DESCRIPCIÓN DEL MÓDULO

# La IA como espejo: sus respuestas dependen de nuestras preguntas

Módulo adaptativo · ~20 min

Ante la inteligencia artificial (IA), oscilamos entre la fascinación y el miedo, dos posturas que le atribuyen una inteligencia que no tiene. Entenderla se vuelve decisivo para usarla. Con la idea de «espejo» de Shannon Vallor, este módulo muestra que la IA no piensa: nos devuelve nuestros datos, nuestros sesgos e incluso nuestro estado interior. Sus fallos —alucinaciones, complacencia, confusión— hablan de nosotros. Su lección: la calidad de lo que se obtiene depende de lo que se aporta.

## OBJETIVOS PEDAGÓGICOS

01 Comprender qué es la IA: un algoritmo que automatiza, no un cerebro con intención o conciencia.

02 Ver la IA como un espejo que devuelve nuestros datos, nuestros sesgos, nuestra confusión y nuestro estado interior del momento.

03 Mantener el espíritu crítico ante tres reflejos: la alucinación, la complacencia y la respuesta genérica por defecto.

04 Mejorar el diálogo: contexto duradero, permiso para discrepar, razonamiento explicado, relectura de las propias peticiones.

## PÚBLICO DESTINATARIO

Cualquier persona que usa la IA a diario y quiere ganar finura en su forma de preguntarle.

- **Requisitos previos:** ninguno

## DESARROLLO DEL MÓDULO

En autonomía, con un entrenamiento adaptado a cada participante.

1. **Autodiagnóstico**: sobre 9 dimensiones, para situar las propias prácticas y orientar el entrenamiento.
2. **Vídeo de experto**: de 10 minutos, con 3 pausas de reflexión.
3. **Casos prácticos**: 2 primeros casos elegidos a partir del autodiagnóstico y, después, otros 16 opcionales. Cada respuesta recibe un feedback: sus efectos y una pista para mejorar.
4. **Compromiso con fecha**: una acción que probar, redactada por el participante o elegida entre las propuestas.
5. **Ficha resumen descargable**: puntos clave, preguntas, compromiso, notas personales y listas de comprobación.

## UN EJEMPLO ENTRE LOS CASOS PROPUESTOS

Un cliente descontento ha escrito a Kwame, gestor de clientes en una empresa de alquiler, y a su compañera Farah. Farah cree que le toca a Kwame; él cree que le toca a ella. Kwame cuenta su versión al asistente de IA, que le da la razón. ¿Qué debe hacer?

## EVALUACIÓN Y VALIDACIÓN

Umbral: 80 %

La puntuación tiene en cuenta las respuestas a los casos prácticos, el compromiso formulado y, en su caso, las actividades de aprendizaje complementarias. Para validar el módulo se requieren un compromiso y una puntuación de al menos 80 %.

### MÉTODOS Y MODELOS

- La IA como espejo: Shannon Vallor, The AI Mirror (2024)
- Complacencia algorítmica (sycophancy): concepto técnico reconocido
- Dar a la IA permiso para discrepar: método Afthonios

### FUENTES Y REFERENCIAS


- Stanford, estudio sobre la complacencia de las IA (Science, 2026): los modelos dan la razón al usuario un ~49 % más de veces que un humano
- Shannon Vallor, The AI Mirror (2024), profesora de ética de la IA en la Universidad de Edimburgo
- Vectara, «Hallucination Leaderboard» (actualización de mayo de 2026): tasa de alucinación al resumir documentos, del 1,8 % en el mejor modelo al 24,2 % en el menos fiable
- Dhuliawala et al., Findings of ACL (2024): verificación estructurada y fiabilidad de las respuestas de la IA



Módulo creado por Afthonios.com.

Contenido verificado el 15 de octubre de 2026