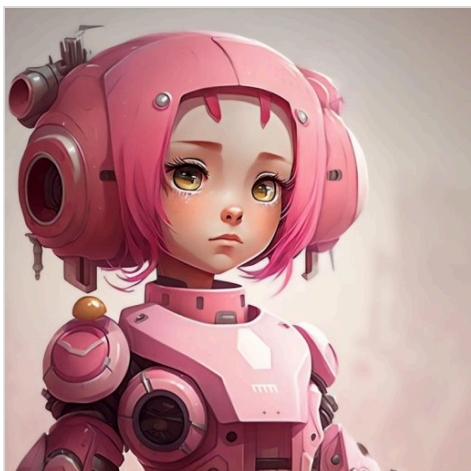


L U M B R E



# Reto: crea un analizador de PDF

**Materia:** IA Aplicada, Vibe Coding y Negocios Digitales

07/10/2026

## Antes de empezar

Tercer reto: subes un PDF, la IA saca los datos que importan y los regresa en una tabla con confianza por campo. Tu primer contacto con IA dentro de la app, no solo en el prompt.

### QUÉ APRENDERÁS

- Construir una herramienta que reciba un PDF (factura, contrato, estado de cuenta), extraiga campos clave con IA y los muestre estructurados con nivel de confianza, entrenando el patrón documento → datos.

## El reto: un PDF entra, una tabla ordenada sale

### Mapa conceptual

Reto: crea un analizador de PDF

El reto: un PDF entra, una tabla ordenada sale

Contexto · Wikipedia

El flujo: documento → texto → JSON → tabla

El mapa mental interactivo es interactivo en la versión online.



### Contexto · Wikipedia

Reto: Reto puede referirse a:

Amikeca Reto, "Red de la amistad" en esperanto.

Competición, desafío.

Los habitantes de Retia.

Wikipedia: Reto <https://es.wikipedia.org/wiki/Reto> Texto consultado y adaptado de Wikipedia (licencia CC BY-SA 4.0).

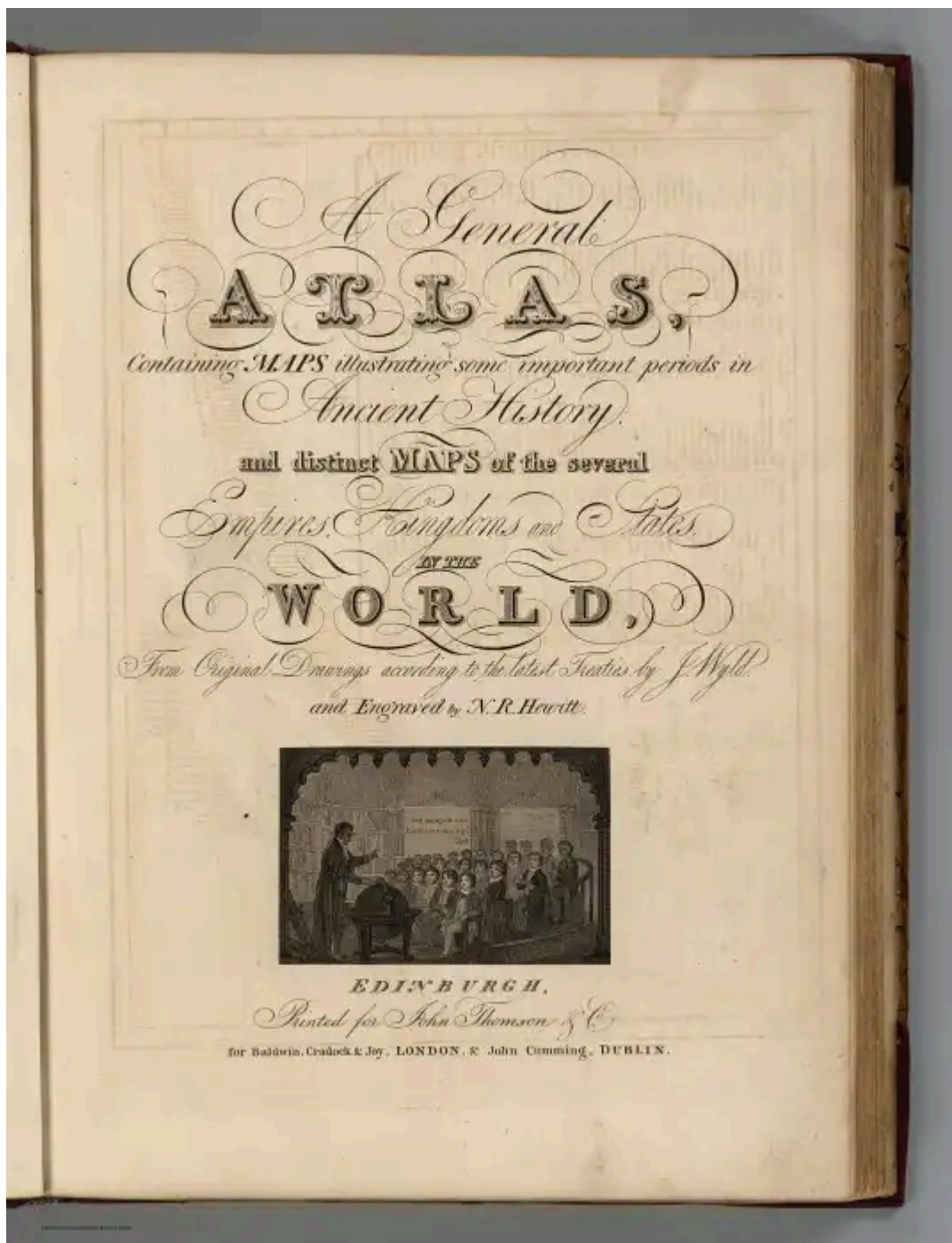


Imagen libre de Wikimedia Commons (Public domain, autor: Thomson, John Wyld, James, 1790-1836  
Lizars, W.H. Hewitt, N.R.)

## VIDEO

**Vibe coding: construir software con IA sin ser programador**

Aquí la IA deja de ser tu copiloto de código y se vuelve el motor dentro de la aplicación: el usuario sube, la IA extrae, el humano valida.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=BAveIPkWaL4>

**Especificación mínima:** (1) pantalla de **subida** de un PDF (factura o contrato, tú eliges el caso); (2) la app extrae texto y pide a la IA un JSON con campos definidos (proveedor, fecha, total, IVA, vencimiento — o partes, objeto, plazo, monto); (3) **tabla de resultados** con un nivel de confianza por campo y semáforo para lo que la IA no encontró; (4) botón de **copiar/corregir** por campo.

**Fuera del alcance:** OCR de fotos, multi-PDF, plantillas por proveedor, historial y usuarios. Este reto es distinto a los anteriores: aquí **la IA corre dentro de tu producto** y su salida no es determinista. Un CRM o un tablero puedes probarlos una vez y ya; un extractor con IA se prueba con 10 documentos distintos, porque el número 11 va a traer el total en otra parte.

**La regla de oro de la extracción: confianza, no verdad**

Todo campo extraído llega con un "¿qué tan seguro eres?". Los que no pasan el umbral se marcan en rojo y los revisa un humano. Así se diseña con IA sin auto-engaño (C15): tu app no "lee facturas", tu app "propone lecturas con evidencia".

**El flujo: documento → texto → JSON → tabla**

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

**1. Cómo atacarlo**

Los cuatro campos que traen de vuelta a todos los extractores. Diseña para el caso en que la IA no lo encuentre.

| Campo         | Por qué se pierde                                              | Defensa en el diseño                                                  |
|---------------|----------------------------------------------------------------|-----------------------------------------------------------------------|
| Total         | Aparece "Total", "Gran total", "Total MXN" y a veces dos veces | Pedir también la frase textual donde lo encontró; humano valida       |
| Fecha         | Formatos: 12/03, 03-12, "12 de marzo"                          | Normalizar a AAAA-MM-DD en la salida, no en la interpretación         |
| RFC/proveedor | Razones sociales con S.A. de C.V. en 4 variantes               | Lista de alias permitidos o dejar en amarillo siempre                 |
| IVA           | No todas las facturas lo desglosan                             | null es respuesta válida: el semáforo rojo es información, no fracaso |

## 2. Pruebas tu analizador con 10 facturas y va bien. Un amigo te pasa un PDF escaneado (foto del documento) y la app revienta. ¿Qué concluyes?

- ☐ Que la IA no sirve para extraer datos de facturas.
- ☐ Que el alcance del reto se cumplió: PDFs con texto extraíble. El documento escaneado requiere OCR, que quedó fuera del alcance a propósito; se registra como siguiente reto, no como fracaso.
- ☐ Que hay que agregar OCR ahora mismo para que quede completo.
- ☐ Que debiste probar con más de 10 facturas.

## 3. En un extractor con IA, mostrar el campo como "no encontrado" (rojo) cuando la confianza es baja es un fracaso de la herramienta.

- ☐ Verdadero
- ☐ Falso

## 4. Escribe el prompt de extracción de TU caso (documento elegido, los 5-8 campos exactos, la instrucción de JSON con confianza y null) y pégalo en el foro. Compáralo con el de dos compañeros: ¿quién definió mejor qué pasa cuando el dato NO está?

Si tu prompt no dice qué hacer cuando un campo no aparece, tu app va a improvisar con tu reputación.

---



---



---



---

**Siguiente reto:** Conecta dos APIs — de manejar un documento a manejar dos sistemas que no sabían el uno del otro: el músculo de la integración real.

## Referencia curricular

|                 |                                               |
|-----------------|-----------------------------------------------|
| <b>Materia:</b> | IA Aplicada, Vibe Coding y Negocios Digitales |
|-----------------|-----------------------------------------------|

---

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://empresa.lumbre.lat/object/iav-lab-r01-0a04>

Material educativo generado desde Lumbre · [lumbre.lat](https://lumbre.lat)