TRABAJO DOCUMENTAL BT-01

IA que puedes comprobar



BeautyThings

GUÍA PRÁCTICA DE IA

IA que puedes comprobar

Sistemas prácticos para trabajar con documentos, elegir herramientas y automatizar sin perder el control

Para quién es: autónomos y personas de administración, operaciones, consultoría y marketing en micropymes de servicios. **Qué necesitas:** una herramienta de IA que ya uses, Windows con PowerShell y el paquete descargable del libro. Para el capítulo 15, Python 3.10 o posterior. **Cómo se probó:** todas las prácticas se ejecutaron y midieron con datos inventados, los mismos que tienes en el paquete.

Qué vas a conseguir

Usas la IA para actas, correos, facturas o informes, pero no sabes cuándo fiarte. Este libro no depende de una herramienta concreta: te enseña a montar trabajos con IA que puedes revisar, con una entrada conocida, un encargo con versión, una salida con formato claro, criterios para aceptarla y una persona que responde por ella.

Al terminar tendrás tres sistemas diseñados: de solicitud a borrador aprobado, de expediente a informe con referencias y de reunión a seguimiento. El tercero lo tendrás funcionando con la receta del paquete. Cada uno incluye errores provocados a propósito y una forma de retirarlo.

Cómo leerlo

Cada capítulo empieza con el problema, sigue con pasos y cómo saber si han ido bien, y termina con el porqué y las pruebas («Lo probamos»).

Los capítulos 01–06 te dan el método. Los capítulos 07–11 lo aplican a documentos. Los capítulos 12–14 cubren la elección de herramientas y el control. Los capítulos 15–16 lo juntan todo y explican cómo mantenerlo.

No hagas las prácticas con información real de clientes. Usa los datos inventados del paquete y cambia a material autorizado solo cuando el flujo ya esté controlado.

Contenido

01 · Delegación y medición **8**

Paso a paso . . . 8
El encargo correcto . . . 9
Cuatro preguntas antes de delegar . . . 9
Lo probamos: el mismo trabajo, dos encargos . . . 9
La línea base . . . 10
Matriz de delegación . . . 10
Ejemplo de salida aceptable . . . 11
Plantilla de decisión . . . 11
Si algo falla . . . 11
Resumen . . . 11

02 · Modelos, contexto, memoria y archivos **12**

Paso a paso . . . 12
El modelo es una pieza, no el sistema . . . 12
Lo probamos: la aplicación también escribe . . . 13
Contexto: la mesa de trabajo . . . 13
Memoria: no des por hecha la continuidad . . . 14
Archivos: formato y procedencia . . . 14
Ficha de ejecución, con un caso real . . . 15
Si algo falla . . . 15
Resumen . . . 15

03 · Clasificación de información y privacidad **16**

Paso a paso . . . 16
Cuatro categorías que provocan una acción . . . 17
Lo probamos: todo con datos inventados . . . 17
Lo probamos: órdenes escondidas . . . 17
Qué sale de tu PC . . . 18
Permisos: lista de lo que sí . . . 18
Checklist antes de cada prueba . . . 19
Si algo falla . . . 19
Resumen . . . 19

04 · Brief operativo **20**

Paso a paso . . . 20
El brief v2 completo . . . 21
Las ocho piezas del brief . . . 21
Lo probamos: qué cambió . . . 22

Cómo escribir cada regla . . . 23
La plantilla del paquete . . . 23
Checklist del brief . . . 23
Si algo falla . . . 23
Resumen . . . 24

05 · Descomposición y prompting iterativo 25

Paso a paso . . . 25
Divide por resultados que puedas mirar . . . 25
El contrato de cada etapa . . . 25
Cambia una cosa cada vez . . . 26
Lo probamos: de v1 a v2, fallo por fallo . . . 26
Repite: una ejecución no es una prueba . . . 27
Ejemplos buenos y malos . . . 27
Pide pruebas, no explicaciones . . . 27
Si algo falla . . . 28
Resumen . . . 28

06 · Criterios de aceptación, rúbricas y pruebas 29

Paso a paso . . . 29
Del adjetivo al criterio . . . 29
Bloqueantes y preferencias . . . 30
El control positivo . . . 30
Lo probamos: el corrector también falla . . . 30
Corrige gratis: guarda las respuestas crudas . . . 31
Lo que no necesita IA . . . 31
Corpus mínimo de prueba . . . 31
Corrección humana y automática . . . 32
Registro de evaluación . . . 32
Resumen . . . 32

07 · Investigación con fuentes y trazabilidad 33

Paso a paso . . . 33
Empieza por una pregunta verificable . . . 33
Lo probamos: ¿existen las citas? . . . 34
Lo probamos sin querer: la fuga por PowerShell . . . 35
Qué fuente vale para qué . . . 35
Registro de evidencia . . . 35
Contradicciones y ausencia de evidencia . . . 36
Del registro al texto final . . . 36
Checklist de una investigación . . . 36
Resumen . . . 36

08 · Resumen y comparación de documentos con referencias 37

Paso a paso . . . 37
Prepara el documento . . . 37
Resumen por cobertura . . . 37
Lo probamos: un resumen con turnos . . . 38
Ausencias y omisiones . . . 38
Comparación por correspondencia . . . 39
Lo probamos: línea a línea frente a correspondencia . . . 39
Normaliza sin inventar precisión . . . 40
Checklist . . . 40
Si algo falla . . . 40
Resumen . . . 40

09 · Extracción estructurada y validación de campos 41

Paso a paso . . . 41
Define el esquema antes de extraer . . . 42
Literal, interpretación y valor normalizado . . . 42
Procedencia por campo . . . 42
Lo probamos: cuatro facturas con trampas . . . 42
Lo probamos: ¿el JSON es JSON? . . . 43
Valida por capas . . . 44
Duplicados: candidatos, no fusiones . . . 44
Cuándo basta una hoja de cálculo . . . 44
Checklist antes de dar por buena una extracción . . . 44
Resumen . . . 44

10 · Actas, correos, propuestas e informes 45

Paso a paso . . . 45
Actas: decisiones, acciones y desacuerdos . . . 45
Lo probamos: la misma acta, dos encargos . . . 46
Correos: un propósito y ningún compromiso sin permiso . . . 46
Lo probamos: un borrador que no promete . . . 46
Propuestas: alcance y cambios entre versiones . . . 47
Informes: cada párrafo con su función . . . 47
El estilo, después . . . 47
Revisa en dos lecturas . . . 47
Checklist del acta y su correo . . . 48
Resumen . . . 48

11 · Hojas de cálculo, OCR y entrada multimodal 49

Paso a paso . . . 49
Hojas de cálculo: conserva la estructura . . . 49

Reglas para datos en tabla . . . 50
Lo probamos: la misma factura, tres calidades de imagen . . . 50
Validación por código: qué caza y qué no . . . 51
OCR: mide la lectura antes que la respuesta . . . 51
Imágenes: pregunta por partes . . . 52
Si algo falla . . . 52
Resumen . . . 52

12 · Selección de herramientas por evidencia 53

Paso a paso . . . 53
Lo probamos: misma batería, varias herramientas . . . 53
Lo probamos: investigación con fuentes . . . 54
Disponibilidad y crédito: también se prueban . . . 54
La matriz, con estados honestos . . . 55
La prueba de salida . . . 55
Provoca fallos, solo en laboratorio . . . 55
Decide con fecha de caducidad . . . 56
Resumen . . . 56

13 · Nube, local, híbrido y RAG 57

Paso a paso . . . 57
Tres arquitecturas . . . 57
Lo probamos: velocidad en local . . . 57
Lo probamos: calidad local frente a nube . . . 58
Lo probamos: qué sale a internet desde «local» . . . 59
Arquitectura híbrida con contrato de datos . . . 59
RAG: buscar antes de responder . . . 59
Checklist de arquitectura . . . 60
Resumen . . . 60

14 · Automatización, agentes y control humano 61

Paso a paso . . . 61
Empieza por lo que no necesita IA . . . 61
Niveles de efecto . . . 62
Lo probamos: una orden escondida en un documento . . . 62
Lo probamos sin querer: tres fallos de configuración . . . 63
Permisos mínimos: checklist . . . 63
Aprobación humana con contexto . . . 63
Reintentos y duplicados . . . 64
Registro y reversión . . . 64
Si algo falla . . . 64
Resumen . . . 64

15 · Tres proyectos integradores **65**

Paso a paso . . . 65
Proyecto C · De reunión a seguimiento controlado (ejecutado) . . . 65
Proyecto A · De solicitud a borrador aprobado (para montar) . . . 68
Proyecto B · De expediente a informe trazable (para montar) . . . 68
Si algo falla . . . 68
Resumen . . . 69

16 · Versionado, regresión y retirada **70**

Paso a paso . . . 70
Qué versionar . . . 70
Cómo guardar las respuestas crudas . . . 71
Lo probamos: brief v1 → v2 . . . 71
Lo probamos: el corrector v1 → v4 . . . 71
Regresión: volver a probar lo que antes fallaba . . . 72
Cuándo volver a pasar la batería . . . 72
Retirada gradual . . . 73
Archivo de prácticas retiradas . . . 73
Si algo falla . . . 73
Resumen . . . 73

01 · Delegación y medición

Quieres que la IA te quite trabajo, pero no sabes qué parte darle ni cómo saber si lo hace bien. Si delegas a ciegas, produces texto más rápido y, sin darte cuenta, pasas más tiempo corrigiendo.

Al acabar este capítulo tendrás tres cosas: una forma de decidir qué tarea merece una prueba, una línea base para saber si la prueba mejora algo y tu primera medición con el acta de ejemplo del paquete.

Paso a paso

Vas a trabajar con `corpus/ACTA-SINTETICA-07.md`, un acta inventada de diez turnos (cada turno es una intervención numerada) entre Marta, Lucía y Sergio. Tiene trampas a propósito: una acción sin fecha concreta, un envío sin responsable, un desacuerdo y una excepción.

1. **Prepara el acta.** Copia el archivo y borra la sección final «Propiedades esperadas». Es la clave de corrección y el modelo no debe verla.

Cómo sabes que ha ido bien: el archivo termina en el turno 10.

2. **Lanza el encargo sencillo.** Abre la herramienta de IA que uses y pega el encargo de abajo (en el libro lo llamamos «brief v1», el ingenuo), seguido del acta. JSON es un formato de texto con campos y valores entre llaves, fácil de leer para un programa.

```
Extrae las tareas de esta acta. Devuelve solo JSON con este formato:
{"tareas": [{"tarea": "...", "responsable": "...", "fecha": "..."}]}
```

Cómo sabes que ha ido bien: recibes una lista de tareas con esos tres campos.

3. **Repite tres veces.** Cada vez en una conversación nueva. Guarda las tres respuestas en archivos de texto.

Cómo sabes que ha ido bien: tienes tres archivos y ninguno «recuerda» al anterior.

4. **Corrige cada respuesta con estas ocho preguntas de sí o no.** Son los «ocho criterios del acta» que usarás en todo el libro.

- ¿Es JSON válido?
- ¿El inventario está asignado a Marta?
- ¿El inventario no tiene una fecha concreta inventada?
- ¿Aparece el envío al cliente?
- ¿El envío no tiene un responsable inventado?
- ¿Aparece el desacuerdo del turno 07?
- ¿Aparece la excepción del turno 09?
- ¿Lucía prepara el borrador del acta? Esto sí es explícito: comprueba que se extrae lo que está.

Cómo sabes que ha ido bien: tienes una nota de 0 a 8 por respuesta.

5. **Mira dos detalles.** ¿Aparecen el desacuerdo y la excepción? ¿A quién se asigna «revisar el alcance»?

Cómo sabes que ha ido bien: puedes señalar en el acta el turno de cada error.

6. **Rellena la plantilla de decisión** (al final del capítulo) para la tarea «extraer acciones de actas».

Guarda las respuestas. En el capítulo 04 repetirás la prueba con un encargo mejor y compararás.

El encargo correcto

No delegues «hacer informes» o «llevar el correo». Son ámbitos, no tareas. Un encargo útil describe una transformación acotada:

> A partir de un acta aprobada, prepara una lista de acciones con responsable, fecha y turno de origen. No completes los campos ausentes. Marca cualquier ambigüedad para revisión.

Así puedes observar la entrada, la salida y los errores. Y separas el borrador de la decisión: el sistema organiza y una persona aprueba.

Cuatro preguntas antes de delegar

Responde sí o no. Si alguna es «no», la tarea no está lista para una prueba.

1. ¿La entrada está disponible y puedo prepararla sin exponer datos innecesarios?
2. ¿El resultado se puede comprobar contra esa entrada o contra una regla clara?
3. ¿Un error tiene efecto limitado porque hay revisión antes de usar el resultado?
4. ¿La frecuencia o la carga de revisión justifica diseñar el flujo?

Una tarea puede ser valiosa y mala candidata a la vez. Responder a una reclamación excepcional o decidir qué compromiso asumir con un cliente exige juicio. La IA puede ordenar antecedentes, pero no debe cerrar la decisión.

Lo probamos: el mismo trabajo, dos encargos

Dimos esa misma acta a cinco modelos de nube con dos encargos distintos. Corregimos cada respuesta con los ocho criterios de arriba, escritos en código, no a ojo. Tres repeticiones por modelo y encargo.

Los modelos fueron Claude Haiku 5.5 y Claude Sonnet 5.5 (a través de Claude Code), GPT-6 luna, DeepSeek V4.1 Flash y DeepSeek V4.1 Flash 1017 (a través de una API compatible con OpenAI; una API es la puerta por la que un programa habla con el modelo).

El segundo encargo, el «brief v2», tiene las mismas tareas pero con reglas explícitas y un esquema con sitio para todo. Lo tienes completo en el capítulo 04.

Modelo	Brief v1 (ingenuo)	Brief v2
Claude Haiku 5.5	16/24	24/24
Claude Sonnet 5.5	15/24	24/24
GPT-6 luna	18/24	24/24
DeepSeek V4.1 Flash 1017	18/24	24/24
DeepSeek V4.1 Flash	18/24	24/24
Total	**85/120**	**120/120**

Qué salió mal con el encargo sencillo:

- **El desacuerdo y la excepción se perdieron en las 15 ejecuciones.** No es culpa del modelo: el formato pedido (`tarea`, `responsable`, `fecha`) no tenía sitio para ellos. Lo que no cabe en el formato desaparece.
- **Asignación por proximidad.** Haiku (2 de 3 veces) y Sonnet (3 de 3) asignaron a Lucía «revisar el alcance» solo porque lo dice ella en el turno 10. Esta es una línea real de Sonnet:

```
{"tarea": "Revisar el alcance antes de cualquier envío", "responsable": "Lucía", "fecha": null}
```

El acta dice «Lucía: el siguiente paso es revisar el alcance antes de cualquier envío». Nadie se compromete a hacerlo. Con el brief v2, ningún modelo cometió ese error.

La lección. La mayor mejora no vino de cambiar de modelo, sino de reescribir el encargo: de 85/120 a 120/120.

Lo que esta prueba no demuestra. Es un acta, tres repeticiones y una tarea sencilla. No es una clasificación de modelos ni garantiza lo mismo con tus documentos. Sí demuestra algo que puedes repetir en una tarde: antes de comparar herramientas, compara encargos.

La línea base

Antes de cambiar el proceso, describe cómo lo haces ahora. No inventes duraciones: escribe «sin medir» hasta que las registres.

Campo	Registro inicial
Tarea	Extraer acciones de un acta
Entrada	Acta aprobada en formato editable
Salida actual	Lista en el documento de seguimiento
Revisión actual	Comparación manual con el acta
Tiempo de preparación	Sin medir
Tiempo de revisión	Sin medir
Errores habituales	Responsable asignado por proximidad, fecha interpretada, desacuerdos perdidos
Información sensible	Por clasificar
Responsable del resultado	Por asignar

Mide en casos parecidos, anota la variación y cuenta también preparación y corrección. Incluye desde el principio casos imperfectos, como el acta del paquete.

La línea base no sirve para demostrar que la IA es mejor. Sirve para que una impresión agradable no se confunda con una mejora.

Matriz de delegación

Clasifica cada parte del trabajo con dos preguntas: ¿se puede verificar con claridad? ¿Qué pasa si falla?

	Efecto limitado	Efecto relevante
Verificable	Borrador automatizado con revisión	Solo con controles, trazabilidad y aprobación explícita
Difícil de verificar	Exploración, no resultado final	No delegar la decisión; como mucho, apoyo documental

Aplicada al acta: extraer acciones con su turno es verificable y de efecto limitado si alguien revisa antes de enviar. Decidir quién hace el envío al cliente (turno 06) no se puede extraer: el acta dice que está pendiente. Esa decisión vuelve a una persona.

Ejemplo de salida aceptable

Estos dos turnos del acta:

> 02 · Marta: yo preparo una versión nueva del inventario la próxima semana.
>
> 06 · Marta: el envío al cliente queda pendiente de confirmar quién lo realiza.

Una salida aceptable:

Acción	Responsable	Fecha	Literal	Turno
Preparar versión nueva del inventario	Marta	PENDIENTE	«la próxima semana»	02
Enviar al cliente	PENDIENTE	PENDIENTE	—	06

No es aceptable asignar el envío a Marta porque habla en ese turno, ni convertir «la próxima semana» en una fecha sin conocer el calendario.

Plantilla de decisión

```
Tarea:
Entrada:
Salida necesaria:
Decisión que NO se delega:
Información implicada:
Error de mayor impacto:
Control previo al uso:
Responsable:
Línea base: sin medir
Decisión: probar / explorar / no delegar
Motivo:
```

Si no puedes rellenarla sin suposiciones importantes, todavía no hay nada que automatizar. Hay una idea que necesita más diseño.

Si algo falla

- **Tus resultados no se parecen a la tabla.** Es normal: cambia la herramienta, la versión y el día. Lo que importa es comparar tus dos encargos entre sí.
- **El modelo acierta todo a la primera.** Comprueba que borraste «Propiedades esperadas». Si la dejaste, el modelo tenía las respuestas.
- **Una ejecución sale bien y otra mal.** Por eso se repite tres veces. Haiku asignó por proximidad 2 veces de 3: con una sola prueba podrías haber concluido lo contrario.
- **Quieres automatizar todo el ámbito.** Divide. Extraer datos, redactar un borrador y aprobar un compromiso no tienen el mismo riesgo.

Resumen

- Delega transformaciones acotadas y comprobables, no ámbitos de trabajo.
- Registra una línea base y escribe «sin medir» en lo que aún no hayas medido.
- Lo probamos: reescribir el encargo subió el resultado de 85/120 a 120/120 con los mismos cinco modelos.
- Lo que el formato no prevé, se pierde: el desacuerdo y la excepción desaparecieron en las 15 ejecuciones del encargo sencillo.
- Repite cada prueba al menos tres veces antes de sacar conclusiones.