



Universidad Politécnica
de Madrid

**Escuela Técnica Superior de
Ingenieros Informáticos**



Grado en Matemáticas e Informática

Trabajo Fin de Grado

**Análisis y Diseño de un Sistema RAG
Multiagente para Soporte a la Decisión
Clínica**

Autor: Iván Hernández Pérez
Tutor(a): Ernestina Menasalvas Ruiz

Madrid, Enero 2026

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Matemáticas e Informática

Título: Análisis y Diseño de un Sistema RAG Multiagente para Soporte
a la Decisión Clínica

Enero 2026

Autor: Iván Hernández Pérez

Tutor: Ernestina Menasalvas Ruiz

Departamento de Lenguajes y Sistemas Informáticos e Ingeniería del Software

Escuela Técnica Superior de Ingenieros Informáticos

Universidad Politécnica de Madrid

Agradecimientos

Antes de dar comienzo a la introducción del trabajo, quiero mostrar mi más profundo agradecimiento a todas las personas que me han hecho llegar aquí.

En primer lugar, gracias a Ernestina Menasalvas, la persona que me ha guiado de forma más directa en la consecución de este trabajo que pone punto y final a una importantísima etapa de mi vida. Estoy realmente agradecido por haberme facilitado la realización de este trabajo y por haber contribuido a mi desarrollo en un ámbito tan enriquecedor como la tecnología biomédica.

También quiero dar gracias a todo el personal docente que me ha acompañado a lo largo de estos años en la escuela y que me ha transmitido todo su conocimiento, haciendo en gran parte posible la realización de este trabajo. Todos han aportado su granito de arena y han contribuido en mi formación tanto académica y profesional, como personal.

Por último, gracias a mis padres por los consejos, por su ejemplo inspirador y por su respaldo; y a mi hermana por el cariño, las risas y el apoyo. Gracias a mi familia y a todo mi entorno más personal por el apoyo incondicional mostrado durante los últimos meses y durante toda la carrera. Estoy profundamente agradecido, y sin todos ellos esta experiencia no habría sido la misma.

Resumen

La creciente incorporación de sistemas de inteligencia artificial en el ámbito sanitario abre nuevas oportunidades para mejorar la eficiencia, la seguridad y la calidad de los procesos asistenciales, especialmente en contextos caracterizados por una elevada demanda, recursos clínicos limitados y una ascendente presión sobre los profesionales sanitarios. En este escenario, uno de los principales retos es la gestión y optimización de las listas de espera, donde una asignación ineficiente de prioridades puede derivar tanto en retrasos clínicamente relevantes como en una sobrecarga adicional de trabajo para el personal médico. En este contexto, el presente Trabajo Fin de Grado aborda el análisis y diseño de un sistema inteligente orientado al soporte a la decisión clínica durante la anamnesis inicial en el ámbito de la reumatología, con el objetivo principal de asistir en la priorización de los casos clínicos para contribuir a la optimización de las listas de espera y, de forma complementaria, automatizar la fase de anamnesis con el fin de optimizar el tiempo efectivo de consulta, y reducir la carga administrativa y la presión asistencial sobre los profesionales sanitarios, favoreciendo indirectamente también una mejora global de los tiempos de espera.

El trabajo propone una arquitectura híbrida basada en Generación por Recuperación Aumentada (RAG) y un Sistema Multiagente (MAS), estructurada en tres fases diferenciadas. En una primera fase de preparación del conocimiento externo, la información clínica de referencia se somete a un proceso de fragmentación semántica, vectorización e inserción en una base de datos vectorial, permitiendo su indexación eficiente. En una segunda fase de recuperación, el sistema identifica y extrae de dicha base el conocimiento más relevante en función de la consulta clínica recibida. Finalmente, en una tercera fase de generación y razonamiento, el sistema automatiza la estructuración de la anamnesis a partir del relato del paciente y, apoyándose en la información recuperada, detecta señales de alarma clínicas y asigna una prioridad asistencial adecuada, generando notas clínicas homogéneas y estructuradas que pueden integrarse directamente en los sistemas de información sanitaria.

Desde el punto de vista metodológico, el trabajo realiza una revisión exhaustiva del estado del arte en modelos de lenguaje y su aplicación en entornos clínicos, así como de los fundamentos teóricos de los sistemas RAG y los enfoques multiagente. Sobre esta base, se diseña e implementa un prototipo funcional que integra ambos paradigmas, justificando las decisiones técnicas adoptadas y delimitando explícitamente el alcance del sistema como una herramienta de apoyo

al juicio clínico, orientada a automatizar tareas administrativas y de documentación sin sustituir la valoración médica.

La evaluación del sistema se lleva a cabo mediante un conjunto de 21 escenarios clínicos reumatológicos simulados, empleando métricas técnicas y criterios cualitativos orientados a analizar la relevancia del contexto recuperado, la coherencia de la información generada y la consistencia en la asignación de prioridades asistenciales. Sobre la muestra evaluada, el sistema alcanza un error absoluto medio (MAE) igual a 0,62 y un error cuadrático medio (RMSE) de 1,1 en la comparación entre la prioridad asignada por el sistema y la proporcionada por el experto clínico, lo que evidencia una concordancia adecuada en el contexto experimental planteado y respalda la viabilidad del enfoque propuesto.

Finalmente, el trabajo discute sobre las limitaciones del sistema y las implicaciones de la integración de modelos de lenguaje generativos en entornos clínicos reales, subrayando la importancia de la supervisión humana, la colaboración interdisciplinar y la investigación continua. El prototipo desarrollado no se plantea como un producto final, sino como un punto de partida para futuras líneas de investigación orientadas a una adopción responsable y sostenible de la inteligencia artificial en la práctica clínica, con el objetivo último de mejorar la eficiencia del sistema sanitario, la equidad en el acceso a la atención y las condiciones de trabajo de los profesionales de la salud.

Abstract

The increasing adoption of artificial intelligence systems in the healthcare domain opens up new opportunities to improve the efficiency, safety, and quality of care processes, especially in contexts characterized by high demand, limited clinical resources, and growing pressure on healthcare professionals. In this scenario, one of the main challenges is the management and optimization of waiting lists, where inefficient prioritization can lead to clinically relevant delays as well as additional working overload for medical staff. Within this context, this Bachelor's Thesis addresses the analysis and design of an intelligent system oriented toward clinical decision support during the initial anamnesis in the rheumatology field, with the main goal of assisting in the prioritization of clinical cases to contribute to the optimization of waiting lists and, complementarily, automating the anamnesis phase in order to optimize effective consultation time, and reduce administrative burden and pressure on healthcare professionals, indirectly favoring too an overall improvement in waiting times.

The work proposes a hybrid architecture based on Retrieval-Augmented Generation (RAG) and a Multi-Agent System (MAS), structured into three differentiated phases. In a first phase of external knowledge preparation, reference clinical information is subjected to a process of semantic fragmentation, vectorization, and insertion into a vector database, enabling efficient indexing. In a second retrieval phase, the system identifies and extracts from this database the most relevant knowledge according to the received clinical query. Finally, in a third generation and reasoning phase, the system automates the structuring of the anamnesis based on the patient's narrative and, relying on the retrieved information, detects clinical alarm signals and assigns an appropriate care priority, generating homogeneous and structured clinical notes that can be directly integrated into healthcare information systems.

From a methodological point of view, the work carries out an exhaustive review of the state of the art in language models and their application in clinical environments, as well as of the theoretical foundations of RAG systems and multi-agent approaches. On this basis, a functional prototype is designed and implemented that integrates both paradigms, justifying the technical decisions adopted and explicitly delimiting the scope of the system as a tool to support clinical judgment, aimed at automating administrative and documentation tasks without replacing medical assessment.

The evaluation of the system is carried out using a set of 21 simulated rheuma-

tological clinical scenarios, employing technical metrics and qualitative criteria aimed at analyzing the relevance of the retrieved context, the coherence of the generated information, and the consistency in the assignment of care priorities. In the evaluated sample, the system achieves a mean absolute error (MAE) equal to 0.62 and a root mean squared error (RMSE) of 1.1 when comparing the priority assigned by the system with that provided by the clinical expert, demonstrating an adequate agreement within the proposed experimental context and supporting the viability of the proposed approach.

Finally, the work discusses the limitations of the system and the implications of integrating generative language models into real clinical environments, highlighting the importance of human supervision, interdisciplinary collaboration, and continuous research. The developed prototype is not conceived as a final product, but rather as a starting point for future research lines oriented toward the responsible and sustainable adoption of artificial intelligence in clinical practice, with the ultimate goal of improving the efficiency of the healthcare system, equity in access to care, and the working conditions of healthcare professionals.

Tabla de contenidos

1. Introducción y Objetivos	1
1.1. Introducción y motivación	1
1.2. Objetivos	6
1.3. Estructura del trabajo	7
2. Estado del arte y fundamentos teóricos	9
2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario	10
2.1.1. Concepto y evolución	10
2.1.2. Aplicaciones generales en medicina y sanidad	15
2.1.3. Sistemas de apoyo a la documentación clínica: AI Medical Scribes	17
2.2. Representación semántica y recuperación de información clínica	22
2.2.1. Embeddings semánticos para la representación de lenguaje natural	22
2.2.2. Recuperación semántica en dominios clínicos	24
2.3. RAG en entornos clínicos	25
2.3.1. Preparación de la información externa: segmentación (<i>chunking</i>)	25
2.3.2. Arquitectura y componentes de los sistemas RAG	26
2.3.3. Flujo de funcionamiento de una arquitectura RAG	27
2.3.4. Ventajas y aplicaciones del enfoque RAG en el ámbito clínico	28
2.4. Sistemas multiagente en entornos clínicos (MAS)	30
2.4.1. Definición, fundamentos y evolución	30
2.4.2. Sistemas multiagente basados en modelos de lenguaje (LLM-MAS)	31
2.4.3. Aplicaciones en el ámbito sanitario	33
2.5. Reumatología como dominio clínico de referencia	34
2.5.1. Complejidad clínica y proceso asistencial en reumatología	34
2.5.2. Guías clínicas ACR y EULAR como conocimiento estructurado	35
2.5.3. Aplicaciones de la inteligencia artificial en reumatología	35
3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema	37
3.1. Visión general del planteamiento de la solución	38
3.2. Análisis de requisitos	39
3.2.1. Requisitos funcionales	39

TABLA DE CONTENIDOS

3.2.2. Requisitos no funcionales	40
3.2.3. Cumplimiento de los requisitos	41
3.3. Arquitectura general del sistema	42
3.4. Preparación del conocimiento externo y generación de embeddings	44
3.5. Recuperación de información clínica durante la ejecución del sistema	46
3.6. Sistema multiagente: agentes y roles	48
3.6.1. Organización de los agentes dentro del flujo de ejecución . .	49
3.6.2. Módulo extractor de parámetros clínicos	51
3.6.3. Descripción de los agentes y sus roles	51
3.6.4. Orquestación y control del razonamiento	54
3.7. Flujo completo de ejecución del sistema	55
3.7.1. Recepción de la consulta y extracción de parámetros clínicos	56
3.7.2. Generación del resumen orientado a recuperación y recuperación de conocimiento externo	56
3.7.3. Integración del contexto clínico y procesamiento multiagente posterior	57
3.7.4. Normalización, construcción de la historia clínica y generación de la salida final	58
3.8. Tecnologías y herramientas empleadas	59
3.8.1. Lenguaje de programación y gestión del proyecto	59
3.8.2. Modelo de lenguaje generativo empleado	60
3.8.3. Segmentación semántica y generación de <i>embeddings</i>	60
3.8.4. Base de datos vectorial	61
3.8.5. Gestión de salidas y trazabilidad	62
4. Experimentos, evaluación y discusión	63
4.1. Introducción	63
4.2. Diseño experimental	64
4.2.1. Escenarios clínicos de evaluación	64
4.2.2. Desarrollo incremental y configuración del sistema evaluado	65
4.2.3. Protocolo de evaluación del sistema completo integrado . . .	67
4.3. Criterios y métricas de evaluación	67
4.3.1. Criterios clínicos cualitativos	67
4.3.2. Evaluación de la relevancia del contexto recuperado	68
4.3.3. Evaluación de la asignación de prioridad clínica	69
4.4. Resultados experimentales	69
4.4.1. Resultados de relevancia contextual	69
4.4.2. Resultados de asignación de prioridad clínica	70
4.4.3. Grado de cumplimiento de los requisitos definidos	70
4.5. Discusión	74
4.5.1. Limitaciones identificadas	75
5. Conclusiones y Trabajo Futuro	77
5.1. Conclusiones	77
6. Análisis de impacto	81
Bibliografía	87

Anexos	101
A. Recibo digital e informe de similitud de <i>Turnitin</i>	101

Índice de figuras

1.1. Porcentaje de empresas que usan tecnologías basadas en inteligencia artificial en la unión europea [1].	2
1.2. Porcentaje de empresas que usan tecnologías basadas en inteligencia artificial en la unión europea según el tamaño de la empresa [1].	3
1.3. Evolución de la inversión del sector privado en IA [3].	4
1.4. Evolución de la inversión del sector público en IA [3].	4
1.5. Evolución del uso de la IA y genAI en las empresas [3].	5
2.1. Evolución de los modelos de lenguaje [18].	11
2.2. Taxonomía de los modelos de lenguaje [18].	12
2.3. Estructura de un modelo neuronal de lenguaje [35].	14
2.4. Evolución de los LLMs en el ámbito sanitario [51].	16
2.5. Porcentaje de médicos estadounidenses que han sentido al menos un síntoma de burnout durante los últimos 15 años [58], [60].	19
2.6. Diagrama de arquitectura RAG estándar [82].	28
2.7. Estructuras de organización básicas de los sistemas MAS [89].	31
2.8. Arquitectura de organización holónica de un sistema MAS [87].	32
2.9. Cronología de los principales lanzamientos de modelos de lenguaje de gran escala (LLM) (2023 – 2025) [90].	33
3.1. Visión general del planteamiento conceptual de la solución y su papel como soporte en el proceso asistencial.	39
3.2. Diagrama de la arquitectura de componentes software propuesta.	43
3.3. Módulo generador de embeddings de datos externo.	44
3.4. Módulo de recuperación de los datos externos integrado con el módulo generador multiagente.	46
3.5. Arquitectura de organización del sistema multiagente.	50
3.6. Esquema interno del módulo extractor de parámetros.	54
3.7. Flujo de ejecución de procesamiento de una consulta.	56

Índice de Tablas

3.1. Agentes del módulo extractor de parámetros clínicos y variables extraídas	52
3.2. Agentes del sistema multiagente y sus roles funcionales	53
4.1. Ejemplos representativos de consultas clínicas en lenguaje natural empleadas en la evaluación	65
4.2. Grado de cumplimiento de los requisitos funcionales del sistema .	72
4.3. Grado de cumplimiento de los requisitos no funcionales del sistema	73

Capítulo 1

Introducción y Objetivos

1.1. Introducción y motivación

En los últimos años, la *Inteligencia Artificial (IA)* ha experimentado una rápida y significativa evolución, convirtiéndose en una herramienta fundamental de trabajo en numerosos sectores. Los avances en el aprendizaje profundo y, principalmente, en los modelos de lenguaje de gran escala (Large Language Models, LLMs) han permitido el desarrollo de sistemas capaces de procesar y generar texto con una fluidez antes impensable. Además, estos últimos han acelerado la adopción de la IA considerablemente; permitiendo a las empresas su uso para automatizar tareas cognitivas.

Véase la Figura 1.1, publicada por *Eurostat* en enero de 2025 [1].

Se puede observar el gran incremento que hubo entre 2023 y 2024 en el uso de estas tecnologías en las empresas de la unión europea. En 2023, un 8% de las empresas usaban tecnologías IA, mientras que en 2024, este dato ascendía a un 13,48%.

No obstante, este dato no parece alinearse plenamente con el notable auge que han experimentado estas tecnologías en el ámbito empresarial. Sin embargo, si se estudia este mismo dato desglosándolo por el tamaño de la empresa tal y como *Eurostat* define en [2], se aprecian porcentajes considerablemente más altos.

Véase la Figura 1.2, publicada en el mismo artículo de la figura anterior [1].

Los datos indican que más de un 41% de las grandes empresas reportaron usar tecnologías basadas en IA. Además, según el informe publicado por *Stanford Institute for Human-Centered Artificial Intelligence* en [3], la inversión mundial en IA sigue una tendencia alcista desde 2013 (ver Figuras 1.3 y 1.4). En base a estos datos y a la actualidad, se puede deducir que esta inversión ha seguido incrementándose en 2025, y que por tanto, la tendencia de crecimiento observada en el uso de sistemas de IA en empresas se ha mantenido intacta. Por otra parte, dentro del propio ecosistema de la inteligencia artificial, la llamada inteligencia artificial generativa, representada principalmente por los modelos de lenguaje de gran escala (LLMs), ha experimentado un auge especialmente pronunciado,

Capítulo 1. Introducción y Objetivos

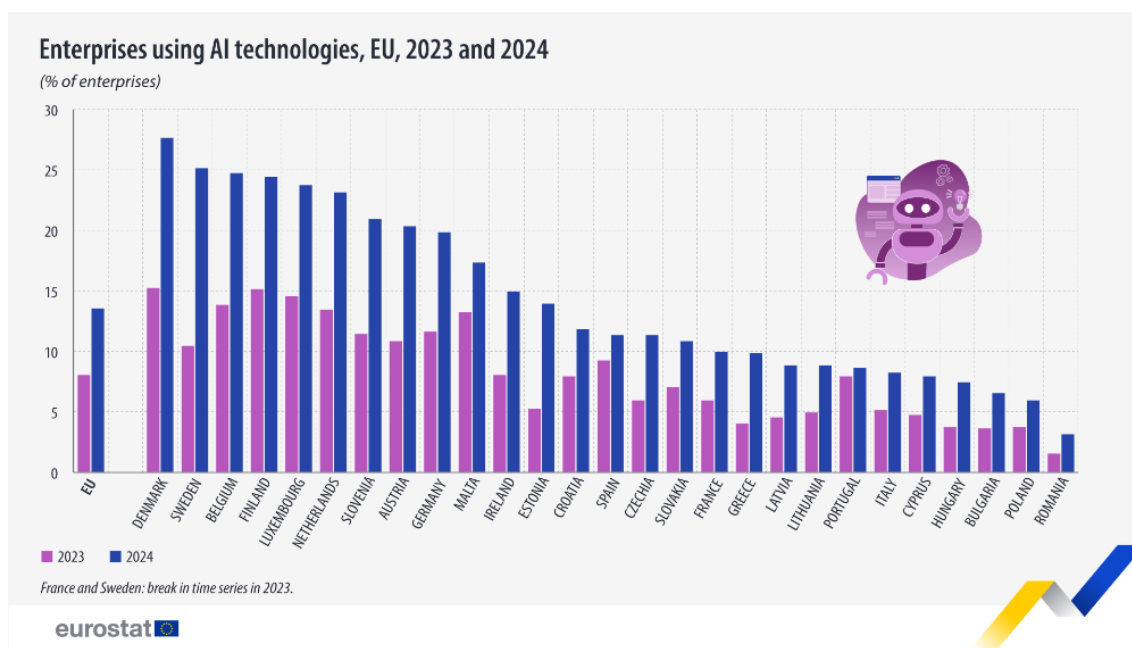


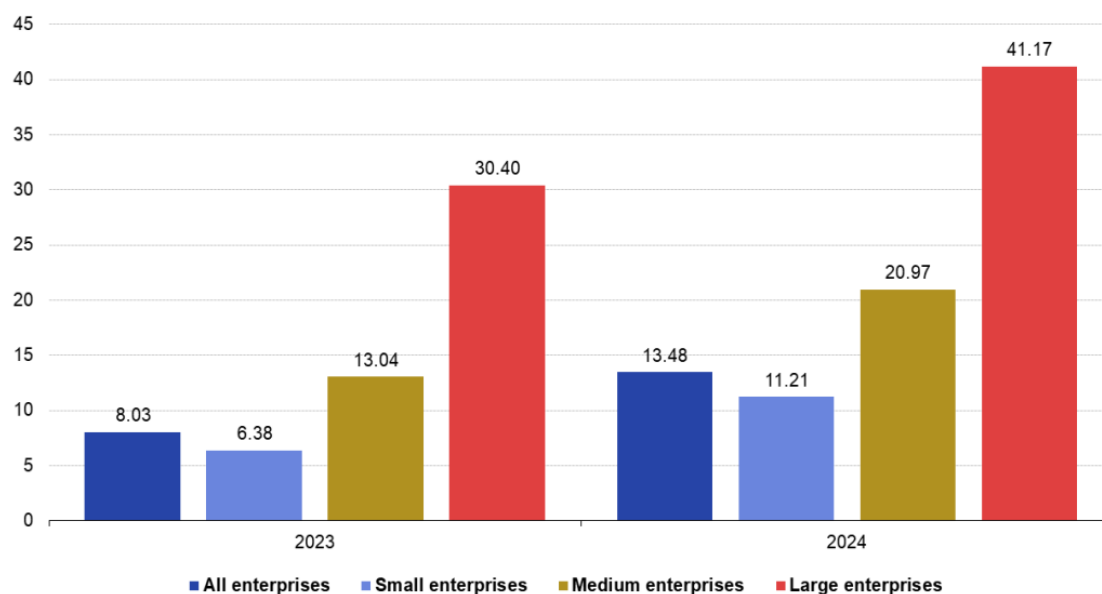
Figura 1.1: Porcentaje de empresas que usan tecnologías basadas en inteligencia artificial en la unión europea [1].

consolidándose como una de las áreas con mayor proyección y adopción a nivel corporativo (Véase la Figura 1.5).

Hoy en día, la IA está muy presente en diversos sectores como la automoción, finanzas o los recursos humanos [4], [5]. El impacto de esta expansión tecnológica también se ha hecho evidente en el ámbito sanitario, mas su adopción ha de realizarse con especial cautela debido a los requisitos de seguridad, privacidad, trazabilidad y riesgo regulatorio. La IA ha comenzado a desempeñar un papel clave en la optimización de procesos clínicos y en el apoyo a la toma de decisiones médicas. Desde sistemas de diagnóstico asistido por imagen hasta plataformas de atención virtual, la IA se ha consolidado como una herramienta para mejorar la eficiencia y la calidad del servicio sanitario [6], [7], [8]. La capacidad de los modelos de lenguaje para interpretar y generar información médica en lenguaje natural abre nuevas oportunidades en la comunicación entre pacientes y profesionales, como la redacción clínica, resumen de historias, búsqueda de evidencia, o la gestión y priorización de consultas.

En este contexto, la reumatología representa un área donde el potencial de estas tecnologías puede tener un impacto especialmente relevante. Esta especialidad médica, según el *Ministerio de Sanidad y Política Social* [9], se ocupa del estudio de las enfermedades reumáticas, también denominadas «enfermedades médicas del aparato locomotor» o «enfermedades musculoesqueléticas». Los profesionales del área, resaltan que las listas de espera en esta especialidad son excesivamente largas, lo que influye negativamente en la calidad y rapidez de la atención. Uno de los principales factores detectados en la práctica clínica que contribuyen

Enterprises using AI technologies by size class, EU, 2023 and 2024
(% of enterprises)



Source: Eurostat (online data code: isoc_eb_ai)

eurostat 

Figura 1.2: Porcentaje de empresas que usan tecnologías basadas en inteligencia artificial en la unión europea según el tamaño de la empresa [1].

a esta situación es la gran extensión y variabilidad de las consultas atendidas: los pacientes narran explicaciones muy extensas, y con frecuencia, estas incluyen información redundante o detalles poco relevantes, lo que dificulta al profesional identificar rápidamente los elementos clave para orientar el caso. Según indican los médicos especialistas, la valoración inicial se basa únicamente en cinco parámetros fundamentales, que permiten caracterizar el estado clínico del paciente, y en la detección de un conjunto limitado de síntomas de alarma que determinan la urgencia o el tipo de actuación requerida. La falta de una estructura homogénea en las consultas y el exceso de información no esencial genera un retraso en la respuesta asistencial.

En este sentido, la anamnesis de calidad se considera un elemento decisivo en el diagnóstico reumatológico. De hecho, los coordinadores de la Escuela SER de Anatomía e Imagen afirman que *“la mayoría de las enfermedades musculoesqueléticas de alta prevalencia pueden resolverse mediante una correcta anamnesis y exploración”* [10]. Sin embargo, en la práctica actual existen barreras, además de la mencionada en el párrafo anterior, que dificultan este proceso: la dispersión de información en distintos canales, la variabilidad en los estilos de redacción de las notas clínicas, el tiempo restringido de atención directa, y la carga de tareas administrativas asociadas a la gestión clínica.

Ante este panorama, surge la motivación de este trabajo: ¿pueden los sistemas

Capítulo 1. Introducción y Objetivos

Global private investment in AI by geographic area, 2013–24

Source: Quid, 2024 | Chart: 2025 AI Index report

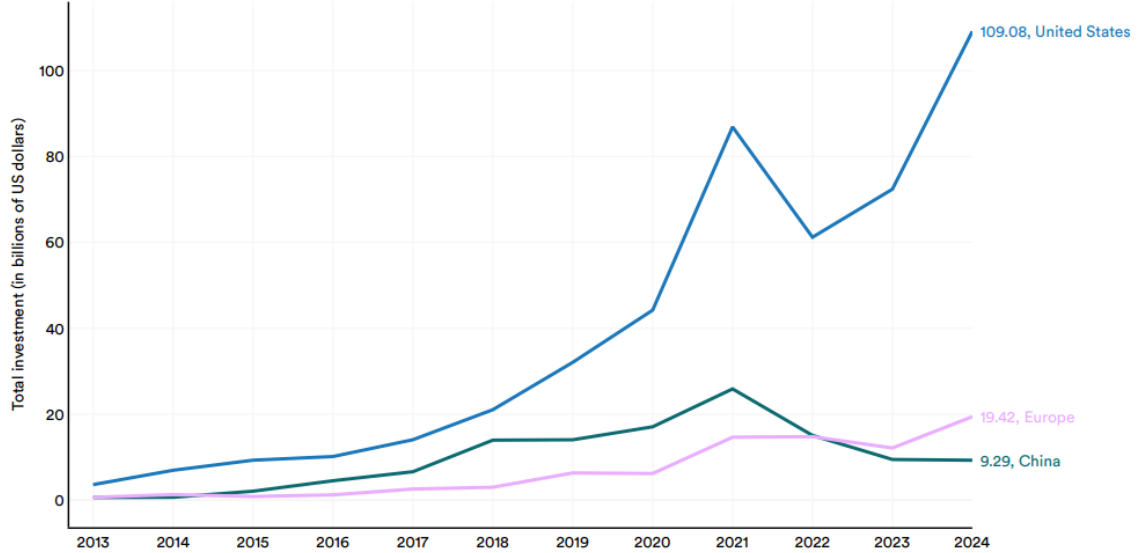


Figura 1.3: Evolución de la inversión del sector privado en IA [3].

Public spending on AI-related contracts in the United States and Europe, 2013–23

Source: AI Index, 2025 | Chart: 2025 AI Index report

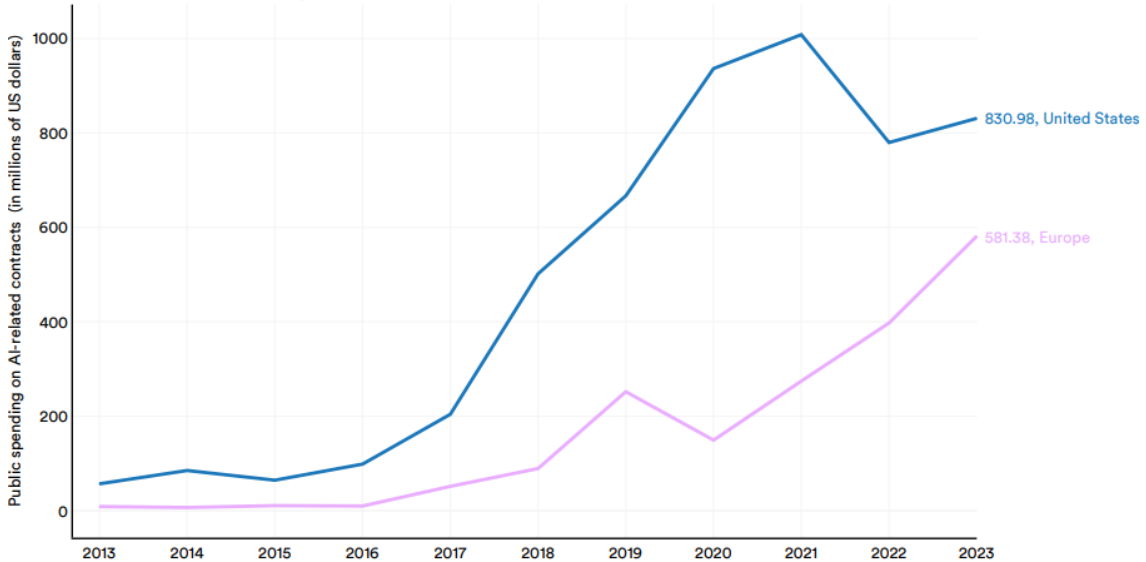


Figura 1.4: Evolución de la inversión del sector público en IA [3].

1.1. Introducción y motivación

Share of respondents who say their organization uses AI in at least one function, 2017–24

Source: McKinsey & Company Survey, 2024 | Chart: 2025 AI Index report

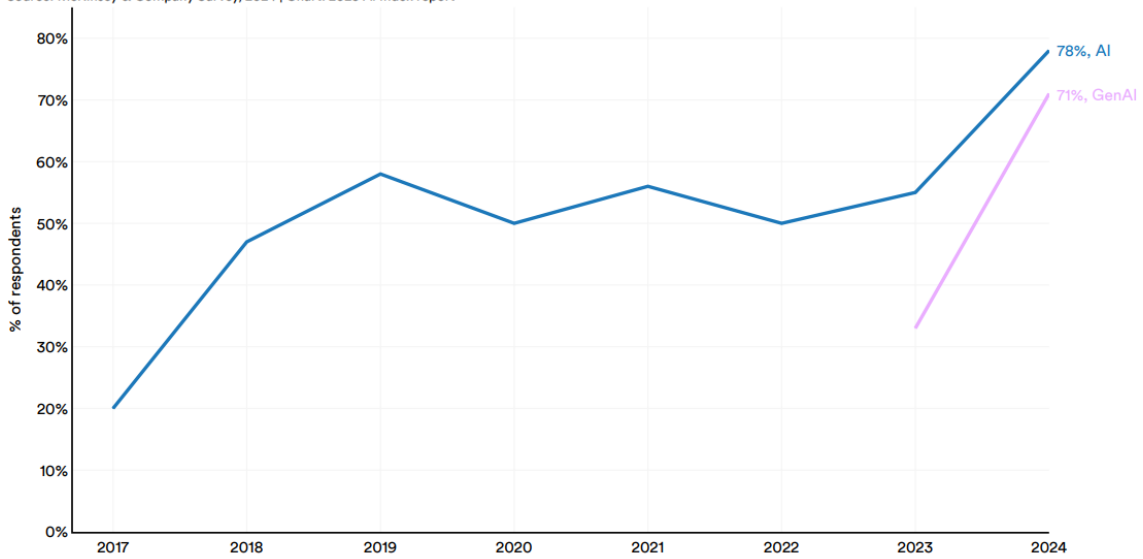


Figura 1.5: Evolución del uso de la IA y genAI en las empresas [3].

basados en recuperación aumentada (RAG, del inglés *Retrieval-Augmented Generation*) y los enfoques multiagente (MAS, del inglés *Multi-Agent System*) contribuir a mejorar la calidad y eficiencia de la anamnesis inicial? ¿Es posible hacerlo sin comprometer la seguridad, la privacidad y la trazabilidad de la información clínica? Para explorar estas cuestiones, este trabajo propone el desarrollo de un sistema inteligente especializado en reumatología, capaz de asistir al profesional sanitario mediante la síntesis y priorización automatizada de las consultas recibidas.

1.2. Objetivos

A partir de las cuestiones expuestas en el apartado anterior, se establecen los siguientes objetivos que guían el desarrollo del presente trabajo.

El **objetivo general** consiste en **diseñar y validar** un prototipo basado en técnicas de inteligencia artificial generativa que asista al profesional sanitario mediante la síntesis y estructuración automática de la información clínica relevante, así como en el apoyo a la asignación de prioridades asistenciales, favoreciendo una toma de decisiones más ágil, eficiente y coherente durante la anamnesis inicial en reumatología.

De este objetivo principal se derivan los siguientes **objetivos específicos**:

- **Profundizar** en el conocimiento teórico y técnico de los LLMs, analizando su aplicabilidad en entornos clínicos y los desafíos asociados a su uso.
- **Comprender** el funcionamiento y los fundamentos arquitectónicos de los sistemas RAG, evaluando sus ventajas para la gestión y recuperación de información médica.
- **Estudiar** el enfoque de los sistemas MAS y su potencial para la coordinación de tareas en entornos complejos basados en inteligencia artificial.
- **Integrar** en el prototipo ambos enfoques (RAG y MAS) con el fin de asistir la anamnesis clínica inicial en reumatología.
- **Evaluar** el rendimiento y la idoneidad de ambos enfoques en un caso de estudio práctico centrado en el ámbito sanitario.
- **Identificar y analizar** las principales tecnologías, técnicas, modelos y enfoques empleados en el desarrollo de sistemas de inteligencia artificial aplicados al procesamiento de información clínica, así como las limitaciones detectadas en el modelo propuesto, con el fin de proponer posibles mejoras y establecer líneas de trabajo futuras.

En conjunto, el cumplimiento de estos objetivos permitirá obtener una visión integral sobre la aplicabilidad de un sistema que integra RAG y MAS en entornos clínicos, aportando evidencia empírica acerca de su potencial para optimizar la eficiencia, la calidad y la trazabilidad en los procesos de atención reumatológica. De este modo, el trabajo no está únicamente orientado al desarrollo técnico de un prototipo funcional, sino también a la generación de conocimiento aplicable en el ámbito de la tecnología médica.

1.3. Estructura del trabajo

El presente documento se organiza en varios capítulos que reflejan las distintas fases del desarrollo del proyecto. De esta forma, se intenta facilitar tanto la lectura y comprensión del mismo, como su realización. A continuación, se describe la estructura de la memoria del trabajo.

El actual **Capítulo 1. Introducción**, introduce el contexto general y la motivación del estudio, destacando la relevancia actual de la inteligencia artificial en el ámbito sanitario y, en particular, en la reumatología. Asimismo, se definen los objetivos del proyecto y estructura de la memoria del trabajo realizado.

El **Capítulo 2, Estado del arte**, recopila los principales avances en el campo de los modelos de lenguaje y su aplicación en entornos clínicos. Se profundiza en los fundamentos teóricos de los sistemas basados en recuperación aumentada (RAG) y en los enfoques multiagente (MAS), analizando sus componentes, funcionamiento e integración. Asimismo, se contextualiza el uso de estas tecnologías en el ámbito de la reumatología como dominio clínico de referencia, abordando la complejidad del proceso asistencial, el papel de las guías clínicas ACR/EULAR como conocimiento estructurado y las aplicaciones actuales de la inteligencia artificial en esta especialidad médica.

El **Capítulo 3. Métodos: Arquitectura propuesta y orquestación de agentes**, describe la arquitectura de componentes software propuesta, detallando la orquestación de los agentes, los módulos que lo componen y el flujo de información entre ellos. Se especifican las herramientas, modelos y técnicas empleadas para la implementación del prototipo, justificando las decisiones de diseño adoptadas.

El **Capítulo 4. Experimentos, evaluación y discusión**, presenta los experimentos realizados para validar el prototipo dentro del caso de estudio reumatológico. Se exponen los métodos y métricas de evaluación, los resultados obtenidos y un análisis crítico de los mismos, discutiendo la eficacia del enfoque integrado RAG+MAS, sus limitaciones y las implicaciones prácticas para la atención clínica.

El **Capítulo 5. Conclusiones y trabajo futuro**, sintetiza los principales hallazgos del proyecto, respondiendo a las preguntas de investigación iniciales. Además, se plantean posibles líneas de desarrollo y mejora para la integración del sistema en entornos sanitarios reales y para la extensión de su uso a otras especialidades médicas.

Finalmente, el **Capítulo 6. Análisis de impacto**, examina las implicaciones potenciales del sistema desarrollado desde distintas perspectivas, incluyendo el impacto personal, empresarial, social, económico, medioambiental y cultural. Asimismo, se estudia la alineación del proyecto con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030.

Capítulo 2

Estado del arte y fundamentos teóricos

El presente capítulo tiene como objetivo revisar el trabajo previo ajeno, y contextualizar y fundamentar las bases conceptuales sobre las que se sustenta el desarrollo del sistema propuesto. Para ello, se revisan los principales avances en el ámbito de los modelos de lenguaje, la recuperación aumentada y los sistemas multiagente, así como su aplicación específica en entornos clínicos y, en particular, en el campo de la reumatología.

En primer lugar, se aborda la evolución de los modelos de lenguaje y su papel en el ámbito sanitario.

A continuación, se presentan los fundamentos teóricos de los sistemas RAG, describiendo los principios de funcionamiento de sus componentes.

Posteriormente, se introduce el concepto de los sistemas MAS como enfoque para la orquestación, control y mejora de sistemas inteligentes. Se analizan las estrategias de coordinación entre agentes y su potencial para garantizar mayor robustez, trazabilidad y adaptabilidad en entornos complejos como el clínico.

Finalmente, se revisan las aplicaciones actuales de la inteligencia artificial en el ámbito de la reumatología. En este contexto, se analizan tanto los procesos clínicos propios de la especialidad como el papel de los criterios diagnósticos y clasificatorios establecidos por el *American College of Rheumatology* (ACR) [11] y la *European Alliance of Associations for Rheumatology* (EULAR) [12], así como las principales líneas de aplicación de técnicas de inteligencia artificial en reumatología. Este análisis permite establecer el contexto clínico y las bases conceptuales necesarias para la correcta comprensión y justificación del sistema desarrollado en el presente trabajo.

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

2.1.1. Concepto y evolución

El lenguaje constituye la herramienta más poderosa del ser humano, el medio a través del cual se construye el conocimiento, se transmite la cultura y se materializa el pensamiento [13], [14]. Por ello, uno de los desafíos y objetivos principales de la investigación en inteligencia artificial reside en la creación de máquinas capaces de leer, escribir y comunicarse como los humanos [15]. Para abordar este desafío, surge la rama de la inteligencia artificial llamada **Procesamiento del Lenguaje Natural** (NLP, del inglés *Natural Language Processing*). Los **modelos de lenguaje** (LM, del inglés *Language Modeling*), definidos en [16] como modelos computacionales diseñados para entender, generar y predecir lenguaje humano, constituyen uno de los principales enfoques adoptados para llevar a cabo el avance dentro de esta rama del NLP.

Sus orígenes se sitúan un siglo atrás, en 1913, con la introducción del **Modelo de Márkov**. A partir de este modelo, aunque de forma más pronunciada en las últimas 4 décadas, los modelos de lenguaje han experimentado una profunda evolución, pasando de enfoques estadísticos basados en la frecuencia de palabras a arquitecturas neuronales capaces de capturar relaciones semánticas complejas. En la actual sección, exploraremos esta progresión tecnológica, culminada con el desarrollo de los LLMs, que constituyen la base de muchos de los sistemas de inteligencia artificial contemporáneos [17]. La Figura 2.1, muestra la evolución de los modelos de lenguaje desde la aparición del Modelo de Márkov. Para mayor comprensión, la Figura 2.2 muestra una taxonomía y categorización de los distintos enfoques usados en LM.

Aunque se podría hablar de numerosos modelos y avances en la historia de los LMs (véase la Figura 2.1), en esta sección nos centraremos en 5 importantes momentos:

- **El origen** de los LMs se sitúa en los trabajos de Andrey Márkov [19] y Claude Shannon [20]. Márkov sentó las bases matemáticas para los **N-gramas** [21], modelos probabilísticos que estiman la probabilidad de aparición de una palabra sabiendo las N-1 anteriores. Como se ha mencionado, Márkov sentó las bases, ya que fue Shannon quien propuso en [20] el uso de las cadenas de Márkov para modelizar el lenguaje natural, creando un modelo que caracterizaba la probabilidad de las secuencias de caracteres de longitud N, es decir, de los N-gramas de caracteres. Los N-gramas son la base de la actual tecnología de reconocimiento de voz. Shannon, además, introdujo un nuevo estilo de experimento que fue clave para posteriores investigaciones, el cual se basaba en obtener conocimiento humano sobre el lenguaje pidiendo a personas que predijesen el siguiente elemento en un texto [22]. Varios experimentos realizados [23], [24] con el objetivo de estudiar potenciales mejoras en la modelización del lenguaje, condujeron a la conclusión de que las personas fácilmente mejoraban el rendimiento de un modelo usando aparentemente sentido común y conocimiento propio sobre

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

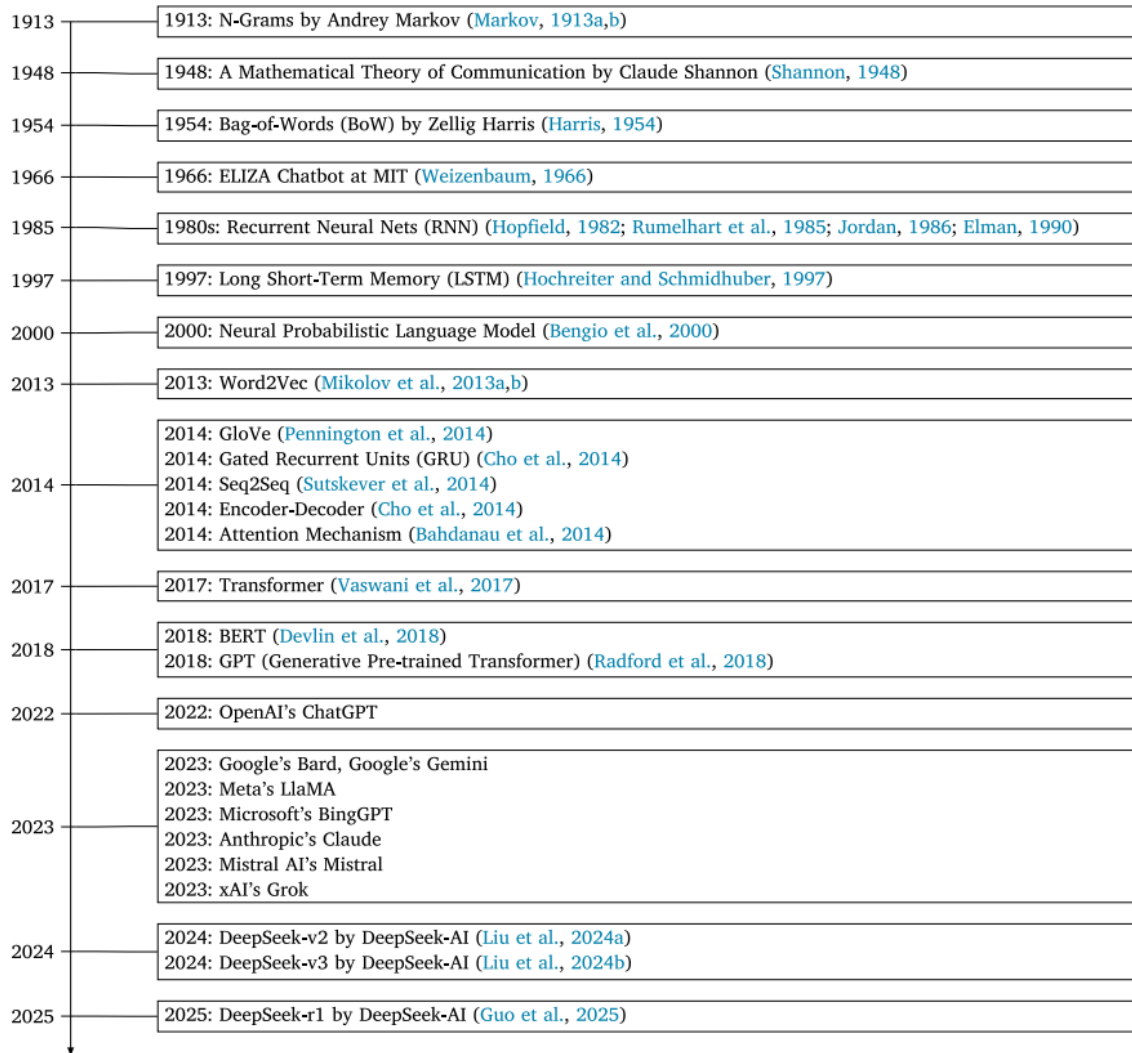


Figura 2.1: Evolución de los modelos de lenguaje [18].

Capítulo 2. Estado del arte y fundamentos teóricos

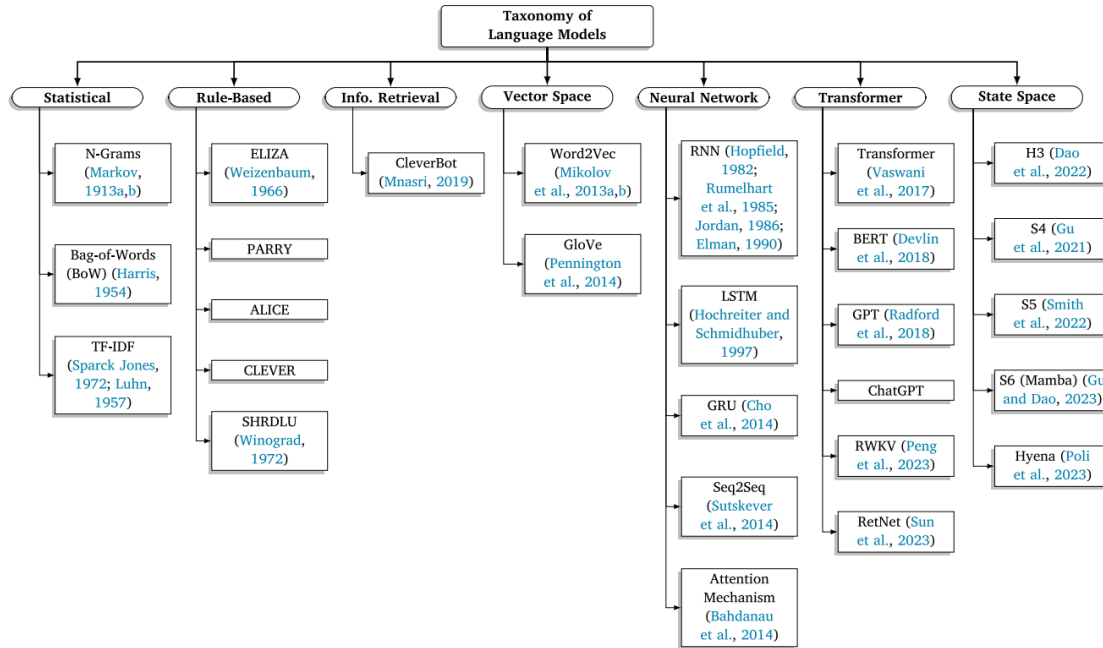


Figura 2.2: Taxonomía de los modelos de lenguaje [18].

el lenguaje.

- Modelos estadísticos (SLM, del inglés *Statistical Language Model*):** dentro de los que se incluyen los N-gramas de caracteres mencionados en el párrafo anterior, están basados en la probabilidad de aparición conjunta de cierta unidad lingüística (caracteres, palabras, oraciones). Los SLMs [25], son distribuciones de probabilidad $P(s)$ sobre todas las posibles oraciones (en general sobre todas las posibles ocurrencias de cierta unidad lingüística) s . Por ejemplo, sea s la oración "*Parece que va a llover*", y w_i la palabra que aparece en la posición i dentro de la oración. Es decir:

$$w_1 = \text{Parece}, w_2 = \text{que}, w_3 = \text{va}, w_4 = \text{a}, w_5 = \text{llover}$$

Entonces, $P(w_i)$ se define como la probabilidad de que la palabra w_i sea la que aparece en la posición i de la oración s . Por tanto, el objetivo es calcular la probabilidad de que s aparezca en un texto:

$$P(s) = P(w_1, w_2, w_3, w_4, w_5) \quad (2.1)$$

Aplicando la probabilidad condicionada obtenemos:

$$P(s) = P(w_1) \cdot P(w_2 | w_1) \cdot P(w_3 | w_1, w_2) \cdot P(w_4 | w_1, w_2, w_3) \cdot P(w_5 | w_1, w_2, w_3, w_4) \quad (2.2)$$

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

Aplicando el método de *Estimación de Máxima Verosimilitud (EMV, ó MLE, del inglés Maximum Likelihood Estimation)* [26], y la interpretación frecuentista de la probabilidad [27], se pueden estimar las probabilidades sustituyéndolas por las frecuencias si el tamaño del corpus de entrenamiento es suficientemente grande y representativo de la siguiente forma:

$$P(w_i | w_1 w_2 \cdots w_{i-1}) = \frac{P(w_1 \cdots w_{i-1} w_i)}{P(w_1 \cdots w_{i-1})} = \frac{C(w_1 w_2 \cdots w_i)}{C(w_1 w_2 \cdots w_{i-1})} \quad (2.3)$$

siendo $C(w_j \cdots w_k)$ el número de ocurrencias de la secuencia $w_j \cdots w_k$ en el corpus de entrenamiento. De esta fórmula se asume que la palabra en la posición i está relacionada con las $i - 1$ anteriores palabras. Este cálculo lleva consigo una importante limitación necesaria para poder obtener las probabilidades condicionadas: calcular y guardar $C(X)$, siendo X una oración de longitud n (o i , si queremos seguir con el índice usado para el ejemplo de la oración s), ya que el número de posibles oraciones crece exponencialmente con el tamaño del vocabulario. Para un vocabulario de k diferentes palabras, existen k^n (ó k^i) posibles secuencias de palabras de longitud $n(i)$. Para abordar este problema, en una primera instancia se acude a la llamada **asunción** o **propiedad de Márkov** [28], que aplicada al contexto de los modelos de lenguaje, indica que la probabilidad de una palabra depende sólo de un número limitado de palabras anteriores (generalmente una), y no de toda la secuencia previa. Es decir, básicamente se acude a los n -gramas con una n pequeña, obteniendo así las siguientes aproximaciones:

$$P(w_i | w_1 \cdots w_{i-1}) \approx P(w_i | w_{i-n+1} \cdots w_{i-1}) \quad (2.4)$$

$$P(X) \approx \prod_{w=1}^I P(w_i | w_{i-n+1} \cdots w_{i-1}) \quad (2.5)$$

Por ejemplo, para la oración $s = \text{"Parece que va a llover"}$, entonces la probabilidad 2.1 sería igual a :

$$P(s) \approx \prod_{w=1}^5 P(w_i | w_{i-n+1} \cdots w_{i-1}) \quad (2.6)$$

El problema de llevar a cabo esta *reducción de dimensionalidad*, es que, aunque hayamos simplificado el problema, todavía existen muchos n -gramas y no todos aparecerán en nuestro corpus de entrenamiento. Luego un n -grama que no se observe durante el entrenamiento tendrá una probabilidad de 0, y la probabilidad condicional resultante de toda la oración en la que aparezca también será 0.

Este comportamiento hace que el modelo subestime la probabilidad de los n -gramas que no aparecen en el corpus de entrenamiento y sobreestime

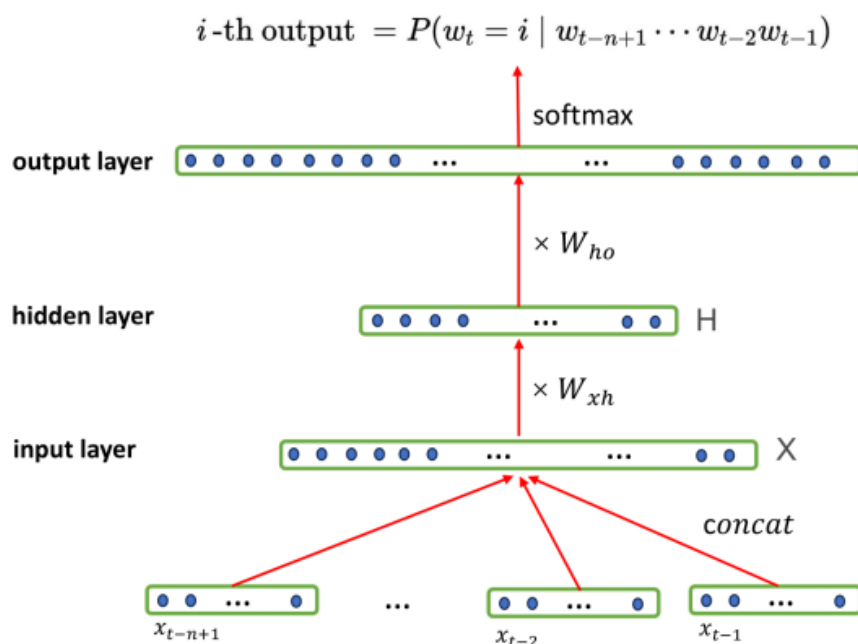


Figura 2.3: Estructura de un modelo neuronal de lenguaje [35].

la probabilidad de aquellos que sí aparecen. Para solventar este asunto, se han desarrollado numerosas técnicas de *smoothing* [29], distribuyendo parte de la masa de probabilidad hacia los *n-gramas* no observados. Algunos ejemplos conocidos de estas técnicas son las estimaciones *back-off* [30] y *good-Turing* [31].

- **Modelos neuronales de lenguaje (NLM, del inglés *Neural Language Models*):** predicen la probabilidad de las secuencias de palabras mediante redes neuronales. Algunos ejemplos de estos modelos son las *Redes Neuronales Recurrentes* (RNN, del inglés *Recurrent Neural Network*) [32], las *Redes de Memoria a Largo-Corto Plazo* (LSTM, del inglés *Long Short-Term Memory*) [33], o las *Modelos de Lenguaje Neuronales Probabilísticos* (NPLM, del inglés *Neural Probabilistic Language Model*) [34], entre otros (Ver Figura 2.2). La Figura 2.3 muestra una arquitectura simple de una NLM.
- **Modelos de lenguaje preentrenados (PLM, del inglés *Pre-trained Language Models*):** se entrenan inicialmente con grandes volúmenes de texto no etiquetado, lo que les permite adquirir de forma generalizada los principios básicos del lenguaje (el vocabulario, la sintaxis, la semántica y las relaciones lógicas). Esta es la fase de *pre-entrenamiento*, mediante la que obtenemos un modelo general que puede utilizarse para realizar diversas tareas de procesamiento de lenguaje natural, como la traducción automática, generación de resúmenes, o sistemas de pregunta-respuesta. Posteriormente, para obtener un rendimiento óptimo, el modelo pasa por una segunda fase de entrenamiento con un conjunto de datos más reducido y especializado. Esta es la fase llamada *fine-tuning* [36]. Numerosos trabajos

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

se han apoyado en este paradigma para desarrollar diferentes arquitecturas [37], [38], entre las que destacan modelos como *BERT* [39] y *GPT-2* [40].

- **Grandes modelos de lenguaje (LLM):** han provocado la creación de una nueva era [41], y fue iniciada con la introducción de *GPT-3* [42], representan una evolución significativa respecto a los modelos preentrenados recién explicados. Estos sistemas, como *GPT-4* [43], o *LLaMA* [44], entre otros, se entrenan con corpus de texto masivos y cuentan con decenas de miles de millones de parámetros, lo que les otorga una mayor capacidad para procesar y generar lenguaje natural. El objetivo fundamental de los LLMs es permitir que las máquinas comprendan las instrucciones humanas y respondan de forma coherente con los valores y la intención del usuario. Su entrenamiento combina dos etapas esenciales: un pre-entrenamiento inicial parecido al realizado con PLMs, con grandes corpus de propósito general, seguido de un proceso de *alineamiento con valores humanos*, que ajusta su comportamiento sin necesidad de cambiar de dominio o tarea (como se hace con los PLMs). Esto hace que los LLMs presenten una notable capacidad de adaptación a contextos diversos, lo que los convierte en modelos de propósito general. Por ejemplo, *GPT-3* es capaz de usar información contextual, algo que *GPT-2* no es capaz de hacer [45]. Este salto se ha logrado debido al incremento exponencial en el número de parámetros, el volumen de datos empleados y la potencia computacional disponible. En términos de magnitud, los LLMs suelen superar los cien mil millones de parámetros, con volúmenes de entrenamiento que oscilan entre cientos de gigabytes y varios terabytes de texto. Siguiendo con el ejemplo comparativo entre *GPT-2* y *GPT-3*, la versión más grande de *GPT-2*, cuenta con 1,5 mil millones de parámetros y fue entrenada con aproximadamente 40 GB de texto, mientras que *GPT-3* alcanza los 175 mil millones de parámetros y utiliza cerca de 570 GB de datos. Esta comparación evidencia la notable diferencia entre ambos paradigmas en cuanto a cantidad de parámetros y volumen de datos de entrenamiento.

2.1.2. Aplicaciones generales en medicina y sanidad

Evolución de los modelos de lenguaje en medicina

La evolución de los LMs en el ámbito médico ha estado marcada por importantes avances en arquitectura de modelos, volumen de datos de entrenamiento y técnicas de *fine-tuning* adaptadas específicamente para aplicaciones sanitarias. En las primeras etapas, los modelos biomédicos se centraron en el preentrenamiento de PLMs, como *BERT* [39], sobre corpus médicos especializados. *BioBERT* [46] y *ClinicalBERT* [47] son dos ejemplos destacables nacidos a partir de este enfoque. El primero fue creado para realizar trabajos de minería de textos biomédicos extensos y el segundo para predecir readmisiones en un periodo menor a 30 días de pacientes en hospitales, y ambos obtuvieron resultados notablemente mejores que el modelo *BERT*.

Posteriormente, surgieron LLMs diseñados específicamente para aplicaciones médicas, entrenados desde cero con grandes volúmenes de literatura científica

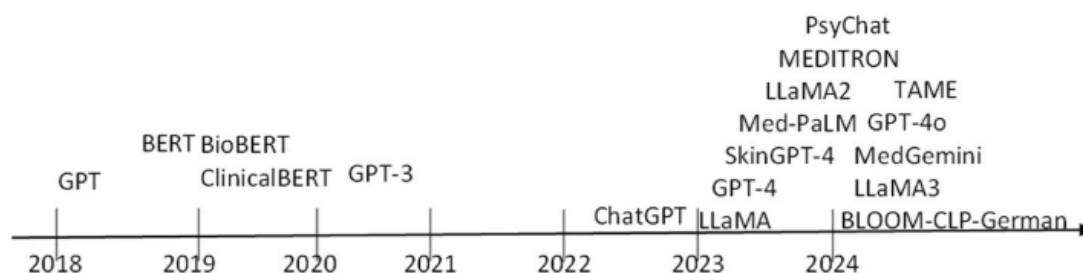


Figura 2.4: Evolución de los LLMs en el ámbito sanitario [51].

y registros clínicos electrónicos, integrando conocimiento especializado directamente en su fase de preentrenamiento (p.e. *Meditron* [48]). Más recientemente, se han desarrollado modelos multimodales capaces de procesar texto e imágenes médicas de manera conjunta, como *Med-PaLM* [49], que ha mostrado resultados prometedores en tareas de respuesta a preguntas médicas basadas en imágenes clínicas y diagnósticos. Otro ejemplo de este tipo de modelos es *MedGemini*, que además de lograr buenos resultados en diálogo médico multimodal (como *Med-PaLM*), ha logrado un desempeño sobresaliente en exámenes de licencia médica, tareas de razonamiento clínico, resumen de textos clínicos e investigación médica [50].

La Figura 2.4, obtenida en [51], evidencia que desde la aparición de *ChatGPT* en noviembre de 2022, se han multiplicado las oportunidades de aplicación de LLMs en medicina.

Aplicaciones de los LLMs en el ámbito sanitario

Como ya se ha ido analizando durante las secciones anteriores, los LLMs han demostrado un potencial significativo para mejorar la eficiencia en diversas tareas de procesamiento de lenguaje natural clínico. Entre las aplicaciones más relevantes destacan:

- **Procesamiento de historias clínicas y extracción de información.** Permiten analizar registros electrónicos de salud para extraer información clave, identificar síntomas relevantes y detectar patrones clínicos. El mencionado *BioBERT* [46] es un ejemplo de modelo con un alto rendimiento obtenido en este tipo de tareas. *GatorTron* [52], por ejemplo, demostró también un alto rendimiento en tareas como la extracción de conceptos clínicos, la extracción de relaciones médicas, la similitud semántica de textos, la inferencia de lenguaje natural y la respuesta a preguntas médicas.
- **Generación de informes y resúmenes clínicos.** Los LLMs facilitan la redacción automática de informes o notas clínicas, reduciendo la carga administrativa y mejorando la consistencia documental. *MedPaLM* [49] es un claro ejemplo de un modelo con buenos resultados en la realización de esta tarea.

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

- **Sistemas conversacionales.** LLMs integrados en chatbots clínicos o asistentes virtuales recopilan información de anamnesis, evalúan síntomas y sugieren derivaciones según criterios predefinidos. Un ejemplo de este tipo de sistemas es *PsyChat* [53], creado en 2024 como asistente para la ayuda psicológica.
- **Razonamiento clínico y apoyo a la toma de decisiones.** Los LLMs han demostrado capacidades destacables en razonamiento clínico y evaluación médica. *ChatGPT*, por ejemplo, durante un estudio de investigación [54], aprobó el examen estandarizado *USMLE* necesario para poder ejercer la medicina en Estados Unidos. Modelos avanzados han sido capaces de integrarse en flujos de trabajo clínicos, combinando datos de historias clínicas, imágenes médicas y contexto clínico específico del paciente, generando sugerencias diagnósticas y recomendaciones terapéuticas mientras mantienen al profesional sanitario como principal decisor. Estudios recientes también resaltan la utilidad de LLMs en la identificación de pacientes de alta prioridad en urgencias [55]. Sin embargo, se ha observado variabilidad en su rendimiento dependiendo del contexto clínico y de la complejidad de las tareas, lo que subraya la necesidad de supervisión experta y la revisión y actualización continua para optimizar su desempeño.

En conjunto, estas evidencias muestran que los LLMs pueden ampliar significativamente la práctica clínica, pero su implementación requiere garantizar seguridad, fiabilidad y adecuada integración con los flujos de trabajo humanos.

Entre estas aplicaciones, han emergido recientemente sistemas orientados específicamente a la automatización de la documentación clínica y la asistencia en la anamnesis, conocidos como **escribas médicos**, o *AI medical scribes* [56]. Dada su similitud con el sistema propuesto, se analizarán en la siguiente sección.

2.1.3. Sistemas de apoyo a la documentación clínica: AI Medical Scribes

En los últimos años, una de las aplicaciones más relevantes de los modelos de lenguaje en el ámbito sanitario ha sido el desarrollo de sistemas orientados a la automatización de la documentación clínica, comúnmente denominados *AI medical scribes* [56]. Estos sistemas tienen como objetivo principal reducir la carga administrativa asociada a la práctica médica mediante la captura, síntesis y estructuración automática de la información clínica generada durante la interacción con el paciente. Es decir, su finalidad es permitir al personal médico no perder tiempo en redactar la consulta de un paciente.

Diversos estudios recientes confirman que la sobrecarga administrativa constituye uno de los principales desafíos en la práctica clínica actual, afectando tanto a los profesionales sanitarios como a los pacientes. Según el estudio *Time to Care, Pan-European Perspective*, realizado por *Tandem Health* en noviembre de 2025, un 81 % de los médicos encuestados señaló que los procesos de documentación clínica afectan negativamente en la calidad de las interacciones con sus pacientes. El 59 % reconoció experimentar retrasos en la elaboración de sus no-

Capítulo 2. Estado del arte y fundamentos teóricos

tas clínicas y el 66 % consideraron que las tareas administrativas consumen una parte excesiva de su tiempo de trabajo, generando una notable preocupación por la precisión y exhaustividad de la documentación elaborada. De hecho, un 70 % de los médicos participantes en el estudio, manifestó esta preocupación. Desde la perspectiva del paciente, un 71 % manifestó haber percibido que el médico se encontraba excesivamente ocupado en realizar tareas administrativas durante la consulta, lo que contribuye a una sensación de desatención, desconfianza y frustración. Estos resultados ponen de relieve que la carga documental no solo impacta en la organización del trabajo médico, sino que también erosiona la calidad percibida de la atención sanitaria y la relación médico-paciente.

Esta situación no solo impacta negativamente en la eficiencia asistencial, sino que también se ha relacionado con un incremento del agotamiento profesional, el famoso *síndrome de burnout*. Según [57], se caracteriza por tres síntomas o dimensiones principales: cansancio emocional, que genera agotamiento ante exigencias laborales que no son consideradas atractivas; despersonalización, que provoca distanciamiento en el trato e incluso rechazo hacia las personas del ámbito laboral; y falta de realización personal, que genera una actitud negativa hacia uno mismo y hacia el trabajo, acompañada de irritabilidad, baja productividad y escasa autoestima. Tal y como se recoge en la Figura 2.5, correspondiente a un estudio realizado por la *Mayo Clinic* [58], una proporción significativa de médicos en Estados Unidos ha experimentado al menos un síntoma de burnout a lo largo de los últimos años, evidenciando la persistencia de este fenómeno en el tiempo. También en España, un estudio realizado por la Escuela Nacional de Sanidad del Instituto de Salud Carlos III [59] ha puesto de manifiesto niveles preocupantes de agotamiento profesional entre los médicos, asociados principalmente a la carga administrativa, la presión asistencial y las exigencias derivadas de la documentación clínica.

En este contexto, los escribas médicos basados en inteligencia artificial emergen como una solución tecnológica orientada a aliviar la carga administrativa y a optimizar los flujos de trabajo clínicos. Diversos estudios, como el ya mencionado de *Time To Care* de *Tandem Health*, muestran que tanto los profesionales sanitarios como los pacientes perciben de forma mayoritariamente positiva la aplicación de la IA en los procesos de documentación clínica. En particular, un 66 % de los médicos consultados considera que la automatización de estas tareas conllevaría unos notables beneficios en el ahorro de tiempo y en la mejora de la interacción médico-paciente. Desde la perspectiva de los pacientes, un 60 % cree que estas herramientas contribuirían a una atención más rápida y segura, mientras que un 50 % espera una menor probabilidad de errores en sus historiales médicos y un mayor grado de personalización en la atención. No obstante, esta aceptación está condicionada a la existencia de garantías claras en materia de privacidad, protección de datos y transparencia en el funcionamiento de los sistemas de IA, garantías exigidas por más de un 70 % de los pacientes europeos.

En el contexto específico español, los datos refuerzan esta percepción favorable hacia el uso de la IA: un 90 % de los médicos españoles consideró que la documentación clínica mediante IA sería beneficiosa en algún grado para ellos.

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

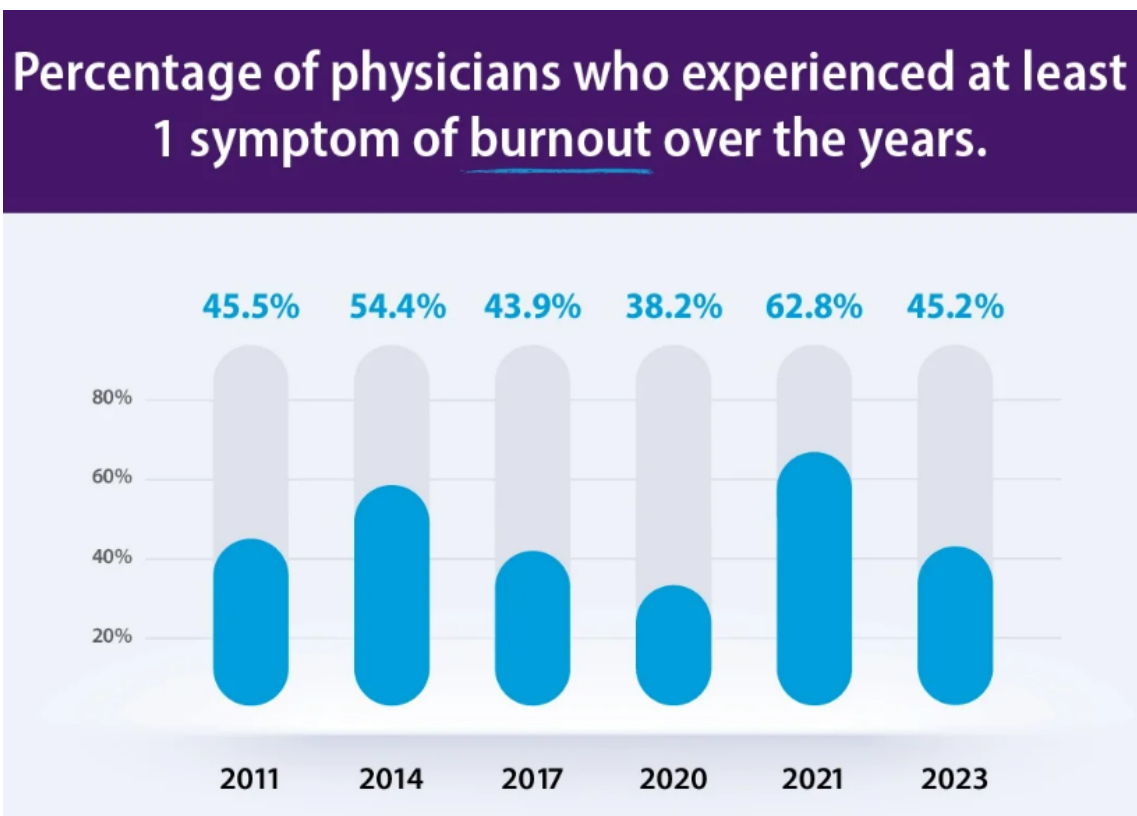


Figura 2.5: Porcentaje de médicos estadounidenses que han sentido al menos un síntoma de burnout durante los últimos 15 años [58], [60].

Capítulo 2. Estado del arte y fundamentos teóricos

Además, el 66% resaltó el impacto positivo que tendría en su calidad de vida. Por último, el 73% contestó que mejoraría las relaciones médico-paciente. Desde la visión de los pacientes, los datos son aún más pronunciados, habiendo más de un 86% que manifestó la posibilidad de mejora en el cuidado que reciben gracias al uso de la IA.

A pesar del elevado interés y de una creciente adopción inicial, el uso rutinario de herramientas de documentación clínica basadas en IA sigue siendo limitado. En Europa, únicamente el 13% de los profesionales sanitarios afirma utilizar de forma regular una herramienta de documentación con IA, y el 30% comenta haberla usado de manera ocasional. En España estas cifras aumentan significativamente, habiendo un 43% de profesionales que afirma utilizarlas activamente. Entre los que aún no lo hacen, la mitad asegura que las incorporaría si se demostrara su fiabilidad y adecuación. Este hecho pone de manifiesto la necesidad de desarrollar soluciones que no solo resulten útiles, sino que también sean fiables, seguras, adaptadas a las necesidades reales del entorno clínico, y aceptadas por los pacientes.

Desde un punto de vista funcional, los escribas emplean LLMs para procesar consultas clínicas en lenguaje natural, ya sea a partir de transcripciones de audio de la consulta o de texto introducido directamente por el paciente o el profesional sanitario. A partir de esta información no estructurada, estos sistemas son capaces de generar resúmenes clínicos, estructurar la anamnesis, identificar síntomas relevantes y completar automáticamente distintos apartados de la historia clínica, siguiendo tanto formatos definidos por el personal médico, como formatos estandarizados utilizados en la práctica clínica (p.e. SOAP [61]).

El uso de estos modelos basados en procesamiento del lenguaje natural permite además una mayor integración con los flujos de trabajo clínicos existentes, facilitando una documentación más rápida y consistente. Frente a los escribas humanos o a soluciones digitales tradicionales, los sistemas basados en IA ofrecen un mayor potencial de escalabilidad y sostenibilidad, así como una reducción de costes operativos, manteniendo niveles elevados de precisión en la captura y estructuración de la información clínica [62], [63]. Estos avances se han visto impulsados por el progreso reciente en técnicas de aprendizaje automático y procesamiento del lenguaje natural, que han mejorado significativamente la capacidad de los sistemas para comprender terminología médica, interpretar el contexto clínico y generar documentación de forma eficiente y coherente [64]. Como consecuencia, los escribas médicos basados en IA se perfilan como herramientas con potencial para reducir la carga cognitiva del profesional sanitario, contribuir a la disminución del agotamiento profesional y mejorar la calidad global de la atención al paciente.

En la práctica, diversas compañías tecnológicas y del sector sanitario han comenzado a desarrollar y desplegar este tipo de soluciones en entornos clínicos reales. Plataformas como *Nuance Dragon Ambient eXperience (DAX)*, *Amazon HealthScribe* o *Heidi* de Google ilustran el creciente interés por integrar asistentes de documentación clínica basados en LLMs dentro de los sistemas de información clínicos. Para mayor profundización sobre estas plataformas, visi-

2.1. Modelos de lenguaje y su aplicación en el ámbito sanitario

tar [65], [66], [67]. Estos enfoques han demostrado una notable capacidad para generar texto clínico coherente y para reducir el tiempo dedicado a tareas de registro, lo que ha impulsado su adopción progresiva en determinados contextos asistenciales.

No obstante, a pesar de su potencial, los escribas médicos basados en inteligencia artificial presentan limitaciones relevantes cuando se analizan desde una perspectiva clínica, técnica y regulatoria. En primer lugar, muchos de estos sistemas funcionan como soluciones de tipo *caja negra* [68], [69], ofreciendo escasa transparencia sobre el razonamiento seguido para generar determinadas conclusiones o resúmenes clínicos. Esta falta de explicabilidad dificulta comprender cómo se prioriza la información, identificar posibles sesgos y anticipar la aparición de errores, lo que plantea desafíos significativos en términos de fiabilidad, trazabilidad y responsabilidad clínica. Además, al tratarse en muchos casos de herramientas generalistas, su adaptación a especialidades médicas con necesidades clínicas específicas, como la reumatología, resulta limitada.

Desde un punto de vista clínico, otro aspecto crítico es la fiabilidad de la documentación generada automáticamente. Aunque algunos estudios reportan tasas relativamente bajas de errores y/o alucinaciones en los sistemas basados en IA, estas adquieren una relevancia considerable en el ámbito sanitario, donde un único error puede tener consecuencias graves para la seguridad del paciente. La experiencia práctica ha ilustrado distintos tipos de fallos en los escribas médicos, incluyendo la generación de contenido ficticio, la omisión de información clínica relevante y la interpretación incorrecta de afirmaciones dependientes del contexto, lo que puede derivar en una documentación inexacta de síntomas, tratamientos o planes terapéuticos. A estas limitaciones se suma la incapacidad inherente de estos sistemas para capturar información no verbal, como el lenguaje corporal, signos visuales de malestar o determinantes sociales y culturales de la salud, elementos que pueden ser clínicamente relevantes y que tradicionalmente son recogidos por el profesional sanitario o por escribas humanos.

Asimismo, en el caso de los sistemas transcritores de audio, los actuales presentan dificultades para gestionar adecuadamente escenarios con múltiples interlocutores, lo que puede dar lugar a errores de atribución del discurso entre paciente y profesional. Estas limitaciones se ven agravadas por desigualdades en el rendimiento de los sistemas de reconocimiento del habla, que han mostrado tasas de error superiores en determinados grupos poblacionales, reflejando sesgos en los datos de entrenamiento y en el diseño algorítmico [70].

En este contexto, a pesar de las mejoras observadas en los flujos de trabajo y de la creciente satisfacción y disposición a la adopción de estos sistemas por parte del profesional sanitario, el componente humano continúa siendo esencial, especialmente para la supervisión de casos clínicos complejos y la detección de inconsistencias o errores potencialmente críticos. Esta necesidad de supervisión pone de manifiesto que la adopción de escribas médicos basados en inteligencia artificial debe realizarse de forma prudente y acompañarse de un análisis riguroso y sistemático de la evidencia científica existente sobre su impacto real en la práctica clínica. Comprender sus capacidades, limitaciones y efectos en en-

tornos asistenciales reales resulta fundamental para garantizar un uso seguro y eficaz, orientado a maximizar sus beneficios sin comprometer la seguridad del paciente.

A partir de estas ventajas y limitaciones, se ha identificado un espacio de investigación y desarrollo para sistemas de documentación clínica más avanzados, capaces de combinar las capacidades generativas de los modelos de lenguaje con mecanismos que permitan recuperar conocimiento médico fiable, contextualizar la información clínica y validar los resultados obtenidos. La incorporación de arquitecturas basadas en recuperación aumentada (Retrieval-Augmented Generation) y de enfoques multiagente orientados a la descomposición, supervisión y validación de tareas permite abordar de forma sistemática muchas de las carencias observadas en los AI medical scribes actuales, dotando a estos sistemas de mayor robustez, transparencia y adaptabilidad al dominio clínico.

El sistema propuesto en este trabajo se sitúa dentro de esta línea de evolución, conceptualizándose como una extensión especializada del paradigma de los AI medical scribes. A diferencia de los enfoques generalistas, el modelo desarrollado integra mecanismos de recuperación de conocimiento clínico relevante junto con una arquitectura multiagente diseñada para asistir de manera estructurada la anamnesis inicial y el triaje en reumatología. Este enfoque no persigue únicamente automatizar la documentación clínica, sino también mejorar la calidad, la priorización y la trazabilidad de la información generada, sentando así las bases para un sistema de apoyo a la decisión clínica más fiable y alineado con las necesidades reales del entorno sanitario. Los fundamentos teóricos de estos enfoques se abordan en las siguientes secciones, dedicadas al estudio de la recuperación aumentada de información y de los sistemas multiagente aplicados a entornos clínicos complejos.

2.2. Representación semántica y recuperación de información clínica

El procesamiento automático de información clínica en lenguaje natural plantea desafíos específicos derivados de la complejidad semántica del dominio médico, la variabilidad en la forma de expresar los síntomas y la frecuente ausencia de estructura en los textos generados por los pacientes. Para abordar estos retos, resulta fundamental disponer de mecanismos que permitan representar el significado del texto de forma computacional y recuperar información relevante de manera precisa. En este contexto, los embeddings semánticos y las técnicas de recuperación basadas en representaciones vectoriales constituyen la base de numerosos sistemas modernos de apoyo clínico.

2.2.1. Embeddings semánticos para la representación de lenguaje natural

Un *embedding* es una representación vectorial (o vectorización) de una unidad lingüística (como una palabra, una frase o un documento) en un espacio vecto-

2.2. Representación semántica y recuperación de información clínica

rial de dimensión finita. A diferencia de las representaciones simbólicas tradicionales, los embeddings permiten codificar relaciones semánticas, y, en el caso de modelos contextuales, dependientes del contexto, entre los distintos elementos del lenguaje, de forma que textos con significados similares se representan mediante vectores cercanos en el espacio vectorial [71], [72], [73].

La capacidad de los *embeddings* para capturar relaciones semánticas se debe a los procesos de entrenamiento de los modelos de *embeddings*, basados en el análisis de grandes corpus textuales y en el aprendizaje de patrones de co-ocurrencia y contexto. De este modo, los modelos aprenden que palabras o expresiones que aparecen en contextos similares tienden a compartir significado, incluso cuando no coinciden exactamente en su forma léxica. Esta propiedad resulta especialmente relevante en el ámbito clínico, donde un mismo concepto puede expresarse mediante múltiples formulaciones.

En los enfoques clásicos de representación textual, como *One-Hot Encoding*, *Bag of Words* (BoW) o los métodos basados en *Term Frequency-Inverse Document Frequency* (TF-IDF) [18], [71], [73], los textos se modelan mediante vectores de alta dimensionalidad y naturaleza dispersa, conocidos en la literatura como modelos de representación dispersa, o *sparse embeddings*. Estas representaciones reflejan la presencia o frecuencia de aparición de los términos en el corpus y han demostrado ser eficaces en determinados contextos. No obstante, presentan limitaciones significativas en dominios complejos como el sanitario, ya que no capturan relaciones semánticas entre términos distintos y dependen de la coincidencia exacta de palabras, lo que dificulta la recuperación de información relevante en presencia de sinónimos, abreviaturas, variaciones expresivas o formulaciones clínicas alternativas.

Por el contrario, los modelos de representación densa o *dense embeddings* permiten superar muchas de estas limitaciones al incorporar información semántica en representaciones vectoriales compactas. Modelos como *Word2Vec*, *Doc2Vec*, o *GloVe* aprenden dichas representaciones a partir de grandes corpus textuales, capturando regularidades estadísticas y relaciones semánticas entre términos que van más allá de la coincidencia léxica directa [18], [71], [73]. Como resultado, fragmentos de texto que expresan un mismo significado mediante formulaciones diferentes pueden representarse de forma cercana en el espacio vectorial, facilitando su comparación y recuperación incluso cuando no comparten palabras exactas.

En el ámbito biomédico, se han desarrollado modelos específicos para la generación de embeddings entrenados sobre literatura científica y registros clínicos, con el objetivo de capturar el vocabulario especializado y las relaciones semánticas propias del dominio sanitario. Dentro de este enfoque, modelos como *BioWordVec* [74] producen representaciones vectoriales estáticas de términos biomédicos a partir de grandes corpus especializados, mientras que arquitecturas basadas en *transformers* [18], como *ClinicalBERT*, o *BioBERT* [46], [47], permiten extraer embeddings contextuales a partir de notas clínicas reales, adaptando la representación del texto al contexto en el que aparece. El uso de estos modelos ha demostrado un rendimiento superior frente a embeddings de propósito

general en tareas como similitud semántica, extracción de información clínica y recuperación de documentos médicos, especialmente en escenarios caracterizados por alta variabilidad lingüística y terminología especializada.

El uso de embeddings semánticos puede resultar especialmente ventajoso en el procesamiento de texto clínico generado por pacientes, que suele caracterizarse por una redacción no estructurada, una elevada variabilidad lingüística y el uso frecuente de expresiones coloquiales o imprecisas. Al permitir representar textos clínicos de forma consistente en un espacio vectorial, estos modelos facilitan la identificación de similitudes semánticas entre descripciones clínicas heterogéneas y la comparación efectiva entre consultas, síntomas o documentos médicos. Esta capacidad constituye un requisito fundamental para el desarrollo de mecanismos de recuperación de información clínica basados en similitud semántica, que se analizan con mayor detalle en la Sección 2.2.2.

2.2.2. Recuperación semántica en dominios clínicos

A partir de las representaciones vectoriales proporcionadas por los embeddings, es posible implementar mecanismos de recuperación semántica que permiten localizar información clínicamente relevante de forma más flexible y precisa que los enfoques basados en coincidencias léxicas. En estos sistemas, tanto las consultas como los documentos se representan como vectores en un espacio de alta dimensión, y la recuperación se realiza mediante la búsqueda de los vectores más cercanos al vector de la consulta.

Una de las métricas más empleadas para medir la similitud entre vectores es la similitud coseno (o *cosine similarity*), que evalúa el ángulo entre dos representaciones vectoriales independientemente de su norma [75]. En combinación con los embeddings semánticos, este enfoque permite identificar textos semánticamente similares incluso cuando presentan diferencias significativas en su forma superficial, lo que resulta especialmente útil en el contexto clínico.

Para soportar este tipo de búsquedas de manera eficiente, se emplean bases de datos vectoriales (p.e. *Qdrant*, *PineCone*, *PgVector*) junto con algoritmos de búsqueda aproximada de los *vecinos* más cercanos (*ANNS*, del inglés *approximate nearest neighbor search*), diseñados para manejar grandes volúmenes de embeddings con baja latencia y complejidad sublineal [76]. Estas infraestructuras constituyen un componente esencial en sistemas clínicos modernos basados en inteligencia artificial.

No obstante, la recuperación semántica en dominios clínicos plantea desafíos relevantes. Uno de los principales es el equilibrio entre las métricas *recall* y *precision* [77]. Un sistema con alto *recall* puede recuperar una gran cantidad de información potencialmente relevante, pero a costa de introducir ruido o resultados poco precisos. Por el contrario, priorizar la *precision* puede reducir el riesgo de errores, pero aumentar la probabilidad de omitir información clínicamente importante. En el ámbito sanitario, donde las decisiones pueden tener consecuencias directas sobre la seguridad del paciente, este equilibrio adquiere una importancia crítica.

Además, cuando la recuperación semántica se integra con modelos generativos, una selección inadecuada de los documentos recuperados puede incrementar el riesgo de generar información incorrecta o descontextualizada, contribuyendo a la aparición de alucinaciones. Por este motivo, la recuperación de información clínica debe diseñarse con mecanismos de control, filtrado y contextualización adecuados, especialmente cuando actúa como entrada para sistemas generativos.

En conjunto, los embeddings semánticos y las técnicas de recuperación vectorial constituyen una base tecnológica esencial para el desarrollo de sistemas clínicos avanzados. Sin embargo, su uso efectivo en entornos sanitarios requiere una integración explícita con arquitecturas que permitan acotar, validar y contextualizar la información recuperada antes de su utilización por modelos generativos, como ocurre en los enfoques **Retrieval-Augmented Generation** (RAG), que se analizan en el siguiente apartado.

2.3. RAG en entornos clínicos

Los LLMs han demostrado una notable capacidad para generar texto coherente y contextualmente relevante a partir de una amplia variedad de entradas. No obstante, estos modelos se entrenan habitualmente sobre *corpus* amplios y heterogéneos, lo que limita su especialización en dominios críticos y altamente regulados como el sanitario. En consecuencia, los LLMs generan respuestas basándose exclusivamente en el conocimiento implícito adquirido durante el entrenamiento, sin acceso directo a fuentes externas actualizadas, lo que puede dar lugar a información desactualizada, incompleta o incorrecta, así como a la aparición de alucinaciones [18], [78]. Estas limitaciones adquieren especial relevancia en contextos clínicos, donde incluso errores menores pueden tener implicaciones significativas para la seguridad del paciente.

La arquitectura *Retrieval-Augmented Generation* (RAG) surge con el objetivo de mitigar estas carencias mediante la integración explícita de mecanismos de recuperación de información con modelos generativos. En lugar de generar respuestas únicamente a partir de los parámetros internos del modelo, es decir, del conocimiento implícito adquirido durante el entrenamiento, las arquitecturas RAG incorporan información relevante recuperada dinámicamente de una base de conocimiento externa, que se utiliza como contexto adicional por el modelo de lenguaje durante el proceso de generación [79].

2.3.1. Preparación de la información externa: segmentación (*chunking*)

Antes de describir los componentes principales de una arquitectura RAG, resulta necesario considerar el proceso de preparación de la información externa sobre la que se apoya el sistema. Dado que los documentos clínicos, guías médicas o artículos científicos suelen presentar una longitud considerable y abarcar múltiples conceptos, estos deben ser segmentados en fragmentos de menor tamaño mediante un proceso conocido como *chunking*. Esta segmentación permite

obtener unidades semánticamente más coherentes, que posteriormente pueden representarse mediante embeddings y almacenarse en una base de datos vectorial.

La estrategia de *chunking* adoptada influye de forma directa en el rendimiento del sistema de recuperación. El tamaño de los *chunks* se ajusta habitualmente con el objetivo de equilibrar la completitud del contexto y la especificidad de la información: los fragmentos deben ser lo suficientemente grandes como para contener contexto útil, pero lo bastante pequeños como para ajustarse de forma precisa a las consultas y poder integrarse dentro de las ventanas de contexto de los modelos. En la literatura se emplean habitualmente técnicas de segmentación basadas en un número fijo de tokens, a menudo combinadas con solapamiento entre fragmentos para evitar pérdidas importantes de contexto, así como enfoques más avanzados basados en límites semánticos o estructurales del texto. En el dominio clínico, donde la información relevante puede encontrarse distribuida a lo largo de distintas secciones y formulaciones, una segmentación adecuada resulta especialmente importante para mejorar la precisión, reducir la recuperación de información irrelevante y facilitar la trazabilidad de los fragmentos utilizados durante la generación.

2.3.2. Arquitectura y componentes de los sistemas RAG

Desde un punto de vista conceptual, una arquitectura RAG se compone de varios elementos funcionales claramente diferenciados:

En primer lugar, el componente de recuperación, o **retriever**, se encarga de identificar y recuperar fragmentos de información relevantes a partir de una consulta dada. Este proceso se apoya habitualmente en técnicas de representación y recuperación semántica basadas en embeddings, mediante las cuales tanto la consulta como los fragmentos previamente segmentados (*chunks*) se representan en un espacio vectorial común y se comparan empleando algoritmos de búsqueda aproximada de los vecinos más cercanos (véase el apartado 2.2).

En segundo lugar, la base de conocimiento, o **knowledge base**, almacena el conocimiento externo sobre el que se apoya el sistema en forma de fragmentos o *chunks* previamente vectorizados. En el contexto de este trabajo, dichos fragmentos derivan de documentos clínicos, guías médicas o literatura científica, entre otras fuentes de información relevantes para el dominio sanitario. En la práctica, esta base de conocimiento se implementa habitualmente mediante bases de datos vectoriales que permiten almacenar y recuperar eficientemente los embeddings asociados a cada fragmento.

Finalmente, el componente generador, o **generator**, habitualmente un LLM, utiliza la consulta original junto con los fragmentos de información recuperados de la base de conocimiento como contexto adicional durante el proceso de generación. De este modo, el modelo combina el conocimiento implícito adquirido durante su entrenamiento con información externa explícita y específica del dominio para producir una respuesta informada, contextualizada y actualizada.

Cabe señalar que, en la mayor parte de la literatura revisada [80], [81], las ar-

arquitecturas RAG se describen como la combinación de dos módulos principales: un módulo de recuperación y un módulo generador, considerando la base de conocimiento como un elemento subyacente o implícito dentro del proceso de recuperación. No obstante, en este trabajo se adopta una descomposición explícita en tres componentes con el fin de facilitar el análisis conceptual del sistema y clarificar el papel que desempeña cada uno de ellos.

2.3.3. Flujo de funcionamiento de una arquitectura RAG

En un flujo de funcionamiento típico de un sistema RAG, el primer paso consiste en la recuperación de información relevante a partir de un corpus externo previamente fragmentado en *chunks* (véase la Sección 2.3.1). Sea la consulta de entrada denotada por q , y \mathcal{C} el conjunto de todos los *chunks* almacenados en forma de vectores en la base de datos vectorial. Entonces el módulo de recuperación vectoriza la consulta q y obtiene un conjunto de *chunks* candidatos $Z = \{c_1, c_2, \dots, c_n\}$, donde n es el número de *chunks* recuperados y $n \ll |\mathcal{C}|$, comúnmente denominados *top- n chunks* o *n chunks más relevantes*.

A continuación, el generador utiliza tanto la consulta q como los *chunks* recuperados Z para producir una salida y (p.e. una respuesta a la consulta). De manera formal, la arquitectura RAG puede interpretarse como un modelo generativo con una variable latente que define una distribución de probabilidad sobre las salidas y mediante la marginalización sobre todos los *chunks* del corpus:

$$P(y | q) = \sum_{c \in \mathcal{C}} P_{\text{ret}}(c | q) P_{\text{gen}}(y | q, c) \quad (2.7)$$

donde $P_{\text{ret}}(c | q)$ representa la probabilidad de recuperar el *chunk* c dada la consulta q (es decir, la distribución de salida del módulo de recuperación), y $P_{\text{gen}}(y | q, c)$ denota la probabilidad condicionada del módulo generador de producir la salida y dado q y un *chunk* recuperado concreto c . En la práctica, $P_{\text{ret}}(c | q)$ suele ser no nula únicamente para los n *chunks* más relevantes recuperados, lo que proporciona una aproximación tratable a la suma completa sobre el corpus [82]:

$$P(y | q) \approx \sum_{i=1}^n P_{\text{ret}}(c_i | q) P_{\text{gen}}(y | q, c_i), \quad (2.8)$$

El módulo recuperador R puede definirse formalmente como una función $R(q, \mathcal{C}) \rightarrow Z$ que recibe una consulta q y devuelve un subconjunto reducido Z del corpus \mathcal{C} , que es probable que contenga información relevante para la consulta planteada.

Por diseño, las arquitecturas RAG mantienen dos tipos de memoria: una memoria paramétrica, correspondiente al conocimiento codificado en los pesos del modelo generativo, y una memoria no paramétrica, constituida por el corpus textual externo al que se accede mediante el proceso de recuperación. Se puede ver ilustrada una arquitectura RAG estándar en la Figura 2.6, obtenida en [82].

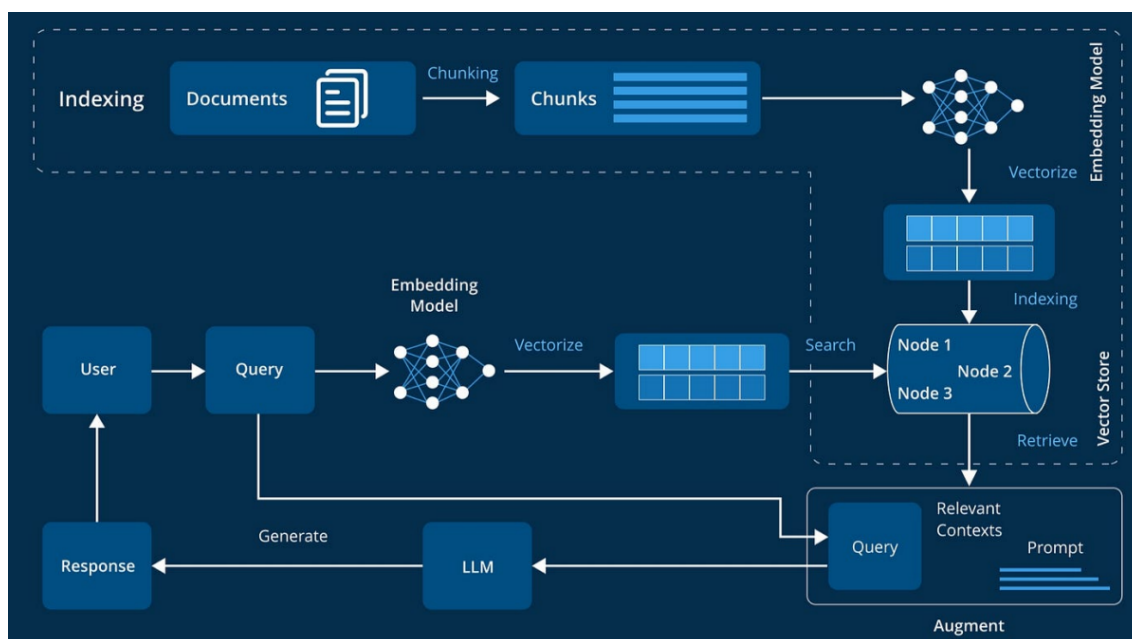


Figura 2.6: Diagrama de arquitectura RAG estándar [82].

2.3.4. Ventajas y aplicaciones del enfoque RAG en el ámbito clínico

La principal ventaja del enfoque RAG frente al uso de modelos generativos puros reside en su capacidad para acotar y fundamentar el proceso de generación en información explícita y verificable. Al condicionar la respuesta del modelo a fragmentos de información recuperados previamente desde una base de conocimiento externa, los sistemas RAG reducen de forma significativa la probabilidad de alucinaciones y facilitan la trazabilidad del contenido generado, permitiendo identificar las fuentes en las que se apoya cada respuesta. Asimismo, la separación entre el modelo generativo y la base de conocimiento posibilita la actualización del conocimiento sin necesidad de reentrenar el modelo, un aspecto especialmente relevante en el ámbito sanitario, donde la evidencia científica evoluciona de forma constante [83].

En el contexto clínico, los enfoques basados en RAG han despertado un interés creciente debido a su potencial para proporcionar acceso contextualizado a información médica fiable. Entre sus aplicaciones más relevantes se encuentran los sistemas de *clinical question answering* [84], que permiten responder consultas clínicas apoyándose en literatura médica o guías de práctica clínica; el acceso asistido a protocolos y recomendaciones clínicas actualizadas; y el soporte a la toma de decisiones clínicas mediante la recuperación y síntesis de información relevante para un caso concreto [85]. Estas aplicaciones resultan especialmente valiosas en entornos donde la precisión, la transparencia y la actualización del conocimiento constituyen requisitos fundamentales.

No obstante, la adopción de arquitecturas RAG en contextos clínicos presenta una serie de desafíos relevantes, derivados tanto de las características del do-

minio sanitario como de las limitaciones inherentes a las tecnologías empleadas [86]. Entre los principales retos identificados en la literatura se encuentran los siguientes:

- **Dominio de uso distinto y ruido en la recuperación:** causados por la heterogeneidad de los sistemas de historia clínica electrónica, los distintos estilos de documentación y la coexistencia de fuentes de calidad desigual, lo que puede afectar a la relevancia de los fragmentos recuperados.
- **Latencia:** derivada de la naturaleza secuencial del proceso de recuperación y generación, especialmente cuando se opera sobre grandes corpus clínicos o bases de conocimiento extensas.
- **Riesgos de privacidad y cumplimiento normativo:** asociados al acceso no controlado a información sanitaria protegida.
- **Debilidad en la recuperación clínica:** provocada por la incapacidad de modelos de recuperación generales para capturar adecuadamente la semántica específica del dominio médico.
- **Texto clínico ruidoso y no estructurado:** consecuencia del uso frecuente de abreviaturas, errores tipográficos, notas incompletas y redacciones inconsistentes en los registros clínicos.
- **Limitaciones en la evaluación:** las métricas tradicionales (p.e. BLEU, ROUGE o BERTScore) enfatizan la similitud superficial entre textos y no logran capturar aspectos críticos como la corrección clínica, la seguridad o la calidad del razonamiento. Tienden a pasar por alto alucinaciones, omisiones de información relevante e inexactitudes específicas del dominio clínico.
- **Limitaciones multimodales:** originadas por enfoques centrados exclusivamente en texto, que ignoran información relevante procedente de imágenes médicas, resultados de laboratorio o datos genómicos.
- **Restricciones de infraestructura:** relacionadas con los elevados requisitos de almacenamiento y computación, así como con posibles limitaciones de conectividad en entornos clínicos reales.
- **Aprendizaje no continuo:** causada por el carácter estático de los sistemas frente a la evolución continua de la evidencia médica.
- **Falta de supervisión humana:** derivada de pipelines completamente automatizados sin revisión clínica profesional o validación por parte de profesionales sanitarios.
- **Sesgos:** asociados a corpus de entrenamiento sesgados y a la infrarrepresentación de determinadas poblaciones minoritarias.

En conjunto, *Retrieval-Augmented Generation* constituye un marco arquitectónico especialmente adecuado para el desarrollo de sistemas clínicos basados en inteligencia artificial, al combinar la capacidad generativa de los modelos de lenguaje con mecanismos explícitos de recuperación de conocimiento externo

actualizado y revisado. No obstante, las limitaciones y desafíos descritos ponen de manifiesto que la adopción de RAG en el ámbito sanitario requiere un diseño cuidadoso y específico del dominio, que incorpore estrategias de control, validación, supervisión humana y adaptación clínica. En este contexto, el presente trabajo toma como punto de partida el paradigma RAG para proponer una arquitectura orientada a mitigar parte de estas limitaciones, alineada con los requisitos de seguridad, trazabilidad y fiabilidad propios del entorno clínico.

2.4. Sistemas multiagente en entornos clínicos (MAS)

2.4.1. Definición, fundamentos y evolución

Los sistemas multiagente (*Multi-Agent Systems*, MAS) [87] constituyen un área dentro de la denominada inteligencia artificial distribuida (*Distributed Artificial Intelligence*, DAI) [88], en la que múltiples **agentes** interactúan entre sí de forma coordinada, ya sea cooperativa o competitiva, para abordar tareas complejas que resultan difíciles de resolver mediante el uso de un único agente con el objetivo de alcanzar determinados fines.

De forma general, no existe una definición única y universalmente aceptada del término *agente* en el ámbito de la inteligencia artificial, debido a la diversidad de formas, niveles de abstracción y dominios de aplicación en los que estos pueden operar. No obstante, una de las definiciones más ampliamente aceptadas es la propuesta por Russell y Norvig en su libro *Artificial Intelligence: A Modern Approach*, donde describen un agente como una entidad autónoma capaz de percibir su entorno y actuar sobre él con el objetivo de alcanzar determinados fines. En el contexto de los sistemas software, esta definición se traduce en agentes capaces de procesar información de entrada, tomar decisiones y producir acciones o salidas orientadas a la realización de tareas concretas.

Estructura de organización de los agentes

Desde el punto de vista del diseño, los sistemas multiagente pueden organizarse siguiendo distintas topologías de coordinación, en función del grado de jerarquía y del tipo de interacción entre agentes. Entre las estructuras más básicas descritas en la literatura se encuentran las arquitecturas planas o también llamadas *equi-level* (véase la Figura 2.7(a)), en las que los agentes interactúan de forma horizontal con todos los agentes con los que se puedan comunicar; las arquitecturas jerárquicas (*hierarchical*) (véase la Figura 2.7(b)), donde uno o varios agentes líderes coordinan el comportamiento de agentes subordinados; y las arquitecturas anidadas (*nested*) (véase la Figura 2.7(c)), que combinan ambos enfoques mediante la combinación de agentes en subestructuras jerárquicas y/o planas en un mismo sistema multiagente. A partir de estas configuraciones fundamentales se han desarrollado arquitecturas de organización multiagente más complejas, como las de *equipo*, las *holónicas* o las *coaliciones*, orientadas a mejorar la escalabilidad, la flexibilidad y la adaptación a entornos complejos y cambiantes. Se puede observar un ejemplo de organización holónica en la Figura 2.8.

2.4. Sistemas multiagente en entornos clínicos (MAS)

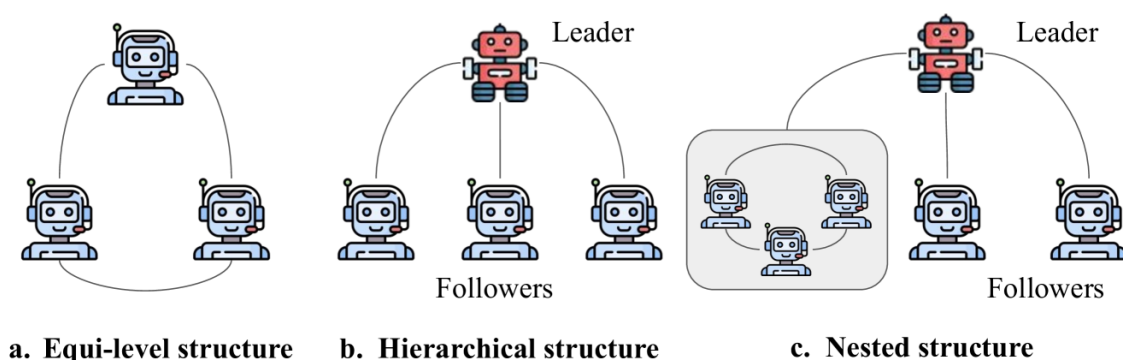


Figura 2.7: Estructuras de organización básicas de los sistemas MAS [89].

Evolución del paradigma

A lo largo de su evolución, el paradigma multiagente ha dado lugar a distintas realizaciones, que abarcan desde agentes basados en reglas y modelos simbólicos hasta enfoques más recientes apoyados en técnicas de aprendizaje automático y modelos neuronales, como los sistemas multiagente basados en modelos de lenguaje introducidos en la siguiente sección. En todos los casos, la característica común reside en la descomposición del problema en tareas específicas y entidades funcionales que cooperan mediante mecanismos y protocolos de comunicación y coordinación [88], ya sea de forma explícita o implícita.

2.4.2. Sistemas multiagente basados en modelos de lenguaje (LLM-MAS)

En los últimos años, la irrupción de los LLMs (véase la Figura 2.9) ha dado lugar a una nueva generación de sistemas multiagente, en los que los agentes utilizan estos modelos como núcleo de razonamiento y generación. A diferencia de los MAS tradicionales, los *LLM-based multi-agent systems* presentan capacidades más generales y un mayor grado de autonomía, al poder interpretar instrucciones en lenguaje natural, razonar sobre información compleja y producir salidas textuales ricas sin depender de reglas explícitas predefinidas.

En el contexto de sistemas multiagente basados en modelos de lenguaje, pueden encontrarse enfoques con distintas estrategias de orquestación. Más allá de los sistemas multiagente distribuidos con agentes autónomos que interactúan de forma dinámica, también existen arquitecturas multiagente orquestadas, en las que agentes especializados operan de manera secuencial dentro de un flujo de procesamiento estructurado diseñado y controlado por el creador del sistema. En este tipo de enfoques, cada agente cumple un rol funcional bien definido y el intercambio de información se realiza a través de las salidas generadas en cada etapa del proceso.

Dentro de este nuevo paradigma, es habitual encontrar agentes especializados que desempeñan funciones diferenciadas dentro de un mismo sistema. Entre

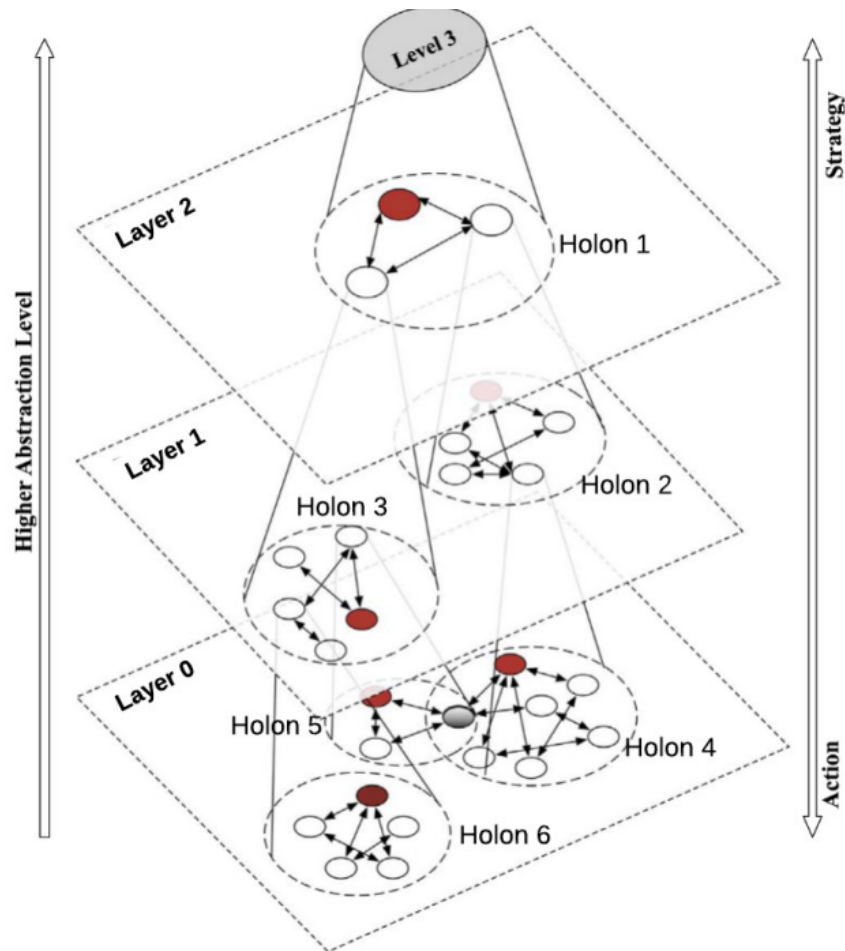


Figura 2.8: Arquitectura de organización holónica de un sistema MAS [87].

los roles más comunes descritos en la literatura se encuentran los *agentes analistas*, encargados de interpretar y descomponer la información de entrada; los *agentes críticos* o revisores, orientados a detectar inconsistencias, errores o problemas de coherencia; los *agentes de validación*, responsables de contrastar resultados con conocimiento externo o criterios predefinidos; y los *agentes orquestadores*, que coordinan la ejecución del sistema y gestionan el flujo de información entre agentes [90]. Esta especialización funcional permite abordar tareas complejas de forma estructurada y controlable.

El uso de arquitecturas multiagente basadas en modelos de lenguaje ofrece ventajas claras frente a enfoques monoagente, especialmente en tareas complejas y dominios críticos. En primer lugar, favorece la modularidad del sistema, al permitir diseñar, evaluar y modificar cada agente de forma independiente según su rol funcional. En segundo lugar, facilita un mayor control del razonamiento al separar explícitamente distintas fases del proceso, como el análisis, la generación y la validación de resultados. Aunque esta flexibilidad de organización no garantiza una robustez intrínseca, sí permite diseñar arquitecturas más robustas y tolerantes a fallos, ya que la especialización funcional y la incorporación de

2.4. Sistemas multiagente en entornos clínicos (MAS)

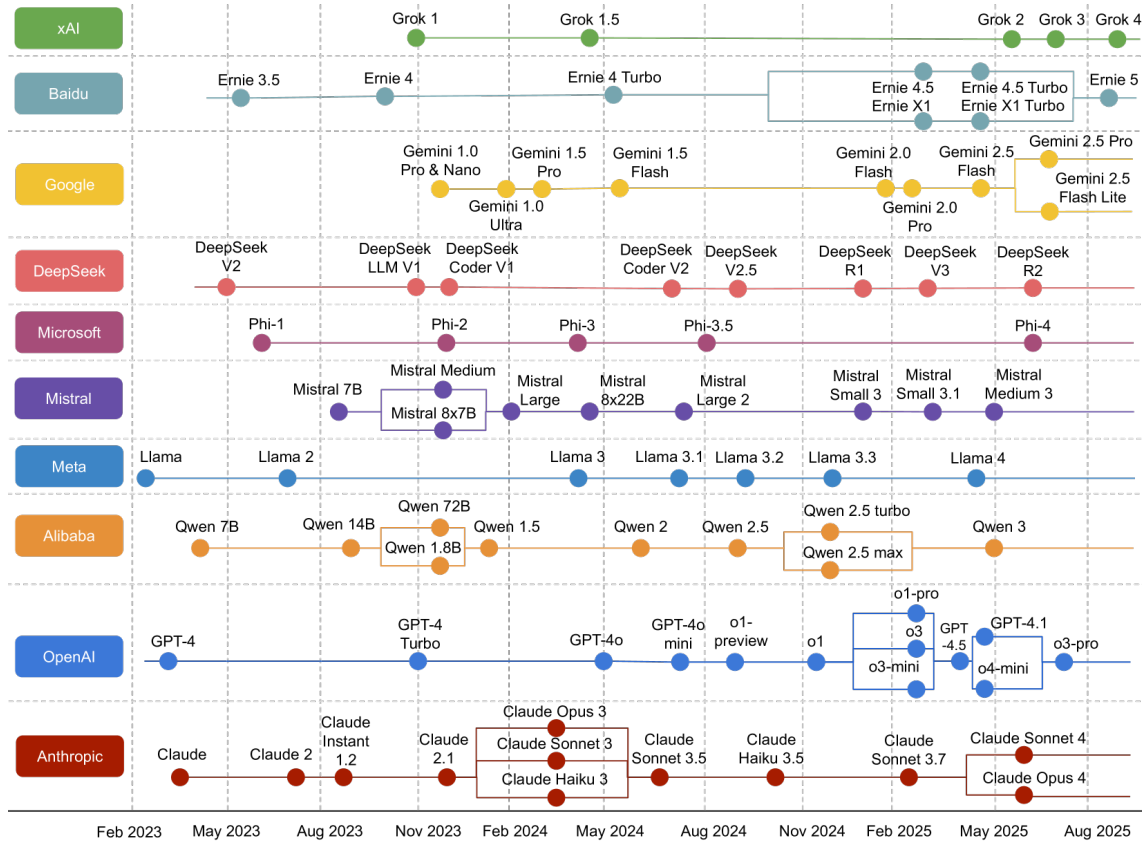


Figura 2.9: Cronología de los principales lanzamientos de modelos de lenguaje de gran escala (LLM) (2023 – 2025) [90].

mecanismos explícitos de revisión, debate o validación contribuyen a detectar y mitigar errores individuales antes de que se propaguen al resultado final. Asimismo, estos enfoques resultan más fácilmente escalables a sistemas complejos, al favorecer la asignación de roles especializados y la coordinación de múltiples agentes dentro de arquitecturas modulares [91].

2.4.3. Aplicaciones en el ámbito sanitario

En el ámbito sanitario, los sistemas multiagente han sido ampliamente estudiados como una estrategia adecuada para abordar la complejidad inherente a los procesos clínicos, que suelen involucrar múltiples fuentes de información, etapas de razonamiento y criterios de decisión [92]. La descomposición funcional en agentes especializados permite mejorar la interpretabilidad de los resultados, facilitar su validación clínica y aumentar la seguridad del sistema, aspectos especialmente relevantes en entornos sanitarios. La reciente incorporación de LLMs como núcleo de razonamiento de los agentes refuerza aún más este enfoque, al permitir integrar de forma flexible y contextualizada información textual no estructurada junto con conocimiento clínico especializado en lenguaje natural. La literatura recoge aplicaciones de enfoques multiagente en tareas como el

soporte a la toma de decisiones clínicas, la interpretación de información médica heterogénea, la monitorización continua de pacientes, la detección temprana de eventos adversos y la coordinación de procesos asistenciales, con beneficios potenciales que incluyen una mayor precisión diagnóstica, tratamientos más personalizados, mejoras en la eficiencia operativa y una optimización más efectiva de los recursos sanitarios [92], [93]. Estudios recientes han mostrado, por ejemplo, que modelos predictivos multiagente para la detección de sepsis pueden superar a sistemas de puntuación clínica tradicionales como *qSOFA* [94], y que en ámbitos como la oncología estos enfoques permiten diseñar tratamientos más ajustados a las características individuales del paciente [95]. Asimismo, la aplicación de sistemas multiagente en tareas de documentación clínica, como los sistemas de dictado ambiental, ha demostrado contribuir a la mejora de los flujos de trabajo y a la reducción de la carga administrativa y del agotamiento profesional del personal sanitario [56], [63], [64].

En conjunto, la evidencia existente pone de manifiesto el potencial de los sistemas multiagente como marco adecuado para el desarrollo de soluciones clínicas avanzadas basadas en inteligencia artificial. No obstante, como ya se ha mencionado en secciones anteriores, la aplicación efectiva de estos enfoques requiere su adaptación a dominios clínicos concretos, con características propias en términos de complejidad, disponibilidad de conocimiento estructurado y necesidades de validación. En este sentido, la reumatología se presenta como un caso de estudio especialmente relevante, que se aborda en el siguiente apartado.

2.5. Reumatología como dominio clínico de referencia

2.5.1. Complejidad clínica y proceso asistencial en reumatología

La reumatología, constituye una especialidad médica caracterizada por una elevada complejidad clínica, derivada de la heterogeneidad de las enfermedades reumáticas, la variabilidad en la presentación de los síntomas y la frecuente superposición de manifestaciones clínicas entre distintas patologías. Muchas de estas enfermedades cursan de forma crónica y presentan síntomas iniciales inespecíficos, lo que dificulta el diagnóstico temprano y convierte a la anamnesis y a la correcta interpretación de la información clínica inicial en elementos clave del proceso asistencial.

En este contexto, la identificación precoz de patrones clínicos relevantes, la detección de posibles señales de alarma y la priorización adecuada de la información recogida durante la consulta resultan fundamentales para una atención segura y eficaz. Estas características hacen de la reumatología un dominio especialmente exigente desde el punto de vista del razonamiento clínico y del manejo de información médica compleja y, en muchos casos, no estructurada.

2.5. Reumatología como dominio clínico de referencia

2.5.2. Guías clínicas ACR y EULAR como conocimiento estructurado

La práctica clínica en reumatología se apoya de manera sistemática en guías de práctica clínica y recomendaciones basadas en la evidencia, desarrolladas por organismos de referencia como el *American College of Rheumatology* (ACR) [11] y la *European Alliance of Associations for Rheumatology* (EULAR) [12]. Estas guías proporcionan criterios diagnósticos, recomendaciones terapéuticas y algoritmos de decisión clínica para un amplio conjunto de enfermedades reumáticas, con el objetivo de homogeneizar la práctica clínica y mejorar la calidad asistencial.

El contenido de estas guías se organiza habitualmente en torno a combinaciones de síntomas, signos clínicos, resultados analíticos y factores de riesgo, lo que las convierte en una fuente de conocimiento explícita, formalizada y basada en reglas clínicas. Este hecho las hace especialmente adecuadas para su integración en sistemas de apoyo a la decisión clínica basados en inteligencia artificial, donde resulta esencial disponer de conocimiento fiable, trazable y alineado con estándares clínicos reconocidos.

2.5.3. Aplicaciones de la inteligencia artificial en reumatología

En los últimos años, la aplicación de técnicas de inteligencia artificial en el ámbito de la reumatología ha experimentado un crecimiento significativo, impulsado por la disponibilidad de grandes volúmenes de datos clínicos y por la necesidad de mejorar el diagnóstico y el manejo de enfermedades reumáticas complejas. La literatura recoge trabajos orientados a la detección precoz de patologías inflamatorias, la predicción de la progresión de la enfermedad, la estratificación de pacientes y la optimización de decisiones terapéuticas, empleando tanto modelos de aprendizaje automático tradicionales como enfoques más recientes basados en aprendizaje profundo. En este sentido, se han explorado modelos aplicados al diagnóstico por imagen, técnicas de análisis genómico y el uso de tecnologías *wearable* para la monitorización continua de la enfermedad. Los resultados reportados indican que estas aproximaciones contribuyen a mejorar la precisión diagnóstica, facilitan una intervención más temprana y permiten aplicar estrategias terapéuticas más personalizadas [96], [97]. Asimismo, se han explorado aplicaciones de procesamiento de lenguaje natural para el análisis de historias clínicas electrónicas, la extracción de información relevante de notas médicas y la identificación automática de síntomas y factores de riesgo descritos en texto libre.

En esta última línea, trabajos recientes han abordado de forma explícita la integración de guías clínicas reumatológicas en sistemas basados en modelos de lenguaje. Un ejemplo representativo es el estudio presentado en [98], en el que se propone un marco basado en *Retrieval-Augmented Generation* para optimizar la aplicación clínica de las recomendaciones de EULAR y ACR. En dicho trabajo, especialistas en reumatología realizan en primer lugar un procesamiento manual y estructuración del contenido de las guías, partiendo de los manuscritos principales de las recomendaciones y excluyendo información suplementaria. Como parte de este proceso, se eliminan secciones no esenciales, como

Capítulo 2. Estado del arte y fundamentos teóricos

referencias bibliográficas, se revisa la jerarquía y títulos de encabezados y se reorganizan tablas y recuadros informativos al final de los documentos con el objetivo de minimizar el ruido y la redundancia. Asimismo, para preservar la información gráfica, los investigadores usan *ChatGPT o1* para transformar las figuras incluidas en descripciones textuales, permitiendo su incorporación en formato textual dentro de la base de conocimiento. Este tipo de estudios pone de manifiesto el potencial de combinar conocimiento clínico estructurado con modelos de lenguaje mediante mecanismos de recuperación de información, al tiempo que evidencia la importancia de la validación y la alineación con guías clínicas reconocidas.

Por todo ello, la reumatología constituye un dominio clínico apropiado para el estudio y desarrollo de sistemas inteligentes de apoyo clínico. Su dependencia de conocimiento especializado, guías clínicas consolidadas y procesos de razonamiento estructurados refuerza la motivación para el enfoque propuesto en este trabajo, que combina modelos de lenguaje, recuperación de información y una arquitectura multiagente con el objetivo de asistir de forma fiable y trazable en las fases iniciales del proceso asistencial.

Capítulo 3

Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

Sobre la base teórica, clínica y tecnológica establecida en el capítulo anterior, en este capítulo se presenta la arquitectura propuesta del sistema, diseñada para asistir de forma estructurada y fiable en la anamnesis clínica en el ámbito de la reumatología. La arquitectura combina modelos de lenguaje de gran escala con mecanismos de recuperación de información y una organización multiagente, con el objetivo de mejorar la calidad, trazabilidad y validación de la información clínica generada, así como de facilitar la priorización de pacientes en función de la gravedad de su situación clínica.

Como punto de partida, y tras presentar una visión general del planteamiento de la solución propuesta, se lleva a cabo un análisis de requisitos que permite traducir las necesidades detectadas en el dominio clínico (entre ellas, la optimización de las listas de espera mediante la identificación temprana de casos prioritarios) y las limitaciones identificadas en el estado del arte, en un conjunto de requisitos funcionales y no funcionales. Este análisis, elaborado en las fases iniciales del proyecto, ha guiado las decisiones de diseño adoptadas, la selección de las tecnologías empleadas y la definición de los roles de los distintos componentes y agentes del sistema. En especial, se han considerado aspectos críticos como la fiabilidad y completitud de los resultados, la trazabilidad de la información clínica y de los resultados generados, la modularidad del sistema, la posibilidad de validación clínica, la adecuación a un posible entorno real y el soporte explícito a la priorización asistencial.

A partir de dichos requisitos, se propone una arquitectura híbrida basada en *Retrieval-Augmented Generation* (RAG) y sistemas multiagente (MAS), diseñada para asistir de forma estructurada la anamnesis inicial en reumatología. En lugar de un sistema monoagente o de un MAS distribuido completamente autónomo, se opta por un enfoque multiagente orquestado por el programador, en el que agentes especializados operan de manera secuencial y coordinada dentro de

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

un flujo de procesamiento controlado. Esta elección responde a la necesidad de mantener un alto grado de control sobre el razonamiento del sistema, facilitar la trazabilidad y validación intermedia de los resultados, y reducir el riesgo de errores en un dominio crítico como el clínico reumatológico.

El capítulo detalla los componentes principales del sistema, incluyendo los módulos de entrada y preprocesamiento de la información clínica, los mecanismos de recuperación de conocimiento externo, los agentes encargados de la estructuración, detección de señales de alarma y validación clínica, así como el modelo generativo subyacente. Asimismo, se describen las herramientas, librerías y modelos de lenguaje empleados en la implementación del prototipo.

Finalmente, se expone el flujo completo de funcionamiento del sistema, desde la recepción de la consulta en lenguaje natural hasta la generación de una salida estructurada y validada, poniendo énfasis en la interacción entre los distintos agentes y en los mecanismos diseñados para garantizar coherencia, robustez y trazabilidad. Este enfoque metodológico sienta las bases para la evaluación experimental que se presenta en el capítulo siguiente, donde se analiza el comportamiento del sistema propuesto en un escenario reumatológico representativo.

3.1. Visión general del planteamiento de la solución

Antes de abordar el análisis detallado de los requisitos del sistema, resulta conveniente presentar una visión general del planteamiento de la solución propuesta desde un punto de vista conceptual. La Figura 3.1 ilustra de manera esquemática el papel del sistema como elemento intermediario entre el paciente y el profesional sanitario, actuando como una capa de apoyo intermedia en el proceso asistencial.

En este planteamiento, la información clínica del paciente, expresada en lenguaje natural, es procesada por el sistema mediante un conjunto de componentes internos que, desde la perspectiva del usuario, operan como una “caja negra”. A partir de dicha información, el sistema analiza la consulta, contrasta el contenido con el conocimiento clínico disponible y genera resultados estructurados y contextualizados que son presentados al médico. De este modo, el profesional sanitario recibe información sintetizada y relevante que le permite centrar su atención en la evaluación clínica, la toma de decisiones y la priorización asistencial, sin que el sistema sustituya en ningún caso su juicio clínico.



Figura 3.1: Visión general del planteamiento conceptual de la solución y su papel como soporte en el proceso asistencial.

3.2. Análisis de requisitos

El diseño del sistema propuesto parte de un análisis de requisitos cuyo objetivo es trasladar las necesidades reales del entorno clínico a un conjunto de especificaciones técnicas que guíen el desarrollo del prototipo. Este análisis se llevó a cabo en las fases iniciales del proyecto y se fundamenta tanto en la revisión del estado del arte presentada en el Capítulo 2 como en un documento de especificación elaborado por una médico especialista en reumatología, en el que se recogen pautas clínicas, requisitos asistenciales y directrices sobre la estructuración y el formato de la información de salida, así como ejemplos utilizados para concretar el contenido esperado y validar el comportamiento del sistema durante su desarrollo.

A partir de estas fuentes, los requisitos se clasificaron en dos grandes categorías: requisitos funcionales y requisitos no funcionales. Los requisitos funcionales describen las capacidades que el sistema debe ofrecer desde el punto de vista clínico y operativo, mientras que los requisitos no funcionales recogen propiedades de calidad necesarias para garantizar un uso fiable, seguro y potencialmente integrable en un entorno asistencial real.

3.2.1. Requisitos funcionales

Como partida para el procesamiento de una consulta clínica, el sistema debe cumplir los siguientes requisitos funcionales (RF):

- RF1.** Permitir la entrada de texto con la descripción clínica aportada por el paciente.
- RF2.** Extraer y estructurar un conjunto de seis parámetros clínicos considerados

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

fundamentales por el especialista en el proceso de anamnesis: edad y sexo, inicio y duración de los síntomas, patrón de afectación, presencia de inflamación funcional, síntomas generales o extraarticulares, y tratamientos previos junto con su respuesta.

- RF3.** Indicar explícitamente aquellos parámetros clínicos no referidos por el paciente, marcándolos como “*No referido*”, y añadir un mecanismo que asigne una etiqueta de completitud en función del número de parámetros extraídos, clasificando la consulta como completa, con información parcial o con información escasa.
- RF4.** Asignar un nivel de prioridad asistencial numérico comprendido entre 0 y 10 para cada consulta, acompañado de una breve justificación textual, con el objetivo de facilitar la priorización asistencial y optimizar la gestión de las listas de espera. Indicar si la información extraída es incompleta.
- RF5.** Calcular el nivel de prioridad en base a un conjunto de reglas clínicas proporcionadas por el especialista, que asignan distintos pesos a la información recogida en la consulta, tales como la presencia de señales de alarma, el patrón inflamatorio identificado o el grado de limitación funcional.
- RF6.** Identificar y listar las posibles señales de alarma clínicas detectadas en la consulta, indicando explícitamente cuando no se identifica ninguna con “*Ninguna referida*”.
- RF7.** Generar un resumen clínico de la consulta, con una extensión máxima de 200 palabras, que sintetice de forma clara y técnica, desde un punto de vista médico, la información clínica aportada por el paciente.
- RF8.** Generar una nota clínica muy breve, de dos a tres líneas, orientada a su incorporación directa en la historia clínica del paciente.
- RF9.** Sugerir hasta 3 preguntas de aclaración que permitan completar la información clínica relevante en consultas posteriores.
- RF10.** No diagnosticar ni pautar tratamientos en ningún caso.
- RF11.** No inferir información no proporcionada por el paciente o no respaldada por las guías clínicas.
- RF12.** Emplear de forma sistemática y obligatoria las guías clínicas de la European Alliance of Associations for Rheumatology (EULAR) y del American College of Rheumatology (ACR) [11], [12], y de forma opcional las de la Sociedad Española de Reumatología (SER).
- RF13.** La salida del sistema debe seguir una plantilla estructurada previamente especificada por la médico especialista.

3.2.2. Requisitos no funcionales

El sistema debe cumplir los siguientes requisitos no funcionales (RNF):

- RNF1.** Uso de terminología clínica homogénea.

- RNF2.** Preservar expresiones vagas utilizadas por los pacientes (por ejemplo, “desde hace tiempo”), evitando la inferencia, reinterpretación o normalización automática de dichas expresiones cuando ello pueda conducir a alucinaciones o a una subestimación del riesgo clínico.
- RNF3.** En materia de seguridad y privacidad, cumplir con la normativa vigente en protección de datos, en particular el Reglamento General de Protección de Datos (RGPD) [99] y la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD) [100]. El sistema debe aplicar principios de minimización y pseudonimización de datos, evitando el almacenamiento de información clínica identificable salvo que exista una necesidad asistencial debidamente documentada.
- RNF4.** El sistema debe garantizar el cifrado de la información tanto en tránsito como en reposo.
- RNF5.** Minimizar la aparición de alucinaciones, limitando sus respuestas a información respaldada por la consulta del paciente y por conocimiento recuperado de fuentes médicas fiables, en particular de las guías clínicas empleadas.
- RNF6.** Incluir referencias explícitas a las guías clínicas utilizadas, indicando la versión y el fragmento relevante, con el fin de facilitar la validación de la información por parte del profesional sanitario.
- RNF7.** Presentar tolerancia a fallos, de modo que un fallo de razonamiento de uno de los agentes no comprometa de forma global el funcionamiento del sistema, permitiendo una degradación controlada de la salida o la mitigación del error.
- RNF8.** En términos de rendimiento, el sistema debe mantener tiempos de respuesta compatibles con un uso práctico, estableciendo umbrales de latencia aceptables tanto para la extracción estructurada de la información como para la generación completa de la salida.
- RNF9.** Garantizar una disponibilidad continua de los servicios de inferencia del modelo de lenguaje subyacente, de forma que sea accesible las 24 horas del día.
- RNF10.** Desde el punto de vista del mantenimiento, la arquitectura del sistema debe ser modular, facilitando su escalabilidad, evolución y actualización, contemplando estrategias de despliegue que permitan validar nuevas versiones sin interrumpir el servicio.
- RNF11.** Incorporar mecanismos que garanticen el control, la supervisión y la posibilidad de auditoría, complementando los requisitos de seguridad, trazabilidad y fiabilidad clínica.

3.2.3. Cumplimiento de los requisitos

Cabe señalar que los requisitos presentados en esta sección corresponden al conjunto final de requisitos considerados para el desarrollo del prototipo. Si

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

bien el análisis inicial sirvió como punto de partida, algunos requisitos fueron ajustados o refinados de forma incremental a lo largo del desarrollo del proyecto, a medida que se evaluaban resultados parciales y la viabilidad técnica de determinadas funcionalidades. Este proceso se llevó a cabo de manera controlada y coherente con los objetivos clínicos del sistema, sin alterar los requisitos fundamentales definidos en las fases iniciales.

Dado el carácter exploratorio y académico de este trabajo, no todos los requisitos definidos han podido ser implementados completamente en el prototipo desarrollado. No obstante, el análisis de requisitos ha servido como marco de referencia para priorizar las funcionalidades y propiedades de calidad más relevantes desde el punto de vista clínico, en particular la estructuración de la información y la priorización asistencial.

El cumplimiento concreto de los requisitos definidos se evalúa en el Capítulo 4, analizando tanto aquellos que han sido implementados satisfactoriamente como los que han sido abordados de forma parcial o no han podido ser implementados. Estos últimos se discuten identificándose como limitaciones del sistema y como líneas claras de mejora y trabajo futuro.

3.3. Arquitectura general del sistema

A partir del análisis de requisitos presentado en el apartado anterior, se define una arquitectura software orientada a dar soporte a la anamnesis clínica en reumatología mediante la combinación de modelos de lenguaje de gran escala, mecanismos de recuperación de conocimiento externo y un sistema multiagente orquestado. La arquitectura ha sido diseñada con el objetivo de garantizar modularidad, trazabilidad y control del razonamiento, permitiendo la asignación de prioridad asistencial de los pacientes en función de la gravedad de su situación clínica. La Figura 3.2 presenta una visión general de la arquitectura de componentes software propuesta.

Desde un punto de vista conceptual, el sistema adopta un enfoque híbrido basado en RAG integrado dentro de una arquitectura multiagente. El componente RAG se apoya en un módulo específico de preparación y almacenamiento de conocimiento externo, encargado de procesar las guías clínicas utilizadas por el sistema (ACR y EULAR). Este módulo, identificado en el diagrama de la Figura 3.2 como **Módulo generador de embeddings de datos externos**, realiza tareas de segmentación (*chunking*) de los documentos, generación de representaciones vectoriales mediante *embeddings*, así como su indexación y almacenamiento en una base de datos vectorial.

La ejecución de este módulo constituye una fase previa necesaria al funcionamiento del sistema, ya que permite preparar el conocimiento clínico que será posteriormente recuperado durante el procesamiento de las consultas. De este modo, la recuperación de información clínica relevante durante la ejecución del sistema se apoya en una base de datos vectorial previamente construida, garantizando la coherencia del proceso de generación.

3.3. Arquitectura general del sistema

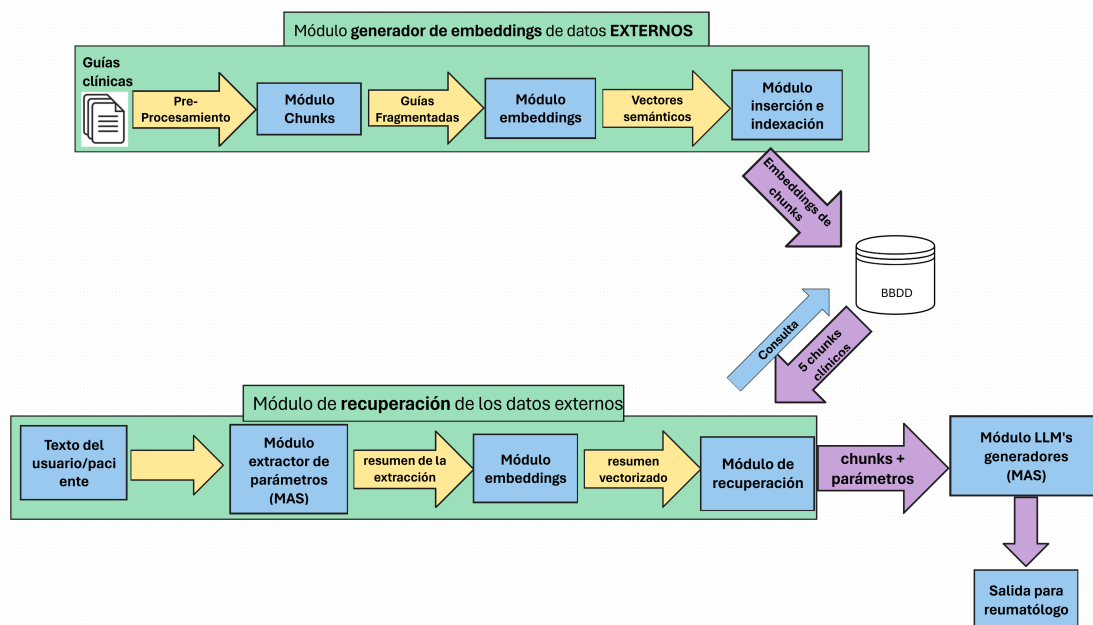


Figura 3.2: Diagrama de la arquitectura de componentes software propuesta.

Una vez preparado y almacenado el conocimiento externo, el funcionamiento del sistema sigue un flujo de procesamiento secuencial y orquestado. El sistema recibe como entrada una consulta clínica en lenguaje natural, que es procesada de forma progresiva por los distintos agentes especializados definidos en la arquitectura. En las fases iniciales del flujo, se recuperan fragmentos de conocimiento clínico relevantes desde la base de datos vectorial, con el objetivo de contextualizar el razonamiento posterior de los agentes. Estas tareas son realizadas por el **Módulo de recuperación de los datos externos** (véase la Figura 3.2).

A lo largo de la ejecución, los agentes son activados de manera ordenada para llevar a cabo tareas específicas, como la estructuración de la información clínica, la detección de señales de alarma, la evaluación de la completitud de los datos disponibles y la asignación de una prioridad asistencial. Los resultados intermedios generados por cada agente se integran de forma incremental, dando lugar a una salida final estructurada, orientada a su uso por el profesional sanitario. Las tareas de generación de la salida son realizadas por los agentes del módulo generador, denominado **Módulo LLMs generadores** en la Figura 3.2, que se apoyan tanto en la información estructurada previamente extraída por el **Módulo extractor de parámetros** contenido en el módulo de recuperación en la figura, como en el conocimiento clínico recuperado.

La organización multiagente permite descomponer el proceso clínico en tareas especializadas, alineadas con los requisitos funcionales definidos, mientras que la orquestación explícita de los agentes, definida por el programador, garantiza

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

un alto grado de control sobre el razonamiento del sistema. Esta decisión facilita la trazabilidad de las decisiones adoptadas y permite la detección y gestión de posibles errores o inconsistencias durante el procesamiento.

En los apartados siguientes se describen con mayor detalle los distintos componentes que conforman la arquitectura, comenzando por los módulos de entrada y preparación del conocimiento externo, y continuando con la definición de los agentes especializados y el flujo completo de ejecución del sistema.

3.4. Preparación del conocimiento externo y generación de embeddings

Como se ha introducido en el apartado anterior, el sistema se apoya en un enfoque de RAG que requiere la disponibilidad previa de conocimiento clínico externo estructurado en forma vectorial. Para ello, se define un módulo específico de preparación y almacenamiento de conocimiento externo, responsable de procesar las fuentes clínicas utilizadas por el sistema y de generar la base de datos vectorial que será posteriormente consultada durante la ejecución del sistema. La Figura 3.3 muestra el detalle de este módulo dentro de la arquitectura propuesta.

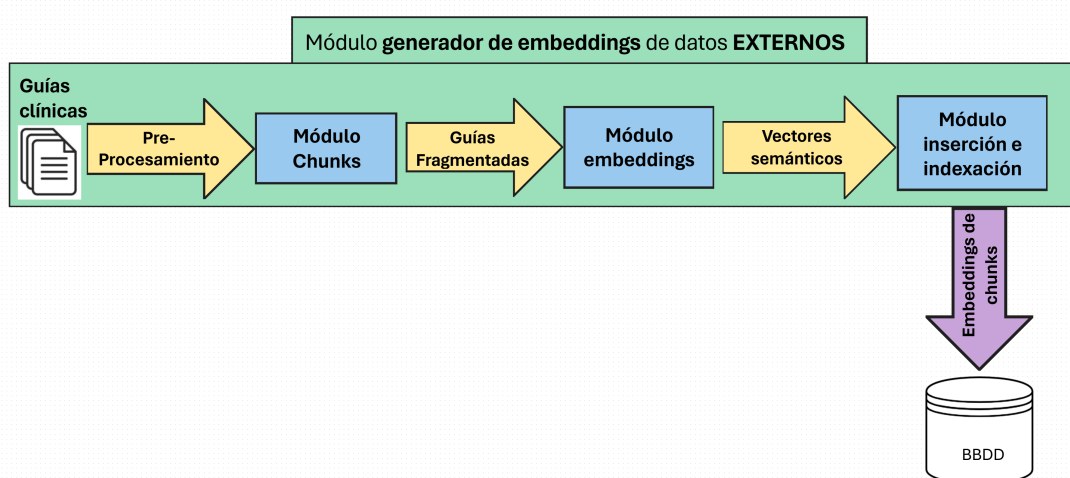


Figura 3.3: Módulo generador de embeddings de datos externo.

Las fuentes de conocimiento empleadas corresponden a guías clínicas reconocidas en el ámbito de la reumatología, en particular las guías de la European Alliance of Associations for Rheumatology (EULAR) y del American College of Rheumatology (ACR). En este trabajo no se parte directamente de los documentos originales en bruto, sino de versiones optimizadas de dichas guías previamente preprocesadas a partir del trabajo [98], descrito en la Sección 2.5.3, en el que se realiza una limpieza y estructuración del contenido clínico. Se recuerda

3.4. Preparación del conocimiento externo y generación de embeddings

que este preprocesado incluye, entre otros aspectos, la revisión y normalización de encabezados, la eliminación de referencias bibliográficas y la descripción textual de figuras, con el objetivo de facilitar su posterior explotación por arquitecturas RAG.

El procesamiento de estas guías preprocesadas se realiza de forma previa a la ejecución del sistema y constituye una fase independiente del flujo de procesamiento de las consultas clínicas. En primer lugar, los documentos optimizados son sometidos a un proceso de segmentación (*chunking*), cuyo objetivo es dividir el contenido textual en fragmentos de tamaño manejable y semánticamente coherentes. Para ello, se emplean mecanismos de división semántica del texto, diseñados para preservar la coherencia clínica de los fragmentos generados. Este paso resulta clave para mejorar la calidad de la recuperación posterior, ya que permite asociar cada fragmento a una unidad de significado clínico concreta y reducir la pérdida de contexto durante el proceso de búsqueda.

Una vez segmentados los documentos en *chunks*, cada fragmento es transformado en una representación vectorial densa mediante un modelo de *embeddings*. Estas representaciones vectoriales permiten capturar relaciones semánticas entre los fragmentos de texto y las consultas clínicas, facilitando la recuperación de información relevante incluso cuando no existe una coincidencia literal entre los términos empleados (véase el Apartado 2.2). En el dominio clínico, es habitual el uso de sinónimos, abreviaturas y expresiones diversas para referirse a un mismo concepto. Además, el proceso de recuperación se basa en consultas formuladas a partir de la información proporcionada por el paciente, que puede ser imprecisa o poco técnica, aunque durante el procesamiento de la consulta se trate de normalizar y especializar dicha información. Por el contrario, las guías clínicas emplean un lenguaje especializado y formal. El uso de representaciones vectoriales semánticas permite reducir esta brecha terminológica, facilitando la recuperación de información relevante incluso cuando existe una discrepancia entre el lenguaje del paciente y el de las fuentes clínicas.

Los vectores generados, junto con la información asociada a cada fragmento, es decir, los metadatos, como el identificador del *chunk* que representa y el nombre del documento clínico original del que proviene, son posteriormente indexados e insertados en una base de datos vectorial. Este proceso permite construir una estructura persistente que puede ser consultada de manera eficiente durante la ejecución del sistema y que actúa como repositorio de conocimiento clínico externo preparado para su uso en el componente RAG.

Cabe destacar que la correcta ejecución de este módulo constituye una condición necesaria para el funcionamiento del sistema. La calidad del preprocesado inicial, de los fragmentos generados, de los *embeddings* asociados y de la indexación realizada influye directamente en la relevancia y utilidad de la información recuperada durante la ejecución del sistema. Por este motivo, la preparación del conocimiento externo se concibe como una fase claramente diferenciada y controlada, alineada con los requisitos de fiabilidad clínica, trazabilidad y control del razonamiento definidos en el análisis de requisitos.

En el apartado siguiente se describe el mecanismo de recuperación de informa-

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

ción clínica durante la ejecución del sistema, detallando cómo la base de datos vectorial generada en esta fase es consultada para enriquecer el razonamiento de los agentes y apoyar la generación de salidas clínicas estructuradas.

3.5. Recuperación de información clínica durante la ejecución del sistema

Una vez preparado el conocimiento clínico externo y generada la base de datos vectorial descrita en el apartado anterior, el sistema incorpora un mecanismo de recuperación de información que se activa durante la ejecución de cada consulta clínica. Este mecanismo forma parte del enfoque RAG y tiene como objetivo proporcionar contexto clínico relevante que apoye el razonamiento posterior del sistema. La Figura 3.4 ilustra este módulo de recuperación (el rectángulo verde) dentro de la arquitectura propuesta, mostrando cómo el proceso de recuperación se integra e intercala con el sistema multiagente a lo largo del flujo de ejecución.

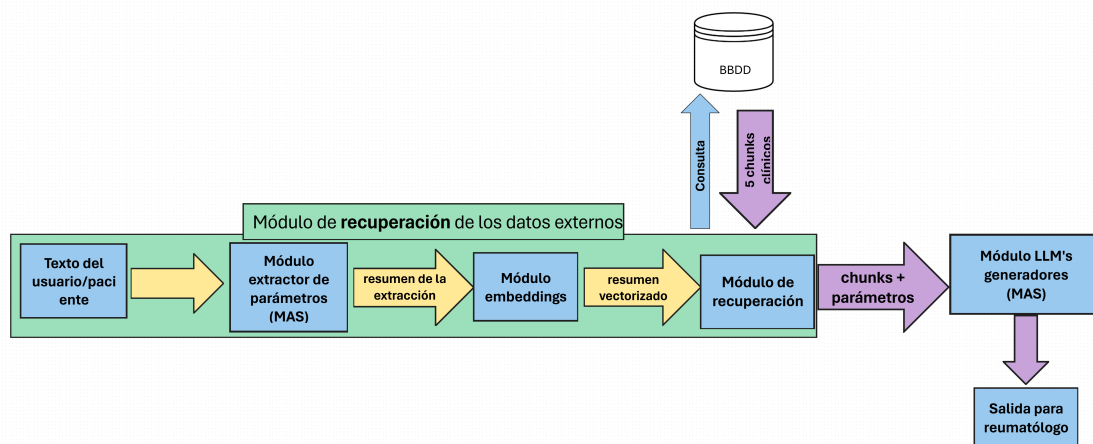


Figura 3.4: Módulo de recuperación de los datos externos integrado con el módulo generador multiagente.

La recuperación de información se realiza de forma controlada y explícitamente orquestada por el programador dentro del flujo de ejecución del sistema. En concreto, la recuperación no es ejecutada de manera autónoma por ningún agente, sino que se activa en un punto bien definido del *pipeline*, una vez que el sistema ha procesado la consulta del paciente hasta obtener una representación estructurada de la información clínica relevante. Esta decisión responde a la necesidad de mantener un alto grado de control sobre el flujo de razonamiento y garantizar la trazabilidad del proceso de recuperación.

La integración del mecanismo de recuperación con el sistema multiagente no se concibe como una etapa aislada, sino como un proceso intercalado dentro del flujo de ejecución. En este sentido, el sistema multiagente puede entenderse

3.5. Recuperación de información clínica durante la ejecución del sistema

conceptualmente como organizado en dos fases diferenciadas que enmarcan el proceso de recuperación de conocimiento externo.

En una primera fase, un agente especializado del sistema multiagente se encarga de generar una descripción breve y focalizada de la consulta clínica, orientada específicamente a la recuperación de información. Esta descripción se construye a partir de la información estructurada extraída previamente de la consulta del paciente, siguiendo los criterios clínicos definidos por la médico especialista. A partir de este resumen, el sistema formula la consulta de recuperación que será utilizada para interrogar a la base de datos vectorial.

A continuación, el sistema ejecuta el proceso de recuperación de información utilizando dicha descripción orientada a *retrieval*. Para ello, la consulta formulada es transformada en una representación vectorial mediante un modelo de *embeddings*. Cabe destacar que el modelo de *embeddings* utilizado para vectorizar la consulta debe coincidir con el modelo empleado previamente para generar los vectores asociados a los fragmentos almacenados en la base de datos; sin embargo, no tiene por qué coincidir con el utilizado durante la fase de segmentación semántica de las guías clínicas. Esto se debe a que la coherencia del espacio (vectorial) semántico es un requisito indispensable para la recuperación semántica, mientras que la segmentación constituye un proceso previo e independiente del modelado de *embeddings*.

La recuperación se ejecuta siempre para cada consulta clínica y se realiza una única vez durante el flujo de procesamiento. El sistema recupera un número fijo de fragmentos de conocimiento reumatológico externo, concretamente cinco fragmentos (*top-five chunks*). Esta elección responde a una heurística de diseño orientada a proporcionar al modelo generativo un contexto clínico suficientemente amplio como para asegurar una sensibilidad adecuada en la recuperación de información relevante, evitando al mismo tiempo la saturación del *prompt* y la degradación del razonamiento del modelo derivada de la inclusión de un volumen excesivo de contexto irrelevante o redundante.

Además, durante el proceso de segmentación semántica se observó una variabilidad notable en la extensión de los fragmentos generados, coexistiendo fragmentos particularmente extensos con otros de carácter más escueto. En este contexto, la recuperación de cinco fragmentos se consideró una elección equilibrada, capaz de adaptarse a dicha variabilidad sin comprometer la calidad del contexto proporcionado al modelo. Asimismo, durante las pruebas realizadas se constató que, para distintas consultas clínicas, el contexto recuperado contenía, en promedio, información directamente relacionada con el caso planteado. Por lo tanto, se decidió mantener este valor como una heurística adecuada para el caso de estudio considerado. La búsqueda se basa en la similitud semántica entre la representación vectorial de la consulta y los vectores almacenados en la base de datos, utilizando como medida de proximidad la métrica del coseno (*cosine similarity*) [75].

Cada resultado recuperado incluye, además de su representación vectorial, los metadatos asociados que permiten identificar tanto el *chunk* original como la guía clínica de procedencia (los nombres de los ficheros correspondientes). A

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

partir de estos metadatos, el sistema recupera el contenido textual literal de los fragmentos seleccionados, que constituye el contexto clínico externo utilizado en las fases posteriores del procesamiento. Es decir, que en lugar de almacenar directamente el contenido textual completo de cada fragmento en la base de datos vectorial, el sistema conserva referencias y metadatos que permiten identificar de forma unívoca los fragmentos y su fuente clínica. El texto literal asociado a cada fragmento se recupera posteriormente a partir de estas referencias con el objetivo de mantener una organización clara y separada entre el índice de búsqueda semántica y el contenido clínico legible. Esta decisión de diseño favorece la trazabilidad del conocimiento empleado, facilita la auditoría de las fuentes clínicas utilizadas y evita la duplicación innecesaria de información dentro del índice vectorial.

Adicionalmente, con fines de trazabilidad y análisis, el contenido textual de los fragmentos recuperados, junto con sus metadatos asociados, así como la puntuación de similitud obtenida con respecto a la consulta, son volcados en ficheros intermedios durante la ejecución del sistema. Este registro permite inspeccionar de forma explícita qué información clínica ha sido recuperada para cada consulta, con qué grado de similitud semántica y a partir de qué fuentes, facilitando el análisis del comportamiento del mecanismo de recuperación y la validación de su adecuación clínica.

En una segunda fase, el sistema multiagente vuelve a intervenir para procesar de forma integrada la información original de la consulta, los parámetros clínicos estructurados y el conocimiento clínico recuperado. Los fragmentos recuperados son inicialmente integrados y resumidos por un agente específico, con el objetivo de generar un único texto coherente que sintetice la información clínica relevante. Este texto es posteriormente utilizado únicamente por aquellos agentes para los que el conocimiento externo resulta necesario, en particular los agentes encargados de la mejora y contextualización de la consulta clínica, la detección y revisión de señales de alarma, y el cálculo de la prioridad asistencial.

Esta integración explícita de información procedente de guías clínicas contribuye a mejorar la coherencia y alineación de las salidas generadas con el conocimiento médico de referencia, reforzando la fiabilidad clínica del sistema. Asimismo, al incorporar conocimiento clínico externo de forma selectiva y controlada, se evita la sobrecarga de información y la reducción del riesgo de interferencias innecesarias en el razonamiento de los agentes.

En el siguiente apartado se describe con mayor detalle la organización del sistema multiagente, así como el papel específico de cada agente dentro del flujo de ejecución, incluyendo aquellos que hacen uso del contexto clínico recuperado.

3.6. Sistema multiagente: agentes y roles

El sistema propuesto se apoya en una organización multiagente orquestada, en la que distintos agentes especializados cooperan de forma secuencial para llevar a cabo tareas clínicas bien definidas dentro del proceso de anamnesis asistida. Cada agente está diseñado para cumplir un rol específico, alineado con los

3.6. Sistema multiagente: agentes y roles

requisitos funcionales identificados, y opera sobre distintas representaciones intermedias de la consulta clínica, como listas de síntomas estructurados, parámetros clínicos extraídos o versiones contextualizadas de la propia consulta.

La Figura 3.5 muestra la arquitectura de organización del sistema multiagente y el flujo de información entre los distintos agentes definidos. En ella se observa cómo el texto original proporcionado por el paciente actúa como entrada inicial del sistema y es procesado progresivamente por distintos agentes especializados, cuyos resultados intermedios se encadenan y reutilizan a lo largo del flujo de ejecución, dando lugar a una salida final que cumple los requisitos definidos.

A diferencia de sistemas multiagente con un mayor grado de autonomía y coordinación emergente, en este trabajo la ejecución de los agentes se encuentra explícitamente orquestada por el flujo de control definido en el código. El orden de activación de los agentes y el intercambio de información entre ellos están determinados de forma explícita, evitando mecanismos de planificación o negociación autónoma entre agentes. Este enfoque permite mantener un control estricto del flujo de razonamiento, aportando un comportamiento predecible y reproducible al sistema y facilitando la trazabilidad y la inspección de los resultados intermedios generados por cada agente, aspectos especialmente relevantes en un dominio clínico sensible como el reumatológico. Asimismo, este diseño facilita la identificación de posibles inconsistencias durante el procesamiento y la localización del punto del flujo en el que se producen.

Todos los agentes definidos en el sistema utilizan LLMs con *prompts* específicos adaptados a su función. No obstante, algunos de ellos incorporan además reglas clínicas explícitas, como listas predefinidas de señales de alarma o criterios ponderados para el cálculo de la prioridad asistencial, combinando razonamiento generativo con conocimiento estructurado y favoreciendo un comportamiento más determinista en aquellos agentes para los que esta cualidad resulta conveniente o necesaria.

3.6.1. Organización de los agentes dentro del flujo de ejecución

Tal y como se ilustra en la Figura 3.5, el sistema multiagente se integra de forma intercalada con el mecanismo de recuperación de información clínica, organizándose en un flujo secuencial y orquestado que combina la extracción estructurada de información, la recuperación de conocimiento externo y la generación progresiva de las salidas clínicas requeridas.

En una primera fase, el texto original de la consulta del paciente es procesado por el **Módulo extractor de parámetros**, compuesto por varios agentes especializados que operan de manera independiente sobre el mismo texto de entrada. Estos agentes extraen de forma estructurada los distintos parámetros clínicos considerados relevantes para la anamnesis inicial, tal y como se especificó en la fase de definición de requisitos. Este módulo se concibe como una unidad funcional dentro del flujo global del sistema; la descripción detallada de su estructura interna y de los agentes que lo componen se presenta en la Sección 3.6.2.

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

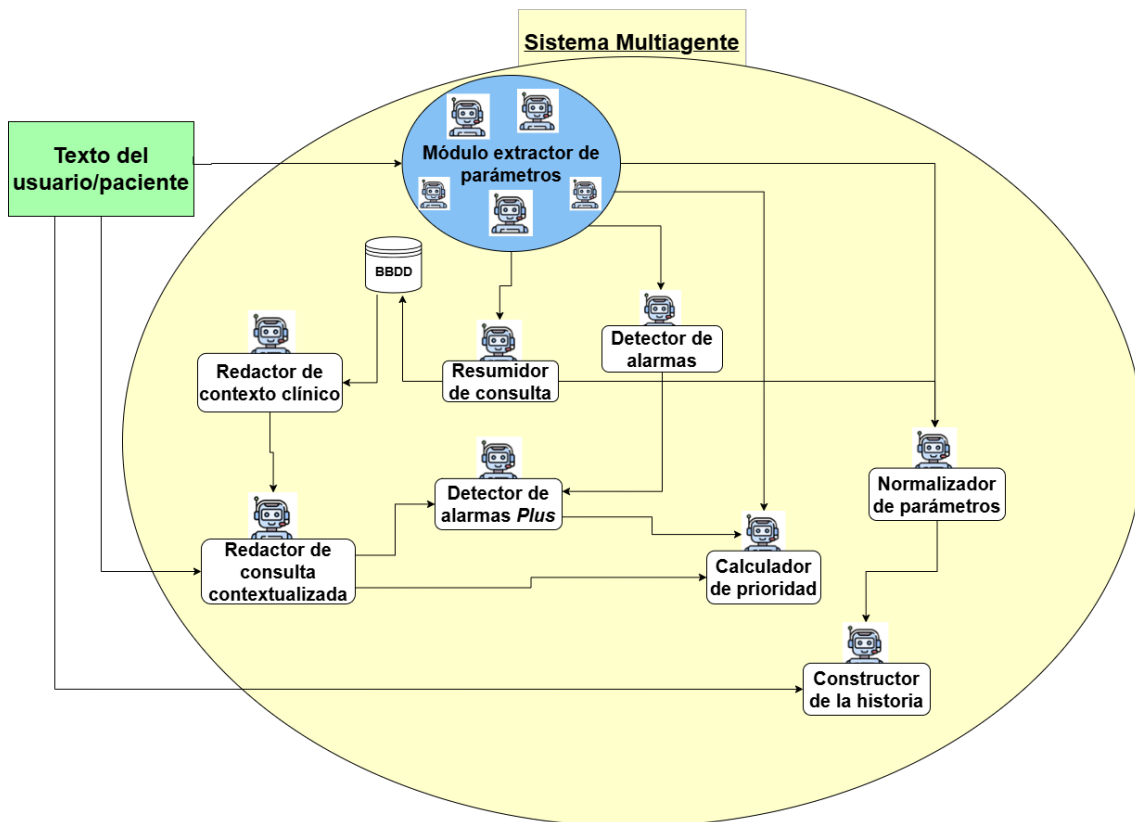


Figura 3.5: Arquitectura de organización del sistema multiagente.

A partir de los parámetros extraídos, el sistema activa un agente **Resumidor de consulta** orientado a la recuperación de información clínica relevante, que genera una descripción sintética de la consulta en inglés. Esta decisión responde a que las guías clínicas empleadas por el sistema se encuentran redactadas en dicho idioma y permite maximizar la coherencia semántica durante el proceso de recuperación. Esta descripción se utiliza como entrada para el proceso de recuperación de conocimiento externo, el cual es ejecutado de forma explícita por el programador, y no por un agente autónomo. El resultado de dicha recuperación es posteriormente integrado por el agente **Redactor de contexto clínico**.

En una segunda fase, los agentes posteriores continúan el procesamiento de la consulta combinando, según el caso, la información original del paciente, los parámetros clínicos estructurados y el conocimiento clínico recuperado. En concreto, el **Detector de alarmas** inicial opera exclusivamente a partir de los parámetros estructurados extraídos por el módulo extractor de parámetros, identificando posibles señales de alarma y sugiriendo preguntas de aclaración sin hacer uso directo de la información recuperada de las guías clínicas.

El resto de agentes de esta fase incorporan de forma selectiva el conocimiento clínico externo recuperado. Tal y como se aprecia en la Figura 3.5, esta información se emplea en la revisión y refinamiento de las señales de alarma mediante

3.6. Sistema multiagente: agentes y roles

el agente **Detector de alarmas plus**, en la generación de una consulta clínica contextualizada por el agente **Redactor de consulta contextualizada** y en el cálculo de la prioridad asistencial por el agente **Calculador de prioridad**.

Finalmente, los agentes **Normalizador de parámetros** y **Constructor de la historia clínica** operan sin utilizar conocimiento externo recuperado. El primero se limita a normalizar y traducir los parámetros estructurados con el fin de cumplir el formato especificado en los requisitos. El segundo genera una historia clínica breve que resume y relaciona los parámetros extraídos de la consulta, manteniéndose deliberadamente cercano a la consulta original. Esta proximidad reduce el riesgo de introducir errores adicionales y permite, además, facilitar la identificación de posibles inconsistencias en las salidas generadas por otros agentes.

3.6.2. Módulo extractor de parámetros clínicos

Uno de los componentes clave del sistema multiagente es el *Módulo extractor de parámetros*, encargado de analizar al inicio del *pipeline* la consulta clínica original del paciente y extraer de forma estructurada los parámetros considerados fundamentales para la anamnesis inicial en reumatología. Este módulo se compone de seis agentes especializados, cada uno de los cuales recibe exclusivamente como entrada el texto libre proporcionado por el paciente y se centra en la extracción de un conjunto concreto de parámetros clínicos.

La Figura 3.6 muestra el esquema interno de este módulo, ilustrando cómo la consulta del paciente actúa como entrada común para los distintos agentes extractores, que operan de forma paralela y especializada sobre el mismo texto. Esta organización permite descomponer la tarea de extracción clínica en sub-tareas bien definidas, reduciendo la complejidad de cada agente y mejorando la robustez del proceso de estructuración.

Cabe destacar que, en una fase inicial de la implementación del proyecto, se intentó abordar esta tarea mediante el uso de técnicas de **reconocimiento de entidades nombradas** (*Named-Entity Recognition*, NER) [101], [102], empleando las librerías *SpaCy* y *MedSpaCy*. Sin embargo, debido a la complejidad superior y al rendimiento inferior de este enfoque en comparación con el módulo extractor basado en agentes descrito, dicha aproximación fue finalmente descartada.

La Tabla 3.1 resume los agentes que conforman este módulo, así como los parámetros clínicos específicos que extrae cada uno de ellos.

3.6.3. Descripción de los agentes y sus roles

La Tabla 3.2 resume los agentes definidos en el sistema, sus entradas principales, salidas y el rol funcional que desempeñan dentro del flujo de ejecución.

Cabe señalar que, además de los agentes que componen el sistema multiagente operativo, se desarrolló un agente auxiliar basado también en un modelo de lenguaje con fines evaluativos. Este componente no participa en la interacción

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

Tabla 3.1: Agentes del módulo extractor de parámetros clínicos y variables extraídas

Agente extractor	Parámetros clínicos extraídos
Edad y género	Edad del paciente y género (masculino: "M"/femenino: "F").
Inicio y tiempo	Inicio del cuadro clínico (agudo, subagudo o crónico) y duración temporal expresada por el paciente (p. ej., "3 meses", "10 días").
Patrón de afectación	Tipo de afectación articular (periférica o axial), tamaño de las articulaciones (pequeñas/grandes), simetría/asimetría, distribución (mono-, oligo- o poliarticular) y presencia de entesitis o afectación tendinosa.
Inflamación funcional	Presencia de rigidez matutina (sí/no), duración de la rigidez (en minutos), dolor nocturno (sí/no) y mejoría con el movimiento (sí/no).
Síntomas generales y extraarticulares	Síntomas sistémicos o extraarticulares como fiebre, pérdida de peso, afectación cutánea o de mucosas, ocular, neurológica, torácica/respiratoria, renal/urinaria, fenómeno de Raynaud, fotosensibilidad y otros síntomas detectados (p.e., astenia, insomnio..).
Tratamientos previos y respuesta	Tratamientos previos recibidos (AINEs/corticoides/FAMEs/tratamientos biológicos) y respuesta percibida por el paciente (alivio parcial/sin efecto).

3.6. Sistema multiagente: agentes y roles

Tabla 3.2: Agentes del sistema multiagente y sus roles funcionales

Agente	Entrada	Salida	Rol principal
Agentes extractores de parámetros	Consulta del paciente	Parámetros clínicos estructurados	Extracción estructurada de información clínica relevante (véase La Tabla 3.1)
Resumidor de consulta	Parámetros estructurados	Resumen de los parámetros extraídos de la consulta	Generación de una consulta en inglés optimizada para la recuperación de información clínica
Redactor de contexto clínico	Fragmentos recuperados	Contexto clínico unificado	Integración y síntesis del conocimiento clínico externo recuperado
Redactor de consulta contextualizada	Consulta original + contexto clínico unificado	Consulta clínica enriquecida, contextualizada	Integración del conocimiento externo en la descripción clínica del paciente para obtener un resumen clínico avanzado
Detector de alarmas	Parámetros estructurados	Lista inicial de alarmas y preguntas	Identificación preliminar de señales de alarma y preguntas de aclaración
Detector de alarmas <i>plus</i>	Lista inicial de alarmas y preguntas + consulta contextualizada	Alarmas y preguntas revisadas	Revisión y validación clínica de alarmas y preguntas sugeridas
Agente calculador de prioridad	Parámetros + alarmas + consulta contextualizada	Nivel de prioridad y justificación	Cálculo de la prioridad asistencial mediante reglas clínicas ponderadas
Agente normalizador de parámetros	Parámetros estructurados+ resumen de los parámetros	Parámetros normalizados	Normalización de valores, marcado de campos no referidos y traducción al español del resumen
Constructor de historia clínica	Consulta original + parámetros normalizados	Historia clínica breve	Construcción de una nota clínica muy breve que recoja principalmente los parámetros extraídos

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

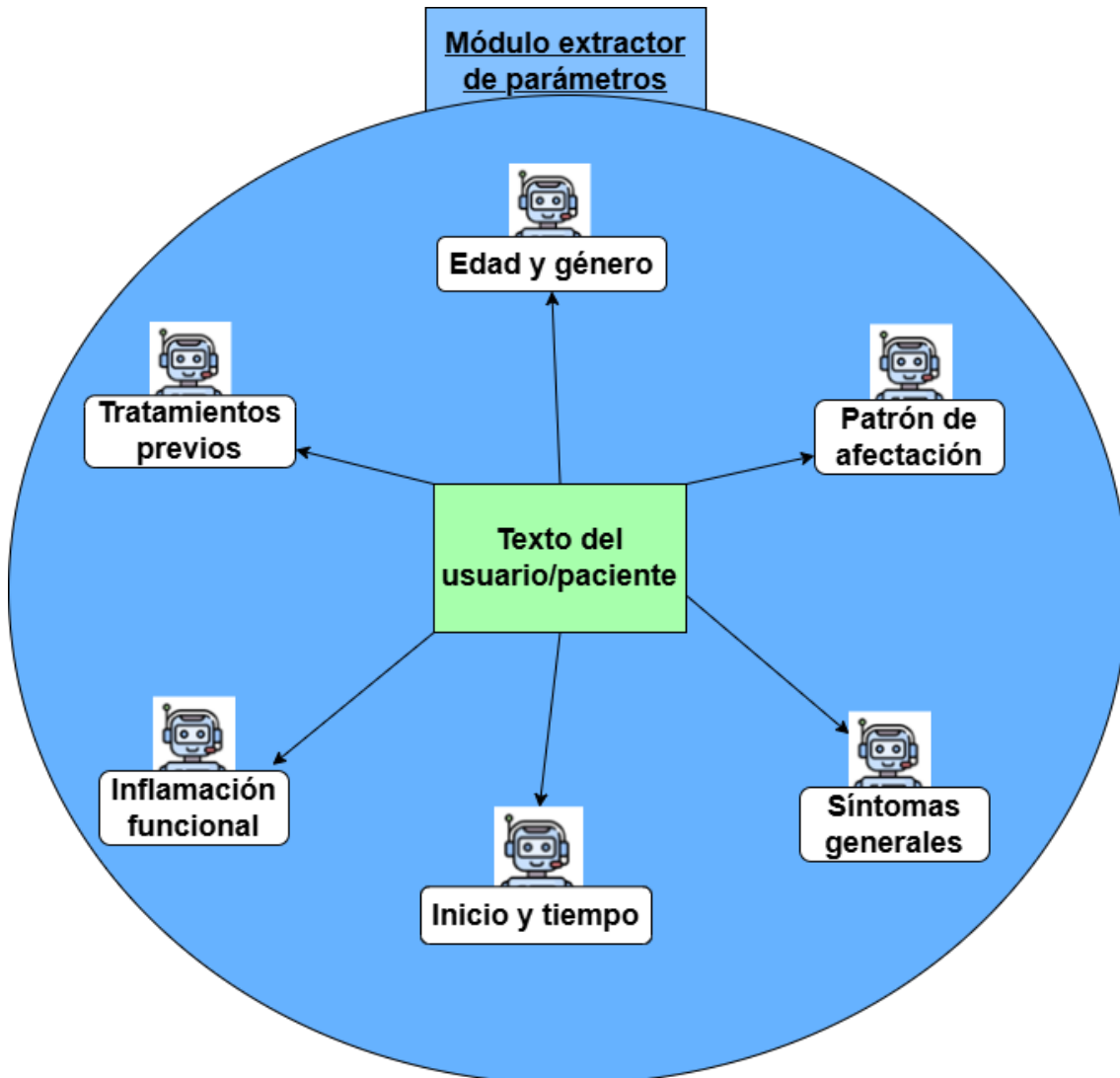


Figura 3.6: Esquema interno del módulo extractor de parámetros.

entre agentes ni influye en la generación de respuestas, y se describe en detalle en el Capítulo 4.

3.6.4. Orquestación y control del razonamiento

La ejecución de los agentes se encuentra estrictamente orquestada por el flujo de control definido en el código del sistema. No existe un agente final encargado de ensamblar la salida completa; en su lugar, los resultados generados por los distintos agentes se recogen y combinan de forma explícita durante la ejecución del *pipeline*, dando lugar a una salida estructurada que incluye los parámetros clínicos relevantes, las señales de alarma, las preguntas sugeridas, la prioridad asistencial y los distintos resúmenes clínicos generados.

3.7. Flujo completo de ejecución del sistema

Como parte de esta orquestación, las salidas producidas por el sistema se vuelcan de forma sistemática en ficheros de salida. Por un lado, se genera un fichero final que recoge la salida completa requerida por el sistema, conforme a los requisitos definidos. Por otro lado, se almacenan ficheros intermedios que contienen las salidas parciales generadas por cada uno de los agentes, así como información asociada a su proceso de razonamiento durante la ejecución.

En este contexto, una arquitectura multiagente explícitamente orquestada resulta especialmente adecuada para facilitar futuras extensiones del sistema, ya que las nuevas funcionalidades pueden incorporarse como modificaciones controladas del flujo de ejecución, evitando efectos emergentes difíciles de aislar y validar en un dominio clínico.

Además, este mecanismo permite inspeccionar y analizar de manera individual las contribuciones de cada agente, facilitando la validación clínica de los resultados intermedios y el análisis del comportamiento del sistema. En caso de resultados incorrectos o inesperados, el acceso a estas salidas intermedias y al rastro del razonamiento seguido por los agentes permite identificar de forma precisa en qué punto del flujo se ha producido el fallo o la desviación.

Asimismo, la separación de responsabilidades entre agentes y el registro explícito de sus salidas contribuyen a mejorar la modularidad del sistema y a facilitar futuras extensiones, ajustes o refinamientos de forma controlada. En conjunto, la organización multiagente propuesta proporciona un equilibrio entre especialización, control del razonamiento y trazabilidad, aspectos clave para la aplicación de modelos generativos en un contexto clínico como el de la reumatología.

3.7. Flujo completo de ejecución del sistema

En este apartado se describe de forma integrada y secuencial el flujo completo de ejecución del sistema propuesto, desde la recepción de la consulta clínica en lenguaje natural hasta la generación de la salida estructurada final, pasando por la generación de ficheros intermedios. El objetivo de este apartado es unificar los distintos componentes y mecanismos descritos en secciones anteriores, proporcionando una visión global y coherente del comportamiento del sistema durante su ejecución.

Conviene señalar que el flujo descrito durante el apartado, corresponde exclusivamente al procesamiento de una consulta clínica individual. El sistema incorpora un segundo flujo independiente, descrito en el Apartado 3.4, dedicado a la preparación y almacenamiento del conocimiento clínico externo (segmentación, generación de *embeddings*, indexación e inserción), que se ejecuta de forma previa y desacoplada del procesamiento de las consultas.

La Figura 3.7 muestra de forma esquemática dicho flujo de ejecución, ilustrando las principales fases por las que atraviesa la consulta clínica, incluyendo la extracción estructurada de información, la recuperación de conocimiento clínico externo mediante RAG y el procesamiento multiagente posterior hasta la generación de la salida final.

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

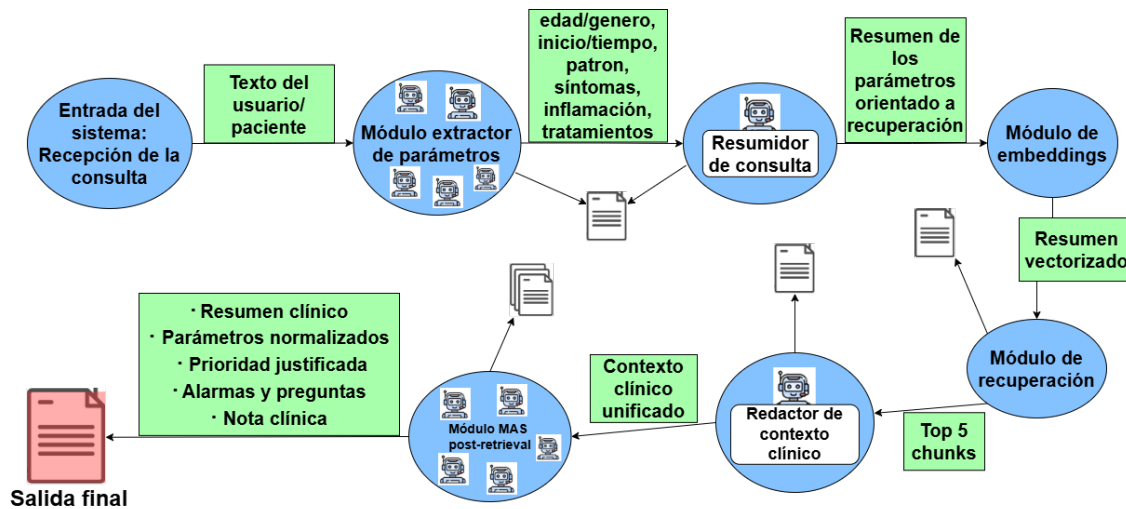


Figura 3.7: Flujo de ejecución de procesamiento de una consulta.

Tal y como se refleja en la arquitectura de componentes software presentada previamente y en el diagrama de flujo correspondiente, el sistema sigue un flujo de procesamiento determinista y reproducible. Todos los pasos se ejecutan de manera secuencial y explícitamente orquestada para cada consulta clínica, sin distinciones condicionales ni activación dinámica de componentes.

3.7.1. Recepción de la consulta y extracción de parámetros clínicos

El flujo comienza con la recepción de la consulta clínica exacta proporcionada por el paciente en lenguaje natural. Este texto constituye la entrada inicial del sistema y es procesado, en una primera fase, por el *Módulo extractor de parámetros* (véase Sección 3.6.2). Dicho módulo activa a los agentes especializados independientes pasándoles el mismo texto de entrada, los cuales extraen los parámetros clínicos recogidos en la Tabla 3.1: la edad y el sexo, inicio y tiempo/duración de los síntomas, el patrón de afectación articular, los síntomas generales o extraarticulares, la inflamación funcional y los tratamientos previos.

Las salidas generadas por estos agentes se integran en una estructura de datos y se registran en un fichero intermedio, permitiendo conservar tanto los parámetros estructurados extraídos como la información asociada al proceso de razonamiento seguido por cada agente extractor durante esta fase inicial.

3.7.2. Generación del resumen orientado a recuperación y recuperación de conocimiento externo

A partir de los parámetros clínicos estructurados, el sistema activa un agente *Resumidor de consulta*, cuyo cometido es generar una descripción breve y focalizada de la información clínica relevante, orientada específicamente al proceso de recuperación de conocimiento externo. Este resumen se construye a partir de los parámetros extraídos de la consulta y se genera en inglés, dado que las

3.7. Flujo completo de ejecución del sistema

guías clínicas empleadas como fuente de conocimiento externo se encuentran redactadas en dicho idioma. En este punto del flujo, se integra el resumen en el mismo fichero intermedio en el que se registraron los parámetros extraídos por el *Módulo extractor de parámetros*, así como la información asociada al proceso de razonamiento seguido por el agente resumidor.

Una vez generado este resumen, el sistema ejecuta el proceso de recuperación de conocimiento externo de forma explícita desde el código, en un punto bien definido del *pipeline*. La consulta resumida se vectoriza mediante un modelo de *embeddings* y se utiliza para interrogar la base de datos vectorial, recuperando un número fijo de fragmentos clínicos relevantes, concretamente los cinco fragmentos más similares desde el punto de vista semántico. Para cada fragmento recuperado, se registra en un fichero intermedio adicional tanto su contenido textual como los metadatos asociados y la puntuación de similitud obtenida con respecto a la consulta.

El diseño y funcionamiento detallado de este mecanismo de recuperación se describen con mayor profundidad en el Apartado 3.5.

3.7.3. Integración del contexto clínico y procesamiento multiagente posterior

Los fragmentos clínicos recuperados son integrados y sintetizados por el agente *Redactor de contexto clínico*, que genera un texto unificado con la información relevante procedente de las guías clínicas. Este contexto clínico sintetizado queda disponible para su incorporación selectiva en fases posteriores del procesamiento, en función de las necesidades de cada agente.

A continuación, el agente *Redactor de consulta contextualizada* integra la consulta original proporcionada por el paciente con el contexto clínico sintetizado, generando un resumen clínico técnico de la consulta con una extensión máxima de 200 palabras. Este resumen está orientado a su uso por el profesional sanitario y constituye una de las salidas principales del sistema. El resultado de esta etapa se almacena en un fichero intermedio dedicado, que recoge tanto la consulta original como su versión contextualizada.

Inmediatamente después, el sistema activa el *Detector de alarmas* inicial, que identifica posibles señales de alarma y propone preguntas de aclaración a partir únicamente de los parámetros clínicos extraídos en fases previas y una lista base de alarmas predefinida, sin hacer uso directo del conocimiento clínico externo recuperado. Las señales de alarma detectadas y las preguntas sugeridas, junto con la información asociada al razonamiento seguido por este agente, se registran en un fichero intermedio específico, que permite conservar el resultado de esta fase inicial de detección.

Sobre la base de la consulta clínica contextualizada, el *Detector de alarmas plus* revisa y refina las señales de alarma y las preguntas de aclaración propuestas inicialmente, pudiendo añadir, eliminar o modificar dichos elementos utilizando la información clínica procedente de las guías. El resultado de esta revisión se

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

registra en un fichero intermedio adicional, manteniendo así un rastro explícito de la evolución de las señales de alarma a lo largo del flujo de ejecución.

Finalmente, el agente *Calculador de prioridad* utiliza de forma conjunta los parámetros clínicos estructurados, las señales de alarma revisadas y la consulta clínica contextualizada enriquecida con conocimiento clínico externo para asignar un nivel de prioridad asistencial, acompañado de una justificación textual que explica los principales factores clínicos considerados. Este cálculo se apoya en un conjunto de reglas clínicas explícitas definidas por el especialista, que asignan distintos pesos a la presencia de determinadas señales de alarma, patrones de afectación, síntomas y grados de limitación funcional, combinando dichas reglas con el razonamiento generativo del modelo de lenguaje. La prioridad calculada y su justificación se almacenan en un fichero intermedio específico, que constituye la salida de esta fase previa a la generación final del resultado del sistema.

3.7.4. Normalización, construcción de la historia clínica y generación de la salida final

En las etapas finales del flujo, el agente *Normalizador de parámetros* se encarga de homogeneizar la terminología empleada y de completar explícitamente aquellos campos no referidos, conforme al formato definido en los requisitos. Este agente no utiliza conocimiento externo recuperado, limitándose a operar sobre los parámetros estructurados previamente obtenidos. Asimismo, se encarga de traducir al español el resumen generado por el agente *Resumidor de consulta*, originalmente redactado en inglés con fines de recuperación de información, de modo que pueda ser utilizado posteriormente en la generación de la salida clínica.

Finalmente, a partir de los parámetros normalizados, el resumen ya traducido al español y la consulta original, el agente *Constructor de la historia clínica* genera una nota clínica breve, orientada a su incorporación directa en la historia del paciente.

El resultado completo del procesamiento se vuelca en un fichero final que recoge la salida estructurada del sistema en el siguiente orden:

1. Consulta original aportada por el paciente
2. Resumen clínico técnico contextualizado
3. Parámetros estructurados normalizados
4. Prioridad asistencial numérica (0–10) con su justificación
5. Señales de alarma identificadas
6. Preguntas sugeridas(máx. 3)
7. Nota clínica para la historia

Como consecuencia del proceso descrito, el sistema produce no solo una salida clínica final conforme a los requisitos definidos por la especialista, sino tam-

3.8. Tecnologías y herramientas empleadas

bién un conjunto de documentos intermedios que recogen de forma explícita las distintas etapas del razonamiento seguido durante la ejecución. Estos ficheros constituyen un rastro detallado del comportamiento del sistema, permitiendo analizar de manera precisa, modular y exhaustiva tanto las decisiones adoptadas por cada agente como el uso del conocimiento clínico externo, lo que resulta fundamental para la evaluación y validación del sistema propuesto.

Este flujo de ejecución integrado completa la descripción metodológica del funcionamiento del sistema durante el procesamiento de una consulta clínica individual. En el siguiente apartado se detallan las tecnologías y herramientas empleadas para la implementación del prototipo descrito, incluyendo los modelos, librerías y componentes software utilizados.

3.8. Tecnologías y herramientas empleadas

En este apartado se describen las principales tecnologías, librerías y herramientas software empleadas para la implementación del prototipo propuesto.

3.8.1. Lenguaje de programación y gestión del proyecto

El sistema ha sido implementado íntegramente en **Python 3.12.2**, como proyecto local, sin capa de *backend web*, es decir, sin exposición de servicios, ni interfaz gráfica. Esta elección responde tanto a la experiencia previa del autor con este lenguaje como a la amplia disponibilidad de librerías para el desarrollo de sistemas basados en modelos de lenguaje, recuperación semántica y bases de datos vectoriales, así como a la facilidad de prototipado y experimentación que ofrece.

La gestión de dependencias se ha realizado mediante el gestor de paquetes **uv**, utilizando un archivo `pyproject.toml` como configuración principal del proyecto y un archivo `uv.lock` para fijar versiones exactas de las dependencias. Este enfoque garantiza la reproducibilidad del entorno de ejecución y facilita la replicación del sistema en otros entornos.

Desde el directorio raíz del proyecto, el punto de entrada principal del sistema corresponde al módulo `src.mas.main`, que orquesta la ejecución completa del *pipeline* descrito en los apartados anteriores. El sistema se ejecuta desde la terminal mediante el siguiente comando:

```
py -m src.mas.main --consulta "texto de la consulta"
```

Este módulo constituye el punto único de entrada del sistema y garantiza una ejecución reproducible del flujo completo descrito en este capítulo, encargándose de inicializar los distintos componentes, coordinar la ejecución secuencial de los agentes y gestionar la generación de los ficheros intermedios y de la salida final del sistema.

El código fuente completo del sistema se encuentra disponible en un repositorio público alojado en *GitHub*, que recoge la implementación íntegra del prototipo, la estructura del proyecto y los ficheros de configuración. El uso de un sistema

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

de control de versiones facilita la trazabilidad de los cambios realizados durante el desarrollo, así como la reproducibilidad de los experimentos y la revisión del código asociado a los resultados presentados.

3.8.2. Modelo de lenguaje generativo empleado

Todos los agentes definidos en el sistema utilizan como modelo generativo subyacente el modelo ***gpt-oss:120b*** en su versión *cloud*, accedido a través de la API de *Ollama*. Este modelo se emplea de forma homogénea en todos los agentes del sistema, diferenciándose su comportamiento únicamente a través de los *prompts* y de las restricciones funcionales asociadas a cada rol.

La elección de *gpt-oss:120b* responde a evidencias recogidas en la literatura reciente y en evaluaciones comparativas de modelos de pesos abiertos (*open-weight*), en las que múltiples fuentes coinciden en señalarlo como el modelo de pesos abiertos más capaz para tareas de razonamiento complejo. En particular, diversos trabajos lo describen como “*the most capable open-weight model for reasoning tasks and specifically the best open model for medical reasoning*” [103], [104], [105], destacando su rendimiento en escenarios que requieren razonamiento clínico, integración de conocimiento estructurado y generación de explicaciones coherentes.

Estas capacidades resultan especialmente relevantes en el contexto del presente trabajo, donde el sistema debe interpretar consultas clínicas potencialmente imprecisas, integrar conocimiento procedente de guías médicas y generar salidas estructuradas y justificadas desde un punto de vista clínico. El uso de un modelo con alta capacidad de razonamiento contribuye a mejorar la coherencia y calidad de las salidas generadas por los distintos agentes, manteniendo al mismo tiempo un enfoque basado en modelos abiertos y reproducibles.

Adicionalmente, la elección de *gpt-oss:120b* viene motivada por su amplia ventana de contexto, de hasta *128 000 tokens*, que resulta especialmente adecuada para el escenario planteado en este trabajo. Esta capacidad permite manejar de forma simultánea *prompts* extensos, consultas clínicas largas, información estructurada generada por múltiples agentes y fragmentos de conocimiento clínico recuperados de guías médicas, sin necesidad de realizar truncamientos agresivos que puedan comprometer la coherencia global del razonamiento.

Cabe señalar que, aunque el modelo se utiliza en modalidad *cloud*, su carácter *open-weight* permite, en principio, su despliegue en entornos controlados. Esta característica resulta coherente con los requisitos de gobernanza, trazabilidad y posible adaptación a entornos clínicos reales considerados en el presente trabajo.

3.8.3. Segmentación semántica y generación de *embeddings*

El proceso de segmentación semántica de las guías clínicas se implementa mediante utilidades de la librería *langchain_experimental*, apoyándose en modelos de *embeddings* disponibles en el ecosistema de *Hugging Face*, que permiten estimar similitudes semánticas entre fragmentos de texto consecutivos. Este enfo-

3.8. Tecnologías y herramientas empleadas

que permite identificar puntos de ruptura semántica entre secciones del texto, generando fragmentos coherentes desde el punto de vista clínico y reduciendo la probabilidad de pérdida de contexto durante la recuperación posterior.

Cabe destacar que, aunque se emplean modelos de *embeddings* tanto en la segmentación como en la vectorización, ambos procesos cumplen funciones conceptualmente distintas dentro del sistema: la segmentación se orienta a preservar coherencia semántica local, mientras que la vectorización se utiliza para habilitar la recuperación semántica. Por este motivo, ambos procesos pueden apoyarse en configuraciones de modelos diferentes.

A lo largo del desarrollo del sistema se han experimentado dos configuraciones distintas de modelos de *embeddings*, materializadas en dos colecciones independientes dentro de la base de datos vectorial:

▪ Colección 1:

- Segmentación semántica: `intfloat/multilingual-e5-large`
- Vectorización: `BAAI/BGE-M3`
- Nombre de la colección: `multilingual-e5-large__and__BGE-M3`

▪ Colección 2:

- Segmentación semántica: `Qwen/Qwen3-Embedding-0.6B`
- Vectorización : `Qwen/Qwen3-Embedding-0.6B`
- Nombre de la colección: `Qwen3-Embedding-0.6B__and__Qwen0.6B`

La elección de estas configuraciones responde, en el primer caso, al uso de modelos ampliamente citados como adecuados para recuperación semántica multilingüe, y en el segundo caso, a resultados reportados en la *Massive Text Embedding Benchmark LeaderBoard* (MTEB) en tareas de *retrieval*. Se intentó inicialmente emplear una variante de mayor tamaño del modelo Qwen; sin embargo, se descartó debido a problemas de asignación de memoria durante su ejecución con aceleración CUDA en el entorno disponible.

3.8.4. Base de datos vectorial

Para el almacenamiento e indexación de los *embeddings* generados se ha utilizado **Qdrant** como base de datos vectorial, mediante la librería `qdrant_client`. Para ello, se desarrolló una clase específica propia llamada `MyQdrantClient`, encargada de gestionar la conexión con Qdrant y de encapsular las operaciones de creación, actualización, eliminación, inserción y consulta de colecciones. Qdrant se ejecuta en un contenedor **Docker** con un volumen persistente montado en el entorno local, lo que permite conservar las colecciones entre ejecuciones del sistema.

Cada colección almacena los vectores asociados a los fragmentos clínicos junto con los metadatos identificativos mencionados anteriormente. Durante la recu-

Capítulo 3. Desarrollo: Planteamiento y desarrollo de la solución propuesta al problema

peración, se emplea la métrica del coseno para medir la proximidad semántica entre la consulta vectorizada y los fragmentos almacenados.

3.8.5. Gestión de salidas y trazabilidad

La implementación del sistema incorpora mecanismos explícitos de registro de resultados mediante la escritura de ficheros de texto durante la ejecución. Estos ficheros incluyen tanto la salida final del sistema como los resultados intermedios generados por los distintos agentes y por el mecanismo de recuperación de información.

Este enfoque, basado en operaciones de escritura secuencial en ficheros y orientado a la construcción de un sistema trazable, permite conservar un rastro detallado del comportamiento del prototipo, facilitando el análisis *a posteriori* de los resultados y constituyendo la base sobre la que se apoyan las herramientas auxiliares de evaluación descritas en el Capítulo 4.

Capítulo 4

Experimentos, evaluación y discusión

4.1. Introducción

El presente capítulo describe el proceso de evaluación del sistema desarrollado, así como los criterios empleados para analizar la calidad de las respuestas generadas y la adecuación del conocimiento clínico recuperado durante su ejecución, con el objetivo de analizar su comportamiento ante un conjunto representativo de escenarios clínicos simulados en el ámbito de la reumatología. El objetivo de la evaluación se orienta a validar la capacidad del sistema para recuperar información clínica relevante, generar respuestas coherentes y seguras, y asignar una prioridad clínica adecuada a cada consulta. Este sistema se ha implementado siguiendo un desarrollo incremental, en el que los distintos módulos fueron diseñados, probados y validados de manera progresiva antes de su integración final.

La evaluación experimental presentada en este capítulo se centra exclusivamente en el sistema completamente integrado, ejecutado sobre un conjunto de escenarios clínicos simulados en reumatología. El objetivo principal es analizar y validar la capacidad del sistema para recuperar información clínica relevante, generar respuestas coherentes y seguras, y asignar una prioridad clínica adecuada a cada consulta, siendo esta última el resultado central del sistema.

Dado el carácter clínico del dominio abordado y la naturaleza generativa del sistema, la evaluación no se apoya exclusivamente en métricas automáticas, sino que combina validaciones funcionales durante el desarrollo con métricas semicuantitativas y análisis cualitativos aplicados en la fase final de evaluación.

4.2. Diseño experimental

4.2.1. Escenarios clínicos de evaluación

La evaluación se realizó sobre un conjunto de veintiún escenarios clínicos simulados, diseñados para representar diversas consultas en el ámbito de la reumatología. Estos escenarios fueron proporcionados por una médico especialista en reumatología y consisten en descripciones de consultas formuladas en lenguaje natural, plausibles en un contexto clínico real y potencialmente expresables por pacientes durante una consulta asistencial. Los escenarios no corresponden a pacientes reales y se generaron con fines exclusivamente académicos.

Para cada caso clínico se proporcionó, además de la consulta en lenguaje natural, una solución de referencia definida por la especialista, que especifica los elementos que el sistema debía identificar o generar correctamente. En concreto, esta solución de referencia incluye la extracción de parámetros clínicos relevantes, la detección de posibles signos de alarma, la prioridad asistencial asignada, una justificación clínica de dicha prioridad y una nota breve orientada a su inclusión en la historia clínica.

Cabe señalar que otros elementos generados por el sistema, como el resumen clínico contextualizado o las preguntas sugeridas para ampliar la anamnesis, no formaban parte de la solución de referencia proporcionada por la especialista. Estos elementos fueron objeto de una evaluación cualitativa por parte del autor, centrada en su coherencia clínica, adecuación al contexto de la consulta y utilidad como apoyo informativo. Sin embargo, la relevancia del contexto clínico recuperado fue evaluada de forma cuantitativa mediante un agente auxiliar específico, y complementada posteriormente con revisión manual por parte del autor. Dado que esta evaluación se apoya en un modelo generativo y no constituye el objetivo principal del análisis cuantitativo del sistema, sus resultados se presentan de forma separada en las secciones posteriores.

Las consultas evaluadas cubren distintos tipos de situaciones clínicas, entre las que se incluyen:

- consultas con síntomas inespecíficos o compartidos entre varias patologías, que requieren un análisis diferencial de posibles causas,
- consultas con presentación clara y síntomas característicos, compatibles de forma directa con una patología reumatológica concreta,
- escenarios de seguimiento clínico,
- y consultas que requieren una priorización asistencial elevada.

Este conjunto de casos permite evaluar tanto la calidad informativa del sistema como su capacidad para identificar situaciones clínicas relevantes y priorizarlas de forma adecuada.

A modo ilustrativo, la Tabla 4.1 recoge algunos ejemplos representativos de las consultas formuladas en lenguaje natural empleadas durante la evaluación.

Tabla 4.1: Ejemplos representativos de consultas clínicas en lenguaje natural empleadas en la evaluación

Tipo de consulta	Consulta formulada en lenguaje natural
Análisis diferencial	Llevo más de un año con dolor por todo el cuerpo; duermo fatal y me levanto agotada. Mis últimas analíticas son normales. El estrés lo empeora.
Presentación clara	Desde hace 10 meses dolor lumbar que me despierta por la noche y mejora al moverme. Rigidez por la mañana. A veces duelen los talones.
Seguimiento clínico	Tuve uveítis anterior hace 2 meses. Desde hace tiempo dolor lumbar que mejora al moverme y me despierta por la noche.
Alta prioridad	Desde hace semanas tengo un dolor fuerte en la sien derecha; al peinarme molesta. Últimamente veo borroso por ese ojo.

4.2.2. Desarrollo incremental y configuración del sistema evaluado

La evaluación se realizó principalmente sobre el sistema completo descrito en el Capítulo 3, una vez finalizado su desarrollo e integración. No obstante, el diseño e implementación del sistema se llevaron a cabo siguiendo un enfoque incremental, en el que los distintos módulos fueron desarrollados y validados de manera progresiva.

En una primera fase se desarrolló el módulo de preparación de conocimiento externo (véase la Figura 3.3). Este proceso comenzó con la implementación del módulo de segmentación semántica, que se validó generando fragmentos textuales (ficheros en formato *.txt*) a partir del conjunto completo de guías clínicas (ficheros en formato *markdown*), obteniéndose un total de **1053 fragmentos**. Posteriormente, se implementó el módulo de generación de *embeddings*, que fue probado de forma aislada para verificar la correcta vectorización de los fragmentos y la coherencia dimensional de los vectores generados.

A continuación, se desarrolló la clase `MyQdrantClient` explicada en la Sección 3.8.4 y se validó inicialmente mediante la inserción y recuperación de fragmentos de texto sintéticos, con el objetivo de verificar tanto los métodos de clase desarrollados, como la correcta persistencia de los datos y la configuración del volumen asociado al contenedor. Una vez validado este comportamiento, se procedió a la inserción del conjunto completo de los fragmentos clínicos generados.

En esta fase se fijó también la configuración definitiva de los modelos de *embeddings* empleados en el sistema. En concreto, se utilizó el modelo *multilingual-e5-large* para el proceso de segmentación semántica de las guías clínicas y el

Capítulo 4. Experimentos, evaluación y discusión

modelo *BGE-M3* para la vectorización de los fragmentos y de las consultas clínicas. Esta combinación fue seleccionada tras una fase preliminar de pruebas funcionales y exploratorias no cuantitativas, al mostrar de forma consistente una mayor coherencia semántica en los fragmentos recuperados frente a configuraciones alternativas. La evaluación presentada en este capítulo se realizó exclusivamente sobre esta configuración.

En una segunda fase se desarrolló el módulo de recuperación de información clínica (Véase la Figura 3.4). El primer componente implementado fue el extractor de parámetros clínicos, cuyo desarrollo implicó la creación y validación individual de los distintos agentes que lo componen (véase la Figura 3.6). Durante esta etapa se realizaron pruebas funcionales iniciales utilizando modelos de menor tamaño para verificar la correcta ejecución del flujo, decidiéndose posteriormente el uso del modelo *gpt-oss:120b* (con parámetro de temperatura igual a 0.5) tras estudiar y comprobar su mayor adecuación para tareas de razonamiento clínico.

Una vez validados los agentes de extracción, se diseñó y probó el *pipeline* conjunto de extracción de parámetros, así como el agente *Resumidor de consulta*. Esto permitió validar el módulo completo de recuperación de información externa, comprobando que el sistema era capaz de recuperar los fragmentos clínicos más relevantes mediante consultas vectoriales. En esta fase se verificó la funcionalidad del proceso de recuperación, sin evaluar aún su calidad semántica.

En relación con el número de fragmentos clínicos recuperados en cada consulta, el valor de $k = 5$ utilizado para la recuperación de los *top-k* fragmentos se mantuvo fijo durante todo el proceso de evaluación. Dicho valor corresponde al definido previamente durante la fase de diseño del sistema y no fue objeto de ajuste ni optimización en esta etapa. De este modo, la evaluación se centró en la validación del flujo completo de procesamiento y en la coherencia de los resultados generados, asumiendo como constante el número de fragmentos recuperados.

Finalmente, se desarrolló el sistema multiagente generador de respuestas (véase la Figura 3.5). Este módulo, explicado en el capítulo anterior, incluye agentes para la detección de señales de alarma, la redacción de contexto clínico, la generación de consultas contextualizadas, la asignación y justificación de prioridad asistencial, la normalización de parámetros y la construcción de una nota breve para la historia clínica. Cada agente fue probado de forma individual cuando la complejidad de la tarea lo requirió y, posteriormente, validado de manera conjunta mediante su integración en un *pipeline* principal (*main_pipeline*), encargado de orquestar la ejecución completa del sistema y registrar los resultados intermedios.

De forma adicional al sistema operativo descrito, se implementó un agente evaluador auxiliar denominado *Agente evaluador de relevancia del contexto*, basado también en el modelo *gpt-oss:120b*. Este agente no forma parte del flujo de ejecución del sistema multiagente ni interviene en la recuperación de información o en la generación de respuestas clínicas.

El agente evaluador se diseñó como una herramienta externa de evaluación,

4.3. Criterios y métricas de evaluación

ejecutada mediante un *script* independiente una vez finalizada la ejecución del `main_pipeline`. Como entrada, recibe el resumen de la consulta clínica empleado para la recuperación en la base de datos vectorial y el resumen del conjunto de fragmentos clínicos recuperados durante la ejecución del sistema. A partir de estas, el agente asigna una puntuación de relevancia en una escala discreta de 0 a 5, donde valores mayores indican una mayor adecuación del contexto recuperado con respecto a la consulta planteada.

La evaluación experimental presentada en este capítulo se realizó sobre este sistema completamente integrado, ejecutando el `main_pipeline` sobre los escenarios clínicos definidos por la especialista reumatóloga.

4.2.3. Protocolo de evaluación del sistema completo integrado

De los veintidós escenarios clínicos, cada uno se ejecutó de manera independiente, manteniendo constantes los parámetros de ejecución del sistema y utilizando en todos los casos la misma configuración previamente fijada para posibilitar la agrupación de los resultados. Cada ejecución correspondió a una pasada completa del `main_pipeline`.

Durante cada ejecución se registraron los fragmentos clínicos recuperados, los resultados intermedios generados por los distintos agentes y la respuesta final generada. Estos registros constituyen la base para la evaluación posterior del sistema.

A partir de los resultados intermedios y finales registrados, se aplicaron los distintos mecanismos de evaluación definidos en este trabajo. En particular, la relevancia del contexto clínico recuperado se evaluó mediante el agente auxiliar descrito en la sección anterior, que se ejecutó de forma independiente tras la finalización del `main_pipeline`.

Las puntuaciones generadas por el agente evaluador fueron posteriormente revisadas por el autor. En aquellos casos en los que se detectaron discrepancias evidentes entre el criterio humano y el del agente, las puntuaciones fueron ajustadas. Cabe señalar que, en la práctica, dichas variaciones no superaron en ningún caso un punto respecto a la puntuación original asignada por el agente.

Finalmente, fue en esta fase cuando se calcularon las métricas cuantitativas asociadas a la asignación de prioridad clínica, utilizando como referencia la prioridad definida por criterio experto.

4.3. Criterios y métricas de evaluación

4.3.1. Criterios clínicos cualitativos

La evaluación clínica de las respuestas generadas por el sistema se llevó a cabo mediante un análisis cualitativo, realizado por el autor, orientado a valorar la coherencia y adecuación de las salidas desde un punto de vista clínico. Dado que no existía una solución de referencia formal para todos los elementos ge-

Capítulo 4. Experimentos, evaluación y discusión

nerados, esta evaluación se centró en aspectos generales de calidad clínica y comunicativa.

En particular, se consideraron los siguientes criterios:

- coherencia clínica global de la respuesta en relación con la consulta planteada,
- adecuación del contenido a las guías clínicas empleadas como fuente de conocimiento,
- identificación correcta de los parámetros clínicos extraídos por el módulo extractor de parámetros,
- identificación correcta de posibles señales de alarma cuando procede,
- justificación acorde a la prioridad asistencial asignada,
- sugerencia de preguntas adecuadas al caso para ampliar la anamnesis,
- nota breve representativa del caso clínico y adecuada para su inclusión en la historia del paciente,
- utilización de terminología clínica concreta,
- uso de un lenguaje prudente, no categórico y no alarmista,
- inclusión de recomendaciones de consulta médica en aquellos casos en los que resulta clínicamente pertinente.

Estos criterios permiten analizar la utilidad potencial del sistema como herramienta de apoyo informativo en un contexto clínico, sin sustituir en ningún caso el juicio clínico profesional ni la valoración individual realizada por personal sanitario cualificado.

4.3.2. Evaluación de la relevancia del contexto recuperado

La calidad del contexto clínico recuperado durante el proceso de búsqueda semántica se evaluó mediante el *Agente evaluador de relevancia del contexto*, introducido en secciones anteriores. Esta evaluación tiene como objetivo analizar en qué medida el conjunto de fragmentos clínicos recuperados resulta adecuado para abordar la consulta clínica planteada.

Para cada escenario clínico, el agente asignó una puntuación de relevancia en una escala discreta de 0 a 5, donde 0 indica un contexto completamente irrelevante y 5 un contexto que aborda de forma completa la consulta. Esta puntuación constituye una medida cuantitativa del grado de alineación semántica entre la consulta y el contexto recuperado.

Las puntuaciones empleadas en el análisis final corresponden a las valoraciones resultantes tras la aplicación del agente evaluador y su posterior revisión por parte del autor, conforme al procedimiento descrito en el protocolo de evaluación del sistema.

Cabe recordar que la evaluación de la relevancia del contexto clínico recuperado se realizó exclusivamente sobre la configuración de *embeddings* basada en la combinación *multilingual-e5-large* para la segmentación semántica y *BGE-M3* para la vectorización, al haber mostrado un comportamiento más consistente durante las pruebas exploratorias preliminares.

4.3.3. Evaluación de la asignación de prioridad clínica

Además de la calidad del contexto recuperado, se evaluó explícitamente la capacidad del sistema para asignar una prioridad clínica adecuada a cada consulta. La prioridad asignada por el sistema se comparó con una prioridad de referencia establecida por criterio experto en reumatología, ambas expresadas en una escala discreta de 0 a 10.

Para cuantificar el grado de ajuste entre ambas prioridades se emplearon dos métricas complementarias:

- **Error absoluto medio** (MAE), que proporciona una medida fácilmente interpretable del error medio,
- **Error cuadrático medio** (RMSE), que penaliza de forma más severa las desviaciones grandes, clínicamente más relevantes.

Sea p_i la prioridad asignada por el sistema y p_i^* la prioridad de referencia para el caso i , las métricas se definen como:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |p_i - p_i^*|$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - p_i^*)^2}$$

Con el fin de facilitar la interpretación de los resultados, el valor del RMSE se normalizó en una escala de 0 a 10 mediante la siguiente expresión:

$$\text{Precisión}_{\text{prioridad}} = 10 \cdot \left(1 - \frac{\text{RMSE}}{10}\right)$$

4.4. Resultados experimentales

4.4.1. Resultados de relevancia contextual

Este resultado indica que, en la mayoría de los casos evaluados, el sistema recupera un contexto clínico parcialmente relevante y útil para abordar la consulta, aunque no siempre cubre la totalidad de los aspectos clínicos implicados. En los escenarios con puntuaciones más altas, los fragmentos recuperados abordaban de forma directa la patología o el problema clínico planteado, mientras que las

Capítulo 4. Experimentos, evaluación y discusión

puntuaciones más bajas se asociaron a consultas especialmente ambiguas o con múltiples dimensiones clínicas.

La evaluación de la relevancia del contexto clínico recuperado se realizó sobre todos los escenarios clínicos definidos. Tras la aplicación del agente evaluador y la revisión posterior de las puntuaciones, se obtuvo una puntuación **media de 3,0 sobre 5**, con una **mediana** también de **3,0** y una **desviación estándar de 1,5**.

La distribución de las puntuaciones refleja una variabilidad moderada entre escenarios. En particular, seis de los veintiún casos evaluados (aproximadamente un 29 %) obtuvieron puntuaciones altas, comprendidas entre 4 y 5, lo que indica una recuperación de contexto clínico altamente relevante. Por el contrario, únicamente dos casos (alrededor de un 10 %) presentaron puntuaciones muy bajas, inferiores a 1, correspondientes a escenarios en los que el contexto recuperado resultó claramente poco alineado con la consulta planteada o demasiado generalista.

Estos resultados sugieren que el sistema es capaz de recuperar, de forma general, un contexto clínico parcialmente relevante que permite apoyar el razonamiento posterior de los agentes generativos, aunque con una variabilidad significativa dependiente del tipo y la complejidad de la consulta. La presencia de casos con baja relevancia pone de manifiesto la necesidad de seguir investigando estrategias de mejora del proceso de recuperación semántica, aspecto que se aborda como línea de trabajo futuro.

4.4.2. Resultados de asignación de prioridad clínica

En relación con la asignación de prioridad clínica, el sistema asignó exactamente la misma prioridad que la definida por criterio experto en **trece de los veintiún** casos evaluados. En los casos restantes, la desviación observada fue en general reducida, situándose mayoritariamente en uno o dos puntos sobre la escala de prioridad, si bien se observó un caso aislado con una desviación de tres puntos.

El error absoluto medio (MAE) obtenido fue de **0,62 puntos**, lo que indica que, de media, el sistema se desvía menos de un punto respecto a la prioridad de referencia definida por la especialista.

Por su parte, el error cuadrático medio (RMSE) fue de **1,1 puntos**. A partir de este valor, se calculó la métrica de precisión normalizada sobre una escala de 0 a 10, obteniéndose una puntuación media de **8,9 sobre 10**. Este resultado refleja una elevada capacidad del sistema para estimar de forma adecuada la urgencia relativa de las consultas clínicas simuladas, incluso en presencia de algunos casos atípicos con desviaciones mayores.

4.4.3. Grado de cumplimiento de los requisitos definidos

Con el fin de evaluar de forma estructurada el cumplimiento de los requisitos definidos en el Capítulo 3, se presenta a continuación un resumen del grado de

4.4. Resultados experimentales

cumplimiento de los requisitos funcionales y no funcionales del sistema, evaluados a partir de los resultados obtenidos y del comportamiento observado durante la fase experimental. La Tabla 4.2 recoge el grado de cumplimiento de los requisitos funcionales, mientras que la Tabla 4.3 el de los requisitos no funcionales.

Capítulo 4. Experimentos, evaluación y discusión

Tabla 4.2: Grado de cumplimiento de los requisitos funcionales del sistema

ID	Requisito funcional	Cumplimiento
RF1	Permitir la entrada de texto con la descripción aportada por el paciente	Cumplido
RF2	Extraer parámetros clínicos relevantes (edad, sexo, inicio y tiempo, patrón, inflamación funcional, síntomas generales/extraarticulares, tratamientos previos)	Cumplido
RF3	Indicar explícitamente “No referido” cuando falte información y añadir una etiqueta de completitud de los parámetros	Parcial ¹
RF4	Asignar una prioridad asistencial entre 0 y 10 con justificación clínica, indicando información incompleta cuando procede	Cumplido
RF5	Calcular la prioridad en base a las reglas clínicas proporcionadas por el especialista	Cumplido
RF6	Generar una lista de señales de alarma, indicando “Ninguna referida” cuando no existan	Cumplido
RF7	Generar un resumen clínico breve (máx. 200 palabras)	Cumplido
RF8	Generar una nota breve (2–3 líneas) orientada a la historia clínica	Cumplido
RF9	Sugerir hasta tres preguntas de aclaración para ampliar la anamnesis	Cumplido
RF10	No realizar diagnósticos ni pautar tratamientos	Cumplido
RF11	No inferir datos no proporcionados explícitamente por el paciente	Cumplido
RF12	Utilizar guías clínicas como fuente de conocimiento (EULAR, ACR; SER opcional)	Cumplido ²
RF13	Proporcionar una plantilla de salida estructurada	Cumplido

¹No se ha añadido etiqueta numérica de completitud.

²Las guías de la SER no se han incluido en el sistema.

4.4. Resultados experimentales

Tabla 4.3: Grado de cumplimiento de los requisitos no funcionales del sistema

ID	Requisito no funcional	Cumplimiento
RNF1	Uso de terminología clínica homogénea	Parcial ¹
RNF2	Mantener expresiones vagas del paciente sin inferir información adicional	Cumplido
RNF3	Cumplimiento de normativa de seguridad y privacidad (RGPD, LOPDGDD)	No cumplido
RNF4	Cifrado de información en tránsito y en reposo.	No cumplido
RNF5	Confiabilidad clínica y minimización de alucinaciones, restringiendo respuestas a evidencia recuperada	Cumplido según lo probado
RNF6	Referenciar las guías utilizadas para el procesamiento de una consulta	Cumplido
RNF7	Tolerancia a fallos en el sistema multiagente	Parcial ²
RNF8	Rendimiento del sistema (latencia inferior a umbrales definidos)	Parcial. ³
RNF9	Disponibilidad continua del sistema	Cumplido
RNF10	Mantenibilidad y actualización mediante diseño modular y estrategia de despliegue sin interrupción del servicio	Cumplido
RNF11	Registro de actividad y trazabilidad para auditoría	Cumplido

¹ Diferencias puntuales entre iteraciones (p.e., “astenia” y “fatiga”)

² Existen casos en los que no se ha corregido algún fallo.

³ Cumplido para extracción de parámetros; no para la generación final.

4.5. Discusión

Los resultados obtenidos permiten realizar una valoración global del comportamiento del sistema desarrollado en un contexto clínico simulado. En primer lugar, la evaluación de la relevancia del contexto recuperado muestra que el sistema es capaz de proporcionar, en términos generales, un conocimiento clínico parcialmente relevante y útil para apoyar el razonamiento posterior de los agentes generativos. La puntuación media de 3.0 sobre 5, junto con una mediana coincidente, indica un comportamiento razonablemente consistente, aunque con una variabilidad significativa dependiente del tipo y la complejidad de la consulta. En este sentido, la puntuación media de relevancia contextual sugiere un equilibrio razonable entre cobertura y precisión, teniendo en cuenta la complejidad del dominio y la ambigüedad inherente a muchas consultas clínicas.

La variabilidad de 1.5 observada en las puntuaciones de relevancia contextual no se manifiesta únicamente en forma de desviaciones negativas respecto a la media, sino también mediante la presencia de escenarios con valoraciones claramente superiores, lo que refleja un comportamiento heterogéneo del proceso de recuperación. En general, las puntuaciones intermedias y bajas suelen corresponder a consultas con sintomatología inespecífica, múltiples dimensiones clínicas o varias patologías plausibles, en las que la recuperación semántica tiende a priorizar fragmentos relacionados pero no siempre completamente alineados con el foco principal de la consulta. No obstante, también se identificaron casos puntuales de baja relevancia contextual en consultas con patrones clínicos bien definidos, en los que el sistema no logró recuperar información claramente relacionada con la sintomatología descrita.

Estos resultados evidencian que, si bien el sistema es capaz de proporcionar un contexto clínico generalmente útil y, en algunos casos, altamente relevante, la fase de recuperación semántica constituye uno de los principales puntos de variabilidad del enfoque propuesto. Este comportamiento sugiere que la eficacia de la recuperación no depende únicamente de la arquitectura empleada, sino de factores críticos de la implementación técnica; en particular, la granularidad de la segmentación empleada sobre las guías clínicas y el rendimiento de los modelos de *embeddings* seleccionados influyen directamente en la precisión del proceso. Por lo tanto, estos elementos involucrados en la recuperación semántica constituyen uno de los principales focos de estudio para la mejora del prototipo.

El contraste entre la calidad moderada de la recuperación contextual y el buen rendimiento en la asignación de prioridad sugiere que el sistema multiagente es capaz de compensar, al menos parcialmente, las limitaciones del contexto recuperado mediante mecanismos de razonamiento, detección de señales de alarma y control estructurado de la generación. Esta descomposición del proceso de generación en agentes especializados, parece contribuir a un mayor control del razonamiento y de la estructura de las respuestas, favoreciendo salidas más coherentes, justificadas y seguras desde un punto de vista clínico. Este resultado refuerza la idoneidad del enfoque multiagente como capa de control y mejora sobre sistemas de recuperación aumentada por generación tradicionales.

Asimismo, el uso combinado de agentes evaluadores y revisión humana refuerza la robustez del enfoque combinado propuesto, y ha permitido identificar la importancia de integrar capas adicionales de supervisión y validación del conocimiento empleado por el sistema. La evaluación explícita de la relevancia del contexto recuperado ha permitido no solo analizar el comportamiento del módulo de recuperación, sino también identificar escenarios en los que resulta necesario reforzar los mecanismos de control, más allá de una arquitectura basada exclusivamente en RAG.

Finalmente, el análisis del grado de cumplimiento de los requisitos confirma que el sistema satisface de forma global los objetivos funcionales planteados. No obstante, los resultados obtenidos también ponen de manifiesto una serie de limitaciones funcionales relevantes como la sensibilidad del proceso de recuperación a la formulación de la consulta y al proceso de fragmentación. Asimismo, algunos requisitos no funcionales asociados a seguridad, privacidad y optimización de costes no han sido abordados en el prototipo actual. Estas limitaciones no invalidan los resultados alcanzados, pero delimitan claramente el alcance del sistema desarrollado y orientan de forma directa las líneas de trabajo futuro.

4.5.1. Limitaciones identificadas

El presente trabajo presenta una serie de limitaciones que deben ser consideradas a la hora de interpretar los resultados obtenidos. En primer lugar, en lo relativo a la validez interna, parte de la evaluación se apoya en mecanismos basados en modelos de lenguaje, como el agente evaluador de relevancia contextual, cuyas valoraciones, aunque son revisadas por el autor, pueden introducir sesgos inherentes a este tipo de modelos.

Desde el punto de vista de la validez externa, la evaluación se ha realizado sobre un conjunto limitado de veintiún escenarios reumatológicos clínicos simulados. Aunque estos escenarios fueron definidos por una especialista y diseñados para ser representativos, el tamaño de la muestra y su enfoque en una única especialidad limitan la generalización de los resultados a otros dominios clínicos o a la variabilidad extrema de un entorno asistencial real.

Una limitación técnica específica de la implementación viene dada por el proceso de traducción realizado por los agentes. La síntesis de la consulta realizada por el agente resumidor en inglés y la posterior normalización al español introducen puntos críticos donde podrían producirse pérdidas de matices. No obstante, es reseñable que en ninguno de los veintiún casos evaluados se detectaron errores de traducción o interpretación que afectaran negativamente al razonamiento clínico o a la asignación de prioridad, lo que valida la robustez del flujo diseñado para este conjunto de datos.

Por otro lado, es necesario reconocer que el proceso de decisión médica constituye una actividad multidimensional que excede la capacidad de procesamiento de cualquier sistema basado exclusivamente en representaciones textuales. Al operar sobre datos escritos, el sistema carece de la facultad para integrar factores cualitativos críticos para el diagnóstico, tales como la interpretación del lenguaje

Capítulo 4. Experimentos, evaluación y discusión

no verbal del paciente durante el acto clínico. Por consiguiente, el alcance del prototipo se limita a una interpretación técnica de la evidencia documentada, sin pretender capturar la totalidad de los matices y la profundidad inherentes a una evaluación médica presencial.

En cuanto a la validez de las métricas empleadas, como la puntuación de relevancia contextual o la prioridad clínica discretizada, son aproximaciones diseñadas exclusivamente para permitir una evaluación consistente del sistema en un entorno controlado. Estas medidas facilitan la cuantificación del comportamiento del sistema y su capacidad para estimar la urgencia relativa de las consultas dentro del marco del experimento, pero no pretenden equipararse a los estándares de evaluación multidimensional de la práctica clínica real. Por tanto, dichas puntuaciones deben interpretarse como indicadores de rendimiento técnico específicos para este prototipo, cuya validez se circunscribe al contexto experimental planteado y no a una medida definitiva de eficacia en un entorno asistencial no simulado.

Este enfoque resulta coherente con el objetivo del presente trabajo, que no persigue la automatización completa del proceso asistencial, sino la investigación y validación de un sistema capaz de integrarse de forma supervisada en flujos clínicos reales, con el fin de optimizar fases concretas del proceso clínico, como la clasificación y priorización inicial de consultas. En este sentido, el sistema desarrollado se concibe como un prototipo de investigación orientado a la evaluación conceptual y funcional del enfoque propuesto. Determinados aspectos necesarios para un despliegue clínico completo, como la integración con sistemas de información sanitaria, el cumplimiento de normativas de seguridad y privacidad o la evaluación de rendimiento en condiciones reales de uso, no han sido abordados en esta fase del trabajo. Estas consideraciones no afectan a la validez de los resultados obtenidos en el contexto experimental planteado, pero delimitan el alcance del estudio y se identifican como líneas de trabajo futuro.

Los resultados experimentales y el análisis crítico presentados a lo largo de este capítulo han permitido evaluar de forma global el comportamiento del sistema propuesto, así como identificar sus principales fortalezas, limitaciones y puntos de mejora. Sobre la base de estos resultados, el capítulo siguiente presenta las conclusiones generales del trabajo, valorando el cumplimiento de los objetivos planteados y las líneas de trabajo futuro.

Capítulo 5

Conclusiones y Trabajo Futuro

5.1. Conclusiones

El desarrollo de este Trabajo Fin de Grado ha permitido constatar la viabilidad y el potencial de las arquitecturas híbridas RAG+MAS para revolucionar procesos clínicos complejos, como la anamnesis y el triaje en reumatología. Los resultados alcanzados confirman la satisfacción de los objetivos definidos al inicio del proyecto, habiéndose logrado un prototipo funcional capaz de estructurar parámetros clínicos, detectar señales de alarma y generar notas clínicas coherentes mediante la integración efectiva de las guías de referencia. Además, el éxito alcanzado en la asignación de prioridad asistencial, con una precisión normalizada de 8,9 sobre 10, demuestra que la descomposición del razonamiento en agentes especializados no solo aporta modularidad, sino que actúa como una capa de control crítica capaz de filtrar imprecisiones en la recuperación de información externa.

A pesar de los resultados positivos, se han identificado diversas líneas de investigación que permitirán evolucionar el prototipo actual hacia un sistema más robusto y escalable:

- **Evolución de la arquitectura de recuperación:** Se plantea la incorporación de una capa de reevaluación (*Reranking RAG*) [106]) para optimizar la selección de fragmentos recuperados y la exploración de enfoques *GraphRAG* [107]. Estos permitirían capturar relaciones semánticas complejas entre síntomas y patologías en documentos extensos (como son las guías médicas) que la búsqueda vectorial tradicional no logra identificar con la misma precisión.
- **Integración del agente evaluador en un bucle de realimentación:** Se podría integrar el agente evaluador de relevancia del contexto como parte activa del sistema, en lugar de emplearlo exclusivamente con fines evaluativos. En este escenario, la puntuación de relevancia contextual podría utilizarse como señal de control para detectar automáticamente recuperaciones de baja calidad y activar mecanismos de realimentación, como la repetición del proceso de recuperación, o la reformulación de la consulta empleada.

Esta integración permitiría avanzar hacia un sistema más adaptativo y robusto, capaz de supervisar de forma continua la calidad del conocimiento utilizado durante la generación y de mitigar de manera automática algunos de los fallos identificados en la fase de recuperación semántica.

- **Optimización del rendimiento y reducción de latencia:** Debido al alto tiempo de latencia obtenido para la respuesta final del sistema, resulta prioritario estudiar estrategias de optimización de tiempos de respuesta. Esto incluye la exploración de técnicas de computación asíncrona en el flujo de agentes, el uso de modelos cuantizados o de menor escala para tareas de extracción sencillas, y la implementación de mecanismos de *streaming* de respuestas para mejorar la experiencia de uso en el entorno asistencial.
- **Evaluación del impacto en la eficiencia asistencial y administrativa.** Resultaría de gran interés estudiar el impacto del sistema tanto en la eficiencia asistencial como en la eficiencia administrativa del proceso clínico. En particular, evaluar en qué medida el uso del sistema contribuye a optimizar tareas como la clasificación inicial de consultas, la estructuración automática de la información clínica o la identificación temprana de casos prioritarios, reduciendo la carga administrativa asociada a la atención sanitaria. Asimismo, podría evaluarse el impacto del sistema en la eficiencia de la propia consulta médica, al liberar tiempo del profesional para actividades estrictamente clínicas, así como su influencia en la carga de trabajo y en la calidad de vida laboral de los profesionales sanitarios.
- **Diseño de métricas clínicas asistenciales:** Se propone colaborar con especialistas médicos en el diseño de métricas de evaluación personalizadas que permitan penalizar con mayor peso aquellos fallos que afecten directamente a la seguridad clínica, diferenciándolos de imprecisiones administrativas o de formato.
- **Evaluación con profesionales clínicos:** Otra línea de trabajo futuro fundamental es la evaluación del sistema con la participación directa de profesionales clínicos. La validación por parte de médicos especialistas permitiría analizar de forma cualitativa la utilidad real del sistema, la adecuación clínica de las respuestas generadas y la coherencia de la prioridad asistencial sugerida. Este tipo de evaluación aportaría una perspectiva experta imprescindible para refinar tanto los criterios de priorización como los mecanismos de generación y control del sistema, contribuyendo a una mejor alineación con las necesidades reales del entorno asistencial.
- **Capa de seguridad y despliegue:** Futuras versiones deberán abordar de forma explícita la implementación de normativas de privacidad (RGPD y LOPDGDD) y seguridad de la información de salud protegida (PHI), así como el estudio de su integración técnica con sistemas de información hospitalaria reales.

Más allá de los logros técnicos alcanzados, este trabajo aporta reflexiones relevantes sobre la integración de la Inteligencia Artificial en la sanidad. La experiencia desarrollada confirma que, si bien los modelos de lenguaje generativos

ofrecen capacidades avanzadas de síntesis y razonamiento, por el momento su valor real en contextos clínicos reside en amplificar y asistir el juicio profesional, no en sustituirlo. De forma análoga a la arquitectura multiagente propuesta, cuya eficacia se basa en la cooperación entre agentes especializados, el futuro de la tecnología médica pasa necesariamente por una colaboración estrecha entre disciplinas y por una integración responsable de sistemas inteligentes en los flujos asistenciales.

En este sentido, el presente trabajo demuestra que es posible diseñar y evaluar sistemas híbridos RAG+MAS capaces de aportar estructura, trazabilidad y control a procesos clínicos complejos, sentando las bases para futuras investigaciones orientadas a su adopción supervisada en entornos reales. El prototipo desarrollado no constituye un punto final, sino un punto de partida que invita a seguir explorando el papel de la Inteligencia Artificial como herramienta para mejorar la eficiencia, la seguridad y la calidad de la atención sanitaria, siempre desde una perspectiva centrada en el profesional clínico y en el beneficio del paciente.

Capítulo 6

Análisis de impacto

En este capítulo se analiza el impacto potencial de los resultados obtenidos en el desarrollo de este Trabajo Fin de Grado. El sistema diseñado no solo representa un trabajo de investigación en la aplicación de arquitecturas RAG y MAS, sino que conlleva implicaciones significativas en diversos ámbitos, desde el personal hasta el medioambiental.

Impacto Personal

La realización de este Trabajo Fin de Grado ha representado un hito fundamental en mi trayectoria académica, suponiendo el reto de diseñar y materializar un proyecto técnico complejo desde su concepción inicial a partir de una idea propuesta por un médico especialista. El desarrollo del sistema ha requerido liderar todas las fases de un ciclo de vida de proyecto completo: desde la planificación temporal y de actividades, hasta la ejecución técnica, el diseño de la metodología de evaluación y la elaboración de la documentación final. Especialmente significativo ha sido el desafío de diseñar desde cero módulos específicos como el de preparación de conocimiento externo y el sistema multiagente (MAS) orquestado, lo que ha permitido aplicar de forma integrada las competencias y técnicas adquiridas durante el Grado en Matemáticas e Informática.

A nivel técnico, el proyecto me ha brindado una experiencia de alto valor en el planteamiento y ejecución de soluciones basadas en tecnologías modernas como RAG y MAS, profundizando en el estudio de modelos de embeddings y en la gestión de bases de datos vectoriales con QDrant. Asimismo, el desarrollo del prototipo ha permitido ganar experiencia en la gestión de recursos de computación de alto rendimiento, incluyendo la administración de procesos con aceleración CUDA en GPU y el uso avanzado de la *shell* de *Linux* para la monitorización de procesos y la interconexión entre máquinas locales y remotas.

Finalmente, la oportunidad de investigar en el sector de la tecnología aplicada al ámbito sanitario ha resultado realmente enriquecedora y motivadora. Esta experiencia no solo ha consolidado mi capacidad para resolver problemas técnicos en entornos críticos, sino que ha despertado un profundo interés por la inves-

tigación y el desarrollo de futuros proyectos en el área, donde el impacto de la ingeniería informática y de las matemáticas tiene un beneficio directo y tangible en la sociedad.

Impacto Empresarial

Desde el punto de vista empresarial, el sistema desarrollado presenta un interés potencial para organizaciones del ámbito sanitario y tecnológico, especialmente aquellas involucradas en el desarrollo de soluciones de apoyo a la decisión clínica, gestión de la información médica o automatización de procesos asistenciales. Aunque el prototipo no está concebido como un producto comercial listo para su despliegue, sí demuestra la viabilidad técnica de arquitecturas híbridas RAG+MAS aplicadas a contextos clínicos complejos, lo que puede servir como base para futuras iniciativas empresariales o proyectos tecnológicos.

No obstante, el impacto empresarial también conlleva una serie de desafíos y posibles efectos adversos. La adopción de una solución de estas características implica una inversión inicial en infraestructura tecnológica de alto rendimiento, así como costes recurrentes asociados al mantenimiento del sistema y a la inferencia de modelos de lenguaje de gran escala. Asimismo, la integración técnica con los sistemas de información hospitalaria existentes (HIS) requiere una planificación cuidadosa y puede suponer, al menos de forma temporal, una alteración de los flujos de trabajo establecidos dentro de la organización.

Con el objetivo de mitigar estos riesgos, a lo largo del diseño del sistema se han tomado decisiones orientadas a maximizar la sostenibilidad empresarial del enfoque propuesto. En este sentido, la adopción de una arquitectura modular y escalable permite una evolución controlada del sistema, ajustando progresivamente su complejidad y capacidad en función de las necesidades reales de la organización. Además, la apuesta por el uso de un modelo de lenguaje de pesos abiertos, como *gpt-oss:120b*, posibilita el despliegue del sistema en infraestructuras propias de la institución sanitaria. A diferencia de los modelos accesibles exclusivamente mediante APIs externas (p.e., GPT-4), este enfoque permite ejecutar el procesamiento de la información clínica de forma local, evitando la transferencia de datos sensibles a terceros y permitiendo un mayor control sobre los costes derivados del uso de servicios comerciales.

Impacto Social y Cultural

El impacto social y cultural del trabajo se enmarca en la progresiva transformación de los procesos asistenciales y en la forma en que la inteligencia artificial comienza a integrarse en ámbitos tradicionalmente dominados por la interacción humana directa, como es la práctica clínica. El sistema desarrollado explora el potencial de la IA como herramienta de apoyo al profesional sanitario, con el objetivo último de mejorar la eficiencia, la seguridad y la estructuración del proceso asistencial, sin alterar el papel central del juicio clínico.

Desde un punto de vista social, este tipo de sistemas puede tener un impacto

positivo al favorecer una atención más ordenada y equitativa, ayudando a identificar de manera temprana situaciones de mayor urgencia y reduciendo la carga administrativa asociada a la gestión inicial de consultas. En contextos sanitarios caracterizados por una elevada presión asistencial, estas herramientas podrían contribuir indirectamente a mejorar la calidad percibida de la atención, al permitir que los profesionales dediquen más tiempo a tareas estrictamente clínicas y menos a labores repetitivas o administrativas, favoreciendo una mejora en la interacción médico-paciente y contribuyendo potencialmente a una mejor calidad de vida laboral de los profesionales sanitarios, al mitigar factores asociados a la sobrecarga de trabajo y al riesgo de desgaste profesional.

No obstante, la introducción de sistemas basados en inteligencia artificial en el ámbito sanitario también plantea retos sociales relevantes. Existe el riesgo de generar una confianza excesiva en sistemas automatizados o de que estos sean percibidos como sustitutos del juicio clínico, lo que podría derivar en un uso inapropiado de la tecnología o en un sentimiento de rechazo general por parte de la sociedad hacia estas tecnologías. Por este motivo, el trabajo adopta explícitamente un enfoque en el que la IA se concibe como un elemento de apoyo y no como un agente decisor autónomo, reforzando la necesidad de supervisión humana y responsabilidad profesional.

En este sentido, el proyecto contribuye a fomentar una visión crítica y responsable de la inteligencia artificial aplicada a la salud. El trabajo pone de relieve tanto las capacidades como las limitaciones de los modelos de lenguaje, promoviendo una cultura tecnológica basada en la cautela, la trazabilidad y el respeto por el conocimiento experto. Además, promueve una cultura de colaboración entre humanos y máquinas, en la que la IA amplifica las capacidades de las personas sin sustituir su criterio. Estos enfoques resultan coherentes con la evolución cultural necesaria para una adopción eficiente y sostenible de la tecnología en la sociedad, especialmente en dominios sensibles como el sanitario.

Asimismo, el carácter interdisciplinar del trabajo refuerza una cultura de colaboración entre perfiles técnicos y clínicos, subrayando la importancia del diálogo entre disciplinas para el diseño de soluciones tecnológicas realmente útiles y socialmente aceptables. Esta visión de la colaboración se refleja también en el propio enfoque del sistema multiagente propuesto, en el que agentes especializados cooperan de forma coordinada para alcanzar un objetivo común, aportando capacidades complementarias dentro de un proceso estructurado, controlado y supervisado.

Finalmente, el carácter académico del proyecto y su orientación investigadora contribuyen a la formación de profesionales capaces de comprender tanto las capacidades como las limitaciones de la inteligencia artificial, lo que resulta esencial para una adecuada integración de estas tecnologías en la sociedad. En este contexto, el trabajo pone en valor la investigación continua como motor para mantener una cultura orientada al progreso, en la que el avance tecnológico se apoye en la evaluación crítica, la validación rigurosa y la mejora incremental de los sistemas desarrollados, favoreciendo una adopción responsable y evolutiva de la inteligencia artificial en la sociedad.

Impacto económico

El impacto económico de este trabajo se encuentra intrínsecamente relacionado con el impacto empresarial previamente analizado, si bien se evalúa aquí en términos de sostenibilidad financiera, optimización de recursos y retorno de inversión para el sistema sanitario. La implementación de un sistema de asistencia a la anamnesis basado en inteligencia artificial tiene el potencial de generar ahorros económicos tanto directos como indirectos dentro del proceso asistencial.

En primer lugar, el beneficio económico más directo se deriva de la mejora en la eficiencia operativa. Al automatizar la síntesis y estructuración de la información clínica, se reduce el tiempo que el personal facultativo debe dedicar a labores administrativas por consulta. Esta redistribución del tiempo de trabajo permitiría un uso más eficiente de los recursos humanos, uno de los factores de mayor coste dentro de los sistemas sanitarios, sin que ello implique necesariamente una reducción de personal, sino una reasignación hacia actividades de mayor valor añadido e impacto asistencial. Además, la capacidad de priorización asistencial del sistema contribuye a una gestión más eficiente de la demanda, lo que puede suponer un ahorro preventivo para el sistema de salud al identificar precozmente casos graves que, de otro modo, podrían derivar en complicaciones clínicas mucho más costosas de tratar en estadios avanzados.

No obstante, es necesario considerar también los costes asociados a la adopción de este tipo de tecnologías. El uso de modelos de lenguaje de gran escala y de infraestructuras computacionales avanzadas implica inversiones significativas en hardware, mantenimiento y consumo energético que deben ser auditadas, así como costes asociados a la validación, integración y supervisión del sistema. Por ello, cualquier evaluación económica exhaustiva deberá realizarse en fases posteriores mediante estudios específicos en entornos reales de uso que permitan cuantificar de forma rigurosa el balance final entre los costes y los beneficios derivados de su adopción.

Impacto medioambiental

En lo relativo al impacto medioambiental, el proyecto presenta un balance entre los beneficios derivados de la transformación digital y los desafíos energéticos inherentes a la inteligencia artificial moderna.

En este sentido, el uso de modelos de gran escala y de aceleración mediante GPU conlleva un consumo energético asociado que no puede ser ignorado. Este consumo representa un efecto adverso directo en términos de emisiones de CO₂ de los centros de datos encargados del procesamiento. No obstante, al tratarse de un prototipo de investigación y no de un sistema desplegado a gran escala, dicho impacto resulta acotado y puntual. Además, al usar una arquitectura RAG, y procesar únicamente los fragmentos de conocimiento clínico más relevantes en lugar de documentos médicos extensos, se reduce la carga computacional por consulta y, por ende, el gasto energético.

De forma indirecta, si sistemas de este tipo llegaran a integrarse en entornos

asistenciales reales, podrían contribuir a una mayor eficiencia global del sistema sanitario, reduciendo duplicidades administrativas, desplazamientos innecesarios o ineficiencias en la gestión de la información clínica. Estos efectos, aunque secundarios y difíciles de cuantificar, podrían tener un impacto medioambiental positivo. No obstante, cualquier valoración significativa del impacto ambiental requeriría estudios específicos en escenarios reales de despliegue, que exceden el alcance de este trabajo.

Alineación con los ODS

El presente trabajo se alinea con la Agenda 2030 de las Naciones Unidas, contribuyendo a varios de los Objetivos de Desarrollo Sostenible (ODS) [108]. Los impactos más significativos se identifican en los siguientes:

- * **ODS 3 Salud y Bienestar:** El prototipo de triaje permite agilizar la recopilación y síntesis de datos médicos, lo que favorece diagnósticos más rápidos y precisos. Esto facilita una respuesta más rápida ante patologías graves, optimizando los resultados de salud para los pacientes y apoyando una toma de decisiones basada en la evidencia clínica.
- * **ODS 8 Trabajo Decente y crecimiento económico:** El sistema desarrollado se alinea indirectamente con este objetivo al abordar el problema del desgaste profesional o *burnout* médico. La automatización parcial de tareas administrativas y documentales contribuye a mejorar las condiciones de trabajo del personal sanitario, favoreciendo un uso más eficiente del tiempo y reduciendo factores asociados a la sobrecarga laboral. Desde una perspectiva económica, estas mejoras pueden repercutir positivamente en la productividad del sistema sanitario sin comprometer la calidad asistencial, promoviendo un crecimiento más sostenible basado en la optimización de recursos humanos y en la mejora del bienestar laboral, en línea con los principios del trabajo decente recogidos en el ODS 8.
- * **ODS 9 Industria, Innovación e Infraestructura:** La aplicación de técnicas avanzadas de IA (RAG+MAS) en el apoyo al triaje clínico inicial contribuye a la innovación en la industria de la salud, fortaleciendo la infraestructura digital clínica y facilitando la toma de decisiones con herramientas inteligentes de asistencia.
- * **ODS 17 Alianzas para lograr los objetivos:** El presente trabajo se alinea con este ODS al poner de manifiesto la importancia de la colaboración entre perfiles técnicos y profesionales sanitarios para el diseño de soluciones tecnológicas realmente útiles y responsables. El desarrollo del sistema ha partido del diálogo con un médico especialista, integrando conocimiento clínico y capacidades técnicas en un enfoque conjunto orientado a mejorar la eficiencia y la sostenibilidad del sistema sanitario. Este modelo de cooperación interdisciplinar favorece una integración más segura y efectiva de la inteligencia artificial en la práctica clínica, contribuyendo de forma indirecta a una atención sanitaria de mayor calidad y, en consecuencia, a una población más segura y con una mejor calidad de vida.

Capítulo 6. Análisis de impacto

Por último, es posible establecer una relación entre el proyecto y el **ODS 12, Producción y consumo responsables**, desde una perspectiva exploratoria. En particular, el enfoque basado en arquitecturas RAG podría presentar, en determinados escenarios, un impacto potencialmente más eficiente en términos de consumo energético frente a estrategias basadas en el entrenamiento o *fine-tuning* masivo de modelos de lenguaje de gran escala.

No obstante, esta posible vinculación debe interpretarse con cautela, ya que una evaluación rigurosa del impacto en términos de consumo energético y sostenibilidad requeriría estudios específicos que comparen de forma cuantitativa el coste computacional de las distintas estrategias. Dicho análisis excede el alcance del presente trabajo, por lo que la vinculación con el ODS 12 se plantea únicamente como una potencial contribución de este sistema.

Bibliografía

- [1] Eurostat. «Use of artificial intelligence in enterprises», visitado 2 de nov. de 2025. dirección: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Use_of_artificial_intelligence_in_enterprises.
- [2] Eurostat. «Glossary: Enterprise Size», visitado 2 de nov. de 2025. dirección: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Enterprise_size.
- [3] S. I. for Human-Centered Artificial Intelligence. «Artificial Intelligence Index Report 2025», visitado 3 de nov. de 2025. dirección: https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf.
- [4] I. T. Institute. «7 ejemplos de uso de inteligencia artificial en nuestro día a día», visitado 3 de nov. de 2025. dirección: <https://immune.institute/blog/7-ejemplos-de-uso-de-inteligencia-artificial-en-nuestro-dia-a-dia/>.
- [5] M. bibinitperiod Company. «Superagency in the workplace: Empowering people to unlock AI's full potential», visitado 3 de nov. de 2025. dirección: <https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work>.
- [6] F. S. Institute. «AI: Revolutionizing the Healthcare Landscape, making it more accessible», visitado 3 de nov. de 2025. dirección: <https://frostandullivaninstitute.org/ai-revolutionizing-the-healthcare-landscape-making-it-more-accessible/>.
- [7] M. Chen y M. Decary, «Artificial intelligence in healthcare: An essential guide for health leaders», *Healthcare Management Forum*, vol. 33, n.º 1, págs. 10-18, 2020. dirección: <https://doi.org/10.1177/0840470419873123>.
- [8] J. Morley et al., «The ethics of AI in health care: A mapping review», *Social Science & Medicine*, vol. 260, pág. 113 172, 2020, ISSN: 0277-9536. DOI: <https://doi.org/10.1016/j.socscimed.2020.113172>. dirección: <https://www.sciencedirect.com/science/article/pii/S0277953620303919>.

- [9] M. de Sanidad y Política Social, *Orden SAS/2855/2009, de 9 de octubre, por la que se aprueba y publica el programa formativo de la especialidad de Reumatología*. Boletín Oficial del Estado (BOE), núm. 258, 26 de octubre de 2009, páginas 89605 a 89621, Consultado el 4 de noviembre de 2025, 2009. dirección: <https://www.boe.es/eli/es/o/2009/10/09/sas2855/dof/spa/pdf>.
- [10] S. E. de Reumatología. «Anamnesis, exploración e imagen: las claves para resolver patologías musculoesqueléticas de alta prevalencia», visitado 4 de nov. de 2025. dirección: <https://www.ser.es/anamnesis-exploracion-e-imagen-las-claves-para-resolver-patologias-musculoesqueleticas-de-alta-prevalencia/>.
- [11] A. C. of Rheumatology. «About ACR/ARP», visitado 5 de nov. de 2025. dirección: <https://rheumatology.org/about>.
- [12] E. A. of Associations for Rheumatology. «About EULAR», visitado 5 de nov. de 2025. dirección: <https://www.eular.org/whoweare>.
- [13] S. H. Ng y F. Deng, *Language and Power*, Oxford University Press, 2017. DOI: 10.1093/acrefore/9780190228613.013.436. dirección: <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-436>.
- [14] M. D. Hauser, N. Chomsky y W. T. Fitch, «The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?», *Science*, vol. 298, n.º 5598, págs. 1569-1579, 2002. DOI: 10.1126/science.298.5598.1569. dirección: <https://www.science.org/doi/abs/10.1126/science.298.5598.1569>.
- [15] A. M. TURING, «I.—COMPUTING MACHINERY AND INTELLIGENCE», *Mind*, vol. LIX, n.º 236, págs. 433-460, oct. de 1950, ISSN: 0026-4423. DOI: 10.1093/mind/LIX.236.433. dirección: <https://doi.org/10.1093/mind/LIX.236.433>.
- [16] A. Palikhe, Z. Wang, Z. Yin y W. Zhang, «Fairness Definitions in Language Models Explained», 2025. arXiv: 2407.18454 [cs.CL]. dirección: <https://arxiv.org/abs/2407.18454>.
- [17] G. Liang y Q. Tong, «LLM-Powered AI Agent Systems and Their Applications in Industry», 2025. arXiv: 2505.16120 [cs.AI]. dirección: <https://arxiv.org/abs/2505.16120>.
- [18] M. Ghaseminejad Raeini, «The evolution of language models: From N-Grams to LLMs, and beyond», *Natural Language Processing Journal*, vol. 12, pág. 100168, 2025, ISSN: 2949-7191. DOI: <https://doi.org/10.1016/j.nlp.2025.100168>. dirección: <https://www.sciencedirect.com/science/article/pii/S2949719125000445>.
- [19] D. Link, «Traces of the Mouth: Andrei Andreyevich Markov's Mathematization of Writing», *History of Science*, vol. 44, págs. 321-348, 2006. dirección: <https://api.semanticscholar.org/CorpusID:17979697>.

- [20] C. E. Shannon, «A mathematical theory of communication», *The Bell System Technical Journal*, vol. 27, n.º 3, págs. 379-423, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.
- [21] D. Jurafsky y J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models*, 3rd. 2025. dirección: <https://web.stanford.edu/~jurafsky/slp3/>.
- [22] C. E. Shannon, «Prediction and entropy of printed English», *The Bell System Technical Journal*, vol. 30, n.º 1, págs. 50-64, 1951. DOI: 10.1002/j.1538-7305.1951.tb01366.x.
- [23] E. Black, «An experiment in computational discrimination of English word senses», *IBM Journal of Research and Development*, vol. 32, n.º 2, págs. 185-194, 1988. DOI: 10.1147/rd.322.0185.
- [24] E. Brill, R. Florian, J. C. Henderson y L. Mangu, «Beyond n-grams: can linguistic sophistication improve language modeling?», en *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, USA: Association for Computational Linguistics, 1998, págs. 186-190. DOI: 10.3115/980845.980875. dirección: <https://doi.org/10.3115/980845.980875>.
- [25] R. Rosenfeld, «Two decades of statistical language modeling: where do we go from here?», *Proceedings of the IEEE*, vol. 88, n.º 8, págs. 1270-1278, 2000. DOI: 10.1109/5.880083.
- [26] L. Yan, *Maximum Likelihood Estimation*, 2020. dirección: https://web.stanford.edu/class/archive/cs/cs109/cs109.1206/lectureNotes/LN20_parameters_mle.pdf.
- [27] H. Friedl y S. Hörmann, «Frequentist Probability Theory», en ene. de 2008, ISBN: 9781412927147. DOI: 10.4135/9781452226620.
- [28] V. N. Gudivada, D. Rao y V. V. Raghavan, «Chapter 9 - Big Data Driven Natural Language Processing Research and Applications», en *Big Data Analytics*, ép. Handbook of Statistics, V. Govindaraju, V. V. Raghavan y C. Rao, eds., vol. 33, Elsevier, 2015, págs. 203-238. DOI: <https://doi.org/10.1016/B978-0-444-63492-4.00009-5>. dirección: <https://www.sciencedirect.com/science/article/pii/B9780444634924000095>.
- [29] S. F. Chen y J. T. Goodman, *An Empirical Study of Smoothing Techniques for Language Modeling*, 1996. arXiv: [cmp-lg/9606011](https://arxiv.org/abs/cmp-lg/9606011) [cmp-lg]. dirección: <https://arxiv.org/abs/cmp-lg/9606011>.
- [30] S. Katz, «Estimation of probabilities from sparse data for the language model component of a speech recognizer», *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, n.º 3, págs. 400-401, 1987. DOI: 10.1109/TASSP.1987.1165125.
- [31] W. Gale, «Good-Turing Smoothing Without Tears», mar. de 2000. dirección: <https://www.d.umn.edu/~tpederse/Courses/CS8761-FALL02/Code/sgt-gale.pdf>.

- [32] J. J. Hopfield, «Neural networks and physical systems with emergent collective computational abilities.», *Proceedings of the National Academy of Sciences*, vol. 79, n.º 8, págs. 2554-2558, 1982. DOI: 10.1073/pnas.79.8.2554. eprint: <https://www.pnas.org/doi/pdf/10.1073/pnas.79.8.2554>. dirección: <https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554>.
- [33] S. Hochreiter y J. Schmidhuber, «Long Short-Term Memory», *Neural Computation*, vol. 9, n.º 8, págs. 1735-1780, nov. de 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. eprint: <https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf>. dirección: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [34] Y. Bengio, R. Ducharme, P. Vincent y C. Janvin, «A neural probabilistic language model», *J. Mach. Learn. Res.*, vol. 3, n.º null, págs. 1137-1155, 2003, ISSN: 1532-4435.
- [35] Z. Wang, Z. Chu, T. V. Doan, S. Ni, M. Yang y W. Zhang, *History, Development, and Principles of Large Language Models-An Introductory Survey*, 2024. arXiv: 2402.06853 [cs.CL]. dirección: <https://arxiv.org/abs/2402.06853>.
- [36] D. M. Anisuzzaman, J. Malins, P. Friedman y Z. Attia, «Fine-Tuning LLMs for Specialized Use Cases», *Mayo Clinic Proceedings: Digital Health*, vol. 3, nov. de 2024. DOI: 10.1016/j.mcpdig.2024.11.005.
- [37] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi y N. Smith, *Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping*, 2020. arXiv: 2002.06305 [cs.CL]. dirección: <https://arxiv.org/abs/2002.06305>.
- [38] W. Fedus, B. Zoph y N. Shazeer, *Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity*, 2022. arXiv: 2101.03961 [cs.LG]. dirección: <https://arxiv.org/abs/2101.03961>.
- [39] J. Devlin, M.-W. Chang, K. Lee y K. Toutanova, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2019. arXiv: 1810.04805 [cs.CL]. dirección: <https://arxiv.org/abs/1810.04805>.
- [40] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei e I. Sutskever, *Language Models are Unsupervised Multitask Learners*, 2019. dirección: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.
- [41] T. Teubner, C. Flath, C. Weinhardt, W. Aalst y O. Hinz, «Welcome to the Era of ChatGPT et al.: The Prospects of Large Language Models», *Business & Information Systems Engineering*, vol. 65, mar. de 2023. DOI: 10.1007/s12599-023-00795-x.
- [42] R. Imamguluyev, «The Rise of GPT-3: Implications for Natural Language Processing and Beyond», *International Journal of Research Publication and Reviews*, vol. 4, págs. 4893-4903, mar. de 2023. DOI: 10.55248/gengpi.2023.4.33987.

- [43] O. (2023), *GPT-4 Technical Report, 2023*. arXiv: 2303.08774 [cs.CL]. dirección: <https://arxiv.org/abs/2303.08774>.
- [44] H. Touvron et al., *LLaMA: Open and Efficient Foundation Language Models, 2023*. arXiv: 2302.13971 [cs.CL]. dirección: <https://arxiv.org/abs/2302.13971>.
- [45] K. S. Kalyan, *A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4, 2023*. arXiv: 2310.12321 [cs.CL]. dirección: <https://arxiv.org/abs/2310.12321>.
- [46] J. Lee et al., «BioBERT: a pre-trained biomedical language representation model for biomedical text mining», *Bioinformatics*, vol. 36, n.º 4, J. Wren, ed., págs. 1234-1240, 2019, ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btz682. dirección: <http://dx.doi.org/10.1093/bioinformatics/btz682>.
- [47] K. Huang, J. Altosaar y R. Ranganath, *ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission, 2020*. arXiv: 1904.05342 [cs.CL]. dirección: <https://arxiv.org/abs/1904.05342>.
- [48] A. Bosselut et al., *MEDITRON: Open Medical Foundation Models Adapted for Clinical Practice*, mar. de 2024. DOI: 10.21203/rs.3.rs-4139743/v1.
- [49] K. Singhal et al., *Large Language Models Encode Clinical Knowledge, 2022*. arXiv: 2212.13138 [cs.CL]. dirección: <https://arxiv.org/abs/2212.13138>.
- [50] K. Saab et al., *Capabilities of Gemini Models in Medicine, 2024*. arXiv: 2404.18416 [cs.AI]. dirección: <https://arxiv.org/abs/2404.18416>.
- [51] C. Lin y C.-F. Kuo, «Roles and Potential of Large Language Models in Healthcare: A Comprehensive Review», *Biomedical Journal*, vol. 48, pág. 100868, abr. de 2025. DOI: 10.1016/j.bj.2025.100868.
- [52] X. Yang et al., *GatorTron: A Large Clinical Language Model to Unlock Patient Information from Unstructured Electronic Health Records, 2022*. arXiv: 2203.03540 [cs.CL]. dirección: <https://arxiv.org/abs/2203.03540>.
- [53] H. Qiu, A. Li, L. Ma y Z. Lan, *PsyChat: A Client-Centric Dialogue System for Mental Health Support, 2024*. arXiv: 2312.04262 [cs.CL]. dirección: <https://arxiv.org/abs/2312.04262>.
- [54] T. H. Kung et al., «Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models», *PLOS Digital Health*, vol. 2, n.º 2, págs. 1-12, feb. de 2023. DOI: 10.1371/journal.pdig.0000198. dirección: <https://doi.org/10.1371/journal.pdig.0000198>.
- [55] C. Williams et al., «Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department», *JAMA Network Open*, vol. 7, e248895, mayo de 2024. DOI: 10.1001/jamanetworkopen.2024.8895.

- [56] T. I. Leung, A. J. Coristine y A. Benis, «AI Scribes in Health Care: Balancing Transformative Potential With Responsible Integration», *JMIR Med Inform*, vol. 13, e80898, ago. de 2025, ISSN: 2291-9694. DOI: 10.2196/80898. dirección: <https://medinform.jmir.org/2025/1/e80898>.
- [57] C. Maslach y S. E. Jackson, «The measurement of experienced burnout», *Journal of Organizational Behavior*, vol. 2, n.º 2, págs. 99-113, 1981. DOI: <https://doi.org/10.1002/job.4030020205>. dirección: <https://onlinelibrary.wiley.com/doi/abs/10.1002/job.4030020205>.
- [58] T. D. Shanafelt et al., «Changes in Burnout and Satisfaction With Work–Life Integration in Physicians and the General US Working Population Between 2011 and 2023», *Mayo Clinic Proceedings*, vol. 100, n.º 7, págs. 1142-1158, 2025, ISSN: 0025-6196. DOI: <https://doi.org/10.1016/j.mayocp.2024.11.031>. dirección: <https://www.sciencedirect.com/science/article/pii/S0025619624006682>.
- [59] A. Pujol-de Castro, G. Valerio-Rao, P. Vaquero-Cepeda y F. Catalá-López, «Prevalencia del síndrome de burnout en médicos que trabajan en España: revisión sistemática y metaanálisis», *Gaceta Sanitaria*, vol. 38, pág. 102384, 2024, ISSN: 0213-9111. DOI: <https://doi.org/10.1016/j.gaceta.2024.102384>. dirección: <https://www.sciencedirect.com/science/article/pii/S0213911124000311>.
- [60] A. M. Association, *What is physician burnout?*, en, <https://www.ama-assn.org/practice-management/physician-health/what-physician-burnout>, mayo de 2025.
- [61] V. Podder, V. Lew y S. Ghassemzadeh. «SOAP notes». dirección: <https://www.ncbi.nlm.nih.gov/books/NBK482263/>.
- [62] K. J. Walker et al., «An economic evaluation of the costs of training a medical scribe to work in Emergency Medicine», *Emergency Medicine Journal*, vol. 33, n.º 12, págs. 865-869, 2016, ISSN: 1472-0205. DOI: 10.1136/emmermed-2016-205934. eprint: <https://emj.bmj.com/content/33/12/865.full.pdf>. dirección: <https://emj.bmj.com/content/33/12/865>.
- [63] M. Sasseville et al., «The Impact of AI Scribes on Streamlining Clinical Documentation: A Systematic Review», *Healthcare*, vol. 13, jun. de 2025. DOI: 10.3390/healthcare13121447.
- [64] A. Biswas y W. Talukdar, «Intelligent Clinical Documentation: Harnessing Generative AI for Patient-Centric Clinical Note Generation», *International Journal of Innovative Science and Research Technology (IJISRT)*, págs. 994-1008, mayo de 2024, ISSN: 2456-2165. DOI: 10.38124/ijisrt/ijisrt24may1483. dirección: <http://dx.doi.org/10.38124/ijisrt/IJISRT24MAY1483>.
- [65] S. J. Wendt, C. T. Dinh, M. Sutcliffe, K. Jones, J. M. Scanlan y J. S. Smitherman, «Deploying ambient clinical intelligence to improve care: A research article assessing the impact of nuance DAX on documentation burden and burnout», *Future Healthcare Journal*, vol. 12, n.º 3, pág. 100450, 2025, ISSN: 2514-6645. DOI: <https://doi.org/10.1016/>

- j.fhj.2025.100450. dirección: <https://www.sciencedirect.com/science/article/pii/S2514664525002292>.
- [66] J. Mark, S. Handa y T. Syed. «Introducing AWS HealthScribe – automatically generate clinical notes from patient-clinician conversations using AWS HealthScribe | Amazon Web Services». dirección: <https://aws.amazon.com/es/blogs/industries/industries-introducing-aws-healthscribe/>.
- [67] J. Mark, S. Handa y T. Syed. «Introducing AWS HealthScribe – automatically generate clinical notes from patient-clinician conversations using AWS HealthScribe | Amazon Web Services». dirección: <https://aws.amazon.com/es/blogs/industries/industries-introducing-aws-healthscribe/>.
- [68] O. Komera y R. Manche, *Black-Box Behavior in Large Language Models: Challenges and Implications*, dic. de 2023.
- [69] Z. Yang, Y. Wu, Y. Shen, W. Dai, M. Backes e Y. Zhang, *The Challenge of Identifying the Origin of Black-Box Large Language Models*, 2025. arXiv: 2503.04332 [cs.CR]. dirección: <https://arxiv.org/abs/2503.04332>.
- [70] S. Mess, A. Mackey y D. Yarowsky, «Artificial Intelligence Scribe and Large Language Model Technology in Healthcare Documentation: Advantages, Limitations, and Recommendations», *Plastic and Reconstructive Surgery - Global Open*, vol. 13, e6450, ene. de 2025. DOI: 10.1097/GOX.0000000000006450.
- [71] R. Egger, «Text Representations and Word Embeddings: Vectorizing Textual Data», en ene. de 2022, págs. 335-361, ISBN: 978-3-030-88388-1. DOI: 10.1007/978-3-030-88389-8_16.
- [72] Z. Yin e Y. Shen, «On the Dimensionality of Word Embedding», en *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi y R. Garnett, eds., vol. 31, Curran Associates, Inc., 2018. dirección: https://proceedings.neurips.cc/paper_files/paper/2018/file/b534ba68236ba543ae44b22bd110ald6-Paper.pdf.
- [73] H. Dauda Abubakar y M. Umar, «Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec», *SLU Journal of Science and Technology*, vol. 4, págs. 27-33, ago. de 2022. DOI: 10.56471/slujst.v4i.266.
- [74] Z. Yijia, Q. Chen, Z. Yang, H. Lin y Z. lu, «BioWordVec, improving biomedical word embeddings with subword information and MeSH», *Scientific Data*, vol. 6, mayo de 2019. DOI: 10.1038/s41597-019-0055-0.
- [75] P. D. Turney y P. Pantel, «From Frequency to Meaning: Vector Space Models of Semantics», *Journal of Artificial Intelligence Research*, vol. 37, págs. 141-188, feb. de 2010, ISSN: 1076-9757. DOI: 10.1613/jair.2934. dirección: <http://dx.doi.org/10.1613/jair.2934>.

- [76] L. Ma et al., *A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge*, 2025. arXiv: 2310.11703 [cs.DB]. dirección: <https://arxiv.org/abs/2310.11703>.
- [77] M. Buckland y F. Gey, «The relationship between recall and precision», *J. Am. Soc. Inf. Sci.*, vol. 45, n.º 1, págs. 12-19, ene. de 1994, ISSN: 0002-8231. dirección: https://escholarship.org/content/qt0g80268t/qt0g80268t_noSplash_aaf9300d2ff022ba6d96f424debf50f0.pdf.
- [78] H. He, H. Zhang y D. Roth, *Rethinking with Retrieval: Faithful Large Language Model Inference*, 2022. arXiv: 2301.00303 [cs.CL]. dirección: <https://arxiv.org/abs/2301.00303>.
- [79] J. Genesis y F. Keane, «Integrating Knowledge Retrieval with Generation: A Comprehensive Survey of RAG Models in NLP», *Preprints*, abr. de 2025. DOI: 10.20944/preprints202504.0351.v1. dirección: <https://doi.org/10.20944/preprints202504.0351.v1>.
- [80] P. Lewis et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, 2021. arXiv: 2005.11401 [cs.CL]. dirección: <https://arxiv.org/abs/2005.11401>.
- [81] S. Gupta, R. Ranjan y S. N. Singh, *A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions*, 2024. arXiv: 2410.12837 [cs.CL]. dirección: <https://arxiv.org/abs/2410.12837>.
- [82] A. J. Oche, A. G. Folashade, T. Ghosal y A. Biswas, *A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems: Progress, Gaps, and Future Directions*, 2025. arXiv: 2507.18910 [cs.CL]. dirección: <https://arxiv.org/abs/2507.18910>.
- [83] X. Meng et al., «The Application of Large Language Models in Medicine: A Scoping Review», *iScience*, vol. 27, pág. 109713, abr. de 2024. DOI: 10.1016/j.isci.2024.109713.
- [84] Z. Wang, E. Khatibi y A. M. Rahmani, *MedCoT-RAG: Causal Chain-of-Thought RAG for Medical Question Answering*, 2025. arXiv: 2508.15849 [cs.CL]. dirección: <https://arxiv.org/abs/2508.15849>.
- [85] B. B. Ozmen y P. Mathur, «Evidence-based artificial intelligence: Implementing retrieval-augmented generation models to enhance clinical decision support in plastic surgery», *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 104, págs. 414-416, 2025, ISSN: 1748-6815. DOI: <https://doi.org/10.1016/j.bjps.2025.03.053>. dirección: <https://www.sciencedirect.com/science/article/pii/S1748681525002256>.
- [86] F. Neha, D. Bhati y D. K. Shukla, «Retrieval-Augmented Generation (RAG) in Healthcare: A Comprehensive Review», *AI*, vol. 6, n.º 9, 2025, ISSN: 2673-2688. DOI: 10.3390/ai6090226. dirección: <https://www.mdpi.com/2673-2688/6/9/226>.

- [87] A. Dorri, S. S. Kanhere y R. Jurdak, «Multi-Agent Systems: A Survey», *IEEE Access*, vol. 6, págs. 28 573-28 593, 2018. DOI: 10.1109/ACCESS.2018.2831228.
- [88] B. Parasumanna Gokulan y D. Srinivasan, «An Introduction to Multi-Agent Systems», en jul. de 2010, vol. 310, págs. 1-27, ISBN: 978-3-642-14434-9. DOI: 10.1007/978-3-642-14435-6_1.
- [89] S. Han, Q. Zhang, Y. Yao, W. Jin y Z. Xu, *LLM Multi-Agent Systems: Challenges and Open Problems*, 2025. arXiv: 2402.03578 [cs.MA]. dirección: <https://arxiv.org/abs/2402.03578>.
- [90] N. L. Le y M.-H. Abel, *How Well Do LLMs Predict Prerequisite Skills? Zero-Shot Comparison to Expert-Defined Concepts*, 2025. arXiv: 2507.18479 [cs.IR]. dirección: <https://arxiv.org/abs/2507.18479>.
- [91] J. He, C. Treude y D. Lo, «LLM-Based Multi-Agent Systems for Software Engineering: Literature Review, Vision, and the Road Ahead», *ACM Trans. Softw. Eng. Methodol.*, vol. 34, n.º 5, mayo de 2025, ISSN: 1049-331X. DOI: 10.1145/3712003. dirección: <https://doi.org/10.1145/3712003>.
- [92] A. Borkowski, «Multiagent AI Systems in Health Care: Envisioning Next-Generation Intelligence», *Federal Practitioner*, vol. 42, mayo de 2025. DOI: 10.12788/fp.0589.
- [93] E. Shakshuki y M. Reid, «Multi-Agent System Applications in Healthcare: Current Technology and Future Roadmap», *Procedia Computer Science*, vol. 52, págs. 252-261, 2015, The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015), ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2015.05.071>. dirección: <https://www.sciencedirect.com/science/article/pii/S1877050915008716>.
- [94] H. S. Yang, «Machine Learning for Sepsis Prediction: Prospects and Challenges», *Clinical Chemistry*, vol. 70, n.º 3, págs. 465-467, mar. de 2024, ISSN: 0009-9147. DOI: 10.1093/clinchem/hvae006. eprint: <https://academic.oup.com/clinchem/article-pdf/70/3/465/56822304/hvae006.pdf>. dirección: <https://doi.org/10.1093/clinchem/hvae006>.
- [95] J. Liao et al., «Artificial intelligence assists precision medicine in cancer treatment», *Frontiers in Oncology*, vol. 12, 2023. dirección: <https://api.semanticscholar.org/CorpusID:255497788>.
- [96] M. A. Zo'ubi, «Review of 2024 publications on the applications of artificial intelligence in rheumatology», *Clinical Rheumatology*, vol. 44, págs. 1427-1438, 2025. dirección: <https://api.semanticscholar.org/CorpusID:276645455>.
- [97] M. Doussiere, A. Aboud, G. Dequen y V. Goëb, «Artificial Intelligence in Rheumatology: From Algorithms to Clinical Impact in Osteoporosis and Chronic Inflammatory Rheumatic Diseases», *Journal of Clinical Medicine*, vol. 15, n.º 2, 2026, ISSN: 2077-0383. DOI: 10.3390/jcm15020491. dirección: <https://www.mdpi.com/2077-0383/15/2/491>.

- [98] A. Madrid-García, D. Benavent, C. Plasencia-Rodríguez, Z. Rosales-Rosado, B. Merino-Barbancho y D. Freites-Núñez, «Optimising the clinical application of rheumatology guidelines using large language models: a retrieval-augmented generation framework integrating EULAR and ACR recommendations», *EULAR Rheumatology Open*, vol. 1, n.º 3, págs. 228-236, 2025, ISSN: 3050-7081. DOI: <https://doi.org/10.1016/j.ero.2025.08.001>. dirección: <https://www.sciencedirect.com/science/article/pii/S3050708125000631>.
- [99] *REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. dirección: <https://www.boe.es/doue/2016/119/L00001-00088.pdf>.
- [100] J. del Estado «BOE» num. y d. 0. d. D. De, *Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales*. dirección: <https://www.boe.es/buscar/pdf/2018/BOE-A-2018-16673-consolidado.pdf>.
- [101] M. Munnangi, *A Brief History of Named Entity Recognition*, 2024. arXiv: 2411.05057 [cs.CL]. dirección: <https://arxiv.org/abs/2411.05057>.
- [102] V. Kocaman y D. Talby, «Accurate Clinical and Biomedical Named Entity Recognition at Scale», *Software Impacts*, vol. 13, pág. 100373, 2022, ISSN: 2665-9638. DOI: <https://doi.org/10.1016/j.simpa.2022.100373>. dirección: <https://www.sciencedirect.com/science/article/pii/S2665963822000793>.
- [103] A. Laurent, *GPT-OSS: A Technical Overview of OpenAI's Open-Weight LLMs*, <https://intuitionlabs.ai/pdfs/gpt-oss-a-technical-overview-of-openai-s-open-weight-llms.pdf>, Accessed: 2026-1-10.
- [104] E. C., *Ultimate guide - the best open source LLM for healthcare in 2025*, en, <https://www.siliconflow.com/articles/en/best-open-source-LLM-for-healthcare>, Accessed: NaN-NaN-NaN.
- [105] OpenAI et al., *gpt-oss-120b gpt-oss-20b Model Card*, 2025. arXiv: 2508.10925 [cs.CL]. dirección: <https://arxiv.org/abs/2508.10925>.
- [106] T. Pandit, S. Mahendru, M. Raval y D. Upadhyay, *The Evolution of Re-ranking Models in Information Retrieval: From Heuristic Methods to Large Language Models*, 2025. arXiv: 2512.16236 [cs.IR]. dirección: <https://arxiv.org/abs/2512.16236>.
- [107] J. Wu et al., *Medical Graph RAG: Towards Safe Medical Large Language Model via Graph Retrieval-Augmented Generation*, 2024. arXiv: 2408.04187 [cs.CV]. dirección: <https://arxiv.org/abs/2408.04187>.
- [108] A. G. de las Naciones Unidas, *Objetivos de desarrollo sostenible*, ago. de 2018. dirección: <https://www.un.org/sustainabledevelopment/es/sustainable-development-goals/>.

- [109] ETSIINF. «Recomendaciones sobre el contenido de la Memoria Final», visitado 1 de jun. de 2023. dirección: <http://www.fi.upm.es/?pagina=1475>.
- [110] A. Feder. «BibTeX.org», visitado 1 de jun. de 2023. dirección: <http://www.bibtex.org>.
- [111] W. Stallings, *Computer Organization and Architecture*. Prentice Hall, 2006.

Anexos

Apéndice A

Recibo digital e informe de similitud de *Turnitin*

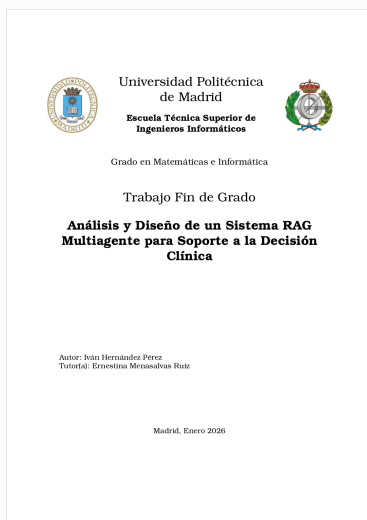


Recibo digital

Este recibo confirma que su trabajo ha sido recibido por Turnitin. A continuación podrá ver la información del recibo con respecto a su entrega.




La primera página de tus entregas se muestra abajo.

Autor de la entrega: IVAN HERNANDEZ PEREZ
Título del ejercicio: Turnitin Memoria Final
Título de la entrega: tfg_etsiinf_IvanHernandezPerez_vFinal.pdf
Nombre del archivo: 18270_IVAN_HERNANDEZ_PEREZ_tfg_etsiinf_IvanHernandezPe...
Tamaño del archivo: 8.8M
Total páginas: 111
Total de palabras: 35,045
Total de caracteres: 206,791
Fecha de entrega: 15-ene-2026 01:35p. m. (UTC+0100)
Identificador de la entrega: 2857242744



IVAN HERNANDEZ PEREZ

tfg_etsiinf_IvanHernandezPerez_vFinal.pdf

-  Turnitin Memoria Final
-  TFG ETSIINF (Moodle PP)
-  Universidad Politecnica de Madrid

Detalles del documento

Identificador de la entrega

trn:oid::1:3457998394

Fecha de entrega

15 ene 2026, 1:35 p.m. GMT+1

Fecha de descarga

15 ene 2026, 1:50 p.m. GMT+1

Nombre del archivo

18270_IVAN_HERNANDEZ_PEREZ_tfg_etsiinf_IvanHernandezPerez_vFinal_530462_2106723374.pdf

Tamaño del archivo

8.8 MB

111 páginas

35.045 palabras

206.791 caracteres

Capítulo A. Recibo digital e informe de similitud de Turnitin






Página 2 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

16% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Fuentes principales

- 15%  Fuentes de Internet
 - 11%  Publicaciones
 - 0%  Trabajos entregados (trabajos del estudiante)
-



Página 2 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

Fuentes principales

- 15% Fuentes de Internet
- 11% Publicaciones
- 0% Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	oa.upm.es	2%
2	Internet	arxiv.org	<1%
3	Internet	hdl.handle.net	<1%
4	Internet	www.coursehero.com	<1%
5	Internet	upcommons.upc.edu	<1%
6	Internet	riunet.upv.es	<1%
7	Internet	uvadoc.uva.es	<1%
8	Internet	www.ser.es	<1%
9	Internet	ntnuopen.ntnu.no	<1%
10	Internet	hal.inria.fr	<1%
11	Internet	www.mdpi.com	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 4 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

12	Internet	spiral.imperial.ac.uk	<1%
13	Internet	www.isciii.es	<1%
14	Internet	riuma.uma.es	<1%
15	Internet	Javier González Huerta. "Derivación, Evaluación y Mejora de la Calidad de Arquite...	<1%
16	Internet	repositorio.comillas.edu	<1%
17	Internet	republicans-science.house.gov	<1%
18	Internet	www.diva-portal.org	<1%
19	Internet	repositorio.grial.eu	<1%
20	Internet	digibuo.uniovi.es	<1%
21	Internet	docplayer.es	<1%
22	Internet	trepo.tuni.fi	<1%
23	Publicación	Sohaib Asif, Fazal Hadi, Qurrat-ul-ain, Yuqi Yan, Vicky Yang Wang, Dong Xu. "The ...	<1%
24	Internet	dspace.cuni.cz	<1%
25	Internet	www.dropbox.com	<1%



Página 4 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

26	Publicación	Leira, Luís Davide Jesus. "Context-Based Caching in ICN for Mobile IoT Environme...	<1%
27	Publicación	Meghna Jani, Jeffrey R. Curtis, Kimme L. Hyrich. "Navigating real-world data sourc...	<1%
28	Publicación	Correia, Alexandre Duarte. "Ferramenta de Pesquisa Avançada para Casos do Tri...	<1%
29	Internet	era.ed.ac.uk	<1%
30	Internet	www.techscience.com	<1%
31	Internet	dehesa.unex.es	<1%
32	Internet	hal.archives-ouvertes.fr	<1%
33	Internet	anm.edu.ar	<1%
34	Internet	edulearn.intelektual.org	<1%
35	Internet	gupea.ub.gu.se	<1%
36	Internet	jxiv.jst.go.jp	<1%
37	Internet	www.researchgate.net	<1%
38	Internet	www.utupub.fi	<1%
39	Internet	www11.urbe.edu	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 6 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

40	Publicación	Jinghui Liu, Daniel Capurro, Anthony Nguyen, Karin Verspoor. ""Note Bloat" impa...	<1%
41	Publicación	Mireia Verdaguer-Tremolosa, Víctor Rodrigues-Gonçalves, Pilar Martínez-López, ...	<1%
42	Internet	bibdigital.epn.edu.ec	<1%
43	Internet	www.boe.es	<1%
44	Internet	docslib.org	<1%
45	Internet	gredos.usal.es	<1%
46	Internet	dspace.ups.edu.ec	<1%
47	Internet	ve.scielo.org	<1%
48	Internet	view.genially.com	<1%
49	Publicación	Diaz Cruz, Rafael. "Methods for Large Context Clinical Question Answering.", Chri...	<1%
50	Internet	assets.researchsquare.com	<1%
51	Internet	tel.archives-ouvertes.fr	<1%
52	Internet	theses.hal.science	<1%
53	Publicación	FRANCISCO JULIÁN ZAMORA MARTÍNEZ. "Aportaciones al modelado conexionista ...	<1%



Página 6 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

54	Publicación	Recalde Varela Pablo Marcel, Bolagay Egas Mauro Fernando, Yanez Velasquez Jor...	<1%
55	Internet	repository.javeriana.edu.co	<1%
56	Internet	seer.uniacademia.edu.br	<1%
57	Internet	shuyuej.com	<1%
58	Publicación	A. Carrasco Laraña, J. Álvarez Troncoso, J.J. Ríos Blanco. "Integración de modelos ...	<1%
59	Internet	dspace.cvut.cz	<1%
60	Internet	www.ijisae.org	<1%
61	Publicación	Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zh...	<1%
62	Internet	jqcsm.qu.edu.iq	<1%
63	Internet	kth.diva-portal.org	<1%
64	Internet	rad.ort.edu.uy	<1%
65	Internet	repository.publisso.de	<1%
66	Internet	www.newmedicaleconomics.es	<1%
67	Publicación	Rui Xing, Jie Luo, Tengwei Song. "BioRel: A Large-Scale Dataset for Biomedical Rel...	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 8 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

68	Publicación	Simão, Raquel Filipa Birra. "Uncertainty-Aware AI for ECG Arrhythmia Multi-Label ..."	<1%
69	Internet	ccc.inaoep.mx	<1%
70	Internet	eu-conf.com	<1%
71	Internet	utoronto.scholaris.ca	<1%
72	Internet	7universum.com	<1%
73	Internet	discovery.researcher.life	<1%
74	Internet	doi.org	<1%
75	Internet	www.xjishu.com	<1%
76	Publicación	"Advances in Information Retrieval", Springer Science and Business Media LLC, 20...	<1%
77	Internet	www.arxiv.org	<1%
78	Publicación	Muhammad Ridho, Alhadi Bustamam, Risman Adnan. "Reconstruction of the Phi-..."	<1%
79	Internet	www.ijfmr.com	<1%
80	Internet	www.medrxiv.org	<1%
81	Publicación	Elena Díez Sánchez. "Aprovechamiento de subproductos de zumos de frutos rojos..."	<1%



Página 8 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

82	Publicación	Giovanni Parodi, Pascual Cantos-Gómez, Chad Howe, Manel Lacorte, Javier Muñoz...	<1%
83	Publicación	Katie Walker, Heather A. Heaton. "The evidence base for scribes and the disruptio...	<1%
84	Publicación	Linning Cui, Kaihong Zhou, Junhua Wang, Zheng Du, Chunying Jiang, Hao Qin. "Re...	<1%
85	Internet	hu.wikipedia.org	<1%
86	Internet	raw.githubusercontent.com	<1%
87	Publicación	Alaa Saleh, Sasu Tarkoma, Anders Lindgren, Praveen Kumar Donta, Schahram Du...	<1%
88	Publicación	J. Mora-Delgado, L. Ramos-Ruperto, M.J. Pardilla, M.Á. Sicilia, A. Rodríguez-Gonzál...	<1%
89	Publicación	Payam Mirfendereski, Alexander Ross Kerr. "Artificial Intelligence and Its Applica...	<1%
90	Publicación	Zichong Wang, Avash Palikhe, Zhipeng Yin, Wenbin Zhang. "Fairness in Language ...	<1%
91	Internet	pmc.ncbi.nlm.nih.gov	<1%
92	Internet	www.globalcaja.es	<1%
93	Internet	www.mercadoelectriconet.com.ar	<1%
94	Publicación	Antonio García Barberá. "Study of the Degradation of New Lubricant Oil Formulat...	<1%
95	Internet	asue.am	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 10 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

96	Internet	ddd.uab.cat	<1%
97	Internet	ejurnal.seminar-id.com	<1%
98	Internet	ijai.iaescore.com	<1%
99	Internet	www.academia.edu	<1%
100	Internet	www.gndiario.com	<1%
101	Internet	www.jetir.org	<1%
102	Internet	www.opos.solutions	<1%
103	Internet	www.rcs.cic.ipn.mx	<1%
104	Publicación	González Leal, Pilar. "Implementación y Perspectivas de la Medicina Integrativa e...	<1%
105	Publicación	Jianzhou Feng, Qin Wang, Huaxiao Qiu, Lirong Liu. "Retrieval In Decoder benefits ...	<1%
106	Internet	centaur.reading.ac.uk	<1%
107	Internet	ebuah.uah.es	<1%
108	Internet	eprints.ucm.es	<1%
109	Internet	es.scribd.com	<1%



Página 10 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

110	Internet	produccioncientificaluz.org	<1%
111	Internet	repositorio.tec.mx	<1%
112	Internet	sinergiejournal.eu	<1%
113	Internet	www.fit.vutbr.cz	<1%
114	Internet	zagan.unizar.es	<1%
115	Internet	JORGE ESTEBAN HERNÁNDEZ HORMAZÁBAL. "PROPUESTA DE UNA ARQUITECTUR...	<1%
116	Internet	vdocuments.es	<1%
117	Internet	www.sciencepub.net	<1%
118	Publicación	Erin Palm, Astrit Manikantan, Herprit Mahal, Srikanth Subramanya Belwadi, Mark...	<1%
119	Internet	ri-ng.uaq.mx	<1%
120	Internet	aaltodoc.aalto.fi	<1%
121	Internet	e-archivo.uc3m.es	<1%
122	Publicación	Cao, Dangli. "Instructional Alignment of Large Language Models: A Framework fo...	<1%
123	Publicación	Ivan Iaroshev, Ramalingam Pillai, Leandro Vaglietti, Thomas Hanne. "Evaluating ...	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 12 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

124	Publicación	Murty, Shikhar. "Building the Learning-From-Interaction Pipeline for Large Langu...	<1%
125	Publicación	Nagaraj, Sujay. "Reconciling With Noise in Machine Learning for Healthcare.", Uni...	<1%
126	Publicación	Omar Dib. "Secure and Incentivized Federated Learning for Resilient Telemedicin...	<1%
127	Internet	ebin.pub	<1%
128	Internet	wiki.freaks-unidos.net	<1%
129	Internet	www.brookings.edu	<1%
130	Publicación	Fernando Ramos-Zaga. "Inteligencia artificial en el abordaje clínico del cáncer de ...	<1%
131	Publicación	Gabriel Quiroz, Diego A. Burgos, Juan Felipe Zuluaga Molina. "Terminología del es...	<1%
132	Publicación	Jordi Martínez Ventura. "Enseñanza de la Arquitectura y Educación para el desarr...	<1%
133	Publicación	Nestor, Bret. "Satisfying Model Validity for Robust Clinical Machine Learning for T...	<1%
134	Internet	e-journal.spa.msu.ru	<1%
135	Internet	repositorio.elpoli.edu.co	<1%
136	Internet	repository.pnj.ac.id	<1%
137	Internet	repository.unab.edu.co	<1%



Página 12 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

138	Internet	speakerdeck.com	<1%
139	Internet	www.bfischool.org	<1%
140	Internet	www.watchguard.com	<1%
141	Internet	1library.co	<1%
142	Publicación	Javier Rodríguez García. "Metodología para la optimización del beneficio de la res..."	<1%
143	Internet	dei.cindoc.cesga.es	<1%
144	Internet	es.slideshare.net	<1%
145	Internet	forum.entitygaming.com	<1%
146	Internet	gestiopolis.com	<1%
147	Internet	live.guttmacher.org	<1%
148	Internet	repositorio.uchile.cl	<1%
149	Internet	repositorio.upeu.edu.pe:8080	<1%
150	Internet	sirio.ugr.es	<1%
151	Internet	suem.edu.ua	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 14 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

152	Internet	systems-analysis.ru	<1%
153	Internet	ughj.unicamillus.org	<1%
154	Internet	www.civitec.mx	<1%
155	Internet	www.ricardoibazeta.com.ar	<1%
156	Internet	www.tdx.cat	<1%
157	Internet	1library.org	<1%
158	Publicación	Chu, Mingi. "Regulation of Health AI Chatbots in Ontario", The University of West...	<1%
159	Publicación	Fernando Checa García, Víctor Núñez Fernández. "Generative AI and advertising c...	<1%
160	Publicación	J & E CONSULTORES GENERALES S.R.L.. "EIA-SD del Proyecto Instalación de la Líne...	<1%
161	Publicación	Jose Luis Isla. "<![CDATA[A Pattern-based Approach for Conceptual Modeling of C...	<1%
162	Publicación	Juan Mora-Delgado, Luis Ramos-Ruperto, María José Pardilla, Miguel Ángel Sicilia ...	<1%
163	Publicación	Mohamed Bahaa, Abdelrahman Hesham, Fady Ashraf, Lamiaa Abdel-Hamid. "A S...	<1%
164	Publicación	Tailia Malloy, Laura Bernardy, Omar El Bachyr, Fred Philippy, Jordan Samhi, Jacqu...	<1%
165	Internet	cassirerstudies.eu	<1%



Página 14 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

166	Internet	confcoast.com	<1%
167	Internet	digibug.ugr.es	<1%
168	Internet	doaj.org	<1%
169	Internet	dspaceapi.live.udesa.edu.ar	<1%
170	Internet	journalwjaets.com	<1%
171	Internet	mmidentity.fmk.sk	<1%
172	Internet	observatorio.campus-virtual.org	<1%
173	Internet	pharmacia.ugr.es	<1%
174	Internet	repositorio.ifgoiano.edu.br	<1%
175	Internet	repositorio.unican.es	<1%
176	Internet	revistas.pucp.edu.pe	<1%
177	Internet	sepd.es	<1%
178	Internet	suprema.stf.jus.br	<1%
179	Internet	test-www.iqvia.com	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 16 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

180	Internet	vz.kneu.ua	<1%
181	Internet	www.cacic2016.unsl.edu.ar	<1%
182	Internet	www.nib.fmed.edu.uy	<1%
183	Internet	www.oepm.es	<1%
184	Internet	www.peterlang.com	<1%
185	Internet	www.pinturadecorativa.atfreeweb.com	<1%
186	Internet	www.scribd.com	<1%
187	Internet	www.slideshare.net	<1%
188	Internet	www.suunto.com	<1%
189	Internet	www.tamps.cinvestav.mx	<1%
190	Internet	www.webpondo.org	<1%
191	Internet	www2.cruzroja.es	<1%
192	Publicación	"Educação e Inteligência Artificial: desafios e diálogos na contemporaneidade", E...	<1%
193	Publicación	"Inter-American Yearbook on Human Rights / Anuario Interamericano de Derech...	<1%



Página 16 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

194	Publicación	Artaza, Nicolás Santelices. "Coastal Planning Comparative Analysis and Recomme...	<1%
195	Publicación	Carabeo, Oresteban. "Integration of AI, Innovative Neurophysiologic Electromag...	<1%
196	Publicación	Edgar Lorenzo Sáez. "Desarrollo de una herramienta integral de gestión de gases...	<1%
197	Publicación	Enrique Soto-Pérez-de-Celis. "El sentido común en oncología", Gaceta Mexicana d...	<1%
198	Publicación	Hernández Albers, Alejandro. "Desarrollo de técnicas inmunoquímicas para la det...	<1%
199	Publicación	Juan Jesús Izquierdo Doménech. "Enhancing Industrial Process Interaction Using ...	<1%
200	Publicación	Quesada, Johan Gregorio Reyes. "Validación de un Instrumento Para Medir el Mie...	<1%
201	Publicación	Yinglong Qiu. "Futura Internacionalización de la educación superior en China y Es...	<1%
202	Internet	aim2flourish.com	<1%
203	Internet	amslaurea.unibo.it	<1%
204	Internet	ansalou.blogspot.com	<1%
205	Internet	bibliotecadigital.exactas.uba.ar	<1%
206	Internet	blog.seidor.com	<1%
207	Internet	cafeanimelair.com	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 18 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

208	Internet	crea.ujaen.es	<1%
209	Internet	eumed.net	<1%
210	Internet	eur-lex.europa.eu	<1%
211	Internet	fp.tm.tue.nl	<1%
212	Internet	habanabierta.bandcamp.com	<1%
213	Internet	happylibnet.com	<1%
214	Internet	jobs.syngenta.com	<1%
215	Internet	jobs.websello.com	<1%
216	Internet	journal.sepln.org	<1%
217	Internet	journals.openedition.org	<1%
218	Internet	managerplus.pl	<1%
219	Internet	patents.google.com	<1%
220	Internet	sinbad2.ujaen.es	<1%
221	Internet	srvsbniisphp-universitatpolit.netdna-ssl.com	<1%



Página 18 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

222	Internet	t4.oecd.org	<1%
223	Internet	theses.gla.ac.uk	<1%
224	Internet	www.aulamedica.es	<1%
225	Internet	www.ceddnet.com	<1%
226	Internet	www.cityofpalestinex.com	<1%
227	Internet	www.ffii.nova.es	<1%
228	Internet	www.iigov.org	<1%
229	Internet	www.mednetpr.com	<1%
230	Internet	www.monografias.com	<1%
231	Internet	www.pondiuni.edu.in	<1%
232	Internet	www.sepln.org	<1%
233	Internet	www.techtitute.com	<1%
234	Internet	www.unicef.org	<1%
235	Publicación	Alzahrani, Esam. "Adapting NLP Techniques in Transfer Learning for Forensic Aut..."	<1%

Capítulo A. Recibo digital e informe de similitud de Turnitin



Página 20 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394


236	Publicación	Dávid Držík, Jozef Kapusta. "Effect of dimension size and window size on word em...	<1%
237	Publicación	Franco Trigo, Lucia. "Participatory Research for the Development of Community P...	<1%
238	Publicación	James R. Langabeer, Karima Lalani. "AI-Driven Innovations in Healthcare - Best Pr...	<1%
239	Publicación	"Proceedings of the International Conference on Information Technology & Syste...	<1%
240	Publicación	A. Castillo Durán, A. Jaureguizar Oriol, J. Rigual Bobillo. "Afectación respiratoria e...	<1%
241	Publicación	Andrew Brown, Muhammad Roman, Barry Devereux. "A Systematic Literature Re...	<1%
242	Publicación	André Barbosa. "Automated essay scoring for Brazilian Portuguese: evidence fro...	<1%
243	Publicación	ChengXiang Zhai. "Statistical Language Models for Information Retrieval", Spring...	<1%
244	Publicación	Hsu, Aliyah. "Toward Trustworthy Language Models: Interpretation Methods and ...	<1%
245	Publicación	Wei, Jing Yi. "Integrating Multi-Modal Artificial Intelligence for Enhanced Cancer ...	<1%



Página 20 de 131 - Descripción general de integridad

Identificador de la entrega trn:oid::1:3457998394

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
Fecha/Hora	Thu Jan 15 15:12:35 CET 2026
Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
Numero de Serie	561
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)