta taula de vectors. Si volen fer preguntes generals, mirarà la taula de vectors amb tots els documents.
* **Eliminar el document**. En cas que hi hagi un document que no creuen que sigui rellevant, es pot eliminar des d’aquesta mateixa pàgina.

## 3.2.5. Visualització dels xats dels alumnes.

Des d’aquesta pàgina, es podrà veure un llistat amb tots els alumnes. De moment, per realitzar la prova de concepte, un professor pot veure tots els alumnes que hi ha registrats a l’eina, però la intenció és que hi hagi un sistema d’invitació a classes per tal d’assegurar que un alumne només estigui registrat a un professor específic. A partir d’aquesta pàgina, podran accedir als xats que han tingut els alumnes amb l’eina, presentats en el mateix format que el xat de la pantalla principal.

## 3.3 Resultats dels tests unitaris

Abans d’unir tots els mòduls que s’han desenvolupat per separat, era important fer decisions respecte als algoritmes d’intel·ligència artificial i les llibreries de Python utilitzades a l’hora de crear l’apliació. Teníem diverses opcions respecte als models a implementar, i per aquest motiu fer proves empíriques ha resultat crític en el procés de decisió.

Per **àudio**, a més de Whisper, també existeix **Dragon Naturally Speaking**. Aquesta eina tradueix text a temps real, i permet l’ús de comandes parlades per tal de modificar la transcripció. En ser un software propietari, amb una llicència que s’ha de pagar, i no té cap API amb la que poder fer crides i integrar el programa fàcilment en la nostra prova de concepte. Per aquests motius, no s’ha implementat.

Tot i així, s’han hagut de fer proves amb WhisperAI. Whisper s’ofereix de dues formes: com a **API** (SaaS) i com a programa de codi obert. En un principi, semblava superior la opció d’executar-ho en local, però els resultats de les proves han demostrat el contrari.

Taula 1. Resultats dels models de Whisper

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Idioma  correcte? | Tasa  d’errors | Temps de transcripció |
| Local - *tiny* | No | 32.46% | 2m 15s |
| Local - *base* | Sí | 22.24% | 4m 29s |
| Local - *small* | Sí | 9.4% | 12m 39s |
| Local - *med* | Sí | 5.22% | 35m 45s |
| Local - *large* | Sí | 3.89% | 1h 4m 28s |
| *API* | Sí | 3.25% | 7s |

El càlcul de la tasa d’errors es realitza amb un algorisme que compara la transcripció real amb l’obtinguda pel model, trobant el nombre d’insercions i esborraments que caldrien per tal que els dos texts fossin equivalents.

Com podem observar, el resultats que obtenim amb la API son molt millors que els que obtenim executant l’arxiu en local. No només utilitza la API el model large, sinó que el temps de resposta és molt major al que obtenim executant el model en local, a causa de les diferències en els recursos disponibles. El que ens podria aportar el model local, respecte a la API, és l’ús de timestamps per les transcripcions, cosa que la API no aporta.

També ha sigut important fer proves amb els models de GPT. Les respostes que retornen ambdós models són bastant diferents.

* El model de **GPT-3.5**, freqüentment, **ignorava la indicació de només utilitzar informació del context proporcionat**. Per exemple, quan només aportàvem apunts d’una classe de Gestió de Projectes, seguia responent a preguntes com “Com es fa una derivada?”, cosa que no hauria de fer. Per tant, GPT-3.5 era més propens a **al·lucinar dades**.
* El model de **GPT-4** segueix les indicacions especificades en el promp inicial. Tot i això, hi ha prompts que han sobreescrit el prompt inicial i han fet que fos possible que l’usuari intentés manipular al model per saltar-se les restriccions imposades pel prompt inicial. Això no era freqüent, i es pot evitar amb prompts més estrictes o una capa de filtratge posterior.

Per tant, a causa de la importància que té que el model no al·lucini les dades, es troba indispensable l’ús de GPT-4 com a model per l’aplicació. Aquest té un cost més elevat que GPT-3.5, però es pot justificar amb l’increment de la qualitat dels resultats.

## 3.4 Proves d’utilització

Un cop finalitzada la prova de concepte, es va utilitzar de forma personal en un grup reduït de persones per tal d’obtenir *feedback* respecte al seu ús. Jo, com a desenvolupador, també la vaig utilitzar per així poder saber quins aspectes es podrien millorar.

* Generalment, la gent que ha provat l’aplicació creu que podria ser útil pels estudis d’una persona.
* Tot i això, creuen que la interfície web s’hauria de netejar i hauria d’aportar més informació respecte als documents. Ara mateix, l’input de documents ocupa massa.
* Les respostes del model tenen sentit.
* Caldria més funcionalitats, especialment dirigides a la gestió de deures i de projectes per poder tenir-ho tot en un sol lloc.
* Creuen que seria interessant si l’eina pogués proposar activitats per tal que l’alumne pugui practicar amb el coneixement.
* La gent creu que la part de gestió de documents és massa confusa.

Idealment, aniria bé realitzar unes proves a major escala un cop s’hagi millorat la eina per a poder tenir una idea major de la seva utilitat en un cas real.

# 4 Conclusió

Amb aquest treball, hem pogut realitzar un estudi sobre el paper que tenen les TIC en l’educació, i hem analitzat les tecnologies d’intel·ligència artificial de NLP actuals per observar un possible cas d’ús que podrien tenir en el sector educatiu. També, amb els annexos, hem vist la importància d’analitzar de forma crítica la intel·ligència artificial, ja que presenta dilemes morals que no podem ignorar si en volem fer un ús responsable. Tot i que trobo que les eines com GPT-4, Whisper i altres eines han aparegut per quedar-se, i que es poden aplicar a tota mena de casos d’ús gràcies a la gran quantitat de camps d’estudi que poden cobrir, no per això hem d’oblidar que els sistemes de IA no són perfectes.

La IA és útil només quan s’utilitza de forma responsable, i tota aplicació que busqui implementar-la per solucionar problemes hauria de preguntar-se per què es vol implementar i quins riscos hi poden haver a l’hora d’implementar-la. Sense aquest anàlisi, ens arrisquem a cometre errors i a la ignorància. Només cal veure els casos que hem exposat en els annexos, en que l’ús de la intel·ligència artificial va perjudicar als que la utilitzaven en comptes d’ajudar-los.

Per concloure aquest treball, vull fer ènfasis en aquest fet: el desenvolupament d’eines d’intel·ligència artificial i l’evolució tecnològica sempre han de tenir una fase d’estudi que explori els possibles casos d’ús que se’ls hi pot donar, i a actuar amb precaució. Només en aquest cas, quan tingui un objectiu clar i definit, serà útil la IA.

En el context d’aquest treball, els nostres objectius eren clars: hem vist la importància de la tecnologia en el sector educatiu, i hem utilitzat la nostra investigació de l’estat d’art per a fer una Prova de Concepte que serveix com a base per una aplicació que, si es desenvolupa més, podria ajudar tant a alumnes com a professors.

# 5 Plans futurs de desenvolupament

Gràcies a l’estructura modular de l’aplicació, es poden afegir noves funcions de forma senzilla. Aquest treball ha mostrat com podem fer una eina que no només realitzi resums de documents que un alumne passi a l’eina, i que permeti fer preguntes sobre el seu contingut per tal de fer que estudiar sigui una cosa més **interactiva,** sinó que també ajuda al professorat a entendre com interactuen els seus alumnes amb els materials d’estudi. A partir d’aquí, trobo que hi ha una sèrie de funcions que es podrien afegir per tal de millorar l’experiència dels alumnes.

* Una funcionalitat de **planificació de tasques** a l’estil de Trello, per tal de poder dividir projectes en tasques més petites. Podem utilitzar l’eina d’intel·ligència artificial per a ajudar als alumnes a saber com podria dividir les tasques. Aquesta és una de les funcions que la gent que ha provat l’eina ha demanat més, ja que ajudaria molt a la gent que té dificultats amb això.
* Integració de **veu** per permetre que l’usuari pugui fer preguntes al xat.
* Afegir **grups de classes** on es puguin fer anuncis o penjar documents. Això permet que hi hagi una major interacció entre professorat i alumnat, de forma similar a **Google Classroom**.
* Afegir interacció amb un OCR com Amazon Textract per tractar documents escrits a mà.

Aquesta eina es pot expandir de moltes formes, i també es podria considerar afegir l’opció d’enviar retroalimentació respecte a l’eina per tal de poder millorar aquesta eina en un futur.

**Agraïments**

Vull agrair al meu tutor per ajudar-me durant la realització d’aquest treball, i al professorat de la Autònoma per donar-me l’oportunitat de fer-lo. També vull agrair al meu professor de Filosofia, Marc Oriol Crespi, per despertar el meu interès en l’escriptura i l’ética. Agraeixo a la meva cosina, Mireia Martínez Gili, per despertar el meu amor per la literatura i per la traducció i localització. I també agraeixo a la meva família per ajudar-me en tot moment, i per permetre’m arribar a on he arribat.

**Bibliografia**

1. C. Belloch, “Las Tecnologías de la Información y Comunicación (T.I.C.)”. Universitat de València. [Online]. Available: <https://www.uv.es/~bellochc/pdf/pwtic1.pdf>
2. “¿Cuáles son las herramientas TIC en la educación?” IFP. [Online]. Available: <https://www.ifp.es/blog/las-herramientas-tic-en-la-educacion>
3. J. Hutchins, “The History of Machine Translation in a Nutshell”. [Online]. Available: <https://aclanthology.org/www.mt-archive.info/10/Hutchins-2014.pdf>
4. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gozem, L. Kaiser, I. Polosukhin, “Attention Is All You Need”. [Online]. Available: <https://doi.org/10.48550/arXiv.1706.03762>
5. OpenAI, “GPT-4 Technical Report.” [Online]. Available: <https://cdn.openai.com/papers/gpt-4.pdf>
6. E. F. Buchner, "Review of The Psychology and Pedagogy of Reading, with a Review of the History of Reading and Writing and of Methods, Texts, and Hygiene in Reading [Review of the book The psychology and pedagogy of reading, with a review of the history of reading and writing and of methods, texts, and hygiene in reading, by E. B. Huey]," Psychological Bulletin, vol. 6, no. 4, pp. 147-150, Apr. 1909. doi: 10.1037/h0066540.
7. Y. Hirn, “The Origins of Art: A psychological & sociological inquiry”, Macmillan & Co, New York.
8. G. A. Plimpton, "The Hornbook and Its Use in America," Proceedings of the American Antiquarian Society, vol. 26, pp. 264-272, Oct. 1916.
9. E. Shepard, "The magic lantern slide in entertainment and education, 1860-1920," History of Photography, vol. 11, no. 2, pp. 91-108, 1987, doi: 10.1080/03087298.1987.10443777.
10. “The Evolution of Educational Innovation,” University of Massachusetts Global. [Online]. Available: <https://www.umassglobal.edu/news-and-events/blog/the-evolution-of-educational-innovation>.
11. “Italy Bans ChatGPT over Data Privacy Concerns”, Cookie Script. [Online]. Available: <https://cookie-script.com/blog/italy-bans-chatgpt>
12. E. Wyatt, “Writers Sue Google, Accusing It of Copyright Violation”, The New York Times. [Online]. Available: <https://www.nytimes.com/2005/09/21/technology/writers-sue-google-accusing-it-of-copyright-violation.html>
13. N. Anderson “Court flunks high schoolers’ appeal on plagiarism data base”, Ars Technica. [Online]. Available: <https://arstechnica.com/tech-policy/2009/04/headline-here/>
14. “Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market – Outcome of proceedings”, Unió Europea. [Online]. Available: <https://data.consilium.europa.eu/doc/document/ST-6637-2019-INIT/en/pdf>
15. N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramèr, B. Balle, D. Ippolito, E. Wallace. “Extracting Training Data from Diffusion Models”. [Online]. Available: <https://arxiv.org/pdf/2301.13188.pdf>
16. LegalEagle, “How to Use ChatGPT to Ruin Your Legal Career”. [Online]. Available: <https://www.youtube.com/watch?v=oqSYljRYDEM>
17. M. O. Crespí Jiménez, “De la impossibilitat d’una etologia digital”, Filosofía, ara [Online]. Available: <http://www.filoara.cat/revista/index.php/FA/article/view/316>

**APÈNDIX**

**A1. La influència de la tecnologia en la història de la educació**

La integració de la tecnologia en l’àmbit educatiu es tracta d’una activitat que molts consideren com a recent. Tot i que és cert que les **Tecnologies de la Informació i la Comunicació** (TIC) són un fenomen més recent, no són les úniques que han revolucionat com ensenyem en la història. És important conèixer com han influenciat les invencions de l’ésser humà en l’educació per **entendre tant els canvis que ha experimentat el sector** com el **motiu pel qual es duen a terme aquests canvis**. Tot canvi està necessàriament impulsat per les motivacions d’uns individus, i aquests motius són els que haurem de conèixer per saber què és el que es busca a l’hora d’integrar la tecnologia en l’educació.

Buchner, en el seu llibre *Review of the Pyschology and Pedagogy of Reading, with a Review of the History of Reading and Writing and of Methods, Texts and Hygiene in Reading* [6] exposa que **l’escriptura sempre ha aparegut amb les diferents civilitzacions de la humanitat**, amb texts que ara tenen més de sis mil anys d’antiguitat. Més vital encara és la importància que li dona al llenguatge no verbal. Citant a Hirn, i el seu *Origins of Art* [7], assenyala que els signes del llenguatge escrit, originalment, es tractaven d’una **representació dels gestos que utilitzava l’home** primitiu per **comunicar-se**; reben el nom d’ideogrames per les seves característiques visuals, i tenen una gran influència en els alfabets actuals (especialment en idiomes asiàtics).

Aquest és el primer exemple de com les invencions de la humanitat han influenciat l’educació. L’escriptura i la seva popularització van permetre que es **transmetés informació de forma més senzilla**, **detallada** i **sense la natura efímera del llenguatge parlat**.

Un cop apareix l’alfabet i la informació escrita es torna més popular, apareix una necessitat d’ensenyar a l’alumne com llegir per tal que pugui accedir a tot el coneixement escrit. George A. Plimpton descriu en l’article *The Hornbook and its use in America* [8] l’educació en l’Edat Mitjana i en l’Edat Moderna. Es comença a popularitzar una forma d’educació que es basa en primer ensenyar l’alfabet, després la gramàtica, després la resta d’estudis sobre el llenguatge natural i després la filosofia. S’ensenya l’alfabet de forma memorística per tal que l’alumne pugui després interpretar tots els texts que se li presenten posteriorment. S’utilitzen **llibres, tauletes** i altres **invencions** per tal d’ajudar a l’alumne aconseguir aquest objectiu. Així, no només hem vist com una invenció pot ajudar als alumnes a aprendre, sinó que **també permet que apareguin noves formes d’ensenyament**.

Si avancem més en el temps, trobarem invencions com les **llanternes màgiques** del segle XVII. Elizabeth Shepard explica en *The màgic lantern slide in entertainment and education, 1860-1920* [9] com aquesta invenció que inicialment servia només per entretindre a la gent es va convertir en una eina que permetia presentar la informació de forma **visual** i **entretinguda**. Es tracta de l’exemple d’un dels primers projectors que s’han fet en la història de la humanitat, permetent que l’alumne pugui no només **llegir i escoltar** el que se li vol ensenyar, sinó que també es busca que ho pugui **observar**. La llanterna màgica s’il·lustra en la Figura 8.

A picture containing indoor, black and white

Description automatically generated with medium confidence

Figura 8. Exemple d’una llanterna màgica.

Al voltant del segle XIX i el segle XX, es pot observar un augment en la invenció de noves tecnologies i mètodes. Una universitat d’Illinois, als Estats Units, va començar els primers programes **d’estudi a distància**. Sobre els anys 20, van començar a aparèixer **programes de ràdio** que tenien la intenció d’ensenyar als alumnes mitjançant la transmissió de classes en viu. El segle XX també va experimentar l’aparició de la calculadora, de la fotocopiadora, i de la computadora moderna i de la **World Wide Web** (www). [10] Aquesta última ha sigut una de les invencions més importants, ja que permet que els alumnes puguin accedir a una gran quantitat d’informació disponible de forma pública.

Aquest context històric ens ha permès observar com d’unides estan la tecnologia i l’educació. Molts dels canvis que han experimentat les formes en què ensenyem venen impulsats per noves tecnologies que permeten mètodes que abans eren impossibles. És crític tenir en compte aquesta relació per adonar-nos de què la relació entre la tecnologia i l’educació s’hauria de veure de forma positiva, i que la tecnologia s’hauria d’explotar per tal de millorar les formes en què ensenyem amb la creació de nous mètodes d’ensenyament. L’escriptura va permetre que la informació es pogués **arxivar** i presentar de forma **detallada i organitzada**, els llibres i les tauletes van promoure un model **d’educació incremental** basat en aprendre a llegir per accedir a la resta del coneixement, les llanternes màgiques van fer aparèixer les primeres formes de projectar imatges per tal d’**afegir elements visuals a les lliçons**, i les noves tecnologies estan **posant a disposició** **de tothom** informació que abans no era fàcilment accessible.

En altres paraules, les grans revolucions de l’educació tenen una tendència a presentar la informació de forma més immersiva i a fer-la més accessible a tothom, i la tecnologia és l’eina clau per aconseguir això.

**A2. Ramificacions ètiques de les eines d’intel·ligència artificial**

Tot i que la utilitat de GPT i altres models de llenguatge massius sigui indiscutible, i el fet que són una eina que es continuarà desenvolupant per millorar la forma en què treballem sigui inevitable, també ha fet aparèixer una sèrie de preguntes ètiques que tenim la responsabilitat de respondre si pretenem utilitzar-la. En aquest apartat, exposaré les preocupacions que es poden observar sobre **l’entrenament** i sobre **l’ús de les eines de IA.**

L’ús de dades massives per entrenar models d’intel·ligència artificial sempre ha sigut un tema delicat. Molts cops, la informació que s’utilitza per entrenar models d’intel·ligència artificial conté dades amb drets d’autor. Això ha posat en alerta a governs com els d’Itàlia, que va bloquejar l’accés a ChatGPT durant el mes d’abril a causa de preocupacions sobre l’ús de dades dels seus usuaris [11]. Tal com menciona un article de Cookie Script, es va observar que ChatGPT incomplia regulacions estipulades al RGPD: no es controlava **l’edat** dels usuaris de ChatGPT, pot aportar informació **privada** de la gent a causa de les dades que apareixen a l’entrenament, i no **informava** els usuaris del fet que les seves dades estaven sent utilitzades per entrenar el model.

Aquest és només un entre molts altres casos en què la intel·ligència artificial ha sigut criticada pel públic a causa de com s’han entrenat. S’hi han produït crítiques similars amb IA de generació d’imatges (ja que molt de l’art que s’ha utilitzat per entrenar-les estava protegit pels drets d’autor), i amb IA de generació de música, entre d’altres. A aquestes crítiques, la resposta és una de les següents:

* Hi ha gent que **nega que s’estan infringint els drets d’autor** de les obres originals, ja que estan disponibles a Internet de forma oberta. Com els algoritmes d’intel·ligència artificial només utilitzen les dades per aprendre, aquest grup de gent argumenta que aquest aprenentatge és equivalent al que realitzaria un ésser humà. Per tant, si un ésser humà pot accedir a les dades i aprendre d’elles, també s’haurien de poder utilitzar per entrenar models d’IA.
* Hi ha gent que **accepta que s’estan infringint els drets d’autor**, però que consideren que l’ús està justificat a causa de la utilitat que tenen els models d’IA. Si ens aporten valor i milloren la nostra qualitat de vida, llavors la felicitat que ens aporten justifica l’ús no autoritzat de les dades que es poden trobar a Internet.
* Per últim, hi ha gent que accepta que s’estan infringint els drets d’autor i **condemna l’entrenament de la intel·ligència artificial** amb aquestes dades. Utilitzar el treball d’altres persones suposa perjudicar a l’autor original, que no rep cap compensació per haver produït material que després s’ha utilitzat sense el seu consentiment per entrenar models de IA. En molts d’aquests casos, també s’apel·la al fet que la IA es pot utilitzar, posteriorment, per substituir-los. En el cas de la IA de generació d’imatges, en molts casos, es pot utilitzar el nom d’un artista com a paràmetre per imitar el seu estil, per exemple.

Com podem veure, totes tres respostes tracten els següents dos temes: si s’infringeixen els **drets d’autor** de les obres originals o no, i si **l’ús de la intel·ligència artificial aporta suficient valor per justificar l’ús de les dades utilitzades en l’entrenament**.

Gràcies a casos com el de *Authors Guild, Inc. v. Google Books, Inc.* [12] i el de TurnItIn [13], podem veure que existeix una jurisprudència en els Estats Units que defensa els algoritmes d’intel·ligència artificial en els quals l’ús sigui **transformatiu** i no **substitueixi l’obra original**. En el cas de Google Books, per exemple, s’utilitzava escanejos de llibres per crear un motor de cerca; i en el cas de TurnItIn, s’utilitzava assajos d’estudiants per crear un detector de plagi. Aquesta jurisprudència, que defensa que aquestes eines són *fair use*, és similar a les excepcions respecte als drets d’autor en la Unió Europea. En el cas de la Unió Europea [14], es permet l’extracció massiva de dades de forma il·limitada si es realitza per organitzacions **sense ànims de lucre**, mentre que les organitzacions que busquin monetitzar aquesta transformació de les dades hauran de respectar les decisions dels autors individuals respecte si volen o no que els seus treballs s’utilitzin per mineria de dades (sistema de *opt-in* u *opt-out*).

Per tant, sabem que, tot i que existeixen restriccions respecte a l’extracció de dades per entrenar models d’intel·ligència artificial, està generalment permès si el treball és **transformatiu, no impacta negativament a l’autor original** i, en el cas de la Unió Europea, **es respecta el criteri dels autors** en cas que no desitgin que les seves obres s’utilitzin. Això ens porta a la segona preocupació: **l’ús de les eines de IA.** No només existeixen casos en què l’ús d’aquestes eines pot impactar negativament als autors originals, ja que existeix la possibilitat de substitució o d’una reproducció inalterada de l’obra original, sinó que també cal considerar la possibilitat que **la intel·ligència artificial al·lucini dades**.

Un article de Google, *Extracting Training Data from Diffusion Models* [15], explora la possibilitat de generar imatges utilitzades en l’entrenament amb diferents models de generació d’imatges com Stable Diffusion, DALL-E i Midjourney. L’article mostra que existeix la possibilitat que això passi, i explora unes possibles solucions que es podrien aplicar durant la fase d’entrenament per tal d’evitar aquestes reproduccions (evitar els exemples duplicats, per exemple). També exposa la reflexió que els **millors models** són més propensos a retornar aquestes reproduccions, ja que tenen millors **capacitats memorístiques**. Això, però, desperta una paradoxa: si els millors models són més propensos a les reproduccions inalterades de les dades d’entrenament, no significa això que són **menys propensos a transformar** i, per tant, suposen un risc a la privacitat? Si continuem millorant els models de forma irresponsable, suposarà un perill pels drets d’autor i per la privacitat. Per aquest motiu, proposo que s’hauria de desenvolupar **sistemes de filtratge automàtics** que **netegin les dades d’entrenament per desduplicarles** i, a més a més, que comprovin que el **resultat generat sigui prou original** respecte a les dades d’entrenament. Això pot resultar costós, però existeixen mètriques (com les exposades en l’article anteriorment mencionat) per realitzar aquestes exploracions.

Per últim, respecte a l’al·lucinació de les dades a la que són propenses eines com GPT, cal destacar el cas de Lodoca i Schwartz [16]. Aquests dos advocats van fer ús de ChatGPT per generar el cas que van presentar per defensar als seus clients, sense revisar el contingut que va ser generat per la intel·ligència artificial. Aquest cas que es va presentar estava ple de citacions a decisions jurídiques de casos que no existeixen, i va causar una sèrie de demandes per part dels jutges. Es va demanar que Lodoca i Schwartz entreguessin els casos que havien referenciat, i també els van generar amb ChatGPT. Un cop es va exposar el que havien fet Lodoca i Schwartz, van explicar que desconeixien que l’eina podia al·lucinar dades, i argumentaven que l’havien estat utilitzant com si fos un **motor de cerca** sense comprovar si les dades que obtenien o no eren certes. Aquest és només un dels molts exemples de casos en què es fa un ús irresponsable de les eines d’intel·ligència artificial a causa d’una manca d’informació. GPT genera text a partir de prediccions, i no pot comprovar si la informació que retorna és certa o no.

Aquest ús irresponsable de les eines d’intel·ligència artificial es produeix a causa d’un fenomen anomenat **metàfora tecnològica**. Marc Oriol Crespí Jimenez, en el seu article *De la impossibilitat d’una etologia digital* [17], exposa com, sovint, atribuïm qualitats inexistents a la intel·ligència artificial i a la tecnologia per la forma en què imiten la realitat. Això és evident amb eines com ChatGPT, que en imitar la forma d’escriure humana provoquen una confusió en els usuaris: els usuaris sense coneixements sobre la intel·ligència artificial confien massa en l’eina i assumeixen que no es pot equivocar, o fins i tot assumeixen que és millor que nosaltres redactant. Això és perillós perquè provoca que l’usuari es descuidi a l’hora d’utilitzar-la, o que acabi depenent de l’eina en comptes de simplement utilitzar-la per millorar la seva feina del dia a dia. La intel·ligència artificial es pot equivocar, pot al·lucinar i pot retornar resultats imperfectes. El nostre deure, com a usuaris, hauria de ser el de revisar els resultats que obtenim per assegurar-nos que no estem fent servir la tecnologia de forma irresponsable, i que no ens limitem a fer només el que l’eina pot fer.

Hem vist que la intel·ligència artificial té mancances, i que s’ha de tenir en consideració les diferents qüestions ètiques que desperta per poder utilitzar-la responsablement. Per la banda dels **desenvolupadors**, hem de tenir clars els **objectius de les eines** que desenvolupem per assegurar que són útils i **transformatives**, i hem d’implementar mesures per assegurar que **els resultats generats no siguin simples reproduccions** de les dades d’entrenament. Per la banda dels usuaris, hem de ser **crítics** amb les dades que obtenim quan utilitzem aquestes eines i **no confiar-hi cegament**. La intel·ligència artificial només serà útil quan aquests criteris es compleixin, i ignorar qualsevol d’aquestes indicacions només ens perjudicarà.

**A3. Planificació**

|  |  |
| --- | --- |
| **Setmana** | **Tasques** |
| **1** | * Reunió inicial, concretar el treball que volem desenvolupar. |
| **2** | * Fer una recerca molt superficial sobre el tema del TFG, per tal de saber quins aspectes serien els més interessants per desenvolupar. |
| **3** | * Fer una llista dels objectius i de les tasques a realitzar durant el desenvolupament del projecte en si. |
| **4** | * **Realitzar una reunió amb el tutor i lliurar l’informe inicial amb la proposta inicial del TFG, amb les tasques i els objectius ben definits.** |
| **5** | * Investigar l’impacte que la tecnologia ha tingut en l’àmbit de l’educació. Observar les tecnologies existents, plataformes online… |
| **6** | * Investigació dels mètodes d’aprenentatge alternatiu, de les diferents filosofies respecte a l’aprenentatge. Fer una comparativa entre aquestes. |
| **7** | * Investigació de les tecnologies que s'utilitzen durant el projecte per observar les seves capacitats, els casos que se’ls hi dona actualment, etc. |
| **8** | * Estudi de les ramificacions ètiques de la intel·ligència artificial, observar les qüestions que apareixen amb les AGIs, els NLPs, les tecnologies com GPT, etc. |
| **9** | * **Elaboració de l’informe de progrés 1.** |
| **10** | * Continuar amb la investigació centrada en el sector educatiu per identificar problemàtiques. * Buscar la forma en que les eines poden solucionar aquestes problemàtiques, i planejar la metodologia a seguir pel projecte. |
| **11** | * Prototipatge d’una eina que combini diferents tecnologies per intentar sol·lucionar aquestes problemàtiques. Aquest prototip es presenta com a un concepte, no com a una implementació feta. * Tot i així, provar les diferents tecnologies de forma separada (API de GPT, ús de Whisper en local, etc.) per estudiar la viabilitat de la eina. |
| **12** | * Fer un anàlisi de cost sobre el prototipatje pensat per la eina. * Analitzar els resultats obtinguts amb la eina inicialment per tal de així poder acabar de concretar idees per la eina que estem pensant. |
| **13** | * Fer un intent de connexió de les parts individuals per continuar desenvolupant la eina i satisfer les necessitats observades. * Continuar investigant per fer més reflexions sobre l’impacte de la tecnologia en l’àmbit educatiu. |
| **14** | * **Realització de l’informe de progrés II.** * En cas d’estar preparada, fer proves a escala petita (amb gravacions, converses casuals, etc.) de la eina. |
| **15** | * Continuar fent proves i documentant els resultats obtinguts durant aquestes. |
| **16** | * Intentar fer una prova en un cas real, com bé ara seria una classe. |
| **17** | * Finalització de l’informe del TFG, per presentar-lo. * Acabar el prototip de la eina per poder-lo presentar. |
| **18** | * Començar a preparar la presentació. |
| **19** | * Finalitzar la presentació. |
| **20** | * Creació del póster del TFG. |