



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DISEÑO Y DESARROLLO DE UN CHATBOT PARA FACILITAR LA BÚSQUEDA Y ANÁLISIS DE INFORMACIÓN RELEVANTE PARA DIRECTORES EMPRESARIALES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL
INDUSTRIAL

CLAUDIO FELIPE MONTIEL GUZMAN

PROFESOR GUÍA:
FELIPE VILDOSO CASTILLO

MIEMBROS DE LA COMISIÓN:
JOSÉ NALDA REYES
BLAS DUARTE ALLEUY

SANTIAGO DE CHILE
2024

Diseño y desarrollo de un chatbot para facilitar la búsqueda y análisis de información relevante para directores empresariales

El presente informe aborda el diseño y desarrollo de un chatbot basado en Inteligencia Artificial Generativa (GenAI), orientado a facilitar el acceso y análisis de información para directores empresariales.

El objetivo general del proyecto es generar conocimiento interno para la empresa Brain Food con respecto a GenAI. Esto se lleva a cabo por medio de la implementación de un prototipo de chatbot que utilice GenAI para procesar y sintetizar información de documentos empresariales, permitiendo a los directores acceder rápidamente a datos relevantes y análisis precisos.

La metodología utilizada se inspira en la metodología CRISP-DM, la cual ha sido adaptada a las necesidades específicas del proyecto. Esta conlleva el levantamiento de necesidades, diseño de la plataforma, prototipado rápido, varias iteraciones de desarrollo/pruebas y elaboración de una documentación de todo el trabajo realizado, junto con los aprendizajes, limitaciones y desafíos encontrados.

Las tecnologías clave incluyen el uso de modelos de lenguaje grande (LLM), específicamente GPT-3.5 Turbo y GPT-4 para tareas avanzadas como la interpretación de tablas. También se emplean técnicas de embeddings y bases vectoriales para mejorar la recuperación y procesamiento de datos a través del método Retrieval Augmented Generation (RAG).

El proyecto valida la viabilidad y los beneficios de aplicar GenAI en el contexto de la alta dirección empresarial, destacando el potencial de la inteligencia artificial para facilitar la gestión de una gran cantidad de información de manera ágil y confiable. Los resultados muestran la capacidad del chatbot para proporcionar análisis rápidos y precisos de textos principalmente planos y estructurados.

Ante documentos no estructurados, es decir, con texto en varias columnas, presencia de tablas, imágenes, gráficos y esquemas, el chatbot presenta serias dificultades para analizar e interpretar correctamente la información. Esto último lleva a respuestas poco precisas, incorrectas o completamente inventadas.

La experiencia obtenida proporciona una base sólida para futuros desarrollos que solucionen diversas problemáticas que están enfrentando los actuales y potenciales clientes de Brain Food. Esto propicia, además, la consideración de formar un área dedicada exclusivamente a GenAI.

Agradecimientos

En primer lugar, quiero agradecer profundamente a mi familia, a mis padres por darme un apoyo incondicional a lo largo de mi vida, en todos los aspectos. Sin ellos no hubiera sido posible. Todo el esfuerzo que han hecho conmigo tanto emocionalmente como económicamente es sin dudas algo que jamás podré pagar completamente y estoy orgulloso de tenerlos como padres. Por otro lado, a mi hermano por ser un a compañía crucial durante toda mi vida, por brindarme momentos de juegos, risas, consejos y cariño.

A mis amigos, mis *panas* del colegio. Constantemente soy un agradecido de tenerlos en vida desde séptimo básico hasta el día de hoy, esperando que durante muchos años más. El hecho de que aún cuando cada uno podamos estar con sus propios asuntos y no tener contacto seguido, cuando nos juntemos todo siga igual, es algo maravilloso y que nos caracteriza. *Mi panita y gymbro Robert (un macizo), mi compa y personalidad única Maier alias el Malibú, mi pana Shiki el que sigue siempre, my g Melanita; mi bestie y gymsis (le falta disciplina si a veces), mi socia Tamalitra (la maravillosa), mi lokita Alejandra (jugo de naranja con YELO), mi amiga internacional Michellelindísima y a mi rusia Javierita que está en otra etapa de su vida con la Trini.*

A las personas que conocí en la universidad y a los cuales, pese a que no nos vemos tan seguido, puedo llamarlos amigos y les tengo mucho cariño y admiración. *Carlitro y Catita*, muchas gracias por todos los momentos compartidos.

Finalmente, a mis profesores guía y co-guía por su acompañamiento durante el desarrollode este trabajo, gracias por su apoyo y correcciones.

Tabla de Contenido

<i>Introducción</i>	1
<i>Capítulo 1: Antecedentes generales</i>	2
1.1. Contextualización de la industria	2
1.2.1. Información de la empresa	2
1.2.2. Productos o servicios y procesos	3
1.2.3. Organigrama.....	4
<i>Capítulo 2: Descripción del problema u oportunidad</i>	5
2.1. Oportunidad para la empresa	5
2.2. Beneficios potenciales	6
<i>Capítulo 3: Descripción y justificación del proyecto</i>	8
3.1. Descripción del proyecto	8
3.2. Justificación del proyecto.....	9
<i>Capítulo 4: Objetivos</i>	10
4.1. Objetivo General.....	10
4.2. Objetivos Específicos	10
<i>Capítulo 5: Alcances</i>	11
<i>Capítulo 6: Marco conceptual</i>	12
6.1. Inteligencia Artificial.....	12
6.2. Inteligencia Artificial Generativa	13
6.3. Large Language Models	14
6.4. Retrieved Augmented Generation	16
6.5. Componentes del método RAG	16
<i>Capítulo 7: Metodología</i>	18
<i>Capítulo 8: Desarrollo y Resultados</i>	21
8.1. Desarrollo	21
8.2. Resultados	23
<i>Capítulo 9: Discusiones</i>	28
<i>Capítulo 10: Conclusiones</i>	30
<i>Bibliografía</i>	32
<i>Anexos</i>	34

Índice de Ilustraciones

Ilustración 1: Organigrama Brain Food (2024)	4
Ilustración 2: Clasificación de la Inteligencia Artificial Generativa (GenAI)	13
Ilustración 3: Modelos fundacionales	14
Ilustración 4: Flujo de pregunta-respuesta con un LLM.....	15
Ilustración 5: Flujo del método RAG	17
Ilustración 6: Ejemplo de uno de los blogs públicos.....	24
Ilustración 7: Extracto de la Memoria Anual de Falabella 2022.....	25
Ilustración 8: Fragmento de un Estado de Resultados	25
Ilustración 9: Mockup de la interfaz.....	26

Introducción

La inteligencia artificial generativa (GenAI) ha experimentado un crecimiento exponencial en los últimos años, gracias a avances significativos en el poder computacional, la disponibilidad de grandes volúmenes de datos y la sofisticación de algoritmos de aprendizaje profundo. Este progreso ha permitido desarrollar modelos de lenguaje grandes (LLMs) como GPT-3.5 y GPT-4, capaces de generar texto coherente, simular conversaciones humanas y analizar información compleja. En el contexto empresarial, estas capacidades abren un vasto horizonte de oportunidades para optimizar procesos, mejorar la toma de decisiones y personalizar la interacción con los clientes [1].

La adopción de la inteligencia artificial generativa (GenAI) en productos y procesos empresariales ofrece un potencial significativo para mejorar la productividad y la eficiencia a gran escala. A través de GenAI es posible automatizar tareas repetitivas y complejas, lo que permite a las empresas liberar recursos humanos para enfocarse en actividades estratégicas de mayor valor añadido. Además, esta tecnología facilita la personalización de productos y servicios, proporcionando experiencias más alineadas con las necesidades y preferencias de los clientes. La capacidad de GenAI para procesar grandes volúmenes de datos y generar insights precisos y accionables permite a las empresas tomar decisiones más informadas y basadas en datos, mejorando así la competitividad en un mercado cada vez más digital y dinámico. En conjunto, estas capacidades transforman a GenAI en un motor clave para impulsar la innovación y el crecimiento económico en diversos sectores [2].

Este trabajo de título se lleva a cabo bajo la modalidad de práctica extendida en la empresa chilena Brain Food, una consultora de analítica avanzada y transformación digital. Se centra en el diseño y desarrollo de un piloto de producto, un chatbot basado en GenAI, utilizando la metodología de Retrieval-Augmented Generation (RAG) para responder preguntas sobre documentos empresariales. Esta aproximación permite combinar las fortalezas de los LLMs en la generación de lenguaje natural con la precisión de la recuperación de información de documentos propios del usuario, proporcionando respuestas actualizadas y específicas.

El desarrollo de este prototipo no solo demuestra la viabilidad de aplicar GenAI en el ámbito de la alta dirección empresarial, sino que también resalta el potencial de esta tecnología para facilitar y agilizar la forma en que los directores empresariales gestionan y acceden a grandes volúmenes de información. Este proyecto genera un conocimiento valioso para Brain Food en la materia, con el objetivo de utilizarlo en posibles soluciones posteriores a problemáticas que presenten sus clientes.

Capítulo 1: Antecedentes generales

1.1. Contextualización de la industria

La industria de la consultoría es un campo amplio y dinámico que abarca una variedad de servicios especializados dirigidos a ayudar a las organizaciones a mejorar su desempeño, eficiencia y capacidad de innovación. Se caracteriza por su enfoque en ofrecer asesoramiento experto y soluciones personalizadas a problemas complejos en múltiples sectores, incluyendo, pero no limitado a, estrategia corporativa, gestión, operaciones, tecnología de la información, recursos humanos y finanzas [3].

El tamaño del mercado de servicios de consultoría se estima en 323,88 mil millones de dólares en 2024, y se espera que alcance los 431,89 mil millones de dólares en 2029, creciendo a una tasa compuesta anual del 4,96% durante el período previsto (2024-2029) [4]. En el contexto nacional, el ambiente competitivo está concentrado por las consultoras: Deloitte, EY, KPMG, PwC, Boston Consulting Group, McKinsey y Bain & Co. Las empresas mencionadas anteriormente se centran en la dirección estratégica, a diferencia de Brain Food, que se distingue por su enfoque en la transformación digital y la analítica avanzada [5].

La industria de la consultoría está experimentando una transformación significativa impulsada por la digitalización, la inteligencia artificial y la creciente demanda de soluciones innovadoras. La pandemia de COVID-19 ha acelerado la adopción de tecnologías digitales y ha obligado a las empresas a adaptarse rápidamente a nuevas formas de trabajo, lo que ha generado una mayor necesidad de servicios de consultoría especializados [6]. Además, la inteligencia artificial y el aprendizaje automático están revolucionando la forma en que las consultoras abordan los desafíos empresariales, permitiendo un análisis más profundo de los datos y la generación de conocimientos accionables [7].

1.2. Caracterización de la empresa

1.2.1. Información de la empresa

Brain Food es una consultora chilena especializada en Transformación Digital, Analítica Avanzada, Ingeniería de Datos, Inteligencia de Negocios (BI¹), Automatización y Software Factory. Está basada en Santiago de Chile y provee servicios a más de 16 países alrededor del mundo. Desarrolla e implementa soluciones y proyectos end-to-end, colaborando, y asegurando una implementación y adopción exitosa.

¹ Business Intelligence.

Brain Food fue fundada el año 2015 por Jacques y David Albagli, quienes son padre e hijo correspondientemente y se constituye como una Sociedad por Acciones (Spa). En el año 2019 y 2023 se integran nuevos accionistas: Andrés Oksenberg y Diego Halffter respectivamente, ambos como socios de la empresa. Actualmente, la empresa tiene un equipo de más de 50 profesionales y ha ejecutado proyectos en una amplia gama de sectores, colaborando con más de 40 clientes de distintas índoles en más de 15 industrias. Algunos de sus clientes son y han sido: Socovesa, Walmart, SMU, Mall Plaza, Visa, Cencosud, CCU, DuocUC, VTR, entre muchos otros.

1.2.2. Productos o servicios y procesos

Brain Food ofrece cuatro áreas principales de servicios, permitiendo que cada proyecto se enfoque principalmente en una de éstas, a la vez que se beneficia de la colaboración interdisciplinaria de profesionales con distintos perfiles dentro de la empresa. Las áreas y sus respectivas funciones se describen a continuación:

- **Ciencia de datos & IA²:** Proporciona a los clientes la capacidad de tomar decisiones más informadas y reducir riesgos, mediante el análisis de datos proporcionados por el cliente. Esto incluye la creación de modelos predictivos y prescriptivos que facilitan la anticipación de eventos futuros y la identificación de patrones significativos en los datos, utilizando técnicas como Machine Learning, Data Mining y Deep Learning.
- **Transformación Digital:** Se enfoca en alinear estratégicamente la visión del cliente con la tecnología que emplea, a través del desarrollo de metodologías y análisis de inteligencia de negocios. Esto ayuda a los clientes a comprender su posición competitiva en la industria.
- **Automatización & BI:** Dirigida a mejorar la eficiencia operativa y gestión de los clientes mediante el desarrollo de herramientas tecnológicas y matemáticas, como dashboards y cubos OLAP, que buscan minimizar errores humanos y acortar los tiempos de ejecución de tareas en cada proyecto.
- **Ingeniería de datos:** Se encarga de evaluar, preparar, limpiar, consolidar y hacer accesibles los datos del cliente, asegurando que puedan usarlos de manera eficiente y segura. Sus tareas incluyen; desarrollo, testeo y mantenimiento de infraestructuras cloud, desarrollo y automatización de procesos de ingesta, integración y consolidación de datos, y finalmente, diagnóstico de gobernabilidad y madurez de los datos.

² Inteligencia artificial.

1.2.3. Organigrama

En la actualidad, la compañía se distingue por no contar con departamentos de trabajo definidos. En su lugar, para cada proyecto se forman equipos multidisciplinarios integrados por profesionales jóvenes de diversas especialidades, incluyendo Data Science, Data Engineering, Consultoría y BI. Este enfoque tiene como objetivo ofrecer respuestas creativas y efectivas a los retos y problemas que enfrentan las diversas organizaciones. El organigrama actual de la empresa se presenta a continuación.

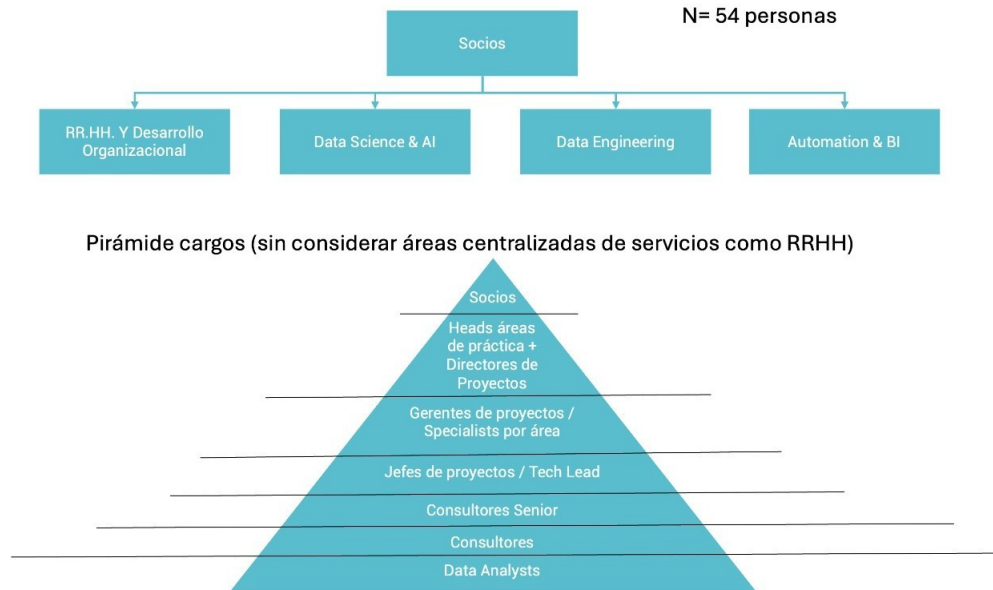


Ilustración 1: Organigrama Brain Food (2024).

Fuente: Departamento RR.HH. Brain Food

Capítulo 2: Descripción del problema u oportunidad

2.1. Oportunidad para la empresa

La inteligencia artificial generativa³ (GenAI) está transformando el panorama tecnológico, ofreciendo nuevas y emocionantes oportunidades para consultoras especializadas como Brain Food. Este auge representa una ventana única para diseñar y crear productos innovadores que puedan ofrecer un valor sin precedentes a los clientes. La capacidad de estas tecnologías para generar contenido, automatizar procesos, mejorar la toma de decisiones y personalizar experiencias abre un abanico de posibilidades para solucionar problemas complejos, optimizar operaciones y descubrir insights ocultos en grandes volúmenes de datos [8].

En este contexto, Brain Food pueden aprovechar la inteligencia artificial generativa para desarrollar soluciones a medida que se alineen con las necesidades específicas de sus clientes. En particular, en sus áreas de Data Science y Business Intelligence, la empresa puede generar lo siguiente:

- Automatización de Reportes (Business Intelligence): Generación automática de informes personalizados para clientes, reduciendo el tiempo de preparación y permitiendo a los analistas concentrarse en tareas de mayor valor agregado.
- Desarrollo de Soluciones Predictivas (Data Science): Uso de LLMs para el desarrollo de modelos predictivos en sectores como ventas, finanzas y operaciones, mejorando la capacidad predictiva y la toma de decisiones estratégicas.
- Interacción Mejorada con el Cliente (Data Science): Implementación de chatbots avanzados que pueden manejar consultas complejas, ofreciendo respuestas más naturales y precisas, lo que mejora la experiencia del cliente.

Considerando lo anterior, el proyecto se constituye como una oportunidad en Brain Food para sumergirse y profundizar en el mundo de la GenAI, en concreto, para experimentar y aprender sobre cómo y cuándo implementar una solución basada en GenAI a problemas de potenciales clientes. Específicamente, por medio de la implementación de un chatbot⁴ que permita responder consultas sobre documentos propios de un potencial tipo de cliente: directores empresariales.

³ La IA generativa se define como una avanzada tecnología de IA capaz de crear contenido nuevo y original, desde texto hasta imágenes y música, aprendiendo de grandes cantidades de datos existentes.

⁴ Un chatbot es un software automatizado que simula conversaciones humanas para asistir a los usuarios en diversas tareas o consultas.

Los directores de empresas, entre sus variados roles, necesitan supervisar y evaluar el desempeño de la empresa, manteniéndose al día con distintas fuentes de información interna y externa. La capacidad de procesar y sistematizar esta información de manera más eficiente permite liberar tiempo para tareas más estratégicas, como analizar la gestión de forma crítica, colaborar en la toma de mejores decisiones, y evaluar y gestionar de mejor forma los riesgos empresariales para proteger los intereses de los accionistas. El desarrollo actual de la GenAI por medio de chatbots basados en LLM's⁵ permite facilitar el análisis de documentos para lograr este objetivo.

Una situación que sucede con frecuencia entre los directores es que pueden recibir los documentos que necesitan para prepararse ante una determinada reunión con muy poca anticipación (casi cercano al inicio de ésta) [9]. En el caso anterior, si se estima el tiempo de preparación de los directores leyendo los diversos textos para extraer la información relevante, en una hora, por medio de la interacción con un chatbot el proceso puede reducirse a unos pocos minutos.

El proyecto se desenvuelve en el área de Data Science de Brain Food, bajo la responsabilidad de la Gerente de Proyectos de tal área. El beneficio de éste corresponde a generar el conocimiento interno para el desarrollo de aplicaciones basadas en GenAI. Esto último, considerando la identificación de problemas que pueden ser solucionados con esta tecnología, su diseño/desarrollo y sus limitaciones/desafíos.

2.2. Beneficios potenciales

Impacto Directo:

- **Identificación de problemáticas:** La realización del proyecto permite a la empresa identificar qué situaciones pueden ser abordadas exitosamente con una solución basada en GenAI, y cuáles no.
- **Generación de conocimiento interno:** Brain Food puede obtener conocimientos teóricos y prácticos sobre una solución basada en GenAI.
- **Diversificación de clientes:** La empresa puede identificar nuevos tipos de clientes que busquen implementar tecnología de GenAI para resolver sus desafíos específicos.

⁵ Un LLM (Large Language Model) es un modelo avanzado de procesamiento del lenguaje natural que utiliza grandes cantidades de datos textuales para generar respuestas coherentes y contextuales, empleando técnicas de aprendizaje profundo.

Externalidades:

- Gestión del cambio: Creación de nuevos roles y competencias laborales centrados en la supervisión, mantenimiento y mejora de sistemas de GenAI, así como en la interpretación de sus outputs.
- Desarrollo de nuevos modelos de negocio: Habilitación de modelos de negocio que no eran posibles antes de la aparición de la IA generativa, como productos personalizados a gran escala.

Capítulo 3: Descripción y justificación del proyecto

3.1. Descripción del proyecto

El proyecto consiste en desarrollar un piloto de producto basado en GenAI que permita a directores de empresas consultar documentos de forma ágil, ordenada y simple, desde una única herramienta. En concreto, una plataforma que disponga de un asistente inteligente con la capacidad de extraer información relevante de documentos, relacionar y realizar inferencias sobre estos. El proyecto se basa en dos niveles de mejora, desde un menor a un mayor nivel de sofisticación:

1. Diseñar POC (Proof of concept)

Contar con una plataforma suficiente para procesar texto plano en documentos con fomarto PDF e interactuar de forma inicial con ella mediante un chatbot.

2. Ampliar MVP (Minimum viable product)

Plataforma con funcionalidades adicionales según prioridad de las necesidades del cliente objetivo. En particular, manejo de tablas en los documentos y realización de reportes en tiempo real.

Lo anteriormente descrito se consolida como el núcleo del proyecto, sin embargo es crucial mencionar que al ser un piloto de producto, pueden surgir variaciones y modificaciones durante el transcurso de este. Tales cambios podrían ocasionarse en función de las necesidades y requerimientos que se analicen y comprendan desde potenciales clientes, así como la factibilidad de los mismos.

La realización del proyecto es liderada por la Gerente de Proyectos del área de Data Science, la cual responde a los socios de la empresa. Por otro lado, diversos empleados de la organización con distintos cargos y expertices, brindan ayuda y asesoramiento en la ejecución del proyecto.

Al finalizar el proyecto, Brain Food obtendrá el conocimiento necesario sobre los casos de uso adecuados, limitaciones y desafíos que conlleva un chatbot para responder consultas sobre documentos empresariales complejos. Este conocimiento se utilizará en proyectos posteriores que soliciten o requieran implementar una solución basada en este tipo de inteligencia artificial generativa.

3.2. Justificación del proyecto

En primer lugar, Brain Food realizó exitosamente en el pasado un proyecto de chatbot para uno de sus clientes: la ACHS (Asociación Chilena de Seguridad). La ACHS identificó un potencial uso de GenAI en sus operaciones diarias, con la cual podrían ayudar a acelerar procesos y liberar horas persona en rutinas diarias. El proyecto consistió en una aplicación web sobre un chatbot que respondiera consultas de los usuarios sobre preguntas frecuentes de recursos humanos.

El trabajo anterior fue un éxito tanto para la ACHS como para Brain Food, puesto que observaron los beneficios tangibles de la inteligencia artificial generativa, no sólo en términos de eficiencia y productividad, sino también en la satisfacción de los empleados, que disfrutaban de una mayor autonomía y de ver a la organización adoptando tecnología de punta, situándose en una posición de vanguardia. Esta experiencia con la ACHS permitió generar un knowhow⁶ inicial sobre el desarrollo de chatbots como alternativas de solución para problemas de extracción y análisis de información en documentos.

Por otro lado, Brain Food aprendió que la inteligencia artificial no es solamente una herramienta tecnológica, sino un facilitador de cambio y crecimiento [10]. Esto último propició que continuaran explorando sus posibilidades. En particular, nuevos usos hacia nuevos clientes. De esta forma, el proyecto en cuestión responde a una oportunidad de negocio que Brain Food pretende aprovechar.

A través de las redes y el entorno laboral de uno de los socios, se identificó una oportunidad única para introducir los servicios de consultoría a directores empresariales que, hasta el momento, no habían colaborado con Brain Food. Dado que los directores de empresas enfrentan desafíos únicos en la gestión de información compleja y toma de decisiones rápida, un chatbot personalizado representa una solución innovadora que puede mejorar significativamente su eficiencia operativa y capacidad de respuesta estratégica. Es importante aclarar que, mediante este Trabajo de Título, se pretende profundizar en el desarrollo e implementación de chatbots mediante el uso de técnicas avanzadas. Lo anterior en comparación con el trabajo realizado para la ACHS.

De esta forma, el proyecto no solo refleja la continua adaptación y respuesta a las necesidades emergentes del mercado por parte de Brain Food, sino que también le permite acceder a una nueva demografía de clientes potenciales. Además, subraya el compromiso con la innovación y la excelencia en el servicio, posicionando a la consultora como socios estratégicos indispensables en el panorama empresarial actual.

⁶ El know-how es el conjunto de conocimientos, habilidades y experiencia práctica necesarios para realizar una tarea específica de manera eficiente. Es fundamental para la ventaja competitiva y la innovación en las organizaciones.

Capítulo 4: Objetivos

4.1. Objetivo General

El objetivo general de este trabajo de título se define como:

Generar conocimiento interno sobre GenAI para Brain Food a través del diseño e implementación de un piloto de una aplicación web basada en un chatbot para directores empresariales, el cual sea capaz de colaborar con el procesamiento rápido de la información contenida en distintos reportes e informes empresariales, investigar los documentos y relacionar ideas entre ellos.

4.2. Objetivos Específicos

- Levantar el estado del arte de la inteligencia artificial generativa para generación de texto.
- Identificar y analizar las necesidades de los potenciales tipos de clientes: los directores empresariales.
- Implementar un prototipo de solución de chatbot que combine el back-end⁷ y el front end⁸.
- Documentar los aspectos técnicos de la implementación del prototipo, junto con los aprendizajes, desafíos y limitaciones de éste.

⁷ El "back end" se refiere a la parte de la aplicación que opera detrás de las escenas, gestionando la lógica, la base de datos y el servidor, esencial para el procesamiento y la gestión de datos.

⁸ El "front end" se refiere a la parte de una aplicación web que los usuarios interactúan directamente, incluyendo la interfaz visual y los elementos con los que interactúa el usuario.

Capítulo 5: Alcances

Los alcances para el proyecto se limitan a tres aspectos fundamentales:

Alcance Tecnológico

Inclusión: Se empleará el método *Retrieval Augmented Generation* alias RAG (el cual se explicará en la sección posterior) para el desarrollo del chatbot, explorando diferentes configuraciones de sus componentes.

Exclusión: No se realizará un *fine tuning*⁹ del *Large Language Model* (conceptos que se van a explicar con más detalle en la sección posterior)

Justificación: Llevar a cabo un fine tuning requiere una cantidad significativa de recursos computacionales, además de conocimientos técnicos especializados. Mientras que el método RAG es intuitivo, simple y eficaz.

Alcance Comercial y Productivo

Inclusión: El proyecto se concibe como una prueba de concepto con el objetivo de explorar las capacidades y limitaciones del método RAG (retrieved augmented generation), el cual será explicado posteriormente.

Exclusión: No se buscará desarrollar un producto comercial ni se realizarán esfuerzos de marketing o comercialización.

Justificación: La naturaleza académica y experimental del proyecto busca generar conocimiento sobre la inteligencia artificial generativa, sin la presión de cumplir con criterios de mercado o rentabilidad comercial.

Alcance de la Plataforma

Inclusión: El chatbot se desarrollará para funcionar en un entorno de servidor local con interacción a través de una interfaz web básica.

Exclusión: No se desarrollarán aplicaciones móviles ni integraciones con plataformas de terceros como redes sociales o sistemas operativos específicos.

Justificación: Mantener el desarrollo en un entorno controlado y simple permite concentrar los recursos en la mejora y evaluación del prototipo en lugar de en la compatibilidad con múltiples plataformas, adecuado para una prueba de concepto.

⁹ El fine tuning es un proceso de ajuste específico de un modelo de aprendizaje automático pre entrenado para optimizarlo en una tarea particular, mejorando su rendimiento al ajustar sus parámetros a partir de un conjunto de datos más focalizado.

Capítulo 6: Marco conceptual

En primer lugar, la disciplina relacionada con la Ingeniería Industrial en el desarrollo de este proyecto corresponde a la Ciencia de datos. Esta combina matemáticas y estadística, programación especializada, análisis avanzado, inteligencia artificial (IA) y aprendizaje automático con conocimientos específicos de la materia para descubrir información útil oculta en los datos de una organización [11]. A continuación, se profundiza en cada una de las sub áreas específicas que el proyecto aborda.

6.1. Inteligencia Artificial

La Inteligencia Artificial (IA) es una rama de las ciencias de la computación que se enfoca en el desarrollo de sistemas y algoritmos capaces de realizar tareas que normalmente requieren de la inteligencia humana. Estas tareas incluyen el aprendizaje, el razonamiento, la resolución de problemas, la percepción y la comprensión del lenguaje natural [12]. El objetivo principal de la IA es crear máquinas que puedan llevar a cabo funciones cognitivas similares a las de los seres humanos, mejorando así la eficiencia y la precisión en una amplia gama de aplicaciones.

La IA se divide en dos categorías principales: la IA débil y la IA fuerte. La IA débil, también conocida como IA estrecha, se refiere a sistemas diseñados para realizar tareas específicas, como el reconocimiento de voz, la detección de fraudes o la clasificación de imágenes. Por otro lado, la IA fuerte, o IA general, apunta a crear sistemas que puedan igualar o superar la inteligencia humana en una amplia variedad de tareas cognitivas [13].

En las últimas décadas, la IA ha experimentado un rápido desarrollo gracias a los avances en el poder computacional, la disponibilidad de grandes conjuntos de datos y la mejora de los algoritmos de aprendizaje automático (Machine Learning). Estos avances han permitido la creación de sistemas cada vez más sofisticados, capaces de aprender y adaptarse a nuevas situaciones sin ser explícitamente programados.

Uno de los campos más prometedores de la IA es el Aprendizaje Profundo (Deep Learning), que utiliza redes neuronales artificiales con múltiples capas para aprender representaciones abstractas de los datos [14]. Esta técnica ha demostrado ser altamente efectiva en tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la traducción automática, entre otras.

6.2. Inteligencia Artificial Generativa

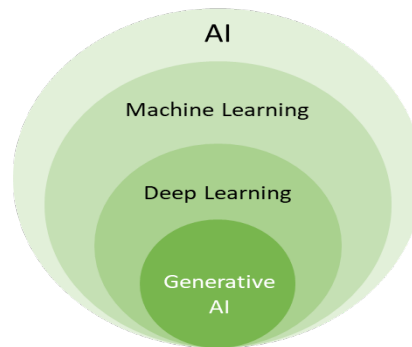


Ilustración 2: Clasificación de la Inteligencia Artificial Generativa (GenAI)

Fuente: [15]

La IA generativa es un subcampo de la IA que se centra en la creación de nuevos contenidos, como imágenes, música, texto o diseños, a partir de modelos entrenados en conjuntos de datos existentes. Estos modelos son capaces de aprender las características y patrones presentes en los datos de entrenamiento y utilizarlos para generar nuevos ejemplos que sean similares, pero no idénticos, a los originales [16].

Esta sub especialidad utiliza arquitecturas de redes neuronales profundas, como las Redes Neuronales Convolucionales (CNN), las Redes Neuronales Recurrentes (RNN) y los Transformers, para aprender las características y patrones subyacentes de los datos de entrenamiento y generar nuevos ejemplos convincentes.

La relación entre el Deep Learning y la IA generativa radica en que esta última aprovecha las capacidades de aprendizaje y representación de las redes neuronales profundas para crear modelos generativos. El Deep Learning proporciona las herramientas y técnicas necesarias para entrenar estos modelos con grandes cantidades de datos, permitiendo que la IA generativa aprenda a generar nuevos ejemplos de manera autónoma.

Las principales aplicaciones de la IA generativa son [17]:

- **Texto:** Dentro de este apartado se encuentran aquellas aplicaciones destinadas a la generación de textos creativos, resúmenes automáticos de textos extensos, corrección de contenido y chatbots que pueden mantener conversaciones con los usuarios.
- **Imágenes:** Aplicaciones para la generación y edición automática de imágenes realistas e, incluso, generación de obras de arte originales de cualquier estilo.
- **Código:** Estas herramientas de IA generativa se utilizan para agilizar el proceso de desarrollo de software, por medio de la generación automática de nuevo código.

- **Audio:** Para la creación de composiciones musicales y mejora de la calidad del audio en las grabaciones.
- **Vídeo:** Este tipo de herramientas de IA generativa están destinadas a la generación de vídeos sintéticos, utilizando algoritmos y técnicas de renderizado en computadora.
- **Diseño:** Aplicaciones enfocadas en el diseño de productos, tanto para la fase de generación de ideas, como para la optimización del diseño y su personalización.

6.3. Large Language Models

Los grandes modelos de lenguaje (o LLM's en inglés) son un tipo específico de modelo de Inteligencia Artificial Generativa diseñados para procesar, entender y generar texto en lenguaje natural. Estos modelos se entrenan en grandes corpus de texto y aprenden a predecir la siguiente palabra o token¹⁰ en una secuencia, basándose en el contexto de las palabras anteriores.

Los LLM's se han entrenado con billones de palabras a lo largo de muchas semanas y meses, junto con una potencia de cálculo masiva. Cuentan con billones de parámetros que presentan propiedades que van mas allá del lenguaje, los cuales pueden interpretarse como el *cerebro* del modelo. Cuantos más parámetros tenga un modelo, tendrá un *cerebro* más desarrollado y por lo tanto, más sofisticadas serán las tareas que puede realizar [18]. A continuación, se muestran los denominados *modelos fundacionales*, los primeros LLM's que se crearon y sus tamaños relativos en término de sus parámetros.

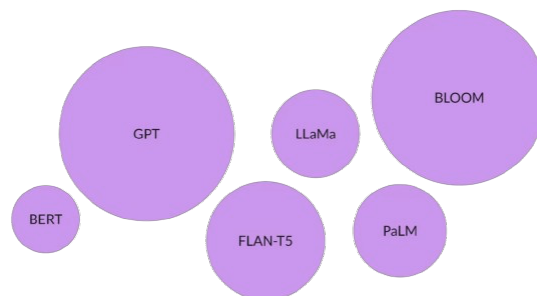


Ilustración 3: Modelos fundacionales

Fuente: [19]

¹⁰ Un token es una unidad básica de texto que se utiliza para representar y procesar el lenguaje de manera eficiente. Los tokens pueden ser frases, palabras, subpalabras, caracteres o incluso símbolos especiales, dependiendo del enfoque utilizado para tokenizar el texto.

Los LLM's son capaces de tomar instrucciones en lenguaje natural o escritas por humanos y realizar tareas de la misma manera que lo haría un humano. El texto que se le pasa a un LLM se conoce como *prompt* y el espacio o memoria disponible para ingresar el prompt se conoce como *ventana de contexto*. Esta última, por lo general es lo suficientemente grande; en el orden de unos miles de palabras, pero varía de un modelo a otro. Finalmente, el output del modelo se denomina *finalización*, y el acto de utilizar el modelo para generar texto se conoce como *inferencia*. La finalización o completion en inglés, esta compuesta del texto proveniente del prompt y la respuesta otorgada por el modelo.

Prompts and completions

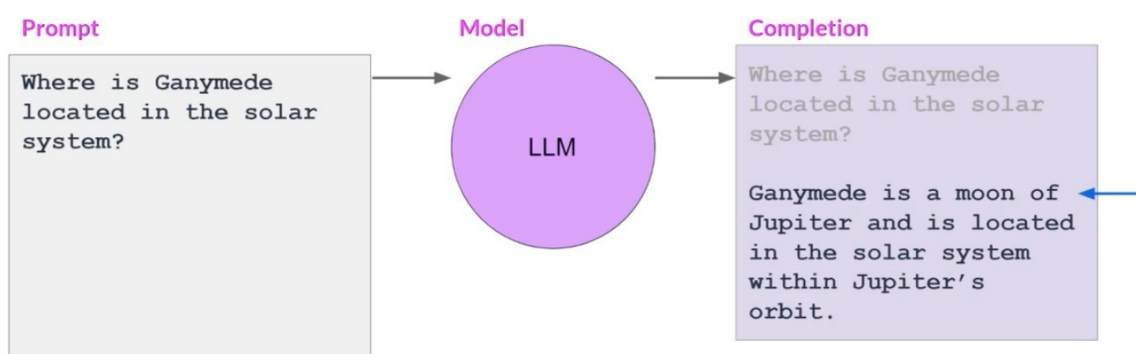


Ilustración 4: Flujo de pregunta-respuesta con un LLM

Fuente: [20]

A pesar de las ventajas de la inteligencia artificial generativa por medio de LLM's, existen riesgos significativos a considerar, en particular para su uso en el contexto empresarial [21]:

- **Alucinación:** Generar contenido inventado, lo que puede llevar a los usuarios a utilizar información falsa.
- **Shallowness:** Tener dificultades para completar tareas de alta complejidad (ej: fotos generadas de personas con manos de 6 dedos).
- **Sostenibilidad:** Consumo de energía y emisión de CO2 aumenta al usar mayores parámetros en los modelos de LLM's.
- **Regulación:** Se anticipan cambios relevantes a nivel regulatorio en el mundo debido al impacto de estas nuevas tecnologías (ej: EU AI Act).
- **Fallas:** Tener sesgos inadvertidos y producir resultados inexactos o insatisfactorios. Además, las decisiones podrían ser cuestionables si el sistema interpreta mal el contexto o las instrucciones.

- **Seguridad:** Hay riesgos de seguridad como pueden ser ataques cibernéticos, manipulación y robo de datos. Donde proteger la información sensible se vuelve cada vez más desafiante.

6.4. Retrieved Augmented Generation

Los LLM's han demostrado habilidades impresionantes en la generación y comprensión del lenguaje natural. Sin embargo, estos modelos dependen de su conocimiento adquirido durante el entrenamiento y por lo tanto, tienen dificultades para acceder y utilizar información específica o actualizada que no estaba presente en sus datos de entrenamiento. Esto puede llevar a respuestas incorrectas, inconsistentes o en el peor de los casos, inventadas.

Para abordar esta limitación, se ha propuesto el enfoque RAG (Retrieved Augmented Generation) [22]. RAG combina la capacidad de generación de lenguaje de los LLM's con la recuperación de información de una base de conocimientos externa. Esto permite al modelo acceder a información específica y actualizada durante el proceso de generación, mejorando la precisión y la coherencia de las respuestas generadas.

6.5. Componentes del método RAG

El método RAG consta de varios componentes principales que trabajan juntos para recuperar información relevante y generar respuestas coherentes. A continuación, se explica brevemente cada componente y su papel en el flujo de trabajo de RAG:

1. Procesamiento de documentos y chunking

- El primer paso es obtener los documentos o fuentes de información relevantes, como archivos PDF, que contienen el conocimiento necesario para responder a las consultas de los usuarios.
- Los documentos se dividen en chunks o segmentos más pequeños, lo que facilita su procesamiento y recuperación. Un chunk puede ser, por ejemplo, párrafos secuenciales.
- El chunking permite manejar documentos largos y extraer información específica y concisa para responder a las consultas.

2. Embeddings

- Los embeddings son representaciones vectoriales numéricas de palabras, frases o documentos en un espacio de alta dimensionalidad.
- Cada chunk de texto se convierte en un embedding utilizando modelos de lenguaje pre-entrenados, como BERT o RoBERTa.
- Los embeddings capturan la semántica y el contexto del texto, permitiendo comparar la similitud entre diferentes chunks y consultas del usuario.
- Al representar los chunks como embeddings, se pueden realizar búsquedas y comparaciones eficientes en el espacio vectorial.

3. Base vectorial

- Los embeddings de los chunks se almacenan en una base vectorial, que es una estructura de datos optimizada para la búsqueda y recuperación de vectores similares.
- Algunas bases vectoriales populares son Chroma, Qdrant, Weaviate, Pinecone, entre otras.
- La base vectorial permite una búsqueda rápida y eficiente de los chunks más relevantes para una consulta dada, basándose en la similitud de sus embeddings.

4. Retriever

- El retriever es el componente encargado de encontrar los chunks más relevantes en la base vectorial para una consulta dada.
- Cuando un usuario realiza una consulta, se genera un embedding de la consulta utilizando el mismo modelo de embedding que se aplicó a los chunks.
- El retriever busca en la base vectorial los chunks cuyos embeddings son más similares al embedding de la consulta, utilizando medidas de similitud entre vectores, como la similitud coseno.
- Los chunks más relevantes se recuperan y se ordenan según su puntuación de similitud.

5. Generación de respuestas con LLM

- Los chunks recuperados se incorporan en el prompt o contexto de un LLM como GPT-3, por ejemplo.
- El prompt se construye concatenando la consulta del usuario y los chunks recuperados, proporcionando al LLM la información necesaria para generar una respuesta coherente y precisa.
- El LLM genera la respuesta final, utilizando su capacidad de generación de lenguaje y el conocimiento adicional proporcionado por los chunks recuperados.
- La respuesta generada se presenta al usuario como la salida final del sistema RAG.

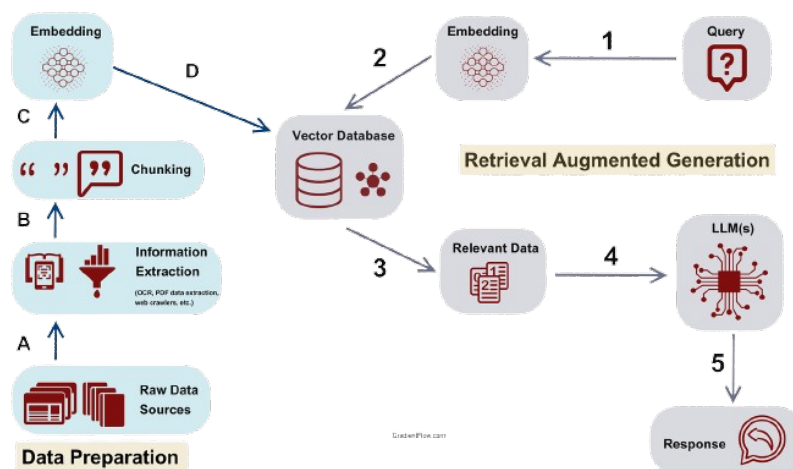


Ilustración 5: Flujo del método RAG

Fuente: [23]

Capítulo 7: Metodología

La metodología adoptada para el desarrollo del proyecto se inspira en la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining). Esta última es ampliamente utilizada en proyectos de ciencia de datos y minería de datos, la cual proporciona una estructura clara y detallada para llevar a cabo proyectos de este tipo. El ciclo de vida de CRISP-DM incluye seis fases: Comprensión del Negocio, Comprensión de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue [24].

La elección de esta metodología está fundamentada en su forma estructurada e iterativa que permite que cada etapa del proyecto se lleve a cabo de manera sistemática, permitiendo ajustes y mejoras continuas a medida que se avanza. Al seguir CRISP-DM, se pueden gestionar eficazmente los riesgos y asegurar que el chatbot desarrollado cumple con los requisitos y objetivos establecidos.

Por otro lado, esta presenta limitaciones como su falta de adaptación específica a proyectos modernos de inteligencia artificial generativa, lo que puede requerir ajustes adicionales [25]. Para el caso de este proyecto en particular, se adaptan los pasos de la metodología CRIPS-DM de la siguiente manera:

Etapas 1: Levantamiento de Necesidades

En esta primera etapa, se realiza un análisis detallado de las necesidades de los potenciales usuarios del chatbot. Se incluye la identificación de los tipos de preguntas que los usuarios podrían querer realizar, así como los documentos sobre los que el chatbot deberá extraer información.

Etapas 2: Levantamiento de Tecnologías Relevantes

Se investigan y evalúan las tecnologías existentes que pueden ser utilizadas para el desarrollo del chatbot. Esto incluye plataformas de desarrollo de chatbots, frameworks de inteligencia artificial, y específicamente tecnologías relacionadas con el método RAG.

Etapas 3: Diseño de la plataforma

En esta etapa, se diseña la arquitectura de la plataforma junto con sus funcionalidades. Esto incluye la definición de la interfaz de usuario, la manera en que se va a interactuar con el usuario, y cómo se integrará el método RAG para la generación de respuestas.

Etapas 4: Prototipado Rápido

Se desarrolla un prototipo inicial que implementa una versión simplificada del chatbot. Este prototipo sirve para validar las tecnologías seleccionadas y la arquitectura diseñada, permitiendo realizar ajustes rápidos basados en los resultados obtenidos.

Etapa 5: Iteraciones de Desarrollo y Pruebas

Siguiendo un ciclo iterativo de desarrollo, se refina el prototipo en dos fases, cada una centrada en mejorar la funcionalidad y la interacción del chatbot. En cada iteración, se realizan pruebas para validar las características del sistema, ajustando el diseño y la implementación según sea necesario. Esta etapa permite asegurar la calidad y la precisión del chatbot.

Etapa 6: Elaborar una documentación de todo el proceso

Para persistir el conocimiento generado a través de la realización del proyecto, se genera una documentación escrita de los aspectos técnicos del trabajo realizado, junto con los aprendizajes, desafíos y limitaciones, además de la documentación del código de programación.

Por último, con respecto a los recursos utilizados para el desarrollo del proyecto se utilizan los siguientes insumos y herramientas relevantes:

- Documentos empresariales (análisis financieros, estados financieros, artículos de retail, memorias anuales).
- Lenguaje de programación Python (librerías como LangChain, Unstructured y Streamlit), sistema de control de versiones Git, software de despliegue de aplicaciones Docker y aplicación de planificación Microsoft Planner.
- Bases vectoriales Chroma y Qdrant.

Por otro lado, se realizan cursos que permiten obtener el conocimiento práctico para llevar a cabo el desarrollo del prototipo planteado. Estos cursos son ofrecidos por la plataforma web *DeepLearning.AI*¹¹, la cual ofrece cursos gratuitos sobre diversos tópicos en Inteligencia Artificial Generativa. Los cursos considerados son:

- ***ChatGPT Prompt Engineering for Developers***: Cómo utilizar la API de la empresa OpenAI para aprovechar sus modelos de LLM, junto con mejores prácticas del *Prompt Engineering*¹².
- ***LangChain for LLM Application Development***: Se enseña como utilizar el framework LangChain para el uso de los LLM's en el desarrollo de aplicaciones.

¹¹ <https://www.deeplearning.ai>.

¹² El prompt engineering es el proceso de diseñar y optimizar los prompts o instrucciones que se proporcionan a los grandes modelos de lenguaje para obtener respuestas o resultados deseados. Implica técnicas como la selección cuidadosa de palabras, el ajuste de la longitud y el formato del prompt, y la incorporación de ejemplos y contexto relevantes.

- ***LangChain with Your Data:*** Se enseña como utilizar el framework LangChain para implementar el método RAG.
- ***Functions, Tools and Agents with LangChain:*** Se muestran nuevas funcionalidades del framework LangChain como herramienta de desarrollador con LLM's.
- ***Understanding and Applying Text Embeddings:*** Cómo funcionan y se utilizan los embeddings de texto para el desarrollo de aplicaciones con LLM's.
- ***Vector Databases: from Embeddings to Applications:*** Cómo funcionan las bases vectoriales y se utilizan para el desarrollo de aplicaciones con LLM's.
- ***Preprocessing Unstructured Data for LLM Applications:*** Cómo mejorar el sistema RAG al preprocesar datos no estructurados proveniente de distintos tipos de documentos.

Capítulo 8: Desarrollo y Resultados

8.1. Desarrollo

A continuación, se detalla el desarrollo para cada uno de los pasos enunciados en la metodología:

1. Levantamiento de necesidades

Se realizan dos reuniones con una potencial cliente, miembro del directorio de cinco empresas nacionales reconocidas, de diversas industrias como inmobiliaria, bancaria y del molibdeno¹³. El objetivo de aquellas reuniones es levantar las necesidades y preocupaciones del segmento de clientes objetivo, vale decir, los directores empresariales.

2. Levantamiento de tecnologías relevantes

Se lleva a cabo una sesión con el equipo de trabajo en el proyecto de la ACHS, para la discusión de las herramientas y tecnologías que emplearon en ese momento y cuales son aplicables a este proyecto en particular. Por otro lado, se investiga en internet sobre el estado del arte de los frameworks existentes para el desarrollo de aplicaciones basadas en chatbots mediante el método RAG. Además, se investiga sobre cursos y documentación que enseñen de forma práctica estos frameworks.

3. Diseño de la plataforma

Se confecciona la plataforma web utilizando Streamlit como interfaz de usuario (front-end) y LangChain (back-end) como orquestador de los componentes del método RAG. Por otro lado, se reúnen documentos públicos para poder realizar las preguntas de prueba con el chatbot y definen funcionalidades de la plataforma a través de pestañas para la interacción del usuario.

4. Prototipado Rápido

Se programa una primera versión del producto siguiendo los pasos del método RAG utilizando el LLM GPT-3.5 Turbo de OpenAI por medio de su API. Este modelo en particular debido a la familiaridad de conocimiento hacia OpenAI por parte de la empresa, su buen performance y por otorgar créditos gratis. El modelo de pricing de OpenAI por medio de su API consiste en cobrar por la cantidad de tokens utilizados tanto en su prompt como en su respuesta de salida en cada llamado a su API. Al utilizar por primera vez la API (y el modelo GPT-3.5 Turbo), OpenAI otorga USD \$5 gratuitos para realizar

¹³ El molibdeno es un elemento natural que se encuentra en la Tierra y se utiliza para el desarrollo de una gran variedad de aplicaciones en el transporte, construcción, generación de energía, agricultura e incluso medicina.

pruebas, los cuales resultaron suficientes para el desarrollo del chatbot ya que la estructura de precios es del orden¹⁴ de USD \$0.5 / 1 millón tokens por input y de USD \$2 / 1 millón tokens por output.

Con respecto a las etapas del sistema RAG, se llevan a cabo como:

- **Pre procesamiento de la data**

Al realizar RAG con documentos que presentan data no estructurada como los PDF's utilizados, los modelos LLM presentan dificultades para interpretarlos correctamente. Por lo tanto, ante un prompt del usuario, se obtiene una respuesta imprecisa, sesgada e insatisfactoria. Debido a esto, se efectúa un pre procesamiento de los documentos anteriores para su posterior utilización. Esto resulta en un archivo *.txt* con solamente el texto de cada uno de los PDF. Para más detalle, se explica lo realizado en el **Anexo A**.

- **Chunking**

Con los archivos de texto puro y plano obtenidos en el pre procesamiento, se crean chunks con una longitud fija de caracteres, debido a su eficiencia computacional y a su facilidad de uso. Se utiliza también un *overlap* entre los chunks, es decir, cada chunk comienza con la última porción del chunk anterior. Lo anterior para mantener el contenido semántico entre los fragmentos.

- **Embeddings**

Se utiliza el modelo de embedding text-embedding-3-large de OpenAI por ser un modelo más complejo que puede capturar un mayor conjunto de características semánticas en los embeddings. Este utiliza 3072 dimensiones en sus vectores.

- **Base vectorial**

Se utiliza la base vectorial **Chroma** para almacenar los embeddings, debido a su característica open source, su facilidad de uso para el desarrollo de prototipo de aplicaciones y a su integración con LangChain.

- **Retrieval**

Se implementa la estrategia **Maximal Marginal Relevance (MMR)**. Para su explicación e ilustración, ir al **Anexo B**.

5. Iteraciones de Desarrollo y Pruebas

Las dos fases de iteración y mejora se constituyen como:

- **Fase 1: Otorgar “memoria” al chatbot.**

¹⁴ <https://openai.com/api/pricing/>

Esta capacidad permite al chatbot recordar las preguntas y respuestas anteriores en la conversación. De esta forma, puede entender mejor el contexto y las necesidades del usuario, y responder de manera más coherente y relevante. Además, ayuda a construir una relación más natural y fluida, similar a la interacción humana. Esta fase se prueba con preguntas genéricas que permitan demostrar que el chatbot “recuerda” a través de sondeos. Por ejemplo, comentar al chatbot el nombre de un usuario y luego preguntar “¿Cómo me llamo?”.

- **Fase 2: Interpretación de tablas en los documentos**

Los documentos pertenecientes al día a día de un director empresarial son archivos altamente cargados en tablas de diversos formatos y estructuras. La lectura, y correcta interpretación de éstas por parte de un LLM no es trivial y es un proceso que requiere un pre procesamiento distinto y más sofisticado. Además, la interpretación de tablas es en sí una tarea que requiere de un modelo más inteligente. Por lo tanto, se opta por utilizar el modelo de vanguardia GPT-4 en vez de GPT-3.5 Turbo.

El testeo en esta fase se lleva a cabo por medio de preguntas a tablas presentes en un documento de Estados Financieros Consolidados a septiembre del 2023, perteneciente a una de las empresas de la cual es parte del directorio la potencial cliente con la cual se tuvo las reuniones para levantar necesidades.

6. Elaborar una documentación digital

Se procede a confeccionar la documentación escrita por medio de un Power Point para su persistencia en Brain Food. Respecto a los códigos de programación, estos se cargan al repositorio propio de la empresa en Bitbucket junto con su *readme*.

8.2. Resultados

De la misma forma que en el apartado anterior, se exponen los resultados para cada paso de la metodología:

1. Levantamiento de necesidades

Las necesidades rescatadas por medio de las dos entrevistas fueron:

- **Acceso a información y análisis confiables:** Los directores requieren datos precisos, actualizados y relevantes para respaldar sus decisiones. Un chatbot basado en inteligencia artificial generativa puede proporcionar respuestas rápidas y precisas a consultas específicas, facilitando el acceso a información crítica.
- **Mejora de la eficiencia y productividad:** Los directores buscan constantemente formas de optimizar los procesos y maximizar la eficiencia en sus organizaciones. Un chatbot puede automatizar tareas rutinarias, como la búsqueda y síntesis de información, liberando tiempo valioso para que los directores se enfoquen en actividades estratégicas.

- **Identificación de oportunidades y riesgos:** Los directores deben estar atentos a las tendencias del mercado, las oportunidades emergentes y los riesgos potenciales. Un chatbot puede analizar grandes volúmenes de datos y destacar insights relevantes, ayudando a los directores a identificar áreas de crecimiento y mitigar riesgos.

2. Levantamiento de tecnologías relevantes

Se identifica y se aprende a utilizar los frameworks LangChain y Streamlit así como la base vectorial Chroma. Por otro lado, se identifican los cursos gratuitos mencionados en la metodología y se completan.

3. Diseño de la plataforma

Se reúnen diversos documentos en formato PDF para poder realizar pruebas con el chatbot, en particular, pertenecientes a:

Blogs públicos sobre retail (artículos, noticias y reportes). Los cuales se caracterizaron por ser de extensión breve y de texto plano principalmente, es decir, documentos tales que su composición fuera de texto estructurado y sin tablas o imágenes.

Competitive Strategies for Retail

Long before you open a retail business, you will need to make some fundamental strategic decisions about where your store will fit in the current marketplace. In other words, you need to decide what you are trying to achieve with your business.

Start by asking yourself this critical question: "Why should a customer shop in my store?" The question is so deceptively simple that you might be surprised by your answer. Most likely your response will be broken down to the ways that you, as a business, are adding value to the products that manufacturers create.

To be successful, every retail business needs to add value to its products by adopting one of three basic strategies and offerings: the **greatest assortment**, the **lowest price** or the **best service**.

Although all of these strategies add value, each one meets the needs of different customers. The challenge you face is to be competitive in all three areas – and to clearly exceed your customers' expectations in one of them.

Assortment and Price

Ilustración 6: Ejemplo de uno de los blogs públicos

Fuente: <https://smallbusinessbc.ca/article/competitive-strategies-retail/>

Memorias Anuales de Falabella (desde el año 2018 hasta el 2022). Este tipo de archivos se escogió por su extensión y estructura altamente compleja, vale decir, documentos de más de 300 páginas con texto no estructurado y gran cantidad de imágenes, tablas y gráficos. Lo anterior con el fin de probar los alcances y limitaciones de la aplicación. El hecho de que fueran de la empresa Falabella se debió a una simple arbitrariedad.



Ilustración 7: Extracto de la Memoria Anual de Falabella 2022

Fuente: Memoria Anual de Falabella 2022

Informes financieros y presentaciones de resultados de una empresa inmobiliaria (documentos privados de la contraparte): Documentos altamente cargados en tablas con información financiera, con una extensión del orden de 50-80 páginas aproximadamente.

8.4 Resumen de la estratificación de la cartera de deudores comerciales y otras cuentas por cobrar corrientes, y derechos por cobrar no corrientes

El resumen de la estratificación de la cartera es el siguiente:

Tramo de morosidad al 30 de septiembre de 2023	Clientes de cartera N°	Total cartera bruta MS	Total cartera neta MS
Al día	208	17.259.019	17.259.019
Entre 1 y 30 días	5	17.687	17.687
Entre 61 y 90 días	2	5.152	5.152
Total	215	17.281.858	17.281.858

Tramo de morosidad al 31 de diciembre de 2022	Clientes de cartera N°	Total cartera bruta MS	Total cartera neta MS
Al día	158	13.940.841	13.940.841
Total	158	13.940.841	13.940.841

La totalidad de la cartera para ambos ejercicios no contiene clientes repactados. Los saldos incluidos en esta cartera consideran deudores comerciales y otras cuentas por cobrar y derechos por cobrar.

8.5 Calidad crediticia

La Sociedad no tiene riesgo de crédito significativo dado que el porcentaje mayor del precio de venta de las viviendas lo recibe de entidades financieras quienes financian a los compradores de viviendas y la diferencia del precio se cobra anticipadamente, previo a la entrega física del bien.

Ilustración 8: Fragmento de un Estado de Resultados

Fuente: Documento privado de la contraparte

Luego, se produce un mockup de la aplicación correspondiente a una página web con cuatro pestañas vacías, pensadas con las siguientes funcionalidades:

- **Empieza aquí:** Mensaje de presentación e introducción a la aplicación
- **Ficha documentos precargados:** Se generaron fichas de texto tipo resumen (autor del documento, fecha de creación, resumen breve, ideas más importantes y keywords) de cada documento que se procesó. El objetivo es poder darle al usuario una primera impresión de cada uno de los documentos precargados y que las dudas o conocimientos que quiera conocer en detalle, las realice en la siguiente sección.
- **Habla con los documentos precargados:** En este apartado es donde se encuentra la interfaz con el chatbot para que el usuario pueda conversar con este sobre cualquiera de los documentos que quiera conocer en profundidad.
- **Prueba con tu propio documento:** Por último, en esta funcionalidad, el usuario puede cargar su propio documento en formato PDF y preguntarle al chatbot lo que desee.

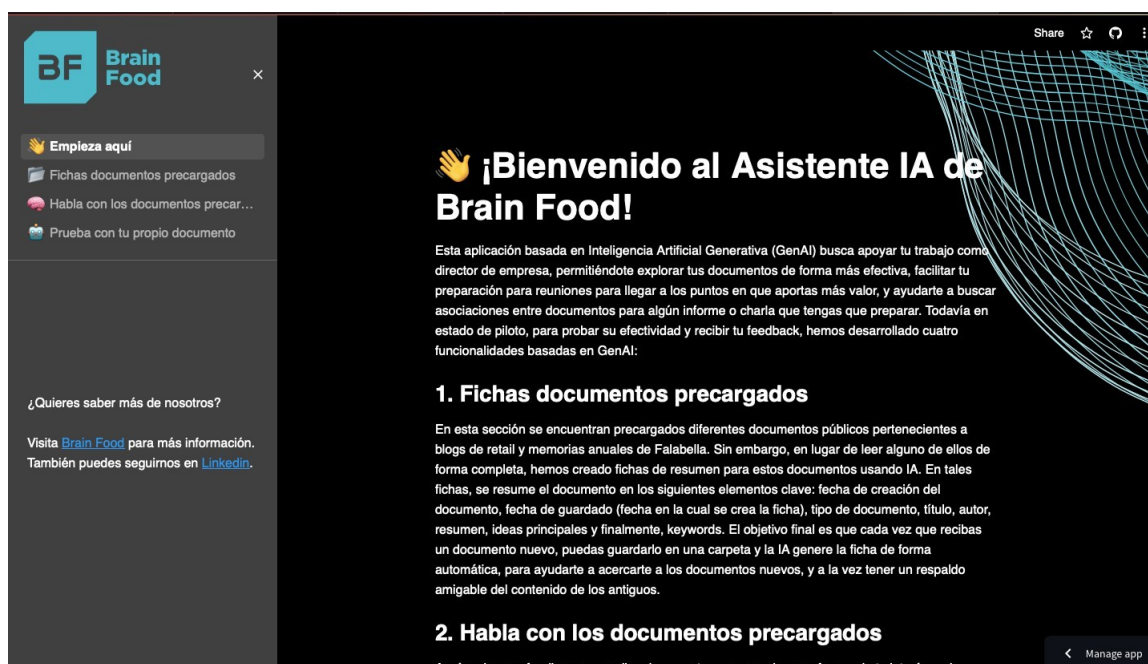


Ilustración 9: Mockup de la interfaz

4. Prototipado Rápido

Se completan las pestañas de la página web descritas anteriormente con las funcionalidades respectivas. Los resultados se muestran en el **Anexo C**.

5. Iteraciones de Desarrollo y Pruebas

Se implementa correctamente la memoria en el chatbot durante la fase 1, logrando recordar preguntas y respuestas de las conversaciones anteriores. Sus resultados se muestran en el **Anexo D**. Respecto a la fase 2, se presentan en el **Anexo E**, dos pares de pregunta-respuesta realizadas al chatbot junto con el resultado verdadero. Las respuestas del chatbot se presentan sin la interfaz de Streamlit.

En relación a los resultados de la fase 2, es crucial destacar que, si bien los resultados fueron mayormente exitosos en las preguntas que se realizaron sobre las tablas, el pre procesamiento de los documentos resultó sumamente ser lento y engorroso. El pre procesamiento se realizó de forma independiente con el flujo del método RAG. Esto implica que hacer una carga en tiempo real del documento para luego realizar preguntas no es factible dado el alto tiempo de espera que se requiere para el tratamiento del documento.

Capítulo 9: Discusiones

Respecto al trabajo realizado durante el proyecto, se presentan a continuación tres cuestionamientos en torno a las decisiones tomadas tanto por el estudiante como por la empresa.

1. Prioridad a documentos complejos (decisión de la empresa)

En los momentos de preguntas de prueba al chatbot, se focaliza en realizar preguntas sobre las Memorias Anuales de Falabella y en mejorar el prototipo para contestarlas correctamente. Este tipo de archivos, son documentos altamente no estructurados. Contienen texto en columnas, gráficos, imágenes, esquemas, tablas y elementos que en conjunto hacen que el proceso de extracción de información por parte del chatbot sea difícil, engorroso y con muchos errores. En concreto, cuando un LLM se confunde con la información que recopila, tiende a *alucinar*¹⁵. Se necesita de un procesamiento avanzado y extenso para poder lidiar con documentos de ese tipo.

Una alternativa a lo anterior es desarrollar las primeras pruebas con documentos más simples, en particular de texto plano y estructurado, como lo son los artículos de retail. Lo anterior permite, en primer lugar, simplificar la validación de las funcionalidades básicas del chatbot, asegurando que puede procesar y responder a consultas sin complicaciones adicionales que podrían introducirse con formatos más complejos como imágenes o tablas.

Por consiguiente, es posible identificar y corregir errores básicos de análisis y comprensión de los fragmentos de texto recuperados por la fase de *retrieval* del proceso RAG, antes de avanzar a desafíos mayores. Además, trabajar con texto plano reduce la carga computacional y los recursos necesarios, facilitando iteraciones más rápidas en las etapas iniciales de desarrollo.

2. Utilización de GPT-3.5 Turbo como LLM para potenciar el chatbot (decisión de la empresa y el estudiante).

Solo en la etapa final del proyecto (en la iteración para considerar el caso de tablas) se opta por utilizar un modelo más avanzado como GPT-4. Aunque GPT-3.5 Turbo ofrece un buen rendimiento con precios económicos, requiere un manejo más sofisticado de todo el proceso RAG para evitar respuestas erróneas o "alucinaciones".

¹⁵ Una "alucinación" en un LLM se refiere a la generación de información incorrecta o ficticia, aunque presentada con convicción. Estas alucinaciones pueden ocurrir debido a limitaciones en el entrenamiento del modelo y su interpretación de los datos.

Lo anterior representa un desafío significativo en términos de las capacidades técnicas requeridas. Es necesario realizar un pre procesamiento más sofisticado de los documentos para poder transformar los PDF's a un texto lo más plano y sencillo posible, para que el LLM logre entenderlo mejor y así responder correctamente. Este pre procesamiento implica utilizar librerías y modelos avanzados y costosos computacionalmente, lo que efectivamente retrasa la realización de iteraciones rápidas.

Como alternativa al uso de GPT-3.5 Turbo, se puede considerar la adopción de modelos más avanzados como GPT-4 u otros modelos de la industria como Gemini 1.5 Pro o Claude 3 Opus. Estos modelos ofrecen mejores capacidades de contextualización e interpretación de textos. Aunque pueden tener un costo inicial más alto, ofrecen beneficios significativos en términos de precisión y adaptabilidad a necesidades específicas del proyecto.

Además, su avanzada capacidad de manejo de lenguaje natural puede reducir los costos a largo plazo asociados con las correcciones de errores y la supervisión humana, así como mejorar la satisfacción del usuario final por la calidad de las interacciones.

3. Sesgo marcado hacia utilizar inicialmente Docker¹⁶ para compartir el proyecto (decisión de la empresa)

Para salir del entorno local de desarrollo y poder compartir el proyecto a otros usuarios para su interacción, se decide *dockerizar*¹⁷ el prototipo del chatbot y cargarlo en la nube de Azure de Brain Food. Esta decisión resulta en diversas problemáticas de errores de código y compilación, dada la complejidad del proyecto con múltiples dependencias y componentes.

Se necesita de conocimientos avanzados en Docker como la gestión de volúmenes y el uso de Docker Compose, lo cual es un obstáculo, particularmente porque no se cuenta con la experiencia técnica requerida para manejar tales conceptos complejos de manera eficiente. Esta complejidad adicional resulta en desviar recursos valiosos de la mejora y desarrollo del prototipo en sí.

Como alternativa a la dockerización y subida a Azure, se propone el uso de la nube de Streamlit. Esta corresponde a una solución mucho más sencilla y directa para compartir el prototipo con usuarios externos. Streamlit Cloud permite cargar y compartir aplicaciones de manera rápida y sin las complicaciones asociadas con Docker. Lo anterior, reduce significativamente la barrera técnica y facilita el proceso de iteración y feedback del prototipo, simplificando el despliegue y concentrando esfuerzos en el desarrollo y perfeccionamiento del chatbot.

¹⁶ Docker es una plataforma de contenedores que permite empaquetar aplicaciones y sus dependencias en unidades estándar para un despliegue eficiente en cualquier sistema, asegurando consistencia en los entornos de desarrollo, prueba y producción.

¹⁷ "Dockerizar" un proyecto implica encapsularlo en un contenedor Docker.

Capítulo 10: Conclusiones

Resultados y cumplimiento de objetivos

Primeramente, se aprende sobre los conceptos de embeddings, bases vectoriales y funcionamiento de los LLM's, además de frameworks para el desarrollo de aplicaciones basadas en inteligencia artificial generativa de texto como LangChain y Streamlit. Todo lo anterior resulta en que se logra levantar información sobre el estado del arte de esta tecnología que permite implementar efectivamente el prototipo deseado.

Se identifica que los potenciales clientes; directores empresariales, necesitan rescatar información confiable y de forma rápida de documentos para poder lidiar con sus actividades de diarias, así como extraer ideas principales y contar con resúmenes concisos de tópicos. De esta forma, se identifican ciertas necesidades de este tipo de clientes.

Se logra diseñar e implementar una plataforma web basada en un chatbot que responde efectivamente consultas sobre documentos de texto plano y estructurado. Las respuestas exitosas permiten colaborar con la búsqueda rápida de información clave. Se utiliza Streamlit como front-end y LangChain (junto con la base vectorial Chroma) como back-end. Por otro lado, se realiza y almacena la documentación de todo el trabajo realizado.

Por último, los resultados que se obtienen con la realización del proyecto permiten a Brain Food obtener un conocimiento valioso en el desarrollo de aplicaciones basadas en inteligencia artificial generativa de texto, en particular, por medio de Large Language Models.

Impacto en la organización

Los resultados obtenidos del proyecto consolidan y expanden los conocimientos existentes que la empresa obtuvo del proyecto anterior con la ACHS. De manera concreta, se aprende sobre la aplicación del método RAG, el cual no había sido utilizado anteriormente. Este último consiste en una técnica sumamente útil para la implementación de chatbots, los cuales son una solución que puede ser aplicada a diversas necesidades que enfrentan sus clientes actuales y potenciales.

Limitaciones y alcances

Ante documentos con estructuras complejas con distintos tipos de elementos, las respuestas son incorrectas o poco precisas, debido a la naturaleza no estructurada de los archivos. Esto último es un punto clave a considerar, puesto que una gran cantidad de documentos que se utilizan en la gestión diaria de los directores empresariales, son textos altamente cargados en tablas, gráficos e imágenes, tales como: reportes financieros, estados de resultados y presentaciones corporativas.

Por lo tanto, existe un desafío en poder lidiar de manera automática y veloz con este tipo de datos. Esto toma especial relevancia cuando se quiere cargar archivos en tiempo real. Al tratarse del diseño de una versión piloto, no se dispone con recursos financieros que permitan tratar con herramientas más avanzadas ni con mayor personal para armar un equipo que trabaje exclusivamente en el proyecto, salvo el estudiante.

Recomendaciones y trabajo futuro

Finalmente, como recomendaciones futuras a la empresa, se propone el seguir profundizando en el desarrollo de aplicaciones de inteligencia artificial generativa por medio de LLM's, en particular, en la puesta en productivo de tales aplicaciones. Esto debido a que el proyecto se desenvuelve en un ambiente en local, lo cual, si bien es un primer paso, no es una solución escalable ni aplicable al mundo real.

El paso a productivo conlleva una serie de variables, conocimientos y recursos que demandan la consolidación de un equipo. Ante esto, surge una segunda recomendación, el de interiorizar la posibilidad de crear un área dentro de la organización propia a Inteligencia Artificial Generativa. Actualmente, se están generando oportunidades de proyectos que pueden ser abordados mediante soluciones de esta tecnología. Por lo tanto, para llegar a nuevos clientes en otras industrias, resulta necesario contar con un departamento exclusivo a este tipo de soluciones. Así, Brain Food puede seguir consolidándose como una consultora de vanguardia.

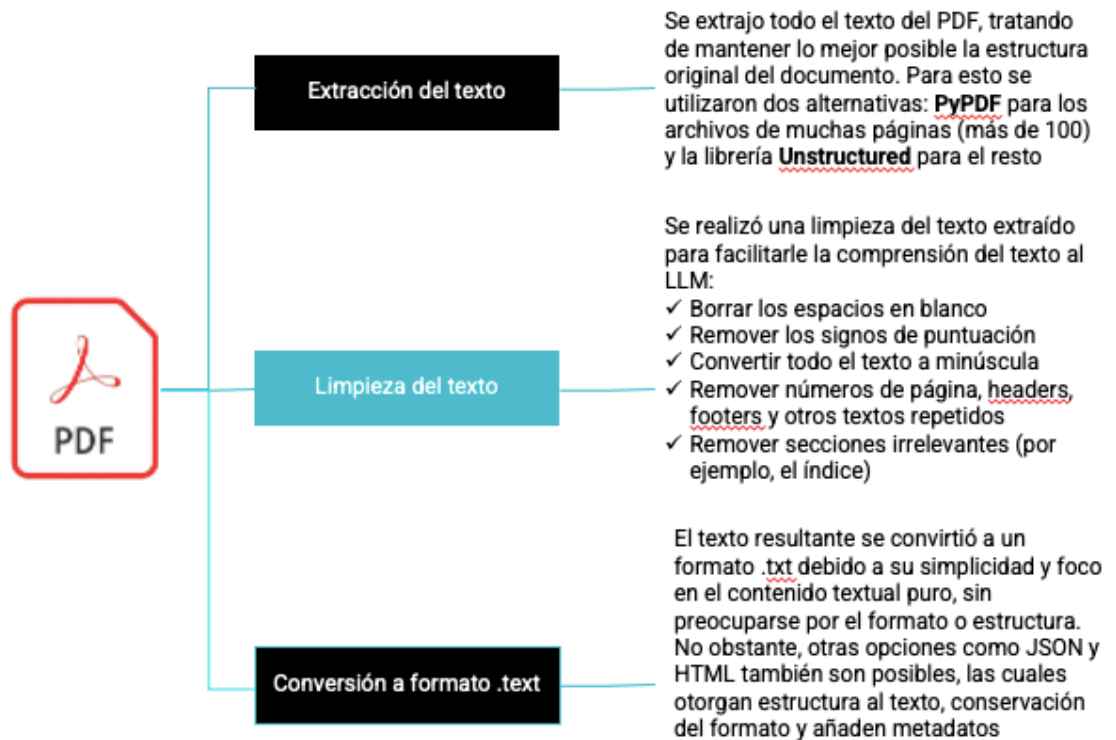
Bibliografía

- [1] Greenstein, B. (n.d.). Your industry's GenAI adoption play. PwC. <https://www.pwc.com/gx/en/issues/technology/genai-adoption-game-changer.html>
- [2] Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023). The economic potential of generative AI: The next productivity frontier. In McKinsey & Company. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier>
- [3] Rasiel, E. M. (1999). The McKinsey Way. McGraw-Hill Education.
- [4] Mercado de servicios de consultoría. (s.f.). Mordor Intelligence. <https://www.mordorintelligence.com/es/industry-reports/consulting-service-market>
- [5] Oksenberg, Andrés. (2022, agosto). Inducción Visión Brain Food / Realizado a los nuevos miembros de la empresa.
- [6] Doherty, N., & Canavan, T. (2022). Consulting in the digital age: How COVID-19 has accelerated the transformation of the consulting industry. *Journal of Management Consulting*, 33(2), 12-19.
- [7] Bansraj, D., & Singh, A. (2021). The role of artificial intelligence in reshaping the consulting industry. *International Journal of Information Management*, 58, 102315.
- [8] Generative AI is here: How tools like ChatGPT could change your business. (2022, 20 de diciembre). McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/generative-ai-is-here-how-tools-like-chatgpt-could-change-your-business>
- [9] Galpin, M. (2023, May 31). Do Directors prepare well for Board meetings? - VUCA. VUCA. <https://www.vuca.com.au/do-directors-prepare-well-for-board-meetings/>
- [10] BrainFood. (2023, 2 de agosto). Potenciando la información automatizada: explorando aplicaciones de modelo de lenguaje LLM - BrainFood. <https://brainfood.cl/blog-caso/potenciando-la-informacion-automatizada-explorando-aplicaciones-de-modelo-de-lenguaje-llm/>
- [11] What is Data Science? | IBM. (n.d.). <https://www.ibm.com/topics/data-science>
- [12] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- [13] Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(3), 417-424.

- [14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [15] Koupanou, N. (2023, 6 noviembre). The AI Boom: Practical Guide to Generative AI for Small Businesses. *Towards AI*. <https://towardsai.net/p/machine-learning/the-ai-boom-practical-guide-to-generative-ai-for-small-businesses>
- [16] *Explained: generative AI*. (2023, 9 noviembre). MIT News | Massachusetts Institute Of Technology. <https://news.mit.edu/2023/explained-generative-ai-1109>.
- [17] Lopez, M. (2023, 20 diciembre). 6 Usos de la inteligencia artificial generativa. *Immune Technology Institute*. <https://immune.institute/blog/usos-inteligencia-artificial-generativa/>.
- [18] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- [19] Chambers, M. (2024). *Generative AI with Large Language Models* [Curso en línea]. Coursera. <https://www.coursera.org/learn/generative-ai-with-llms#modules>.
- [20] Chambers, M. (2024). *Generative AI with Large Language Models* [Curso en línea]. Coursera. <https://www.coursera.org/learn/generative-ai-with-llms#modules>.
- [21] Aboze, B. J. (2024, 30 marzo). Risks of Large Language Models: A comprehensive guide | Deepchecks. *Deepchecks*. <https://deepchecks.com/risks-of-large-language-models/>.
- [22] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *arXiv preprint arXiv:2005.11401*.
- [23] Lorica, B. (2023, 19 octubre). Best Practices in Retrieval Augmented Generation. *Gradient Flow*. <https://gradientflow.substack.com/p/best-practices-in-retrieval-augmented>
- [24] Saltz, J. (2024, 10 abril). Managing generative AI projects. *Data Science Process Alliance*. <https://www.datascience-pm.com/managing-generative-ai-projects/>
- [25] Saltz, J. (2024b, 28 abril). The GenAI life cycle. *Data Science Process Alliance*. <https://www.datascience-pm.com/the-genai-life-cycle/>

Anexos

Anexo A: Procesamiento de los documentos



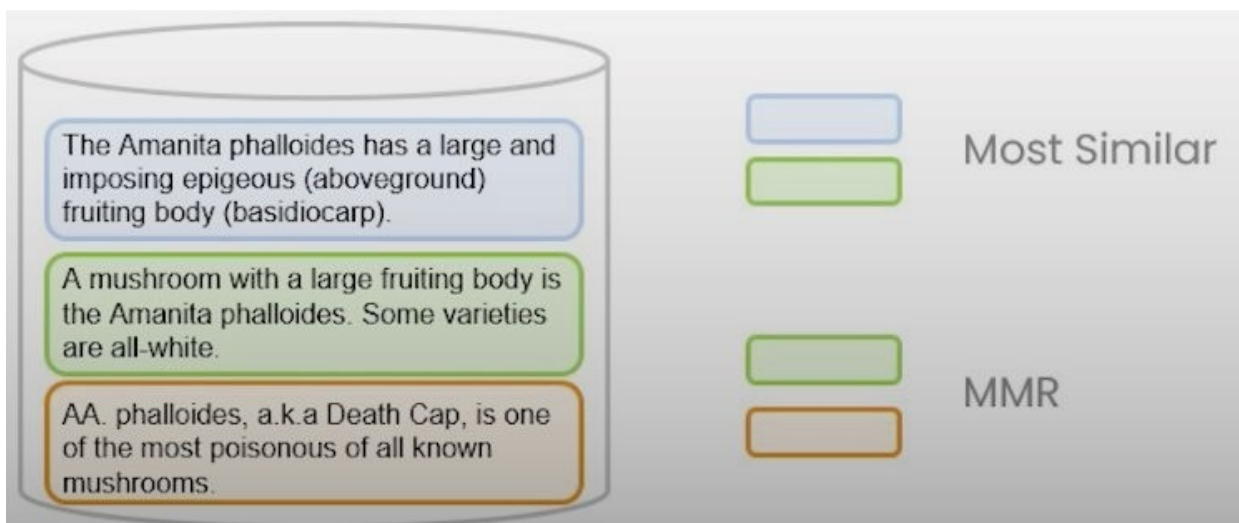
Anexo B: Estrategia Maximal Marginal Relevance

Esta alternativa selecciona los chunks por medio de una combinación entre aquellos que son más relevantes respecto a la query del usuario, pero también más diversos. Esto último quiere decir que selecciona también aquellos chunks que pueden no ser los más cercanos a la consulta en el espacio vectorial de los embeddings, pero que aportan información que puede ser complementaria y útil para responder efectivamente la solicitud evitando la redundancia.

Por ejemplo:

- Query: *Tell me about all white mushrooms with large fruiting bodies*

Se puede observar que los resultados obtenidos por el método *Most Similar* (similarity search, el cual es el método por defecto en la búsqueda de vectores de embeddings) tienen sentido y pueden considerarse correctos. No obstante, MMR recupera un fragmento de texto que aporta una información clave: el hongo Death Cap es uno de los más venenosos.

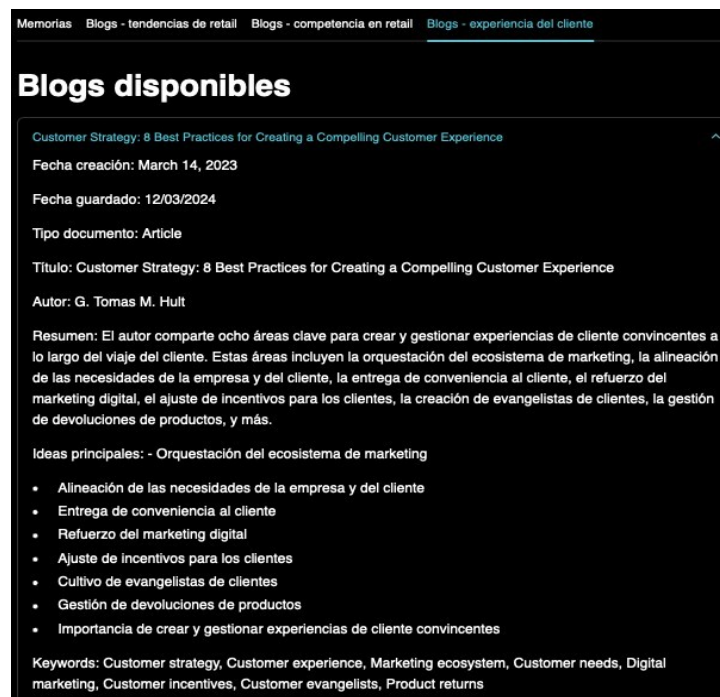


Anexo C: Pestañas de la plataforma

C.1: Pestaña “Fichas documentos precargados”



C.2: Ficha resumen de cada documento precargado



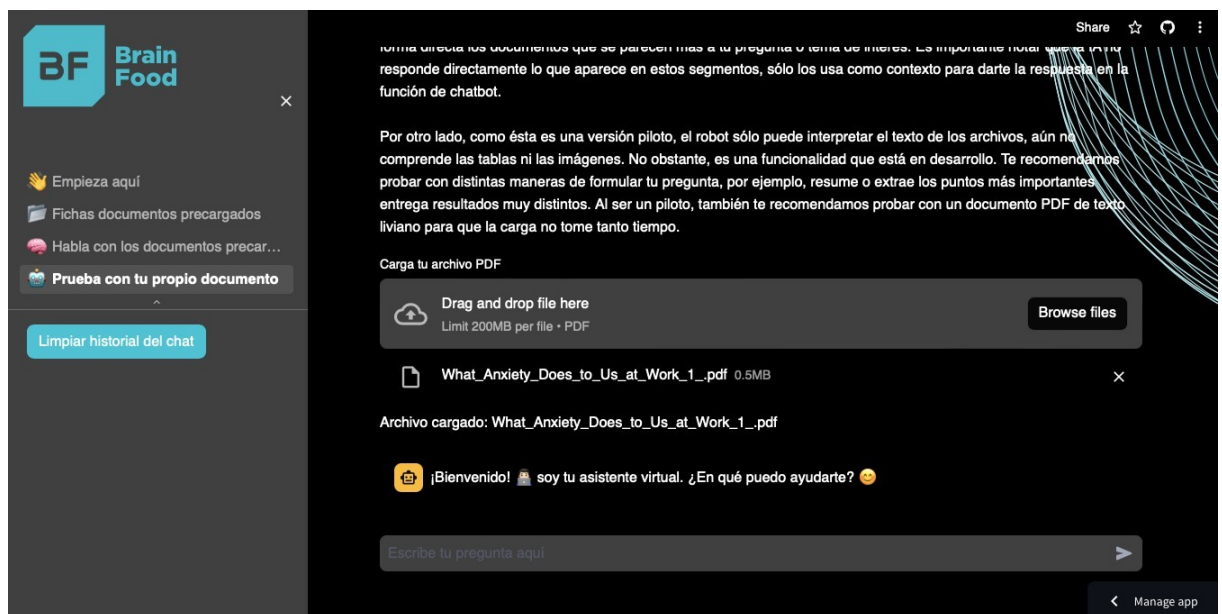
C.3: Pestaña “Habla con los documentos precargados”

The screenshot shows the Brain Food application interface. On the left, a sidebar contains the Brain Food logo and navigation options: 'Empieza aquí', 'Fichas documentos precargados', 'Habla con los documentos precargados' (selected), and 'Prueba con tu propio documento'. Below these is a slider for 'Segmentos de texto a recuperar' set to 3, and a 'Limpiar historial del chat' button. The main chat area has a title 'Habla con los documentos precargados' and an introductory paragraph. A chatbot message says: '¡Bienvenido! soy tu asistente virtual. ¿En qué puedo ayudarte?'. At the bottom is a text input field 'Escribe tu pregunta aquí' with a send button.

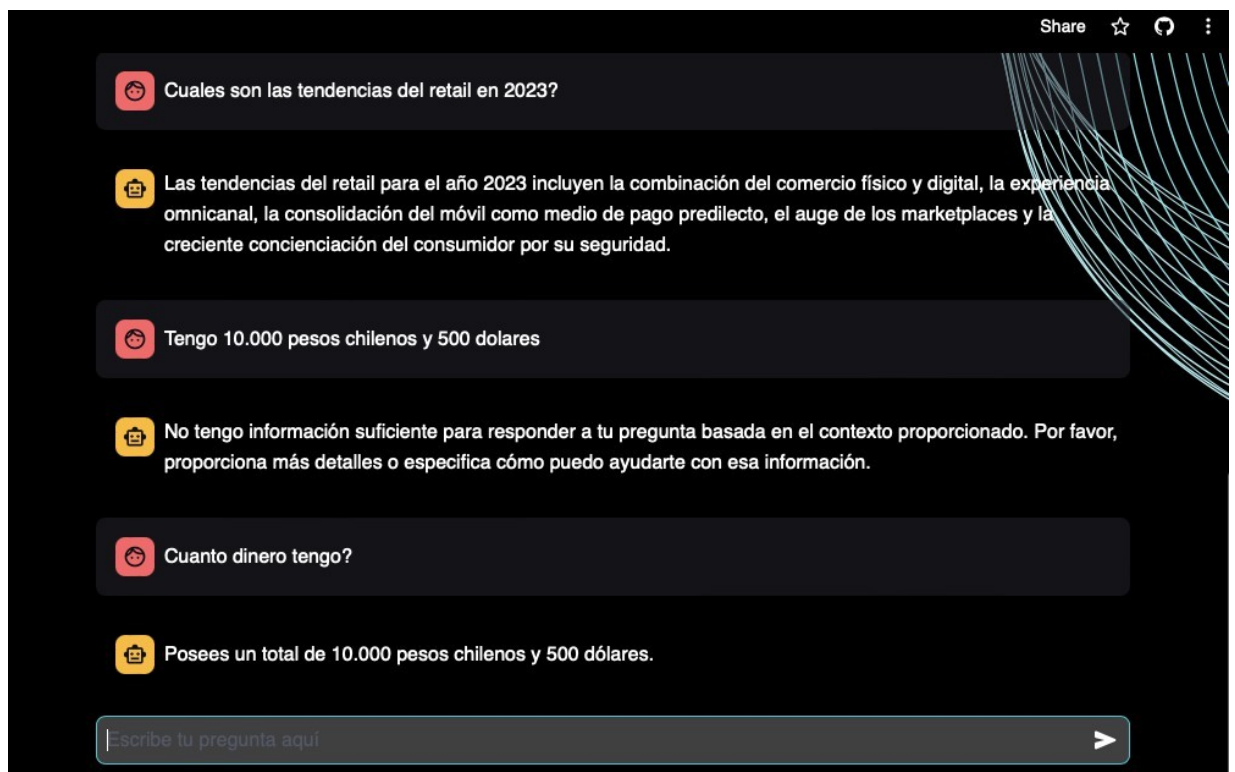
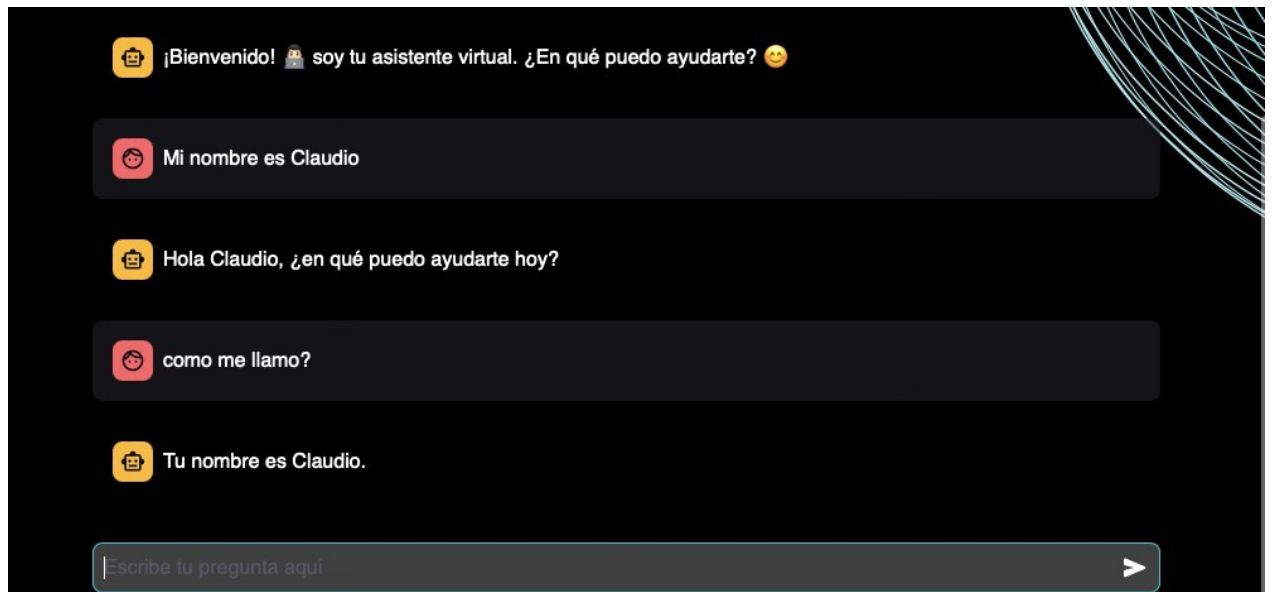
C.4: Interacción pregunta-respuesta con el chatbot

This screenshot shows the same Brain Food interface as C.3, but with a question and answer displayed in the chat. The chatbot's message is: '¡Bienvenido! soy tu asistente virtual. ¿En qué puedo ayudarte?'. The user's question is: 'Cuales son los principales desafíos que enfrentó Falabella en el año 2018. Compáralos con los desafíos que experimentó en el año 2019'. The chatbot's response is: 'En el año 2018, Falabella enfrentó desafíos relacionados con la adquisición del 100% de Linio y la firma de un memorándum de entendimiento para desarrollar y operar Ikea. Estos desafíos implicaban la expansión de la empresa y la gestión de nuevas operaciones y asociaciones. En comparación, en el año 2019, Falabella se vio enfrentada a desafíos relacionados con la pandemia del COVID-19, que generó cuarentenas extensas, cierres de tiendas y malls, así como la necesidad de adaptarse a un nuevo estilo de vida de los clientes, con el aumento del teletrabajo y las clases online. Estos desafíos eran más inesperados y estaban relacionados con cambios en el comportamiento del consumidor y en la forma en que se realizaban las transacciones comerciales.' The input field at the bottom is empty.

C.5: Pestaña “Prueba con tu propio documento”



Anexo D: Prueba de memoria del chatbot



Anexo E: Prueba con tablas

E.1: Prueba 1

Prueba 1 con la pregunta *¿Cuánto es el total de activos corrientes?*

ESTADOS CONSOLIDADOS DE SITUACION FINANCIERA CLASIFICADO
Al 30 de septiembre de 2023 (no auditado) y 31 de diciembre de 2022
(Expresados en miles de pesos)

ACTIVOS	Nota	30-09-2023	31-12-2022
		M\$	M\$
ACTIVOS CORRIENTES			
Efectivo y equivalentes al efectivo	5	8.577.248	23.890.192
Otros activos financieros, corrientes	6	25.222.705	431.893
Otros activos no financieros, corrientes	7	29.358.697	29.923.779
Deudores comerciales y otras cuentas por cobrar, corrientes	8	13.865.869	10.825.330
Cuentas por cobrar a entidades relacionadas, corrientes	9	1.351.098	1.710.938
Inventarios, corrientes	10	248.894.905	222.631.267
Activos por impuestos, corrientes	11	3.277.975	572.413
Total Activos Corrientes		330.548.497	289.985.812

Respuesta del chatbot

```
print(chain.invoke('Cuanto es el total de activos corrientes'))
```

✓ 2.9s

El total de activos corrientes al 30 de septiembre de 2023 es de 330.548.497 M\$.

E.2: Prueba 2

Prueba 2 con la pregunta *¿Cuánto fue el resultado integral de las reservas por diferencias de cambio?*

ESTADOS CONSOLIDADOS DE CAMBIOS EN EL PATRIMONIO Por los períodos de nueve meses terminados al 30 de septiembre de 2023 y 2022 (no auditados) (Expresados en miles de pesos)

2023							
Conceptos	Nota	Capital emitido	Primas de emisión	Reservas por diferencias de cambio por conversión	Otras reservas varias	Otras reservas	C
		M\$	M\$	M\$	M\$	M\$	a
Saldo inicial al 01/01/2023		52.492.611	11.160.579	4.525.101	(1.982.441)	2.542.660	1
Ganancia (pérdida)		0	0	0	0	0	
Otro resultado integral		0	0	1.140.325	0	1.140.325	
Resultado integral		0	0	1.140.325	0	1.140.325	
Dividendos		0	0	0	0	0	
Incremento (disminución) por transferencias y otros cambios		0	0	0	0	0	
Total de cambios en patrimonio		0	0	1.140.325	0	1.140.325	
Saldo final período 30/09/2023	22	52.492.611	11.160.579	5.665.426	(1.982.441)	3.682.985	1

Respuesta del chatbot

```
print(chain.invoke('Cuanto fue el resultado integral de las reservas por diferencias de cambio'))
```

✓ 4.2s Python

El resultado integral de las reservas por diferencias de cambio para el período del 01-01-2023 al 30-09-2023 fue de M\$ 1.140.325