Práctica final — Plan de negocio de una startup AI-first

Asignatura: Deep Learning para Business Analytics — Comillas (ICADE). **Profesor:** Eduardo C. Garrido-Merchán · `ecgarrido@comillas.edu`. **Curso:** 2026–2027. **Modalidad: en grupo. Obligatoria:** sin ella aprobada con un 5, la asignatura no se supera (§7).

1. Qué se entrega

Un **plan de negocio** de una startup cuyo producto necesita IA en el núcleo, con un **prototipo que arranca** y un **pitch de 15 minutos** ante inversores. El plan se escribe para alguien que va a poner dinero: cuánto cuesta servir a un usuario, a qué precio se le vende, a partir de cuántos usuarios la cuenta sale, y qué pasa el día que el proveedor del modelo suba la tarifa. Los tres casos del curso son los tres ensayos de esto: el Caso I compara un modelo contra un baseline honesto, el Caso II dirige un agente y mide la factura, y el Caso III monta el RAG del prototipo.

Las cifras técnicas van **medidas sobre vuestro propio sistema**: el coste por consulta sale de la calculadora del bloque con los tokens que el prototipo gasta de verdad, y el acierto del recuperador, de vuestro conjunto de evaluación. Lo que inventéis de negocio (un precio, una conversión, un tamaño de mercado) va **etiquetado como estimación**, con su supuesto al lado.

2. Las tres piezas técnicas del prototipo

Pieza	Qué tiene que existir y funcionar	Qué se mide y se reporta
RAG	El sistema del Caso III sobre el corpus de vuestro sector, con cita y con abstención.	Acierto@k, fidelidad, abstención y coste por consulta, con el k elegido justificado.
Conectores (MCP)	**Un servidor MCP con dos herramientas** que el agente llame de verdad: una que lee de un sistema vuestro y otra que escribe o ejecuta, con barandilla contra inyección de prompt.	La traza de una sesión: qué herramienta, con qué permiso, cuántos tokens costó la ida y la vuelta, y qué pasa si la herramienta falla.
SFT	La **decisión**, resuelta con el árbol prompt/RAG/SFT del Tema B: qué falla el prompt, por qué RAG no lo arregla y si compensa pagar un fine-tune.	El coste del fine-tune (datos, cómputo, mantenimiento), qué tendría que ser cierto para que saliera a cuenta y con qué medida se comprueba.

De SFT se evalúa la decisión y su aritmética; **entrenar de verdad es bonus** (§6), y `demo_rapida/03_sft.ipynb` lo hace con 160 ejemplos en seis minutos de CPU. El andamiaje de los conectores está en `demo_rapida/mcp_demo/`, con un servidor MCP de 200 líneas sin dependencias.

3. Qué lleva el plan de negocio

La memoria (`memoria.pdf`, 8–12 páginas) es el plan, en este orden: el **problema** y de quién es; la **propuesta de valor**, con el antes y el después del cliente cuantificado; el **mercado y la competencia**, incluido qué pasa si el modelo generalista hace esto gratis dentro de un año; la **arquitectura**, con el diagrama de las tres piezas del §2; el **modelo de ingresos** y el precio; los **costes**, con las seis partidas del Tema B, el coste por usuario y el margen; el **punto de equilibrio** en usuarios y en meses; los **riesgos** técnicos y regulatorios —la clasificación AI Act con su artículo, un riesgo medido sobre datos y la mitigación con su coste—; y el **roadmap** con hitos medibles.

4. Las diez preguntas sobre el modelo

La **ficha del modelo**, una página de la memoria, responde a estas diez, que son también las de la defensa sin IA del §7. La 5 y la 6 piden números propios; la 7, un segundo modelo probado.

1. **Cuál.** Nombre y versión exactos del modelo que lleva dentro el producto. Si hay más de uno, cuál hace qué.
2. **Abierto o cerrado**, con qué licencia, y qué os permite y qué os prohíbe hacer con vuestro producto.
3. **Cuántos parámetros** tiene y **en qué precisión** corre; si está cuantizado, a cuántos bits y qué perdéis.
4. **Qué ventana de contexto** tiene y cuánta gastáis de media, repartida entre instrucciones, fragmentos del RAG e historial.
5. **Cuánto cuesta una consulta**: tokens de entrada, de salida, fracción cacheada y el total en euros.
6. **Cuánto costaría servirlo vosotros**: VRAM, GPU, coste mensual y **a partir de cuántas consultas al mes** compensa frente a la API.
7. **Qué modelo probasteis y descartasteis**, y con qué medida sobre qué conjunto lo descartasteis.
8. **Qué pasa si mañana** sube el precio un 50 % o retiran el modelo: cuánto tardáis en cambiarlo y qué se rompe.
9. **Qué hace el sistema cuando el modelo no sabe**: cuál es la regla de abstención y qué ve el usuario.
10. **Qué dato del cliente ve el modelo**, adónde viaja, quién lo guarda y durante cuánto tiempo.

5. Qué se entrega y cuándo

#	Entregable	Qué es
1	`memoria.pdf`	8–12 páginas: el plan del §3 más la ficha del modelo del §4. Declaración de trazabilidad de IA al principio.
2	`prototipo/`	Las tres piezas del §2 corriendo sobre un caso reducido, con semillas fijadas y un `README.md` de arranque. Tiene que arrancar en el portátil del profesor.
3	`slides.pdf`	El pitch de 15 minutos, con la diapositiva de costes y la de cumplimiento.

Entrega en Moodle en la **semana 14, antes del pitch**; pitch y defensa en la sesión del Día 18; nota tras la última sesión de defensas. Los grupos se cierran en las **semanas 1–2** (3–4 personas) y son los mismos de los Casos II y III. El tema sale del banco de proyectos abiertos (30–40 ideas ancladas en empresas reales, máximo 2 grupos por idea) o lo proponéis vosotros. Hay **4 horas de mentorización**; llegad a cada una con una pregunta concreta.

6. Rúbrica

Componente	Peso	Qué se valora
Plan de negocio	30 %	Problema real, precio y márgenes defendibles, punto de equilibrio calculado, riesgos y roadmap con hitos medibles.
Las tres piezas técnicas	30 %	RAG medido (12), conector MCP que se llama de verdad y con barandilla (10), decisión de SFT con su aritmética (8).
Ficha del modelo	15 %	Las diez preguntas del §4 contestadas con números propios.
Pitch	15 %	Quince minutos, narrativa de inversor, slides y diagramas legibles.
Defensa oral	10 %	Respuestas correctas en la ronda de preguntas, dominio de todo lo entregado.

Bonus, hasta +1 **punto**, sólo si lo obligatorio está completo: un SFT entrenado de verdad con su antes y después medidos (+0,5); un segundo conector MCP contra un sistema externo real (+0,25); el coste medido en una semana de uso real con usuarios de fuera del grupo (+0,25).

Suspende la práctica que el prototipo no arranque, que una cifra de la memoria no se pueda reproducir con lo entregado, o que el plan declare como medido algo que es una estimación.

7. La defensa sin IA y la obligatoriedad

La **ronda de preguntas tras el pitch** se responde en directo, sin asistente y con los ordenadores cerrados. Entran las diez preguntas del §4, la defensa del RAG del Caso III, y dos más: justificad la decisión técnica más discutible del proyecto sin decir «lo dijo el modelo», y defended el margen ante una objeción de inversor. Cualquier miembro del grupo debe poder explicar **cualquier línea, gráfico o párrafo** entregado (regla C, Día 01). No poder defender en directo lo entregado suspende la práctica, por bueno que sea el entregable. La práctica es obligatoria y su nota debe ser ≥ 5 para superar la asignatura, con independencia de Kahoots, casos y examen.

8. Política de IA y ayudas

Permitida y fomentada en todo —código, prototipo, memoria y slides— **con declaración de trazabilidad** al inicio de la memoria y del prototipo: qué herramienta y para qué. Construir una startup AI-first con agentes es exactamente lo que el curso quiere que sepáis hacer; lo que se evalúa es que entendáis y defendáis lo construido. Prohibida en la defensa del §7. Conservad los prompts relevantes. Para dudas, el foro, la mentorización, el *open practicum* de los viernes y los diez minutos después de cada clase; la charla invitada de la semana 8 es la ocasión de validar la idea con alguien del sector.