EXPERIMENTO 01 / COMPROBANTES DE PAGO MÓVIL / 04 NOV 2026

# Un modelo pequeño. Resultados medidos.

GLM-OCR ~0,9B · LoRA · referencia, monto, moneda y fecha

PRUEBA AUTOMÁTICA FILTRADA · n = 200

**20% a 100%**

40 de 200 a 200 de 200 completos

COMPROBACIÓN VISUAL · n = 14

**35,7% a 78,6%**

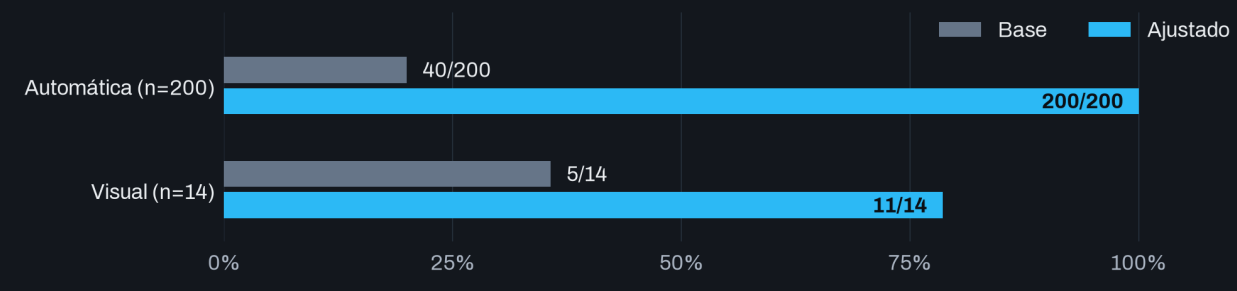
5 de 14 a 11 de 14 completos

Un caso es correcto cuando coinciden los **cuatro campos normalizados**. Cada comparación usa las mismas imágenes antes y después del ajuste.

## 01 / EXACTITUD POR CAMPO

| Métrica | AUTOMÁTICA · 200 | | VISUAL · 14 | |
|---|---|---|---|---|
| | Base | Ajustado | Base | Ajustado |
| Referencia completa | 48,5% | 100% | 71,4% | 92,9% |
| Monto | 99,5% | 100% | 100% | 100% |
| Moneda | 58% | 100% | 50% | 85,7% |
| Fecha | 91% | 100% | 92,9% | 100% |
| 4 datos correctos | 20% | 100% | 35,7% | 78,6% |

## 02 / COMPROBANTES CON LOS CUATRO DATOS CORRECTOS




**Lectura clave:** el 100% corresponde a 200 casos filtrados con etiquetas automáticas contrastadas con banco. En la comprobación visual separada quedaron **3 de 14 casos con algún error**.

MÉTODO / RECURSOS / ALCANCE DE LA EVALUACIÓN

# Qué se midió.

Ajuste local de un modelo abierto para una tarea acotada. Se midieron calidad de extracción, tiempo y memoria.

CICLO DE ENTRENAMIENTO

**4 h 21 min**

Incluye validaciones periódicas

PICO VRAM · PYTORCH

**4,29 GiB**

Memoria asignada a tensores

HARDWARE

**RTX 3060**

12 GiB de VRAM

## 03 / DATOS Y AJUSTE

| Conjunto | Imágenes |
|---|---|
| Descargas válidas | 12.519 |
| Entrenamiento | 7.858 |
| Validación reservada | 1.001 |
| Prueba reservada | 955 |
| Revisión visual | 14 |

| Configuración | Valor |
|---|---|
| Base abierta | GLM-OCR ~0,9B |
| Parámetros ajustados | 2.359.296 |
| Épocas / pasos | 3 / 1.476 |
| Mejor checkpoint | 1.300 |
| Pérdida de validación | 0,000350 |

Separación por pago, conversación y SHA-256, sin grupos compartidos. La selección de checkpoint mide pérdida en **128 imágenes de validación**; la evaluación generativa compara **200 de las 955 de prueba**. Las 14 revisadas visualmente se mantienen separadas.

## 04 / TIEMPO MEDIDO EN LA EVALUACIÓN

### 1,60 s base / 1,87 s ajustado

Promedio por imagen: el ajustado tardó un 17,0% más en esta corrida secuencial. Son tiempos observados, no una prueba de carga concurrente. El tiempo de preparación de datos y esta evaluación final quedan fuera de las 4 h 21 min del ciclo de entrenamiento.

### ALCANCE Y ERRORES PENDIENTES

- Las etiquetas automáticas provienen de OCR de alta confianza contrastado con registros bancarios; no equivalen a una transcripción manual exhaustiva.
- En los 14 casos visuales: una confusión de referencia y dos monedas inferidas como VES cuando la unidad no aparecía.
- Dataset histórico de un banco receptor y varios emisores. Nuevos formatos requieren validación propia.
- El texto bruto puede llevar etiquetas vacías de pensamiento o Markdown; el script extrae y normaliza el objeto JSON.

FUENTES Y ARTEFACTO — Paquete del modelo, scripts y métricas (R2)

Fuentes: comparison.json · training_metrics.json · dataset_summary.json · release_verification.json