

PLAN DOCENTE DE LA ASIGNATURA

Curso académico: 2026/2027

Identificación y características de la asignatura						
Código	402428					
Denominación (español)	Diseño y Construcción de Agentes Inteligentes con LLMs					
Denominación (inglés)	Design and Development of Intelligent Agents with LLMs					
Titulaciones	Máster Universitario en Ingeniería Informática					
Centro	Escuela Politécnica					
Módulo	Inteligencia Artificial					
Materia	Inteligencia Artificial					
Carácter	Optativa	ECTS	3	Semestre	1	
Profesorado						
Nombre		Despacho		Correo-e		
Álvaro Rubio Largo		I.2 (15)		arl@unex.es		
Área de conocimiento	Lenguajes y Sistemas Informáticos					
Departamento	Ingeniería de Sistemas Informáticos y Telemáticos					
Profesor/a coordinador/a (si hay más de uno)	Álvaro Rubio Largo					
Competencias / Resultados de aprendizaje						
COM02. Al finalizar el aprendizaje, el o la estudiante será capaz de adaptar las tecnologías nuevas o emergentes en informática a la resolución de problemas en nuevas áreas o que impliquen el uso de otras disciplinas, contribuyendo al desarrollo de la informática como disciplina.						
COM03. Al finalizar el aprendizaje, el o la estudiante será capaz de concebir soluciones originales, o conocidas adaptadas de manera novedosa, para dar respuesta a problemas en entornos con especificaciones inciertas o incompletas.						
Contenidos						
Diseño y explicación de los componentes clave de un agente basado en modelos de lenguaje de gran escala (LLM): motor de razonamiento, memoria y uso de herramientas. Implementación de patrones de razonamiento y acción como Chain-of-Thought (CoT) y ReAct. Construcción de sistemas de Generación Aumentada por Recuperación (RAG) para fundamentar las respuestas del agente en conocimiento externo. Utilización de frameworks de desarrollo y mecanismos de orquestación para la construcción de aplicaciones agénticas. Evaluación del rendimiento, la fiabilidad y la seguridad de los agentes, incluyendo el análisis de alucinaciones y otros modos de fallo. Estudio de los aspectos jurídicos, éticos y de responsabilidad asociados al desarrollo y despliegue de agentes inteligentes, con especial atención al Reglamento Europeo de Inteligencia Artificial y a la protección de datos personales.						

Temario
<u>Denominación del tema 1:</u> ANATOMÍA DE UN AGENTE BASADO EN LLM <u>Contenidos del tema 1:</u> Del modelo de lenguaje al sistema agéntico. Componentes fundamentales de un agente basado en LLM: modelo, herramientas, estado, memoria y trazabilidad. Uso de herramientas y salidas estructuradas. Contratos de entrada y salida, gestión de errores y permisos. Diferenciación entre tareas que deben resolverse mediante procedimientos deterministas y aquellas en las que resulta apropiado utilizar un LLM. <u>Descripción de las actividades prácticas del tema 1:</u> Análisis y modificación guiada de un sistema agéntico para integrar herramientas, salidas estructuradas, estado y memoria, identificando qué responsabilidades deben recaer en código determinista y cuáles pueden delegarse al modelo de lenguaje.
<u>Denominación del tema 2:</u> PATRONES DE RAZONAMIENTO Y ACCIÓN <u>Contenidos del tema 2:</u> Diseño de instrucciones y contexto para sistemas agénticos. Patrones de razonamiento y acción. Chain-of-Thought (CoT) y ReAct. Ciclo observación-decisión-acción. Ejecución iterativa de herramientas. Gestión de errores, reintentos, bucles y criterios de parada. Control del comportamiento no determinista y diseño de mecanismos de supervisión. <u>Descripción de las actividades prácticas del tema 2:</u> Implementación y análisis de un ciclo ReAct con uso de herramientas, tratamiento de errores y criterios explícitos de parada, estudiando diferentes comportamientos ante una misma observación o fallo.
<u>Denominación del tema 3:</u> RAG Y CONOCIMIENTO EXTERNO <u>Contenidos del tema 3:</u> Fundamentos de la Generación Aumentada por Recuperación (RAG). Documentos, fragmentación y embeddings. Indexación y recuperación de información. Selección de contexto, grounding y generación de respuestas fundamentadas. Citas y trazabilidad de las fuentes. Evaluación de la recuperación y análisis de errores, información desactualizada y contexto insuficiente. Integración de RAG en sistemas agénticos. <u>Descripción de las actividades prácticas del tema 3:</u> Incorporación de un mecanismo de recuperación documental a un sistema agéntico, analizando la fragmentación, la selección de contexto, el grounding y las citas, así como casos en los que la recuperación conduce a respuestas incorrectas, incompletas o no fundamentadas.
<u>Denominación del tema 4:</u> FRAMEWORKS, MCP Y ORQUESTACIÓN MEDIANTE GRAFOS <u>Contenidos del tema 4:</u> Frameworks para el desarrollo y la orquestación de agentes. Ventajas, costes y límites de las abstracciones. Separación entre lógica del dominio e infraestructura de orquestación. Grafos de estado, nodos, transiciones, persistencia e interrupciones. Introducción al Model Context Protocol (MCP) para la integración de herramientas y recursos externos. Criterios para decidir cuándo utilizar un framework y cuándo resulta preferible una implementación directa.

Descripción de las actividades prácticas del tema 4: Transformación guiada de una parte de un sistema agéntico hacia una representación mediante grafo de estado y análisis del uso de MCP, comparando la solución con una implementación directa y valorando qué aporta realmente cada abstracción.

Denominación del tema 5:

EVALUACIÓN, FIABILIDAD Y SEGURIDAD DE AGENTES

Contenidos del tema 5: Evaluación de sistemas agénticos. Diseño de casos de prueba, aserciones, métricas y harnesses de evaluación. Uso de trazas para el diagnóstico de fallos. Fiabilidad, alucinaciones y comportamiento no determinista. Riesgos específicos de los agentes con acceso a herramientas. Prompt injection directa e indirecta. Principio de mínimo privilegio, permisos, confirmaciones y supervisión humana. Diseño de mecanismos de mitigación y auditoría técnica.

Descripción de las actividades prácticas del tema 5: Diseño y ejecución de un pequeño harness de evaluación, análisis de trazas y estudio de casos de alucinación, prompt injection y uso indebido de herramientas, incorporando controles de permisos, mínimo privilegio y supervisión humana.

Denominación del tema 6:

MARCO JURÍDICO, RESPONSABILIDAD Y AUDITORÍA

Contenidos del tema 6: Implicaciones jurídicas y éticas del desarrollo y despliegue de agentes inteligentes. Reglamento Europeo de Inteligencia Artificial (AI Act). Transparencia, supervisión humana y gestión de riesgos. Protección de datos personales y aplicación del Reglamento General de Protección de Datos (RGPD). Responsabilidad asociada al funcionamiento de sistemas basados en inteligencia artificial. Documentación, trazabilidad y auditoría de decisiones y riesgos.

Descripción de las actividades prácticas del tema 6: Análisis de casos breves de aplicación del AI Act, protección de datos, responsabilidad y auditoría de sistemas agénticos, identificando riesgos, obligaciones y medidas de mitigación.

Actividades formativas

		Actividades Presenciales (AP)					Actividades Virtuales (AV)					
TEMA	TOTAL	GG	CH	L	O	S	CST	CSP	CAT	CAP	TP	TA
1	11	1	0	1,5	0	0	0	0	1	1,5	0	6
2	12	1	0	1,5	0	0	0	0	1	1,5	0	7
3	14	1	0	1,5	0	0	0	0	1	1,5	0	9
4	13	1	0	1,5	0	0	0	0	1	1,5	0	8
5	14	1	0	1,5	0	0	0	0	1	1,5	0	9
6	9,5	1	0	0	0	0	0	0	2,5	0	0	6
Evaluación	1,5	1,5	0	0	0	0	0	0	0	0	0	0
Totales	75	7,5	0	7,5	0	0	0	0	7,5	7,5	0	45
		50% Presencialidad					50% Virtualidad					

Actividades Presenciales (AP)	Actividades Virtuales (AV)
Actividades que se desarrollan en un único espacio físico y que implican interacción física entre estudiante y docente: - GG: Grupo Grande (85 estudiantes). - CH: Actividades de prácticas clínicas hospitalarias (7 estudiantes) - L: Actividades de laboratorio o prácticas de campo (15 estudiantes) - O: Actividades en sala de ordenadores o laboratorio de idiomas (20 estudiantes) - S: Actividades de seminario o de problemas en clase (40 estudiantes).	Actividades que no se desarrollan en un espacio físico común. Pueden ser síncronas (implican interacción estudiante / docente) o asíncronas: - CST: Clase síncrona teórica. - CSP: Clase síncrona práctica. - CAT: Clase asíncrona teórica. - CAP: Clase asíncrona práctica.
- TP: Tutorías Programadas (seguimiento docente, tutorías ECTS). - TA: Trabajo autónomo del estudiante.	
Metodologías docentes	
1. Docencia basada en clases expositivas para la presentación y discusión de los contenidos de la asignatura. 2. Actividades prácticas y sesiones de laboratorio guiadas bajo la dirección del profesor, orientadas a la aplicación de los conceptos estudiados y al análisis del comportamiento de sistemas agénticos. 3. Seguimiento síncrono del aprendizaje del estudiante mediante resolución de dudas, análisis de casos, discusión de decisiones de diseño y revisión de resultados. 4. Realización de actividades, trabajos y estudio autónomo por parte del estudiante a partir de los materiales y recursos proporcionados. 5. Metodologías centradas en el alumnado, especialmente Aprendizaje Basado en Proyectos (ABP), mediante el desarrollo progresivo de un proyecto individual que integra los principales conceptos y técnicas tratados en la asignatura.	
Sistemas de evaluación	
Los criterios de evaluación son: • Comprensión de los conceptos, técnicas y arquitecturas fundamentales para el diseño y construcción de sistemas agénticos basados en modelos de lenguaje. • Capacidad para analizar el comportamiento de un sistema agéntico, diagnosticar sus fallos y justificar las decisiones de diseño adoptadas. • Capacidad para aplicar mecanismos de uso de herramientas, memoria, RAG, orquestación, evaluación, fiabilidad y seguridad. • Capacidad para valorar las implicaciones jurídicas, éticas y de responsabilidad asociadas al desarrollo y despliegue de agentes inteligentes. • Resolución clara, concisa, estructurada y razonada de los problemas, casos y trabajos planteados. El sistema de evaluación se rige por la <i>Normativa de Evaluación de las Titulaciones Oficiales de Grado y Máster de la Universidad de Extremadura</i>. De acuerdo con la normativa vigente, los estudiantes pueden elegir entre las siguientes modalidades:	

- Modalidad de evaluación continua (EC): sistema de evaluación constituido por diversas actividades distribuidas a lo largo del semestre de docencia de una asignatura.
- Modalidad de evaluación global (EG): sistema de evaluación constituido exclusivamente por una prueba final o conjunto de pruebas finales que engloban todos los contenidos de la asignatura y que se realizará en la fecha oficial de cada convocatoria.

Durante las primeras cuatro semanas del curso, el estudiante podrá elegir, a través del Campus Virtual, la modalidad con la que quiere ser evaluado para cada convocatoria. En caso de ausencia de solicitud expresa por parte del estudiante, la modalidad asignada será la de evaluación continua (EC).

En ambas modalidades, el estudiante puede obtener la calificación máxima.

EVALUACIÓN CONTINUA (EC)

La evaluación se estructura en dos bloques con igual ponderación:

1. Examen individual (E):

Se realizará una prueba individual, presencial y escrita, de aproximadamente 90 minutos de duración. La prueba estará formada principalmente por preguntas abiertas, análisis de casos y resolución razonada de problemas relacionados con los contenidos de la asignatura.

Podrá incluir, entre otros, interpretación de trazas, análisis de arquitecturas, diagnóstico de fallos, fragmentos breves de código o pseudocódigo, decisiones sobre el uso de herramientas y modelos de lenguaje, sistemas RAG, evaluación, fiabilidad, seguridad y aspectos jurídicos y de responsabilidad.

2. Proyecto individual (P):

El estudiante desarrollará progresivamente durante el semestre un sistema agéntico en el que aplicará los principales conceptos y técnicas estudiados en la asignatura. El proyecto se organizará en tres hitos de desarrollo.

Se valorarán fundamentalmente el funcionamiento del sistema, las pruebas realizadas, los resultados de evaluación, las trazas generadas, las decisiones técnicas adoptadas y la documentación mínima necesaria para justificar el trabajo realizado.

La calificación final de la modalidad de evaluación continua será:

$$\text{Nota Final} = 0,5 \times E + 0,5 \times P$$

Para superar la asignatura será necesario obtener una calificación mínima de 3 puntos sobre 10 tanto en el examen individual como en el proyecto individual, y una nota final igual o superior a 5 puntos sobre 10.

EVALUACIÓN GLOBAL (EG)

En esta modalidad, el estudiante será evaluado exclusivamente mediante actividades realizadas o entregadas en la fecha oficial de cada convocatoria, sin necesidad de participar en las actividades de seguimiento desarrolladas durante el semestre.

La evaluación se estructura en los mismos dos bloques y con la misma ponderación que en la modalidad de evaluación continua:

1. Examen individual (E):

Se realizará una prueba individual, presencial y escrita, basada principalmente en preguntas abiertas, análisis de casos y resolución razonada de problemas relacionados con los contenidos de la asignatura.

Podrá incluir, entre otros, interpretación de trazas, análisis de arquitecturas, diagnóstico de fallos, fragmentos breves de código o pseudocódigo, decisiones sobre el uso de herramientas y modelos de lenguaje, sistemas RAG, evaluación, fiabilidad, seguridad y aspectos jurídicos y de responsabilidad.

2. Proyecto individual (P):

El estudiante deberá entregar el proyecto completo en la fecha oficial de la convocatoria, como máximo al inicio de la prueba presencial. El proyecto deberá integrar los principales conceptos y técnicas trabajados en la asignatura y cumplir los requisitos establecidos para el proyecto de la modalidad de evaluación continua, sin que sea necesaria la realización de los hitos de seguimiento previstos durante el semestre.

Se valorarán fundamentalmente el funcionamiento del sistema, las pruebas realizadas, los resultados de evaluación, las trazas generadas, las decisiones técnicas adoptadas y la documentación mínima necesaria para justificar el trabajo realizado.

La calificación final de la modalidad de evaluación global será:

$$\text{Nota Final} = 0,5 \times E + 0,5 \times P$$

Para superar la asignatura será necesario obtener una calificación mínima de 3 puntos sobre 10 tanto en el examen individual como en el proyecto individual, y una nota final igual o superior a 5 puntos sobre 10.

USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL

El uso de herramientas de inteligencia artificial estará permitido en el desarrollo del proyecto en las condiciones establecidas por el profesorado. El estudiante será responsable del trabajo entregado, de las decisiones adoptadas y de comprender y poder justificar técnicamente el funcionamiento de la solución desarrollada.

Resultados de aprendizaje

Al término del periodo de enseñanza de esta asignatura, el estudiante será capaz de:

- Identificar y explicar los componentes fundamentales de un sistema agéntico basado en modelos de lenguaje, así como las relaciones entre el modelo, las herramientas, el estado, la memoria y los mecanismos de control.
- Diseñar y construir sistemas agénticos que integren herramientas, patrones de razonamiento y acción, memoria y mecanismos de recuperación de conocimiento externo.
- Seleccionar y aplicar de forma razonada técnicas de RAG, frameworks y mecanismos de orquestación en función de las características y necesidades del problema.
- Analizar el comportamiento de un sistema agéntico mediante pruebas y trazas, identificar sus principales modos de fallo y proponer medidas para mejorar su fiabilidad y seguridad.
- Evaluar críticamente qué tareas deben resolverse mediante procedimientos deterministas y cuáles pueden delegarse a modelos de lenguaje, justificando las decisiones de diseño adoptadas.

- Analizar las implicaciones jurídicas, éticas y de responsabilidad asociadas al desarrollo y despliegue de agentes inteligentes, incluyendo los requisitos de protección de datos, transparencia, supervisión humana y gestión de riesgos.

Bibliografía (básica y complementaria)

BIBLIOGRAFÍA BÁSICA

- ❖ Huyen, C. (2025). AI Engineering: Building Applications with Foundation Models. O'Reilly.
- ❖ Albada, M. (2025). Building Applications with AI Agents: Designing and Implementing Multiagent Systems. O'Reilly.

BIBLIOGRAFÍA COMPLEMENTARIA

- ❖ Yao, S. et al. (2023). ReAct: Synergizing Reasoning and Acting in Language Models. ICLR 2023. arXiv:2210.03629.
- ❖ Schick, T. et al. (2023). Toolformer: Language Models Can Teach Themselves to Use Tools. NeurIPS 2023. arXiv:2302.04761.
- ❖ Lewis, P. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS. arXiv:2005.11401.
- ❖ Greshake, K. et al. (2023). Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection. ACM Workshop on Artificial Intelligence and Security (AISeC). arXiv:2302.12173.
- ❖ Anthropic (2024). Building Effective Agents. Documentación de ingeniería.
- ❖ OWASP GenAI Security Project. OWASP Top 10 for LLM Applications.
- ❖ National Institute of Standards and Technology (NIST). Artificial Intelligence Risk Management Framework (AI RMF 1.0).
- ❖ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial).
- ❖ 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 (Reglamento General de Protección de Datos).

Otros recursos y materiales docentes complementarios

Se utilizarán con frecuencia los recursos de la plataforma virtual Moodle/Campus Virtual: materiales docentes, actividades, cuestionarios, casos de estudio, enunciados y recursos del proyecto.

Asimismo, se utilizará documentación técnica oficial y actualizada de los frameworks, protocolos, APIs y herramientas empleadas en la asignatura, así como recursos de referencia sobre evaluación, seguridad y gestión de riesgos de sistemas basados en modelos de lenguaje.