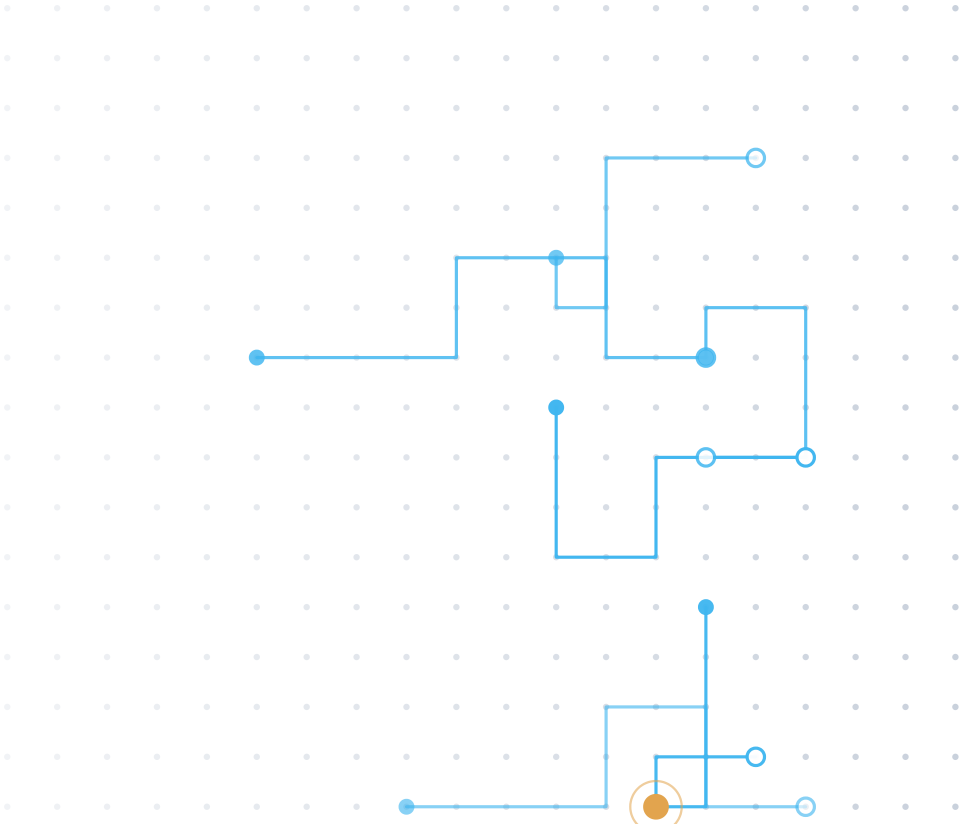


GUÍA PRÁCTICA

Agentes de IA en producción sin fallos silenciosos

Cómo saber cada mañana, con pruebas, qué hicieron tus agentes y qué dejaron sin hacer.

AUTOR
Mario Chasco · Flowara AI

PARA
Dueños de negocio que ya usan agentes de IA y las personas técnicas que contratan para montarlos

LECTURA
9 min

EDICIÓN
v2.0 · noviembre 2026

ANTES DE EMPEZAR

Qué vas a sacar de esta guía

01 Reconocer los fallos que no dan error y la pista que deja cada uno.

02 Poner registro, contador y canario alrededor de cualquier agente.

03 Decidir con pruebas qué tipo de tarea puede salir sin revisión.

04 Exigir un informe de la mañana que diga también lo que no se hizo.

CONTENIDO

01 El silencio no significa que todo vaya bien
02 Los cuatro fallos silenciosos típicos
03 Dónde nacen: cuatro descuidos al construir
04 La regla de tres: registro, contador, canario
05 La autonomía se gana tarea a tarea
06 El informe de la mañana y quién lo lee
07 Ejemplo: una peluquería de barrio con agente de citas
08 Checklist de producción

Para: Dueños de negocio que ya usan agentes de IA y las personas técnicas que contratan para montarlos · 9 min de lectura

01

El silencio no significa que todo vaya bien

Un detector de humo sin pilas está igual de callado que uno que funciona en una casa sin fuego. Desde el sofá no hay forma de distinguirlos. Con los agentes de IA pasa lo mismo: el que no ha hecho nada y el que lo ha hecho todo bien hacen el mismo ruido, ninguno.
Un error que salta en pantalla molesta, pero se arregla esa misma mañana. El peligroso es el que no avisa. El informe dice "cero pendientes" porque no pudo leer el buzón, la tarea programada lleva tres semanas sin ejecutarse o un mensaje salió a medias. Te enteras cuando un cliente se queja, y para entonces el daño ya está hecho.

Esta guía es para el dueño del negocio y para la persona técnica que monta el sistema. Los dos necesitan lo mismo: comprobar cada mañana, en dos minutos, que el agente hizo su trabajo y qué dejó sin hacer. Es lo que dejo montado antes de que un agente trabaje solo, y aquí lo cuento por partes.

> ✦ **IDEA**
>
> Un agente sin registro de errores es como un empleado que dejó de venir sin que nadie lo notara.

02

Los cuatro fallos silenciosos típicos

Casi todos los fallos que pasan desapercibidos encajan en uno de estos cuatro. Ninguno muestra un error, pero todos dejan una pista si sabes dónde mirar.

FALLO	CÓMO SE VE	QUÉ LO DELATA
El informe dice cero y es falso	"Sin mensajes pendientes"	Contador de leídos a cero
La tarea que nunca se ejecutó	Nada: ni error ni informe	Última ejecución más antigua de lo previsto
El mensaje enviado a medias	El cliente recibe media respuesta	Un estado "enviando" que nunca pasa a "enviado"
El contexto caducado	Precios u horarios viejos	Fecha de revisión de cada dato

El primero es el más traicionero, porque "no he encontrado nada" no es lo mismo que "no he podido buscar". Si el acceso al correo caducó, el agente lee cero mensajes y concluye, tan tranquilo, que no hay trabajo. Por eso cada consulta tiene que devolver una de tres respuestas: **encontrado, no encontrado o fallido**. Y "fallido" nunca cuenta como éxito.

El último es lento. Nadie avisa al agente de que el horario de verano ya no vale, así que lo sigue ofreciendo con total seguridad hasta que alguien lo revisa.

EN UNA FRASE

Cero leídos no es cero trabajo: haz que el sistema distinga entre las dos cosas.

03

Dónde nacen: cuatro descuidos al construir

Los fallos de la tabla casi nunca vienen del modelo. Vienen de huecos en el código que lo rodea, y casi siempre son estos cuatro:

01 **Una respuesta que nadie entiende y que no va a nadie.** El modelo devuelve algo que el código no sabe leer, el paso falla y el mensaje se queda en ninguna parte. Toda respuesta ilegible tiene que acabar en la bandeja de una persona.

02 **Una búsqueda con la clave vacía.** Si el teléfono del cliente llega en blanco, la búsqueda puede no traer nada o traer la ficha de otro. Antes de buscar, el código comprueba que la clave existe y tiene el formato correcto.

03 **Una herramienta que no sabe decir "no encontrado".** Si la consulta de agenda solo sabe devolver huecos, un día lleno parece un error, o peor, el agente rellena el vacío con algo inventado. Cada herramienta tiene que poder responder "no hay" de forma explícita.

04 **Dos respuestas automáticas hablando entre sí.** El agente contesta a un "estoy fuera de la oficina", que vuelve a contestar, y así toda la noche. Hay que reconocer las respuestas automáticas y poner un tope de mensajes por conversación.

> ⚠ AVISO
>
> Ninguno de estos cuatro aparece en una demostración con casos bonitos. Salen con mensajes reales, así que entran en las pruebas desde el primer día.

EN UNA FRASE

El modelo redacta; el código que lo rodea es el que tiene que fallar avisando.

04 La regla de tres: registro, contador, canario

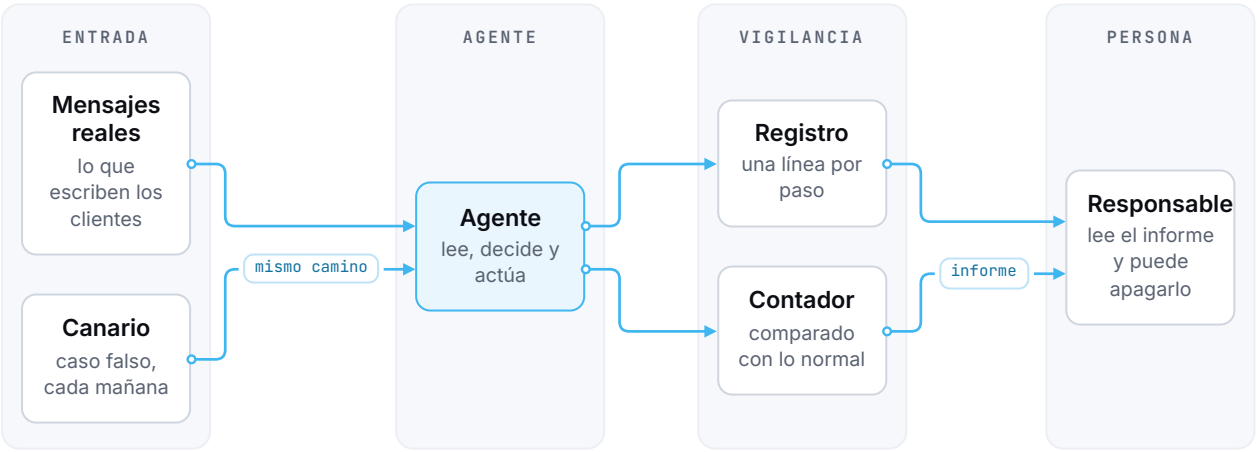
Cada agente que trabaja sin supervisión necesita tres piezas. No son caras ni complicadas, y entre las tres cubren casi todo lo que deja escapar la tabla anterior.

01 **Registro.** Una línea por paso: qué entró, qué se decidió y qué se hizo, con los datos personales enmascarados. Y una persona con nombre que lo lee, porque un registro que nadie mira no vigila nada.

02 **Contador.** Cuántos elementos leyó, cuántos procesó, cuántos pasó a una persona y cuántos fallaron, comparado con lo normal. Un lunes con cero mensajes leídos es una alarma, no un buen día.

03 **Canario.** Un caso falso y conocido que entra cada día por el mismo camino que los reales y tiene un resultado esperado. Si el canario no sale como debe, algo está roto aunque nadie se haya quejado.

FIGURA 1 **Las tres piezas alrededor de un agente**



EJEMPLO

Cada mañana entra un mensaje de prueba de "Cliente Canario" que pide cita para un día de cierre. La respuesta correcta es ofrecer otro día. Si no aparece en el informe, el sistema avisa.

Falta una comprobación más: si la ejecución más reciente es anterior a la hora programada, salta una alerta. Es la que caza la tarea que dejó de ejecutarse sin decir nada.

EN UNA FRASE

Si falta una de las tres piezas, el fallo puede pasar en silencio.

05

La autonomía se gana tarea a tarea

Un agente no es "de confianza" en bloque. Se confía en él **para un tipo de tarea**: contestar preguntas frecuentes, sí; cambiar una cita, todavía no. Esa decisión vive en la configuración y no en las instrucciones del modelo, para que nadie la cambie sin querer al retocar un texto.
Todo empieza en borrador. La IA redacta y una persona aprueba. Cada corrección se guarda como regla escrita y el caso entra en las pruebas. Un tipo de tarea pasa a envío automático solo cuando todas las pruebas pasan y el dueño lo pide.

Las pruebas son un conjunto de **casos de referencia** (en inglés, *golden set*): conversaciones reales anonimizadas, incluidos casos difíciles e intentos de engaño. Cada caso se ejecuta varias veces, por ejemplo cinco, porque un modelo puede acertar una vez y fallar la siguiente con el mismo mensaje. Basta un fallo para que ese tipo siga en borrador.

> *Lo que una regla puede comprobar lo comprueba el código; el modelo solo juzga lo que exige leer.*

PRINCIPIO DE DISEÑO DE FLOWARA AI

En la práctica, el código mira primero si la fecha existe, si el precio coincide con la tarifa y si la hora cae dentro del horario. Lo que exige leer, como el tono o si la respuesta contesta de verdad a la pregunta, lo juzga otro modelo. Ese juez también se calibra: antes de fiarte de él, compara sus notas con casos que una persona ya revisó a mano.

FIGURA 2 **Del borrador al envío automático**



EN UNA FRASE

La autonomía se demuestra con pruebas por tipo de tarea, nunca por intuición.

06

El informe de la mañana y quién lo lee

El informe diario es el contrato entre el agente y su responsable. Para poder creerlo necesita tres cosas:

01 **Contadores:** leídos, procesados, enviados, pasados a una persona y fallidos.

02 **Diferencias:** cada número comparado con lo habitual. "12 fallidos" no dice nada; "12 fallidos, lo normal es 0" sí.

03 **Lo que no hice:** qué no pudo leer, qué dejó en borrador y por qué. Si la lectura fue incompleta, el informe dice "cobertura incompleta", nunca "nada pendiente".

La diferencia se ve mejor con el mismo día contado de dos maneras:

TABLA 1 **El mismo martes, dos informes**

	Sin contadores	Con contadores
BUZÓN INACCESIBLE	"Sin pendientes"	"Leídos 0, lo normal es 30-45"
TAREA NO EJECUTADA	No llega nada	Alerta: no hubo ejecución a las 07:00
CANARIO	No existe	"Canario: no aparece"
LECTURA A MEDIAS	No se menciona	"Cobertura incompleta"
BORRADORES	No se mencionan	"4 cambios de cita esperan aprobación"

Un informe que funciona cabe en tres líneas:

```text
Lun 07:00 · leídos 41 (normal 35-50) · enviados 28 · a persona 9 · fallidos 0
Canario: OK · Cobertura: completa
No hecho: 4 cambios de cita en borrador (requieren aprobación)
```

Además, cada agente tiene **una persona con nombre** que lee ese informe, decide qué hacer con lo que falla y puede apagarlo. No vale "el equipo" ni "el proveedor": tiene que ser un nombre. Puedes delegar el trabajo, pero no la comprensión de lo que pasa.

EN UNA FRASE

Un informe que no dice qué quedó sin hacer no merece tu confianza.

07

Ejemplo: una peluquería de barrio con agente de citas

Una peluquería de barrio usa un agente que contesta los mensajes del negocio: horarios, precios, citas nuevas y cambios. Es un ejemplo inventado para ilustrarlo.

Semana 1. Todo va en borrador. La dueña aprueba cada respuesta desde el móvil y corrige tres veces el mismo error: el agente ofrecía cita los lunes, y los lunes cierran. Esa corrección se convierte en una regla escrita y en un caso de prueba.

Semana 3. Las preguntas frecuentes pasan todas las pruebas, cada caso ejecutado varias veces, y la dueña pide que salgan solas. El reparto queda así:

01 Preguntas frecuentes: envío automático.

02 Citas nuevas: borrador que ella aprueba.

03 Cambios de cita y quejas: siempre a ella.

Un martes. El informe dice "leídos 0 (normal 30-45)" y el canario no aparece. El acceso al canal de mensajes ha caducado. Sin contador, el informe habría dicho "sin pendientes" y la peluquería habría perdido un día de citas sin saberlo. Se renueva el acceso, se apunta su fecha de caducidad y se programa un aviso una semana antes.

Fíjate en que nadie tuvo que reescribir el agente. Bastó con una regla nueva y una fecha apuntada con su aviso.

EN UNA FRASE

Cada error acaba como regla escrita: así mejora un agente sin reescribirlo.

08

Checklist de producción

Antes de dejar un agente trabajando solo, comprueba estos catorce puntos. Si falta alguno, ese agente sigue en borrador:

Antes de dejar un agente solo

14 PUNTOS

- [ ] Una persona con nombre es responsable
  y puede apagarlo
- [ ] Cada paso deja una línea de registro
  con datos personales enmascarados
- [ ] Alguien lee el registro de errores cada día

- [ ] Los contadores se comparan con lo habitual
- [ ] Un canario diario recorre el camino real
- [ ] Una ejecución que no ocurre dispara una alerta
- [ ] "No encontrado" y "fallido" son resultados distintos
  ninguno cuenta como éxito
- [ ] Toda respuesta ilegible acaba en una persona
- [ ] Hay tope de mensajes por conversación
  corta bucles entre respuestas automáticas
- [ ] Ningún mensaje se queda en "enviando" sin revisión
- [ ] Cada dato y cada acceso tienen fecha de caducidad
- [ ] El envío automático va por tipo de tarea
  y vive en la configuración
- [ ] Cada tipo automático pasó todas sus pruebas
  cada caso, varias veces
- [ ] El informe incluye lo no hecho y la cobertura

SIGUIENTE PASO

Diagnóstico gratuito (20 min)

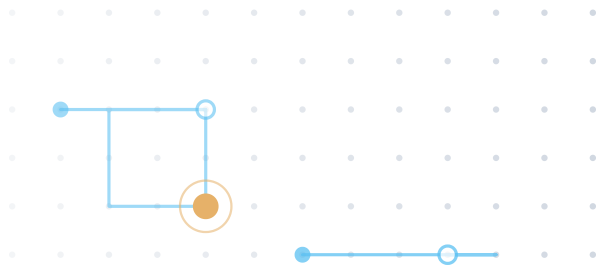
Si quieres ver qué parte de tu negocio se puede automatizar con el riesgo bajo control, reserva un diagnóstico gratuito de 20 minutos.

Reservar diagnóstico ›

flowaraai.com/es/contacto

RECURSOS RELACIONADOS

- Plantilla pública de agente de reservas ↗
- Casos reales de automatización ↗



Mario Chasco · Flowara AI

Sistemas de IA en producción para negocios pequeños.

flowaraai.com
github.com/mariochasco78557
youtube.com/@mariochasco_AI
linkedin.com/in/mario-adrian-chasco-frauca

Puedes compartir esta guía tal cual, citando flowaraai.com.