

De prompt suelto a proceso

Acá el prompt deja de ser tuyo y pasa a ser de la organización, y eso cambia el estándar. **Un prompt que vas a repetir 200 veces es software.** Y el software se prueba antes de ir a producción.

QUÉ VAS A PODER HACER

- 1

Evaluar un prompt con un método, no a ojo.
- 2

Decidir con criterio si una tarea está lista para automatizarse.
- 3

Escribir una plantilla que otra persona pueda usar sin explicaciones.

LA REGLA DEL 10

Prueba el prompt **10 veces** con **CONTENIDOS DISTINTOS**
(10 reportes distintos, 10 reuniones distintas, 10 tickets distintos)

10 de 10 usables sin tocar nada ... listo para automatizar
7 de 10 todavía no: arregla el prompt
Varía muchísimo entre corridas ... baja la temperatura

OJO: **varía el CONTENIDO, no repitas el mismo input 10 veces.**
Repetir el mismo input mide la aleatoriedad del modelo,
no la robustez de tu prompt.

LOS 4 CRITERIOS DE EVALUACIÓN

CRITERIO	LA PREGUNTA QUE TE HACES	CRÍTICO EN
Consistencia	¿Las 10 salidas son del mismo tipo?	Todo lo que se automatiza
Precisión	¿Los datos y afirmaciones son correctos?	Análisis, reportes, cualquier cosa con números
Relevancia	¿Se enfoca en lo pedido o divaga?	Resúmenes y accionables
Claridad	¿Respeto extensión, tono y voz?	Todo lo que sale hacia afuera

Validación cruzada entre modelos. Si un LLM te dice "no encontré esa información", no lo des por cerrado: reformula la pregunta en otro modelo y pide citas exactas en ambos. Si ninguno puede mostrarte el texto literal, casi seguro es alucinación.

ANTES Y DESPUÉS: UNA PLANTILLA QUE OTRO PUEDE USAR

✗ ANTES

Escribe el correo de seguimiento de la reunión con el cliente

Funciona cuando lo escribes tú, porque tienes el contexto en la cabeza. Se lo pasas a alguien del equipo y el resultado es irreconocible.

✓ DESPUÉS

Actúa como asistente de redacción comercial.
Escribe el correo de seguimiento de la reunión.

```
<variables>
Cliente: [nombre y cargo]
Empresa: [empresa]
Notas de la reunión: [pega tus notas crudas]
Acuerdos: [lista]
Próximos pasos y responsables: [lista]
</variables>

<limites>
- Entre 150 y 200 palabras.
- Estructura fija: asunto, saludo, resumen breve, acuerdos en bullets, próximos pasos con responsable y fecha, cierre.
- Sin superlativos ni frases de relleno.
- Devuelve ÚNICAMENTE el correo final. Sin notas, sin explicaciones, sin opciones alternativas.
</limites>
```

Todo lo que cambia entre usos está entre corchetes, y los límites incluyen **qué no hacer**. Cualquiera del equipo la usa igual.

El rol fija el registro antes de que empiece a escribir.

Las variables son el único lugar donde la persona tiene que pensar.

Los límites incluyen la estructura exacta, no solo la extensión.

"Únicamente el correo final" evita que te devuelva tres versiones y un comentario.

EL CICLO DE VIDA

escribir → evaluar (regla del 10) → plantilla → compartir → REEVALUAR

Reevalúa cada semana: los modelos cambian sin avisar y un prompt que andaba puede degradarse solo. Guarda un caso de prueba fijo para detectar cuándo pasó.

Y antes de todo: ¿esta tarea realmente necesita IA?

A veces la respuesta honesta es NO (un filtro de Excel, una regla, una plantilla de texto). Decirlo te da más credibilidad que cualquier demo.

CHECKLIST ANTES DE AUTOMATIZAR

- [] ¿Esta tarea realmente necesita IA?
- [] Probado 10 veces con contenidos DISTINTOS
- [] Consistencia · Precisión · Relevancia · Claridad, ok en las 10
- [] Citas verificadas contra la fuente con Ctrl+F
- [] Definido qué hago cuando falla (fallback manual)
- [] Un caso de prueba fijo guardado para detectar degradación

1. ¿Por qué hay que variar el contenido y no repetir el mismo input?

Porque repetir el mismo input mide cuánto varía el modelo por azar. Lo que quieres medir es si tu prompt aguanta casos distintos, que es lo que va a pasar en producción.

2. Ocho de diez salidas están bien. ¿Automatizas?

No. Dos de cada diez requiriendo intervención manual significa que alguien tiene que revisar las diez, y ahí se fue el ahorro.

3. ¿Cómo detectas una alucinación cruzando modelos?

Reformulas la pregunta en otro modelo y pides citas exactas en ambos. Si ninguno puede mostrarte el texto literal de la fuente, casi seguro es alucinación.

4. ¿Por qué reevaluar cada semana?

Porque los modelos se actualizan sin avisar. Un prompt calibrado puede degradarse sin que cambies nada.

5. ¿Por qué falla una plantilla compartida sin contexto?

Porque la usan tal cual, sin entender qué ajustar. El resultado sale mediocre y concluyen que la IA no sirve.

6. ¿Cuál es la primera pregunta de toda evaluación?

¿Esta tarea necesita IA? A veces se resuelve mejor con una fórmula, una regla o una plantilla de texto.

LOS 4 ERRORES MÁS COMUNES

❗ Evaluar a ojo porque el primer resultado impresionó.

❗ Probar diez veces con el mismo input.

❗ Compartir la plantilla sin los fundamentos que la hacen funcionar.

❗ Dar por terminado el prompt y no volver a mirarlo nunca.