

DISEÑO Y GOBERNANZA DE IA AGÉNTICA

CÓDIGO	AUTOR	DURACIÓN ESTIMADA	NIVEL DE FORMACIÓN
01B06C01	Carlos Ojeda Sánchez	140 h	Avanzado

Dirigido a

Profesionales de cualquier sector que ya estén familiarizados con la IA generativa y quieran avanzar hacia el diseño, configuración, supervisión e integración de agentes de IA, con especial atención a su aplicación responsable en las pymes.

Descripción

Con este contenido de curso profesional el alumnado aprenderá a diseñar e integrar agentes de IA capaces de actuar con distintos niveles de autonomía. La formación abarca los fundamentos de la IA agéntica, la especificación y conexión de agentes, los sistemas multiagente, la gobernanza y la integración organizacional. Mediante las unidades prácticas, se aplicarán estos conocimientos al diseño, configuración y control de soluciones vinculadas a situaciones profesionales.

COMPETENCIAS

1. Analizar la arquitectura de un agente de IA, diferenciando sus componentes y ciclo operativo para valorar su aplicación profesional.
2. Clasificar niveles de autonomía en sistemas de IA según supervisión, riesgo y capacidad de decisión para seleccionar usos adecuados.
3. Identificar patrones agénticos en procesos empresariales, evaluando valor, complejidad y riesgo para priorizar casos de uso viables.
4. Especificar un agente mediante un documento estructurado de requisitos para orientar su diseño, construcción, prueba y supervisión.
5. Diseñar flujos agénticos con pasos, decisiones y ramificaciones para representar procesos ejecutables y controlables.
6. Redactar system prompts para agentes, estructurando rol, instrucciones, límites y salidas para guiar comportamientos consistentes.
7. Interpretar MCP como estándar de conexión entre agentes y herramientas para valorar su papel en arquitecturas interoperables.
8. Configurar capacidades de un agente seleccionando herramientas, permisos y servicios externos para ejecutar tareas con control operativo.
9. Diseñar mecanismos de orquestación y memoria en agentes para coordinar tareas, conservar contexto y controlar decisiones sucesivas.
10. Categorizar patrones multiagente, diferenciando configuraciones y responsabilidades para seleccionar arquitecturas adecuadas a problemas complejos.
11. Diseñar un sistema de orquestación multiagente, definiendo maestro, especialistas e interfaces para coordinar resultados integrados.
12. Especificar protocolos de comunicación entre agentes mediante formatos, mensajes y contratos para asegurar coordinación y trazabilidad.
13. Diseñar mecanismos de supervisión y monitorización para agentes autónomos, definiendo métricas, alertas y niveles de intervención.

DISEÑO Y GOBERNANZA DE IA AGÉNTICA

14. Implementar guardrails de seguridad en agentes, diseñando límites, validaciones y respuestas controladas para reducir riesgos operativos.
15. Aplicar medidas de protección de datos y contención de fallos en agentes para reducir exposición, errores y usos indebidos.
16. Diseñar protocolos de escalado humano, definiendo patrones y puntos de intervención para controlar decisiones sensibles o inciertas.
17. Aplicar el marco ACPC y requisitos de auditoría para garantizar control, trazabilidad y revisión de sistemas agénticos.
18. Planificar la integración de agentes en flujos organizacionales existentes, identificando procesos, sistemas y puntos de automatización viable.
19. Liderar la gestión del cambio asociada a sistemas agénticos, anticipando resistencias y acciones de adopción para facilitar implantaciones sostenibles.
20. Establecer sistemas de medición del rendimiento de agentes, definiendo KPIs y ciclos de mejora para optimizar resultados.
21. Proyectar el rol profesional ante la adopción agéntica, identificando competencias y responsabilidades para adaptarse al nuevo contexto.

CRITERIOS DE EVALUACIÓN (Objetivos)

1. Identificar Definir los componentes principales de un agente de IA y explicar cómo interactúan durante el ciclo operativo. Diferenciar percepción, razonamiento, memoria y acción mediante ejemplos vinculados a procesos empresariales habituales. Representar el ciclo agéntico completo identificando entradas, decisiones, acciones, retroalimentación y posibles puntos de control. Valorar limitaciones y dependencias de la arquitectura agéntica para evitar interpretaciones simplificadas o excesivamente automatistas.
2. Describir los niveles de autonomía agéntica y los diferencia mediante características observables y ejemplos de uso. Clasificar casos empresariales según su grado de autonomía, justificando la decisión con criterios de supervisión y riesgo. Identificar ventajas, límites y condiciones de uso de cada nivel para evitar delegaciones inadecuadas. Proponer el nivel de autonomía más adecuado para un proceso concreto, considerando impacto, control humano y complejidad operativa.
3. Describir los principales patrones agénticos y asociarlos con necesidades profesionales o procesos empresariales concretos. Evaluar procesos propios para detectar oportunidades de automatización agéntica con valor operativo y bajo riesgo. Priorizar casos de uso considerando impacto esperado, dificultad técnica, datos disponibles y necesidad de supervisión. Reconocer patrones poco adecuados cuando el proceso exige juicio humano, alta sensibilidad o control normativo estricto.
4. Identificar las secciones necesarias de una especificación de agente y explicar la función de cada una. Redactar objetivos, límites, entradas, salidas y capacidades de un agente manteniendo coherencia con el caso de uso. Definir restricciones, riesgos y criterios de aceptación que faciliten la prueba y validación posterior del agente. Revisar una especificación agéntica detectando ambigüedades, omisiones o decisiones de diseño insuficientemente justificadas.
5. Representar flujos de trabajo agénticos utilizando pasos, decisiones, condiciones, entradas y salidas de forma ordenada. Diseñar un flujo multi-paso coherente con un objetivo profesional, incorporando validaciones y puntos de control. Identificar bucles, excepciones y dependencias entre tareas para evitar secuencias ambiguas o difíciles de automatizar. Justificar la estructura del flujo relacionando cada fase con el resultado esperado y el nivel de supervisión necesario.
6. Estructurar un system prompt con identidad, objetivo, contexto, instrucciones, restricciones, formato de salida y criterios de actuación. Redactar instrucciones claras y verificables que reduzcan ambigüedades en la ejecución del agente. Incorporar límites, excepciones y pautas de

DISEÑO Y GOBERNANZA DE IA AGÉNTICA

escalado para controlar respuestas inadecuadas o fuera de alcance. Evaluar un system prompt detectando contradicciones, instrucciones insuficientes o riesgos de comportamiento no deseado.

7. Explicar el problema de interoperabilidad que aborda MCP y lo relaciona con la conexión entre agentes y herramientas. Describir la arquitectura cliente-servidor de MCP identificando funciones, responsabilidades y flujo básico de intercambio. Diferenciar recursos, herramientas y prompts dentro de MCP mediante ejemplos aplicables a entornos profesionales. Valorar beneficios y límites de MCP para diseñar agentes conectados con mayor seguridad, trazabilidad y reutilización.
8. Aplicar el principio de mínimo privilegio al seleccionar permisos y herramientas disponibles para un agente. Categorizar herramientas según función, criticidad, datos tratados y nivel de riesgo asociado a su uso. Definir criterios de autorización, registro y supervisión para reducir errores, accesos innecesarios o acciones no deseadas. Justificar la selección de herramientas relacionándola con el objetivo del agente, el contexto y los controles requeridos.
9. Diferenciar patrones de orquestación secuencial, condicional e iterativa, indicando cuándo resulta adecuado cada uno. Explicar el ciclo de planificación, ejecución, observación y replanteamiento en agentes con tareas de varios pasos. Diseñar una estrategia de memoria distinguiendo contexto inmediato, historial útil, datos persistentes y límites de conservación. Identificar riesgos asociados a memoria y orquestación, proponer controles para evitar errores acumulativos o pérdida de contexto.
10. Describir patrones multiagente habituales y diferencia sus estructuras, ventajas, limitaciones y condiciones de aplicación. Identificar el patrón más adecuado para un problema profesional, justificando la elección según complejidad y coordinación requerida. Asignar responsabilidades entre agentes evitando solapamientos, dependencias ambiguas o duplicidades funcionales innecesarias. Valorar riesgos de coordinación, comunicación y control en sistemas multiagente antes de plantear su implementación.
11. Definir responsabilidades del agente maestro y de los agentes especialistas dentro de una arquitectura coordinada. Diseñar interfaces de comunicación entre agentes, especificando entradas, salidas, dependencias y criterios de finalización. Establecer reglas de coordinación, priorización y resolución de conflictos para mantener coherencia en el resultado final. Evaluar la arquitectura propuesta detectando riesgos de sobrecarga, ambigüedad de roles o pérdida de trazabilidad.
12. Diseñar una estructura de mensaje interagente que incluya emisor, destinatario, tipo, contenido, contexto y metadatos. Diferenciar tipos de mensaje según finalidad, urgencia, dependencia funcional y respuesta esperada dentro del sistema. Definir contratos de comunicación que reduzcan ambigüedades, errores de interpretación y pérdida de información relevante. Revisar un protocolo interagente comprobando coherencia, trazabilidad, control de errores y adecuación al flujo diseñado.
13. Diferenciar niveles de supervisión y explica cuándo conviene aplicar revisión previa, seguimiento continuo o control posterior. Definir métricas de rendimiento, calidad, seguridad y uso que permitan evaluar el comportamiento del agente. Diseñar alertas y umbrales de intervención asociados a errores, desviaciones, riesgos o resultados de baja confianza. Proponer un plan de monitorización que combine autonomía operativa con revisión humana proporcionada al nivel de riesgo.
14. Diferenciar tipos de guardrails y explica su función preventiva, correctiva o limitadora en sistemas agénticos. Diseñar reglas de validación y bloqueo para controlar entradas, acciones, salidas y uso de herramientas sensibles. Aplicar criterios de seguridad y proporcionalidad para seleccionar guardrails adecuados al riesgo del caso de uso.
15. Especificar medidas básicas de protección de datos aplicables al diseño y uso de agentes en entornos profesionales. Explicar el riesgo de prompt injection y proponer medidas de mitigación proporcionadas al contexto de uso.
16. Diferenciar patrones de escalado humano y relaciona cada uno con niveles de riesgo,

incertidumbre o impacto. Diseñar un protocolo de escalado completo que indique activadores, responsables, tiempos, evidencias y acciones posteriores.

17. Explicar los componentes del marco ACPC y su función dentro de una arquitectura de control agéntico. Especificar evidencias, registros y trazas necesarios para auditar decisiones, acciones, errores e intervenciones del agente. Proponer revisiones periódicas que permitan mantener controles actualizados ante cambios técnicos, normativos u organizativos.
18. Mapear un proceso organizacional identificando pasos, responsables, sistemas implicados, entradas, salidas y dependencias relevantes. Detectar puntos de integración agéntica valorando impacto, viabilidad, datos disponibles, riesgos y necesidad de control humano. Proponer una integración gradual que minimice interrupciones operativas y facilite validación antes del despliegue completo. Justificar la solución relacionando beneficios esperados, límites técnicos, aceptación de usuarios y continuidad del proceso.
19. Identificar fuentes de resistencia ante la incorporación de agentes y las relaciona con preocupaciones profesionales concretas. Proponer acciones de comunicación, formación y acompañamiento adecuadas a perfiles afectados por el cambio. Definir indicadores de adopción que permitan valorar uso, satisfacción, confianza y barreras durante la implantación.
20. Definir KPIs de eficiencia, calidad, impacto económico y adopción vinculados al objetivo del agente. Establecer líneas base y metas de mejora que permitan comparar resultados antes y después de la implantación. Diseñar un ciclo de revisión continua para analizar resultados, priorizar ajustes y documentar aprendizajes operativos.
21. Identificar cambios de paradigma profesional derivados del paso de ejecutar tareas a diseñar y supervisar agentes. Enumerar competencias técnicas, gestoras y éticas necesarias para trabajar con sistemas agénticos de forma responsable. Proponer un plan personal de adaptación que priorice aprendizajes, práctica guiada y mejora progresiva del desempeño.

CONTENIDOS

MÓDULO 1: FUNDAMENTOS AGÉNTICOS

¿Qué es un agente y en qué se diferencia de lo que ya conoces?

Unidad 1. Anatomía de un agente de IA

1. Definición formal de agente: percibe, razona, actúa
2. 4 componentes: percepción, razonamiento, acción, memoria
3. El ciclo agéntico: percibir → razonar → actuar → actualizar
4. Agente reactivo vs deliberativo
5. El LLM como cerebro: capacidades y limitaciones
6. Diagramas de arquitectura agéntica

Unidad 2. El espectro de la autonomía

1. El espectro: no binario, gradual
2. Nivel 1 Asistente; Nivel 2 Copiloto; Nivel 3 Agente supervisado; Nivel 4 Agente autónomo acotado; Nivel 5 Agente autónomo amplio
3. Criterios: riesgo, reversibilidad, frecuencia, coste del error
4. La autonomía como dial, no como interruptor

Unidad 3. Casos de uso y patrones agénticos

1. 6 patrones: investigación autónoma, procesamiento de documentos, atención y soporte, coordinación y scheduling, monitorización y alerta, creación y publicación
2. Antipatrones: cuándo NO usar un agente

3. Marco: ¿es candidato a agente?

MÓDULO 2: DISEÑO DE AGENTES

¿Cómo se especifica y diseña un agente que funcione?

Unidad 4. Especificación de un agente / DEA

1. Por qué especificar antes de construir
2. El DEA (7 secciones): identidad, objetivos, capacidades, restricciones y guardrails, inputs/outputs, criterios de éxito, interacciones
3. Plantilla DEA completa; buenos vs malos ejemplos; checklist de calidad

Unidad 5. Diseño de flujos multi-paso

1. Pensar en flujos: la mente procesal
2. Elementos: inicio, pasos, decisiones, bucles, fin, excepciones
3. Notación visual: diagramas de flujo agéntico
4. Patrones: secuencial, condicional, paralelo, iterativo, con escalado
5. Diseño de decisiones; manejo de errores: anticipar, recuperar, escalar

Unidad 6. El system prompt agéntico

1. El system prompt como código del agente
2. Diferencia prompt de usuario vs system prompt
3. Estructura de 7 secciones (identidad, contexto, comportamiento, herramientas, restricciones, incertidumbre, formato)
4. Técnicas: few-shot en system, cadenas de pensamiento; errores comunes; testear e iterar

Unidad Práctica 1. Especificar y diseñar un agente

MÓDULO 3: ARQUITECTURA Y CONEXIÓN

¿Cómo se conecta un agente con el mundo real?

Unidad 7. Model Context Protocol (MCP)

1. El problema que resuelve MCP: USB para agentes
2. Arquitectura: cliente, servidor, JSON-RPC
3. Ciclo: descubrimiento → invocación → respuesta
4. 4 tipos de capacidades: Tools, Resources, Prompts, Sampling
5. Servidores disponibles; anatomía de un servidor; MCP en la práctica (Claude Desktop); walkthrough guiado

Unidad 8. Herramientas y capacidades del agente

1. Principio de mínimo privilegio
2. 5 categorías: comunicación, datos, búsqueda, ejecución, especializada
3. Permisos por herramienta; documentar el toolbox
4. Config práctica: agente con email + calendario + búsqueda
5. Gestión de credenciales y secretos; testing antes de producción

Unidad 9. Orquestación y memoria

1. Orquestación: coordinar pasos y herramientas; patrones (secuencial, condicional, paralelo, iterativo)
2. El loop agéntico; planificación (ReAct, CoT agéntico)
3. Memoria de trabajo / corto plazo / largo plazo

4. RAG y bases vectoriales [intro desde cero]
5. Gestión de contexto; estado y persistencia

Unidad Práctica 2. Configurar agente con herramientas reales**MÓDULO 4: SISTEMAS MULTI-AGENTE****¿Cómo colaboran múltiples agentes para tareas complejas?****Unidad 10. Patrones multi-agente**

1. ¿Por qué múltiples agentes? Especialización, escalabilidad, robustez
2. 5 patrones: jerárquico, pipeline, colaborativo, competitivo, especialistas con router
3. Comparativa y combinaciones híbridas

Unidad 11. El Maestro y la Orquesta

1. El agente Maestro: planificador, coordinador, integrador
2. Diseño de especialistas: dominio, interfaz, independencia
3. Protocolo de delegación; integración de resultados; gestión de fallos
4. Ejemplo: arquitectura de referencia

Unidad 12. Comunicación entre agentes

1. El mensaje como contrato; anatomía (from, to, type, content, context, metadata)
2. Tipos de mensaje; canales (directo, broadcast, pub/sub, vía maestro)
3. Sincronización; estado compartido; trazabilidad
4. Ejemplo: protocolo de handoff documentado

Unidad Práctica 3. Diseñar sistema multi-agente**MÓDULO 5: GOBERNANZA AGÉNTICA****¿Cómo se controla lo que no se puede supervisar en tiempo real?****Unidad 13. Supervisión y monitorización**

1. El dilema: más autonomía ≠ menos control
2. 4 niveles de supervisión; métricas (rendimiento, calidad, recursos, comportamiento)
3. Sistema de alertas; dashboard; logs y trazabilidad

Unidad 14. Guardrails: los 4 tipos

1. Filosofía: no confiar, verificar; diseñar para el peor caso
2. 4 tipos de guardrails: entrada (validar inputs, detectar manipulación), proceso (límites de pasos, timeouts, checkpoints), salida (validar output, confirmar irreversibles), recursos (tokens, rate limits, quotas)
3. Lista de comprobación pre-acción

Unidad 15. Seguridad de datos y contención

1. Protección de datos personales
2. Mitigación de prompt injection
3. Contención y respuesta ante fallos

Unidad 16. Escalado a humanos

1. Función del humano en el loop
2. Patrones de escalado operativo

3. Protocolos de intervención humana

Unidad 17. ACPC y auditoría

1. Fundamentos del marco ACPC
2. Capas de control aplicables
3. Evidencias y registros auditables
4. Requisitos de revisión continua

Unidad Práctica 4. Diseñar gobernanza de sistema agéntico

MÓDULO 6: INTEGRACIÓN ORGANIZACIONAL

¿Cómo se llevan los agentes al mundo real de la empresa?

Unidad 18. Agentes en flujos existentes

1. El error de la página en blanco
2. Mapeo de procesos; identificar puntos de inserción (criterios de valor)
3. Tipos de inserción: paralelo, asistido, autónomo, híbrido
4. Gestión de dependencias; plan de transición: piloto, métricas, go/no-go, rollout, rollback

Unidad 19. Gestión del cambio

1. El 70% de transformaciones fracasan por factores humanos
2. 5 fuentes de resistencia y estrategias para cada una
3. Comunicación efectiva; stakeholder management; acompañamiento

Unidad 20. Medición y mejora continua

1. KPIs: eficiencia, calidad, económicos, adopción
2. Establecer baselines; ciclos de revisión
3. Proceso de mejora; reporting a stakeholders

Unidad 21. El futuro del profesional agéntico

1. Cambio de paradigma: ejecutor→diseñador, operador→supervisor, especialista→orquestador
2. Competencias (técnicas, gestión, transversales); nuevos roles
3. Responsabilidades ampliadas; desarrollo continuo; visión personal

Unidad Práctica 5. Diseño completo de sistema agéntico