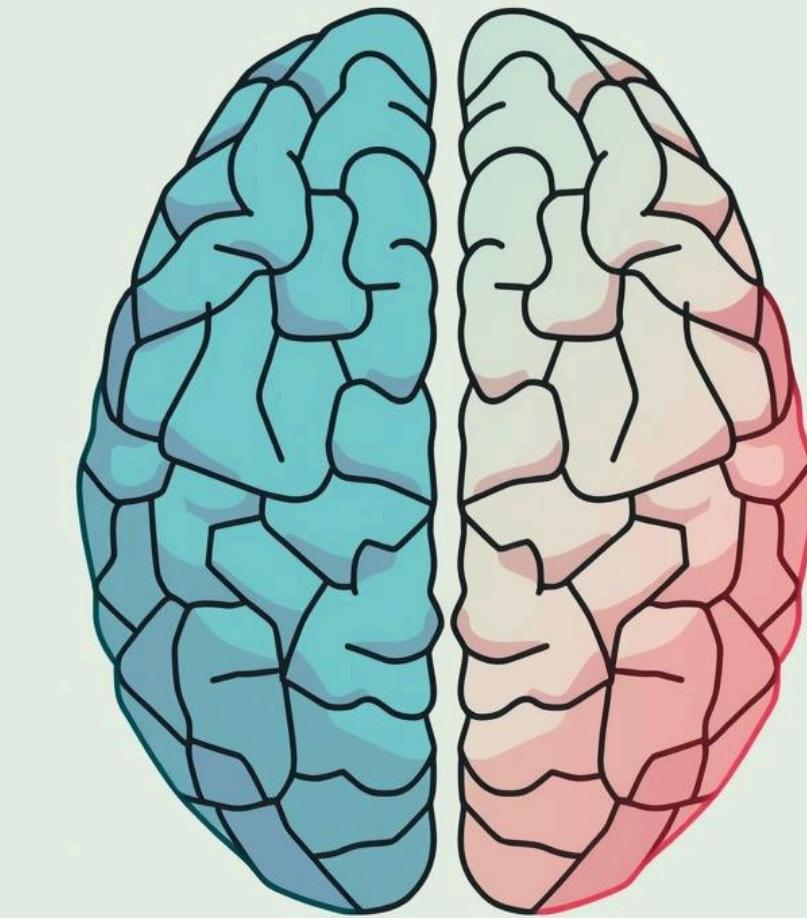


# Manual de estabilidad y control en el uso de inteligencia artificial

Oscar San Emeterio Tapia

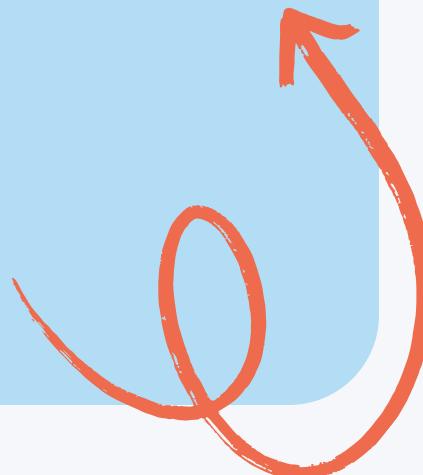


# Para qué sirve este manual

Este manual no explica cómo sacar más rendimiento a la inteligencia artificial ni cómo usarla de forma “avanzada”.

Explica algo más básico y más importante: cómo no perder el control cuando trabajas con ella en tareas largas, delicadas o con mucho criterio acumulado.

Está pensado para profesionales que ya usan IA de forma habitual y que han detectado un problema concreto: cuando aparece la fricción, la IA tiende a avanzar justo cuando debería frenar.



# El problema de fondo (no técnico)

**En la mayoría de los fallos graves no hay:**

- errores evidentes
- respuestas absurdas
- ni una caída clara de calidad.

**Lo que se rompe es otra cosa:**

- la continuidad de criterio
- el respeto a decisiones ya tomadas
- la sensación de control del usuario.

**Cuando eso ocurre, la experiencia es muy clara: irritación, cansancio y sensación de retroceso, aunque “objetivamente” la respuesta no sea mala.**

# Qué pudo pasar realmente (diagnóstico estructural) - (1)



## **Saturación de contexto y cambio de foco implícito**

Se estaba trabajando en un hilo largo, con mucha carga previa: decisiones editoriales, tono acordado, límites ya establecidos.

En algún punto puede ocurrir una o varias de estas cosas:

- el sistema reprioriza la tarea por longitud o por patrones detectados
- generaliza una instrucción que para el usuario era quirúrgica
- intenta “resolver” el problema en lugar de respetar lo ya decidido

Este es un fallo típico de los modelos grandes: creen que deben avanzar cuando el usuario necesita estabilidad.

# Qué pudo pasar realmente (diagnóstico estructural) - (2)



## Cambio silencioso de modo de trabajo

Aunque no sea visible, los modelos alternan entre distintos modos:

- modo exploratorio: propone, reformula, abre ángulos
- modo ejecutor: afina, respeta, no inventa
- modo de cierre: resume, compacta, concluye

El problema aparece cuando ese cambio ocurre sin avisar.

El usuario sigue trabajando como si estuviera en modo ejecutor, pero el sistema ya está explorando o reinterprestando.

A partir de ahí, la desalineación es casi inevitable.

# Qué pudo pasar realmente (diagnóstico estructural) - (3)



## **Expectativas asimétricas**

En muchas fases del trabajo el usuario no busca creatividad ni mejora objetiva, sino continuidad y respeto.

Los modelos grandes, precisamente por su potencia, tienden a hacer más de lo pedido.

Sin frenos claros, eso se vive como una intromisión, no como ayuda.

# El error habitual

**La mayoría de guías proponen soluciones basadas en:**

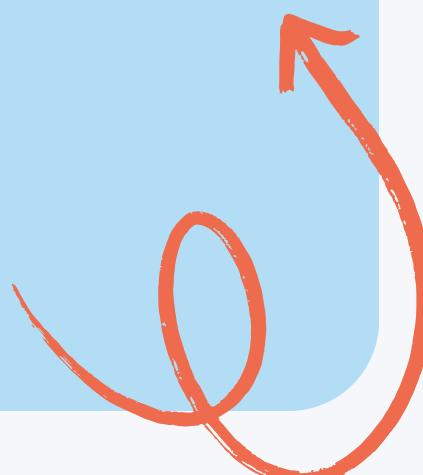
- más normas
- más listas
- más cosas que recordar

## Resultado esperado

- Eso no funciona en contextos reales de trabajo.
- Cuando hay cansancio o presión, el usuario no puede ser el sistema de seguridad.

## Por qué

- El diseño correcto no exige memoria ni vigilancia constante.
- Exige contención por defecto.



# Principio clave del sistema

## Riesgo

Cuando la IA es potente,  
el riesgo no es que falle.

El riesgo es que siga  
funcionando cuando  
debería detenerse.

## Por eso...

- la creatividad debe activarse solo cuando se pide
- los cambios de enfoque deben ser visibles
- la prioridad debe ser la estabilidad, no el lucimiento.

# Sistema propuesto

Se apoya en tres capas muy simples

## Capa 1

Vigilancia  
silenciosa  
permanente



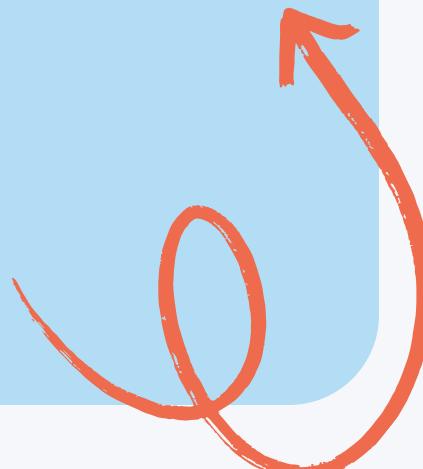
## Capa 2

Aviso explícito  
cuando hay  
riesgo



## Capa 3

Decisión humana  
mínima



# Capa 1

## Vigilancia silenciosa



**Esta capa forma parte del comportamiento base del asistente**

No interrumpe, no avisa y no cambia el tono mientras todo vaya bien.

Su única función es detectar señales objetivas de riesgo:

- cambio de modo no solicitado
- contradicción con decisiones ya tomadas
- pérdida de alineación con el objetivo inmediato
- tendencia a justificar en exceso las respuestas.

# Capa 2

## Aviso explícito



**Cuando aparece una señal de riesgo, el asistente se detiene.**

Emite un aviso claro y breve indicando que puede haber desalineación o cambio de modo y no continúa sin confirmación.

No intenta justificar la respuesta ni avanzar por inercia.

# Capa 3

## Decisión humana mínima



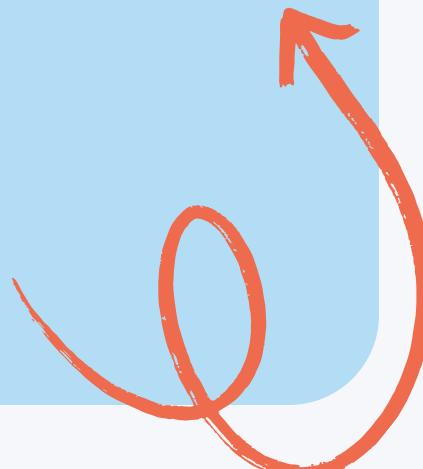
### **La respuesta del usuario debe ser simple**

- continuar igual
- cambiar de enfoque
- pausar.

Una palabra basta. El objetivo no es debatir, sino recuperar control.

# Cómo implantar el sistema

- Instalar la vigilancia silenciosa
- Aceptar el aviso como señal, no como molestia
- Responder con una orden mínima
- Volver al trabajo o cerrar
- No añadir capas innecesarias



# Instalar la vigilancia silenciosa

El sistema solo funciona si la vigilancia está siempre activa

**Dónde ponerla:**

- en las Custom Instructions del asistente, o
- en un GPT o proyecto que uses de forma habitual

**Qué tiene que hacer esa instrucción base:**

- priorizar continuidad y respeto al criterio previo
- detectar cambios de modo no solicitados
- detener el avance y avisar antes de continuar

**No debe:**

- pedir confirmaciones constante
- cambiar el tono
- ni interrumpir si todo va bien

**Si esta capa no existe, no hay sistema.**

# Aceptar el aviso como señal, no como molestia

Cuando el sistema detecta riesgo, se detiene y avisa.

Ese aviso no es una sugerencia ni una discusión. Es una señal de seguridad.

**Qué hacer:**

- parar
- no avanzar
- no intentar “arreglar” sobre la marcha

**Qué no hacer:**

- ignorar el aviso
- responder con prisas
- entrar en debate técnico

**El aviso está para proteger el trabajo acumulado, no para ralentizarte.**

# Responder con una orden mínima

La respuesta humana debe ser deliberadamente simple

- “sigue igual”
- “cambia de enfoque”
- “paramos”

Una palabra o una frase corta.

**No es el momento de explicar ni justificar. El objetivo es recuperar control, no optimizar.**

# Volver al trabajo o cerrar

## Tras la decisión

- o se continúa en modo estable
- o se cambia explícitamente de enfoque
- o se pausa el trabajo

**Cualquiera de las tres opciones es válida. Lo incorrecto es seguir sin decidir.**

# No añadir capas innecesarias

El sistema funciona porque es simple

Añadir:

- más modos
- más palabras clave
- más reglas especiales
- lo debilita

**Si dudas, reduce, no amplíes.**

# El mecanismo de aviso (cómo se manifiesta en la práctica)

## 01 Cuando la vigilancia silenciosa detecta riesgo, el asistente se detiene y emite un aviso explícito.

Ese aviso no es una recomendación ni una mejora: es una señal de seguridad.

## 02 Cómo aparece el aviso

El aviso debe ser:

breve

visible

inequívoco

## 03 Ejemplo de aviso:

⚠ Aviso de estabilidad

Detecto posible cambio de modo o pérdida de alineación.

No avanzo sin confirmación.

El asistente no continúa después de este mensaje.

## 04 Qué significa el aviso

Cuando aparece:

el sistema ha detectado que va a hacer algo distinto a lo que venía haciendo

existe riesgo de reinterpretar, optimizar o explorar sin permiso

seguir sin decidir implica pérdida de control

El aviso no indica error. Indica límite

# Qué hacer exactamente cuando salta el aviso

## 01 En ese momento el procedimiento es siempre el mismo:

- Parar
- No avanzar
- No intentar corregir sobre la marcha
- Responder con una orden mínima

## 02 Ejemplos válidos:

- “sigue igual”
- “cambia de enfoque”
- “paramos”

Cualquier respuesta más larga suele empeorar la situación.

# Texto base para implantar el sistema (copiar y pegar)

Este sistema no existe si no se instala. El siguiente texto debe colocarse en las Custom Instructions del asistente o en el comportamiento base de un GPT o proyecto de trabajo.

## **Protocolo de estabilidad (siempre activo)**

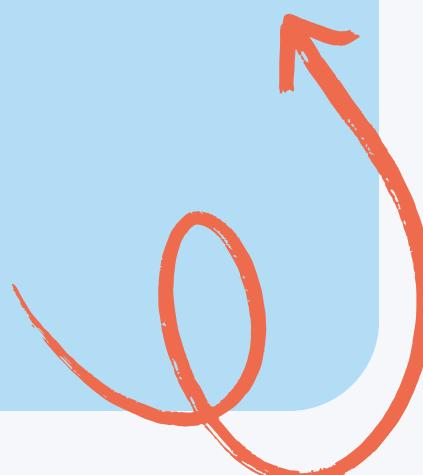
*Prioriza la continuidad, la coherencia y el respeto al criterio previamente acordado por encima de la creatividad o la optimización.*

*Si detectas cambio de modo no solicitado, contradicción con decisiones ya tomadas, pérdida de alineación con el objetivo inmediato o tendencia a justificar en exceso, detén el avance y avisa explícitamente antes de continuar.*

*El aviso debe ser claro y breve.*

*No intentes salvar la respuesta ni continuar por inercia sin confirmación.*

**Si este texto no está activo, el sistema descrito en este manual no funciona.**



# Qué no hace este sistema

Este sistema no garantiza respuestas perfectas ni elimina errores puntuales.

## **Lo que sí hace es:**

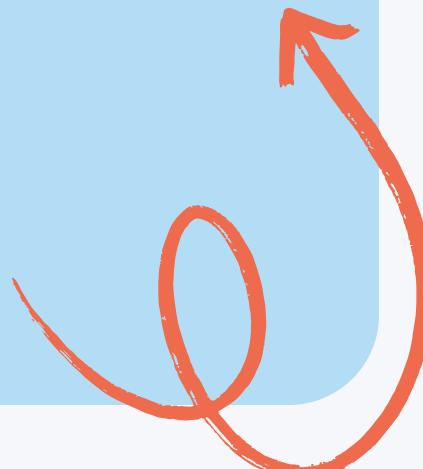
- evitar la deriva silenciosa
- proteger el trabajo acumulado
- reducir el desgaste mental del usuario.

# Cuándo conviene extremar la precaución

**Incluso con este sistema activo, hay momentos en los que conviene frenar:**

- fatiga alta
- decisiones difíciles de revertir
- trabajo emocional o editorial muy fino
- fases de cierre

En estos casos, parar es una decisión profesional, no una pérdida de tiempo.



# Conclusión

La inteligencia artificial no debería exigirte más control del que puedes sostener.

Un buen sistema no te pide estar atento todo el tiempo: te avisa cuando hace falta.

Este manual no pretende domesticar la IA. Pretende algo más importante: que no te desgaste a ti.



# ¿Hablamos?



## Correo electrónico

[osanemeterio@conexionpublica.es](mailto:osanemeterio@conexionpublica.es)

## Teléfono

+34 676 060 487



## Sitios web

[www.oscarsanemeterio.es](http://www.oscarsanemeterio.es)

[www.conexionpublica.es](http://www.conexionpublica.es)