

Universidad Internacional de La Rioja

Escuela Superior de Ingeniería y Tecnología

Máster Universitario en Ingeniería de Software y Sistemas Informáticos

Desarrollo de una Plataforma de IA con control de datos sensibles para el sector empresarial: Aplicación en Open Finance

|  |  |
| --- | --- |
| Trabajo fin de estudio presentado por: | Jonathan Francisco Lema Collaguazo |
| Tipo de trabajo: | Desarrollo |
| Director/a: | David Segurado Moreno |
| Fecha: |  |

Resumen

Se completará en las siguientes etapas

Abstract

Se completará en las siguientes etapas

Índice de contenidos

[1. Introducción 1](#_Toc198755374)

[1.1. Justificación 1](#_Toc198755375)

[1.2. Planteamiento del problema 2](#_Toc198755376)

[1.3. Estructura del trabajo 3](#_Toc198755377)

[2. Contexto y estado del arte 4](#_Toc198755378)

[2.1. Evolución de los modelos de IA hasta la era de la IA Generativa 4](#_Toc198755379)

[2.1.1. Aprendizaje automático clásico (1980 – 2010) 4](#_Toc198755380)

[2.1.2. Explosión del Deep Learning (2012 – 2016) 4](#_Toc198755381)

[2.1.3. Auge de la IA Generativa (2014 – presente) 5](#_Toc198755382)

[2.1.4. Tendencias actuales y adopción empresarial 6](#_Toc198755383)

[2.2. Impacto de la IA generativa en el trabajo cotidiano 6](#_Toc198755384)

[2.3. Regulación y protección de datos 7](#_Toc198755385)

[2.4. tecnologías y metodologías asociadas a la ia generativa 8](#_Toc198755386)

[2.4.1. Vulnerabilidades en el manejo de datos sensibles 8](#_Toc198755387)

[2.4.2. Procesos de filtrado y anonimización en el ámbito empresarial 9](#_Toc198755388)

[2.4.3. Prácticas y propuestas similares 10](#_Toc198755389)

[2.4.4. Retrieval-Augmented Generation (RAG/RAC) y su relación con la IA Generativa 10](#_Toc198755390)

[2.5. Caso de estudio y elección del modelo 10](#_Toc198755391)

[2.5.1. Contexto inicial sobre el tema especializado 11](#_Toc198755392)

[2.5.2. Equipos no técnicos y mala interpretación de la información 11](#_Toc198755393)

[2.5.3. Motivación y enfoque del caso de estudio 11](#_Toc198755394)

[2.5.4. Elección del modelo de IA 12](#_Toc198755395)

[3. Objetivos concretos y metodología de trabajo 12](#_Toc198755396)

[3.1. Objetivo general 12](#_Toc198755397)

[3.2. Objetivos específicos 13](#_Toc198755398)

[3.3. Metodología del trabajo 13](#_Toc198755399)

[4. Desarrollo específico de la contribución 15](#_Toc198755400)

[4.1. Diseño Funcional 15](#_Toc198755401)

[4.1.1. Diseño conceptual 15](#_Toc198755402)

[4.1.2. Identificación de requisitos 17](#_Toc198755403)

[4.1.3. Descripción del sistema software desarrollado 18](#_Toc198755404)

[4.1.4. Evaluación 18](#_Toc198755405)

[4.2. Levantamiento de requisitos 18](#_Toc198755406)

[4.2.1. Identificación de requisitos 18](#_Toc198755407)

[4.2.2. Descripción de la planificación del proyecto de desarrollo de software 18](#_Toc198755408)

[4.3. Tipo 3. Desarrollo de metodología 18](#_Toc198755409)

[4.3.1. Identificación de requisitos 18](#_Toc198755410)

[4.3.2. Descripción de la metodología 19](#_Toc198755411)

[4.3.3. Evaluación 19](#_Toc198755412)

[5. Conclusiones y trabajo futuro 20](#_Toc198755413)

[5.1. Conclusiones 20](#_Toc198755414)

[5.2. Trabajo futuro 20](#_Toc198755415)

[Referencias bibliográficas 21](#_Toc198755416)

[6. Bibliografía 21](#_Toc198755417)

[Anexo A. Artículo 21](#_Toc198755418)

[Anexo B. Título anexo 26](#_Toc198755419)

Índice de figuras

[Ilustración 1 - Flujo funcional plataforma 15](#_Toc198755322)

[Ilustración 2 - Flujo de secuencia Plataforma 17](#_Toc198755323)

Índice de tablas

No se encuentran elementos de tabla de ilustraciones.

# Introducción

Se realizará este apartado en la etapa final del proyecto

## Justificación

El uso de la IA Generativa ha venido creciendo de forma exponencial en el sector empresarial, al ofrecer herramientas capaces de automatizar tareas especializadas y reducir drásticamente las curvas de aprendizaje (OpenIA, 2022). Diversos informes apuntan a que esta tecnología permite mejorar la productividad y agilizar la generación de conocimiento, ya que los usuarios, incluso sin formación técnica, pueden beneficiarse rápidamente de las capacidades de los modelos de lenguaje (McKinsey & Company, 2023) .Sin embargo, este avance conlleva nuevos desafíos de seguridad y confidencialidad cuando se manejan datos sensibles, dado que muchos sistemas de IA actuales no incorporan mecanismos intrínsecos de filtrado o anonimización de la información (Dwork, 2008).

En este sentido, algunas empresas han optado por bloquear por completo el uso de estas herramientas, temiendo posibles fugas de información crítica o incumplimientos de normativas de protección de datos (Sfetcu, 2024). No obstante, el mero acto de prohibir la IA Generativa puede suponer una barrera a la innovación, limitando la competitividad de la organización frente a otras que sí aprovechan las ventajas de estas tecnologías. Es por ello que la solución pasa por habilitar un entorno corporativo controlado, en el cual los modelos puedan ser utilizados con la garantía de que cualquier tipo de dato sensible sea detectado y anonimizado antes de exponerse, salvaguardando así la privacidad y cumpliendo con marcos regulatorios como el RGPD (Kairouz, 2020).

A la vez, se hace cada vez más evidente la necesidad de “entrenar” modelos con conocimiento experto utilizando modelos de lenguaje que han sido pre entrenados y que puedan realizarse al alcance de usuarios no expertos en este campo (Patrick Lewis, 2020). En la actualidad, la popularidad de herramientas como ChatGPT demuestra que la IA Generativa puede democratizar la obtención de respuestas complejas, pero también expone la carencia de filtros adecuados y la falta de verificación de las fuentes. Así, surgió la segunda problemática que se pretende abordar: la importancia de contar con expertos que verifiquen y optimicen la información suministrada al modelo, asegurando exactitud y confiabilidad. De esta manera, se reduce el riesgo de adoptar ciegamente lo que “dice” la IA y se promueve un uso responsable y crítico en el entorno empresarial en cuanto a las grandes LLMs que ya son parte de la realidad de muchos usuarios (Edemacu & Wu, 2024).

Por tanto, la presente investigación propone el desarrollo de una plataforma centrada en dos grandes pilares:

(1) control de datos sensibles, mediante la anonimización o descarte de información crítica antes de su procesamiento por la IA, y

(2) entrenamiento de modelos accesible para personal experto en la materia, aun sin conocimientos técnicos. Con esta aproximación, se abre una vía para que el sector empresarial adopte la IA Generativa de manera confiable, potenciando su competitividad sin poner en riesgo la seguridad de la información y fomentando, a la vez, el desarrollo de criterios críticos y mecanismos de validación en la adopción de estas nuevas tecnologías.

## Planteamiento del problema

El uso de la IA Generativa en entornos empresariales ha evidenciado la necesidad de contar con mecanismos que garanticen el tratamiento seguro de la información, en especial cuando se manejan datos sensibles. Muchas organizaciones, ante el riesgo de exposición de información confidencial, optan por limitar o prohibir su uso, lo que repercute negativamente en la competitividad y en la adopción de soluciones tecnológicas innovadoras.

La forma de solventar esta situación se basa en establecer mecanismos de control de la IA que, por un lado, detecte y anonimice la información crítica o sensible antes de que sea procesada por un modelo y, por otro, facilite la formación y el uso de dichas herramientas sin la necesidad de la prohibición.

El presente trabajo de investigación y desarrollo propone el diseño e implementación de una plataforma que cumpla con dos objetivos generales:

* Control de datos sensibles: implementar un sistema de anonimización o descarte que sirva como “firewall” de la información. Este mecanismo deberá funcionar de forma transparente para el usuario y permitirá a las empresas aprovechar la IA sin riesgos de fuga de datos.
* Entrenamiento de modelos accesible: habilitar un entorno que facilite el entrenamiento y la configuración de la IA por parte de expertos en diversos dominios, aun sin conocimientos técnicos. Esta herramienta debe ser intuitiva, guiando paso a paso el proceso de carga, filtrado y verificación de la información.

Esta plataforma brindaría a las organizaciones una vía intermedia entre la prohibición total y el uso descontrolado de la IA Generativa, permitiendo un aprovechamiento responsable de sus capacidades. Adicionalmente, la plataforma deberá contar con buenas prácticas que permita un correcto control y gestión de accesos para contar con la posibilidad de monitoreo y mejora continua.

## Estructura del trabajo

Este apartado se desarrollará al final

# Contexto y estado del arte

## Evolución de los modelos de IA hasta la era de la IA Generativa

La comprensión de la Inteligencia Artificial (IA) ha transitado por varias fases técnicas y conceptuales que culminan en los actuales modelos generativos de gran tamaño. A continuación, se presenta una síntesis académica de esa evolución con los hitos más relevantes para el ámbito de esta tesis

### Aprendizaje automático clásico (1980 – 2010)

La disponibilidad de datos y de potencia computacional incentivó métodos que aprenden patrones estadísticos:

* Perceptrón y redes neuronales “planas” (Rosenblatt, 1958) sentaron la base neuronal, luego eclipsada por problemas de escalado.
* Algoritmos de propósito general como máquinas de soporte vectorial (Cortes & Vapnik, 1995) o árboles de decisión dominaron la práctica industrial en reconocimiento de texto y tabulación de riesgos.
* El auge del big data provocó interés en pipelines de entrenamiento y validación, pero la representación de conocimiento seguía siendo mayormente discriminativa (clasificar, predecir).

### Explosión del Deep Learning (2012 – 2016)

La deep convolutional network que ganó ImageNet en 2012 (Krizhevsky, Sutskever, & Hinton, 2012) demostró que entrenar redes profundas con GPU podía superar con creces técnicas tradicionales. Dos resultados clave:

* Visión por computador: reducción drástica del error en clasificación de imágenes.
* Procesamiento del lenguaje natural (PLN): aparición de word embeddings continuos (Mikolov, Chen, Corrado, & Dean, 2013)

Estos avances convergieron en la hipótesis de que escalabilidad + datos podrían transferirse a tareas de generación de contenido.

### Auge de la IA Generativa (2014 – presente)

Tabla 1 - Modelos de IA Generativa

|  |  |  |
| --- | --- | --- |
| **Sub-familia** | **Avance técnico** | **Relevancia** |
| **Modelos generativos de imágenes** | *Generative Adversarial Networks* (GAN) (Goodfellow, y otros, 2014) y posteriores modelos de difusión (Ho, Jain, & Abbeel, 2020) | Demostraron la viabilidad de crear datos sintéticos de alta fidelidad. |
| **Variational Auto-Encoders** | Algoritmo de inferencia y aprendizaje varacional (Kingma & Welling, 2013). | Introdujeron la maximización de evidencia para síntesis controlada. |
| **Modelos de lenguaje autoregresivos** | GPT-2 (Radford, y otros, 2019) mostró generación coherente a gran escala. GPT-3 (Brown, y otros, 2020) consolidó el paradigma *few-shot*. | Capacidades lingüísticas de propósito general, accesibles vía API. |
| **Transformers con atención** | Arquitectura Transformer (Vaswani, y otros, 2017)en su paper “Atention is all you need” desplazó a LSTM/GRU en PLN gracias a la atención auto-regresiva, fundamento de los **LLM** actuales. | Facilita paralelización y escalado a billones de parámetros. |

Los modelos generativos hacen algo cualitativamente nuevo: modelan la distribución completa de los datos, lo que permite crear texto o imágenes inéditos y contextuales, en vez de limitarse a clasificarlos.

### Tendencias actuales y adopción empresarial

* APIs comerciales de grandes modelos (p. ej., OpenAI, Anthropic, Google) han democratizado la tecnología para usuarios no técnicos.
* Modelos de código abierto (BLOOM, LLaMA-2) facilitan despliegues privados y adaptaciones específicas.
* Técnicas de alignment y filtrado abordan la seguridad y la protección de datos, requisito central en dominios regulados como Open Finance.
* Paradigma Retrieval-Augmented Generation (RAG/RAC) —tratado en una sección posterior— hibrida grandes modelos con bases documentales especializadas para aumentar precisión y reducir “alucinaciones”.

## Impacto de la IA generativa en el trabajo cotidiano

La adopción de la IA Generativa ha crecido de manera exponencial en los últimos años, impulsada por la proliferación de modelos de lenguaje capaces de producir texto coherente y contextual sin necesidad de una configuración compleja por parte del usuario (OpenIA, 2022). Herramientas como ChatGPT o Google Bard han logrado democratizar la inteligencia artificial, facilitando su implementación en diversos entornos y acelerando la adopción por parte de personas sin perfiles técnicos (McKinsey & Company, 2023).

Esta tendencia, lejos de revertirse, se presenta como un cambio de paradigma que obliga a las empresas a adaptarse para no quedar rezagadas. La productividad y la eficiencia en la creación de contenido, la atención al cliente y la automatización de tareas rutinarias han mejorado significativamente con el uso de sistemas de IA Generativa (Kairouz, 2020). Sin embargo, la facilidad de acceso y el rápido ritmo de implantación superan a menudo la capacidad de las organizaciones para controlar la información que los usuarios introducen en estos servicios (Sfetcu, 2024).

Como consecuencia, se hace imprescindible un enfoque que combine la explotación de las ventajas competitivas de la IA Generativa con la protección de datos sensibles, especialmente cuando las plataformas se nutren de información crítica de la organización. En este sentido, sistemas de filtrado, anonimización y auditoría representan soluciones clave para salvaguardar la confidencialidad y cumplir con las regulaciones vigentes, sin obstaculizar el potencial de innovación ni el crecimiento empresarial. Bajo esta premisa, se justifica la necesidad de plantear propuestas centradas en salvaguardar los datos, al tiempo que se preserva la usabilidad y la efectividad de la IA para todos los perfiles involucrados.

## Regulación y protección de datos

La rápida adopción de la IA Generativa en entornos empresariales ha desencadenado numerosas preocupaciones vinculadas a la protección de datos, especialmente cuando los usuarios introducen información sensible en modelos externos o servicios en la nube. En la Unión Europea, el Reglamento General de Protección de Datos (RGPD) obliga a las organizaciones a implementar medidas que garanticen la confidencialidad, integridad y disponibilidad de la información personal que procesan, con sanciones económicas que pueden llegar hasta el 4% de la facturación global (Comisión Europea, 2016). De manera análoga, en Estados Unidos, la Ley de Privacidad del Consumidor de California (CCPA) y normativas sectoriales como HIPAA, en el ámbito sanitario, exigen salvaguardas para proteger los datos de los usuarios, así como la obligación de notificar brechas de seguridad.

Este escenario se complica por la falta de estandarización en el uso de herramientas de IA Generativa. Algunas empresas operan con plataformas externas cuyas políticas de tratamiento de datos pueden ser ambiguas, incrementando el riesgo de filtración de información confidencial y de incumplir las normas de privacidad (Sfetcu, 2024). Además, el desconocimiento de los empleados sobre la ubicación y forma de almacenamiento de la información que introducen en dichas herramientas agrava el problema (Kairouz, 2020).

En respuesta, han surgido propuestas que contemplan la anonimización o seudonimización de datos antes de su envío a la IA. Estas prácticas, contempladas en el propio RGPD bajo el principio de “privacidad desde el diseño”, permiten a las organizaciones minimizar la exposición de información crítica sin renunciar a los beneficios de la IA Generativa (Dwork, 2008). Sin embargo, la implementación de tales soluciones no siempre es sencilla, pues requiere de herramientas especializadas y del diseño de políticas internas que coordinen los procesos de negocio con el cumplimiento normativo. De ahí surge la necesidad de desarrollar plataformas que unifiquen la protección de datos, el control de acceso y la facilidad de uso, promoviendo una adopción responsable de la IA Generativa en el ámbito empresarial.

## tecnologías y metodologías asociadas a la ia generativa

En la actualidad, muchas de las tecnologías de IA Generativa que han alcanzado un uso masivo se caracterizan por interfaces amigables y accesibles vía web, orientadas a usuarios sin un perfil necesariamente técnico. Ejemplos notorios son ChatGPT, Google Bard y otras plataformas de diálogo con lenguaje natural basadas en grandes modelos de lenguaje (OpenIA, 2022). Su popularidad se fundamenta en la posibilidad de que un empleado de cualquier área pueda redactar consultas en lenguaje natural y obtener respuestas de alta calidad de manera inmediata (McKinsey & Company, 2023). Sin embargo, esta misma accesibilidad expone un punto crítico: el intercambio de datos confidenciales o sensibles sin que medie un proceso previo de filtrado o anonimización (Edemacu & Wu, 2024).

### Vulnerabilidades en el manejo de datos sensibles

Debido a la falta de conocimientos técnicos de gran parte de los usuarios, es frecuente que se compartan con la IA documentos o fragmentos de información que contengan, por ejemplo, nombres de clientes, cifras de facturación o detalles contractuales, sin tomar en cuenta los riesgos de privacidad o los posibles incumplimientos legales (European Union Agency for Cybersecurity (ENISA), 2021). Como resultado, los proveedores externos de IA podrían almacenar e indexar dichos datos, dejando a la organización sin control sobre su destino o forma de utilización (Kairouz, 2020).

Además de ello, diversas plataformas de IA Generativa mantienen políticas de uso y manejo de datos que, si bien estipulan ciertos niveles de protección, no siempre se alinean con la normativa local o las políticas internas de la empresa. Esta disparidad incrementa la probabilidad de filtraciones y dificulta la trazabilidad de lo que ocurre con la información sensible tras su ingreso en el sistema (Dwork, 2008).

### Procesos de filtrado y anonimización en el ámbito empresarial

Para mitigar los riesgos mencionados, han surgido soluciones enfocadas en la detección y tratamiento de datos sensibles antes de su envío a la IA. Entre los mecanismos más empleados destacan:

* Identificación y extracción de PII (Personally Identifiable Information): A través de librerías como Spacy o Presidio, capaces de reconocer nombres, direcciones o números de identificación personal (Arvanitakis & Moustakides, 2020).
* Seudonimización y cifrado selectivo: En lugar de eliminar la información, se la sustituye por equivalentes o se la cifra de modo que su recuperación requiera claves de desencriptación (Ren, Zhang, Hong, & He, 2022).
* Enmascaramiento semántico: Consiste en reemplazar términos sensibles por tokens genéricos (por ejemplo, <CLIENTE\_X>), salvaguardando la coherencia del texto y evitando divulgar datos críticos.

No obstante, la mayoría de estas herramientas o metodologías suelen requerir intervención de personal especializado en NLP (Procesamiento de Lenguaje Natural) y ciberseguridad, generando barreras para equipos de negocio que buscan adoptar la IA Generativa de forma ágil. En la práctica, esto se traduce en procesos manuales o semi-automatizados que ralentizan la adopción de soluciones de IA y desincentivan su uso continuo (Nunes & Abreu, 2022)

### Prácticas y propuestas similares

Existen iniciativas denominadas “privacy firewalls” o “data preprocessors” cuyo objetivo es actuar como una capa intermedia entre el usuario y la plataforma de IA Generativa, analizando el contenido enviado y bloqueando o anonimizando la información sensible. Sin embargo, la mayoría de estos enfoques están centrados únicamente en la protección de datos y no contemplan la necesidad de un entorno en el que usuarios no técnicos puedan también entrenar y validar modelos con información de dominio específico (Lewis, y otros, 2020)

Ante este panorama, se identifica una oportunidad clara para soluciones integrales que reúnan, en una misma plataforma, la filtración y anonimización de información sensible y la facilidad de entrenamiento de modelos por parte de expertos de negocio. La propuesta que se desarrolla en esta investigación busca cubrir justamente este vacío, combinando controles de seguridad con una interfaz intuitiva que permita a los especialistas de área, sector o industria entrenar sus propios modelos y dejarlos disponibles para el uso interno de una empresa—en el caso de este trabajo será el ámbito de Open Finance—extender el conocimiento del modelo sin comprometer la confidencialidad de los datos.

### Retrieval-Augmented Generation (RAG/RAC) y su relación con la IA Generativa

## Caso de estudio y elección del modelo

Se ha elegido al Open Finance como caso de estudio para el entrenamiento del modelo, ya que el autor tiene un grado de expertise alta en ese campo de conocimiento y estará en capacidad de identificar conocimiento especializado para la fase de entrenamiento y criterio suficiente para entender que las respuestas a preguntas serán adecuadas.

### Contexto inicial sobre el tema especializado

Open Finance se deriva de la evolución de Open Banking y se centra en la interconexión y apertura de datos financieros a través de interfaces estandarizadas (APIs) y seguras. Bajo esta premisa, múltiples entidades—bancos, fintech, aseguradoras, entre otros—comparten información financiera de forma controlada para ofrecer servicios y productos más personalizados. Sin embargo, esta apertura de datos también expone a las organizaciones a mayores riesgos de mala manipulación o uso indebido de la información cuando se introducen soluciones tecnológicas sin una adecuada supervisión.

### Equipos no técnicos y mala interpretación de la información

En regiones en proceso de regulación hace que se creen grupos de trabajo mixtos en los que tienen que abordar temas técnicos con el objetivo de aterrizar una normativa legal en un estándar técnico que permita la interoperabilidad de los participantes del ecosistema Open Finance. Ante esta problemática, las herramientas de IA Generativa han permitido que temáticas muy complejas se puedan abordar sin dependencia de expertos en ese campo, lo que ha permitido que el desarrollo y aterrizaje de estos estándares tengan una mayor velocidad. Por otro lado, está el problema de que la IA Generativa no necesariamente llega a dar una respuesta válida y las personas que hacen uso de estas herramientas no tienen el criterio para validar o no las respuestas.

### Motivación y enfoque del caso de estudio

Dada la criticidad de la información manejada y la elevada responsabilidad legal de las instituciones financieras, Open Finance se presenta como un terreno idóneo para evaluar soluciones que combinen el uso de la IA Generativa con mecanismos de control. Por un lado, el sector financiero demanda alta precisión en los modelos para ofrecer recomendaciones y servicios confiables; por otro, exige estrictas garantías de seguridad y cumplimiento normativo (Arner, Barberis, & Buckley, 2020). Esta dualidad resalta la pertinencia de contar con una plataforma y un perfil externo que pueda validar este proceso garantizando el cumplimiento objetivo de la plataforma.

### Elección del modelo de IA

Los grandes modelos poseen conocimiento general, pero pueden carecer de información especializada o actualizada. El paradigma Retrieval-Augmented Generation (RAG)—llamado también Retriever-Augmented Chat (RAC) cuando se orienta a asistentes conversacionales—combina dos componentes:

* Motor de recuperación: busca en una base de documentos interna (índices vectoriales, bases de datos, etc.) los pasajes más relevantes para la consulta del usuario.
* Modelo generativo: emplea esos pasajes como contexto adicional (“evidencia”) para redactar la respuesta.

Frente a otro tipo de modelos este tiene algunas ventajas que permiten una personalización adecuada de los datos de entrenamiento sin crear las alucinaciones, dependencias de fine-tuning y siempre bajo un control de dominio realizado por un experto en la materia.

# Objetivos concretos y metodología de trabajo

## Objetivo general

El objetivo general de este trabajo consiste en diseñar y desarrollar una plataforma tecnológica que facilite a usuarios el acceso a módulos de IA Generativa basados en modelos de RAC/RAG que han sido previamente entrenados por expertos y en la que la información entregada sea adecuada para el uso en un entorno empresarial.

La plataforma adicionalmente tendrá un módulo en el que se podrá entrenar el modelo de forma amigable y sin necesidad de conocimientos técnicos de IA, sino de un conocimiento experto en el área de conocimiento con el objetivo de validar las informaciones recibidas. Para ello este módulo contará con una capa de filtrado que permitirá la anonimización de datos de acuerdo a las leyes de protección de datos vigentes.

## Objetivos específicos

* Definir los requerimientos funcionales y no funcionales de la plataforma, de manera que esta permita a los usuarios acceder a módulos especializados para la generación de modelos entrenados para uso empresarial.
* Diseñar una arquitectura de alto nivel que contemple, al menos, las siguientes capacidades:
* Habilitación de módulos organizados por temáticas y con capacidad de habilitación y restricción por tipos de usuarios.
* Inclusión, dentro de cada módulo, de una interfaz interactiva que permita la captura estructurada de texto.
* Diseño y validación del modelo de IA basados en RACs dentro de plataformas que garanticen que no habrá uso de los datos para entrenamientos del modelo general

## Metodología del trabajo

La metodología empleada para el desarrollo del Trabajo de Fin de Máster (TFM) se basará en el marco ágil Scrum, complementado con elementos de Kanban para el seguimiento continuo del avance.

**Enfoque Ágil**

El desarrollo se estructurará en tres sprints, con una duración estimada de un mes cada uno, alineados con las entregas parciales requeridas por la Universidad Internacional de La Rioja (UNIR). Durante estos sprints se utilizarán los siguientes artefactos de Scrum:

* Product Backlog: conjunto priorizado de funcionalidades y requisitos a desarrollar.
* Sprint Backlog: subconjunto de tareas seleccionadas para cada sprint.
* Sprint Burndown: herramienta de seguimiento del progreso diario respecto a los objetivos del sprint.

Para la gestión visual de tareas e historias de usuario se integrará un tablero Kanban, permitiendo el monitoreo detallado del avance dentro del backlog de cada sprint.

**Ceremonias**

Las ceremonias establecidas para asegurar el cumplimiento de los objetivos del TFM son:

* Daily Meeting interno: revisión diaria del backlog y actualización del estado de las historias de usuario dentro del tablero Kanban, promoviendo la visibilidad y la identificación temprana de bloqueos.
* Retrospectiva mensual con el director del TFM: evaluación del incremento de valor entregado, análisis de impedimentos y recepción de retroalimentación general sobre el desarrollo del trabajo.

Adicionalmente, se invitará a una profesional del sector externo con experiencia en Finanzas Abiertas, quien asumirá el rol de cliente potencial. Esta persona participará en las sesiones de retrospectiva con el objetivo de validar las entregas, proporcionar retroalimentación sobre el diseño de la plataforma y asegurar la alineación con las necesidades del mercado.

**Enfoque hacia un MVP**

Como parte de la estrategia de desarrollo ágil, se priorizará la creación de un Producto Mínimo Viable (MVP) en el menor tiempo posible. Este MVP será una versión inicial funcional de la plataforma, susceptible de ser testada por usuarios reales, lo cual permitirá validar tempranamente los supuestos de diseño y orientar la evolución del producto hacia una versión potencialmente utilizable.

# Desarrollo de la plataforma

## Diseño Funcional

Se ha identificado la necesidad de desarrollar una plataforma del tipo middleware que permita a usuarios no expertos realizar el entrenamiento de un modelo de RAC/RAG, para que con este modelo se pueda habilitar una interfaz de conversación que permita a otros usuarios realizar interacciones con el modelo de forma controlada.

### Flujo Funcional de la plataforma para entrenamiento RAC/RAG

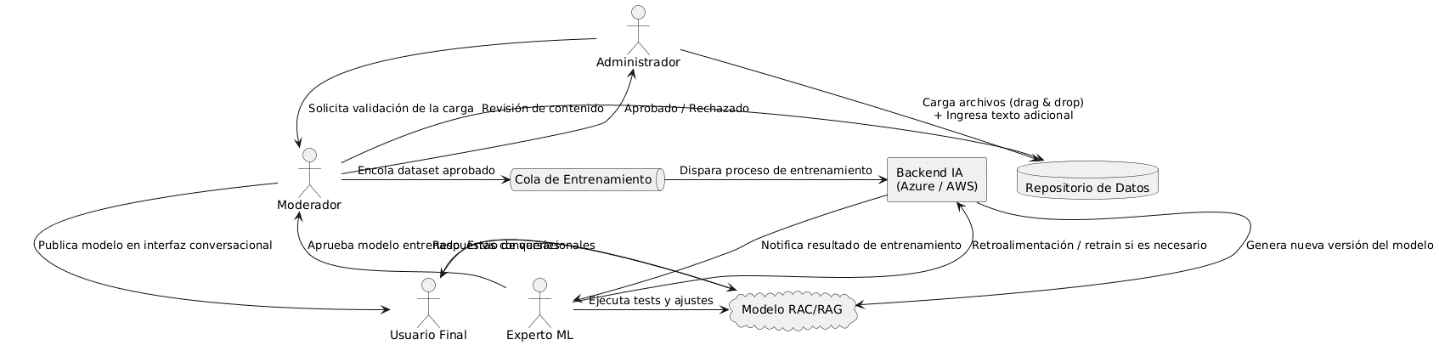


Ilustración 1 - Flujo funcional plataforma

El flujo funcional en la (Ilustración 1) proporciona una visión macro de los procesos que conforman la plataforma y de los actores que intervienen en cada fase. Se distinguen cuatro roles principales—Administrador, Moderador, Experto en ML y Usuario final—y cinco etapas ordenadas de forma lógica y dependiente:

#### Ingesta y preparación de datos

El Administrador incorpora información mediante un mecanismo drag-and-drop para ficheros y un editor de texto para entradas manuales. Ambos métodos convergen en un repositorio de almacenamiento de objetos que preserva la trazabilidad y la versión de cada recurso.

#### Moderación y filtro de calidad

Antes de ser empleada en el entrenamiento, la información es validada por el Moderador. Este control de calidad garantiza la relevancia semántica y la conformidad legal (p. ej., protección de datos personales), evitando sesgos y ruido temprano en el pipeline.

#### Entrenamiento del modelo RAC/RAG

La aprobación del Moderador activa la encolación automática de un job en la cola de entrenamiento. Dicha cola desacopla el plano transaccional del entrenamiento intensivo, delegando la ejecución a servicios gestionados de IA (Azure ML, SageMaker u otros) capaces de aprovisionar GPU bajo demanda y de exponer endpoints de métrica.

#### Evaluación iterativa.

El Experto en ML inspecciona las métricas obtenidas (exactitud, F1, latency, entre otras) y decide, con base en criterios cuantitativos y cualitativos, si se requiere re-entrenamiento, ajuste de hiperparámetros o depuración del corpus. Este ciclo garantiza la convergencia del modelo hacia los objetivos establecidos por la organización.

#### Publicación y explotación.

Una vez que el modelo cumple los umbrales de desempeño, se promueve al estado Production dentro del registro de artefactos. Simultáneamente, la plataforma habilita una interfaz conversacional para el Usuario final, protegida por controles de acceso y sistemas de monitoreo que registran telemetría, alertas y métricas de uso en tiempo real.

En el flujo podemos ver la separación de responsabilidades y la adopción de principios de DevOps y MLOps, lo que facilitará la escalabilidad operativa y de los módulos de desarrollo, así como la gobernanza de los datos y de los ciclos de life-cycle management del modelo.

### Diagrama de secuencia de alto nivel

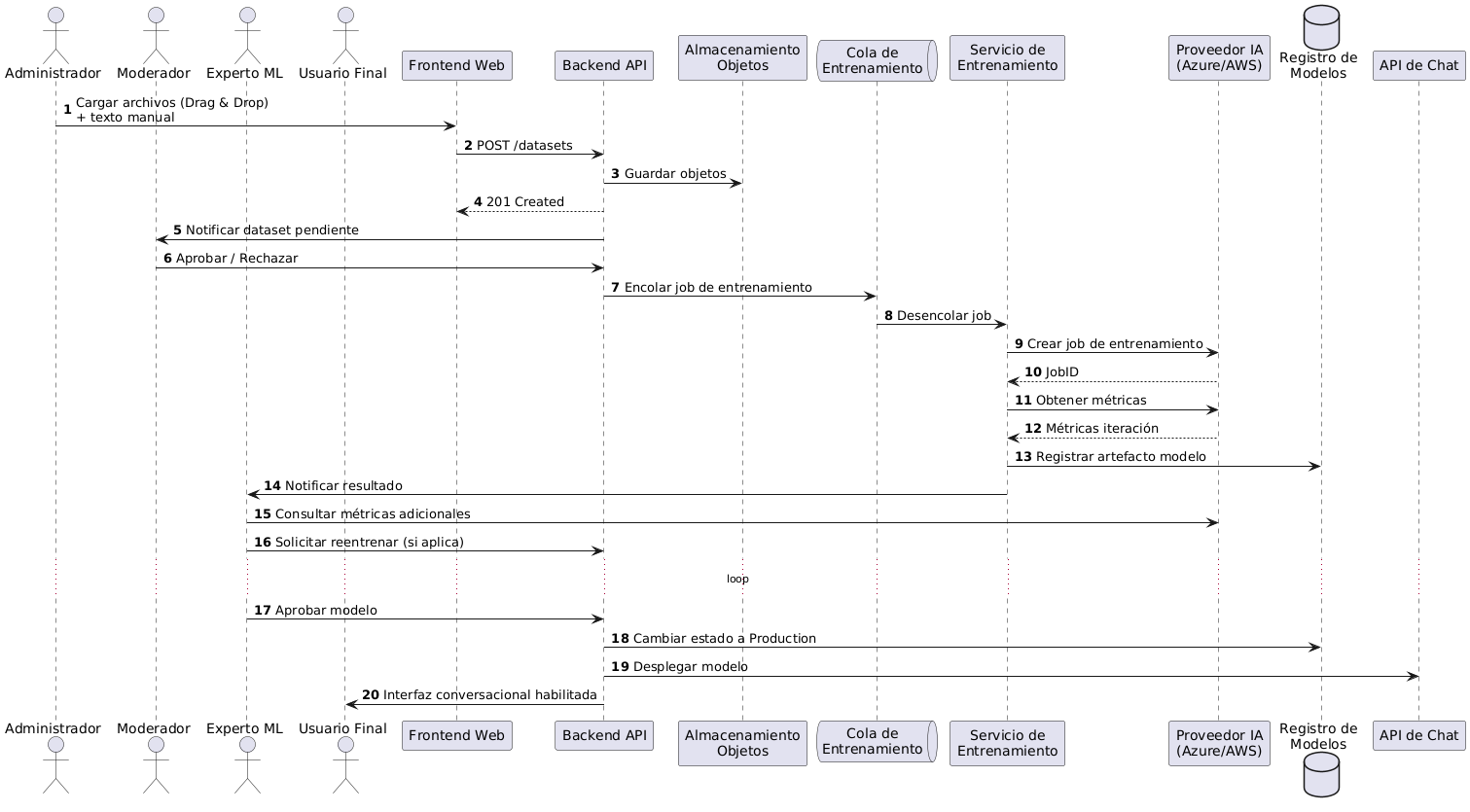


Ilustración 2 - Flujo de secuencia Plataforma

El diagrama de secuencia (Ilustración 2) profundiza en la dinámica temporal de los intercambios entre actores y componentes lógicos de la plataforma. Cada mensaje representa una llamada, notificación o transferencia de estado que ocurre en tiempo de ejecución. A continuación vamos a detallar los siguientes puntos críticos:

#### Desacoplamiento mediante cola de entrenamiento

El Backend API registra datasets y, tras la aprobación del Moderador, inserta un mensaje en la cola (message-driven architecture). Ello evita el bloqueo de recursos de la capa web y permite escalar trabajadores de entrenamiento de manera horizontal.

#### Delegación de cómputo intensivo al proveedor en la nube

El Servicio de Entrenamiento interactúa con el proveedor IA para crear ejecuciones, consultar métricas y recuperar identificadores de trabajo, manteniendo un registro de auditoría en el almacenamiento interno.

#### Gestión de artefactos y versionado

Tras cada iteración, el Servicio de Entrenamiento registra el modelo resultante en el Registro de Modelos con etiquetas de versión y metadatos descriptivos, habilitando la reproducibilidad y la reversión en caso de degradación.

#### Ciclo de validación controlado por el Experto en ML

El Experto puede solicitar nuevas iteraciones vía la API, cerrando el bucle de feedback hasta alcanzar la calidad deseada.

#### Despliegue continuo y activación de la API de Chat

La aprobación final dispara un despliegue automático que expone el modelo a la API conversacional. El Backend confirma la disponibilidad de la interfaz al Usuario final y habilita los mecanismos de observabilidad pertinentes (logs, métricas de uso y alertas).

El flujo tiene como objetivo delimitar los principios de responsabilidad única, event-driven processing y observabilidad integral, que combinan machine learning y servicios de conversación en producción.

### Wireframes

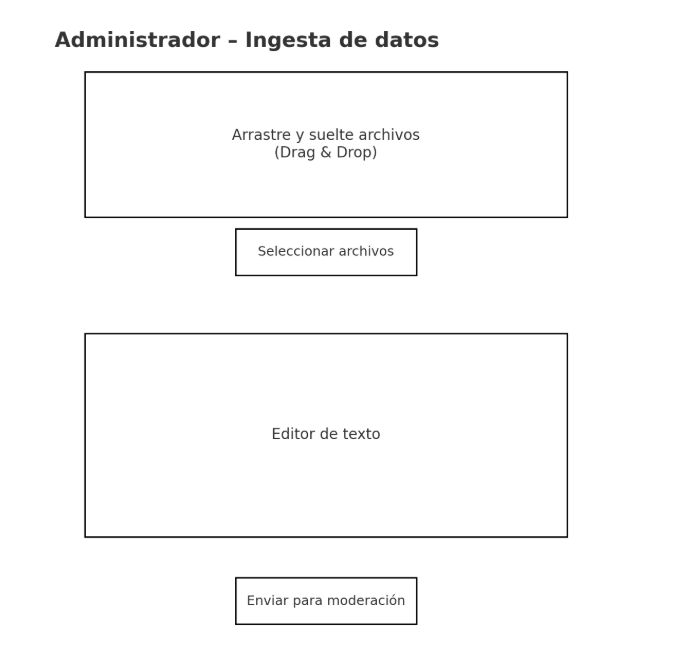


Ilustración 3 - Módulo Administrador - Ingesta de datos



Ilustración 4 - Módulo Administrador - Módulo Moderador

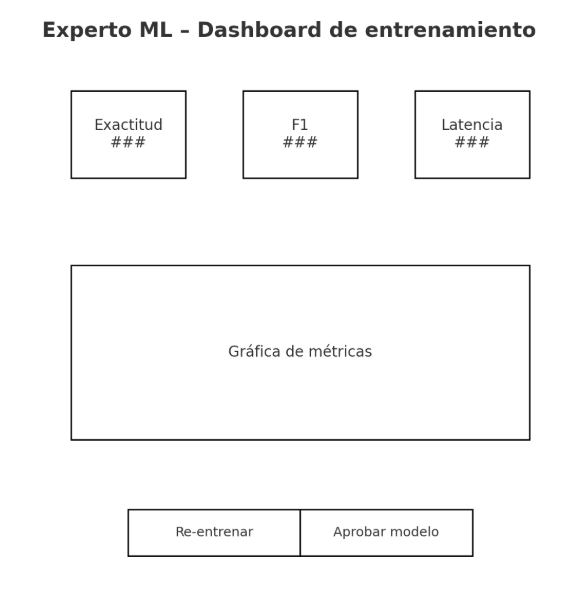


Ilustración 5 - Módulo Dashboard

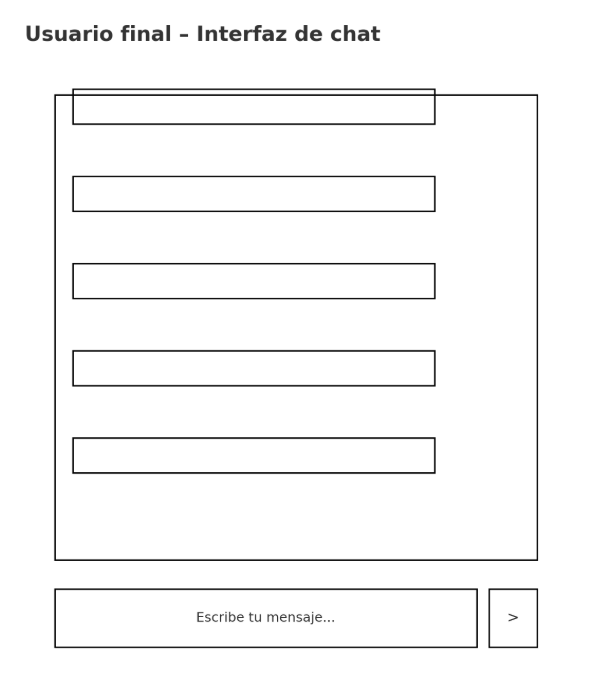


Ilustración 6 - Módulo Interfaz usuario final

## Levantamiento de requisitos

Este tipo de trabajo es similar al tipo anterior, partiendo de la identificación detallada de requisitos y describiendo a fondo el modelo de proceso de desarrollo de software empleado, con la descripción completa de todas las actividades del proceso, pero sin desarrollar su implementación.

### Identificación de requisitos

En este apartado se debe indicar el trabajo previo realizado para guiar el desarrollo del programa. Esto debería incluir la identificación adecuada del problema a tratar, así como del contexto habitual de uso (empresa, institución, etc.). Idealmente, la identificación de requisitos se debería hacer contando con expertos en la materia a tratar. Además, deberás describir las características del proyecto de desarrollo de software. Como mínimo querrás mencionar:

* Qué tecnologías se utilizaron en el proceso de desarrollo (incluyendo justificación de por qué se emplearon y descripciones detalladas de las mismas).
* Cómo se organizó el proceso de desarrollo.
* Qué personas participaron (con datos demográficos, si procede) o qué técnicas de sistemas se emplearon.
* Cómo se planificó el proceso de desarrollo.
* Qué instrumentos de seguimiento y evaluación se utilizaron durante el proceso de desarrollo.

### Descripción de la planificación del proyecto de desarrollo de software

En el caso de una planificación de proyecto de desarrollo de software, deberían aportarse detalles del modelo de proceso de desarrollo, incluyendo las fases e hitos del proceso. También deben incluirse un estudio del dominio, un análisis y modelado de los requisitos del software y un modelado del diseño del software. Se recomienda el uso de diagramas explicativos de la arquitectura o funcionamiento. Al fin y al cabo, en el proyecto tiene que quedar definido todo lo que se va a realizar durante el ciclo de desarrollo del software. Por lo que es importante también describir cómo se va a implementar el software, cómo se va a testear y cómo se va a validar.

## Tipo 3. Desarrollo de metodología

# Conclusiones y trabajo futuro

## Conclusiones

Este último apartado es habitual en todos los tipos de trabajos y presenta el resumen final de tu trabajo y debe servir para informar del alcance y relevancia de tu aportación.

Suele estructurarse empezando con un resumen del problema tratado, de cómo se ha abordado y de por qué la solución sería válida.

Es recomendable que incluya también un resumen de las contribuciones del trabajo, en el que relaciones las contribuciones y los resultados obtenidos con los objetivos que habías planteado para el trabajo, discutiendo hasta qué punto has conseguido resolver los objetivos planteados. Las conclusiones ofrecidas deberán ser consecuencia del trabajo realizado y, por lo tanto, deberán marcar el grado de consecución de los objetivos propuestos (cada objetivo del trabajo se enlazará con una conclusión).

## Trabajo futuro

Finalmente, se suele dedicar un último apartado a hablar de líneas de trabajo futuro que podrían aportar valor añadido al trabajo realizado. La sección debería señalar las perspectivas de futuro que abre el trabajo desarrollado para el campo de estudio definido. En el fondo, debes justificar de qué modo puede emplearse la aportación que has desarrollado y en qué campos.

Referencias bibliográficas

# Bibliografía

Arner, D. W., Barberis, J. N., & Buckley, R. P. (2020). FinTech, RegTech and the Reconceptualization of Financial Regulation. *Northwestern Journal of International Law & Business*, 371-413.

Arvanitakis, D., & Moustakides, G. V. (2020). Adaptive Anonymization for Privacy-Preserving Big Data Analytics. *IEEE Transactions on Information Forensics and Security*, 3716-3728.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., . . . Askell, A. (2020). *Advances in Neural Information Processing Systems 33.* Red Hook, NY: Curran Associates, Inc.

Comisión Europea. (Mayo de 2016). *Reglamento (UE) 2016/679 del Parlamento Europeo.* Obtenido de https://ec.europa.eu/info/law/law-topic/data-protection\_en

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 273-297.

Dwork, C. (2008). Differential Privacy: A Survey of Results. *International Conference on Theory and Applications of Models of Computation* (págs. 1-17). Lecture Notes in Computer Science .

Edemacu, K., & Wu, X. (2024). *Privacy Preserving Prompt Engineering: A Survey.* Arkansas.

European Union Agency for Cybersecurity (ENISA). (2021). *Threat Landscape for Artificial Intelligence.* Athens, Greece.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative adversarial nets. En *Advances in Neural Information Processing Systems 27* (págs. 2672 - 2680). Red Hook, NY: Curran Associates, Inc.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. En *Advances in Neural Information Processing Systems 33* (págs. 6840-6851). Red Hook, NY: Curran Associates, Inc.

Kairouz, P. (2020). *Advances and open problems in federated learning.* Foundations and Trends in Machine Learning.

Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes.* Ithaca, NY: arXiv.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Advances in Neural Information Processing Systems 25.* Red Hook, NY: Curran Associates, Inc.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., . . . Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. En *Advances in Neural Information Processing Systems 33* (págs. 9459-9474). Red Hook, NY: Curran Associates, Inc.

McKinsey & Company. (2023). *The economic potential of generative AI: The next productivity frontier.*

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space.* Ithaca, NY: arXiv.

Nunes, J. M., & Abreu, R. (2022). Designing User-Centric ML-Based Systems: A Survey of Best Practices. *IEEE Access*, 10391-10405.

OpenIA. (30 de Noviembre de 2022). *OpenIA*. Obtenido de https://openai.com/es-ES/index/chatgpt/

Patrick Lewis, E. P.-t. (2020). Retrieval-Augmented Generation for knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models Are Unsupervised Multitask Learners.* San Francisco, CA: OpenAI Technical Report.

Ren, X., Zhang, S., Hong, Y., & He, J. (2022). Privacy-Preserving Techniques in Natural Language Processing: A Survey. *IEEE Access*, 10427-10441.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 386-408.

Sfetcu, N. (2024). Threats of Artificial Intelligence for Cybersecurity. *IT & C*, Volumen 3 - numeral 3.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention Is All You Need. En *Advances in Neural Information Processing Systems 30* (págs. 5998-6008). Red Hook, NY: Curran Associates, Inc.

1. Artículo

Como requisito adicional y obligatorio en el anexo A se deberá incluir un artículo de investigación que resuma el trabajo realizado y los principales resultados obtenidos. Debe incluir lo más importante de cada capítulo y no debe ser una copia literal de lo expuesto en la memoria.

Este artículo, que puede ser en español o en inglés, deberá seguir la plantilla proporcionada a continuación y tendrá una extensión de entre 6 y 8 páginas.

Título

Nombre y Apellidos del Estudiante

Universidad Internacional de la Rioja, Logroño (España)

Fecha

Palabras Clave

Tres a cinco palabras clave ordenadas alfabéticamente y separadas por comas.

Resumen

Breve resumen del trabajo realizado (extensión máxima: 150 palabras). Este resumen debe incluir el objetivo o propósito de la investigación, la metodología, los resultados y las conclusiones.

I. Introducción

I

intoducción en la que debes resumir de forma esquemática pero suficientemente clara lo esencial de cada una de las partes del trabajo.

La lectura de esta introducción ha de dar una primera idea clara de lo que se pretendía, las conclusiones a las que se ha llegado y del procedimiento seguido.

II. Estado del Arte

Estudio a fondo del dominio de aplicación, citando numerosas referencias.

Debe aportar un buen resumen del conocimiento que ya existe en el campo de los problemas habituales identificados.

Numerar las citas de forma consecutiva entre corchetes [1].

III. Objetivos y Metodología

Objetivo general, objetivos específicos y metodología de trabajo aplicada.

IV. Contribución

Desarrollar la descripción de tu contribución.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

V. Resultados

Descripción de los resultados obtenidos.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

*Resultados 1*

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.



Fig. 1. Magnetization as a function of applied field. Note that “Fig.” is abbreviated. There is a period after the figure number, followed by two spaces. It is good practice to explain the significance of the figure in the caption.

En la Figura 1…

En la Tabla I …

Tabla I

Units for Magnetic Properties

|  |  |  |
| --- | --- | --- |
| Symbol | Quantity | Conversion from Gaussian and CGS EMU to SI a |
| F | magnetic flux | 1 Mx ® 10-8 Wb = 10-8 V·s |
| B | magnetic flux density,  magnetic induction | 1 G ® 10-4 T = 10-4 Wb/m2 |
| H | magnetic field strength | 1 Oe ® 103/(4p) A/m |
| m | magnetic moment | 1 erg/G = 1 emu  ® 10-3 A·m2 = 10-3 J/T |
| M | magnetization | 1 erg/(G·cm3) = 1 emu/cm3  ® 103 A/m |
| 4pM | magnetization | 1 G ® 103/(4p) A/m |
| s | specific magnetization | 1 erg/(G·g) = 1 emu/g ® 1 A·m2/kg |
| j | magnetic dipole  moment | 1 erg/G = 1 emu  ® 4p ´ 10-10 Wb·m |
| J | magnetic polarization | 1 erg/(G·cm3) = 1 emu/cm3  ® 4p ´ 10-4 T |
| c*,* k | susceptibility | 1 ® 4p |
| cr | mass susceptibility | 1 cm3/g ® 4p ´ 10-3 m3/kg |
| m | permeability | 1 ® 4p ´ 10-7 H/m  = 4p ´ 10-7 Wb/(A·m) |
| mr | relative permeability | m ® mr |
| w, W | energy density | 1 erg/cm3 ® 10-1 J/m3 |
| N, D | demagnetizing factor | 1 ® 1/(4p) |

Vertical lines are optional in tables. Statements that serve as captions for the entire table do not need footnote letters.

aGaussian units are the same as cgs emu for magnetostatics; Mx = maxwell, G = gauss, Oe = oersted; Wb = weber, V = volt, s = second, T = tesla, m = meter, A = ampere, J = joule, kg = kilogram, H = henry.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

*Resultados 2*

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

VI. Discusión

Tras la presentación objetiva de los resultados, querrás aportar una discusión de los mismos.

VII. Conclusiones

Resumen de las contribuciones del trabajo, en el que relaciones las contribuciones y los resultados obtenidos con los objetivos que habías planteado para el trabajo, discutiendo hasta qué punto has conseguido resolver los objetivos planteados.

Finalmente, hablar de líneas de trabajo futuro que podrían aportar valor añadido al TFM realizado. La sección debería señalar las perspectivas de futuro que abre el trabajo desarrollado para el campo de estudio definido. En el fondo, debes justificar de qué modo puede emplearse la aportación que has desarrollado y en qué campos.

Apéndices

Apéndices, en caso de ser necesario.

Referencias

1. Bermejo, J., Abad, C., Bermejo, J. R., Sicilia, M. A. & Sicilia, J. A. (2020). A Systematic Approach to Malware Analysis (SAMA). *Applied Sciences*, 10(4), 1360.
2. De Vicente, J., Bermejo, J., Bermejo, J. R. & Sicilia, J. A. (2019). The Application of a New Secure Software Development Life Cycle (S-SDLC) with Agile Methodologies. *Electronics*, 8(11), 1218.
3. Sicilia, J. A., Quemada, C., Royo, B. & Escuín, D. (2016). An optimization algorithm for solving the rich vehicle routing problem based on Variable Neighborhood Search and Tabu Search metaheuristics. Jo*urnal of Computational and Applied Mathematics*, 291, pp. 468-477.

1. Título anexo

El resto de los anexos debe de recoger todo aquello que puede ser interesante para el trabajo pero que no es estrictamente esencial, y que distraería la lectura si se colocara en el cuerpo: encuestas, resultados de pilotos, documentos adicionales, capturas de pantalla, código, etc. Pueden incluirse los anexos que se consideren necesarios. Estos no computarán a efectos de extensión del trabajo. Cada parte adicional se numera como un anexo y se ofrece en una página diferente.