

El prompt de Auditoría Forense para ChatGPT

@StephanoBrooke · Palabra clave: PROTEGE

Antes de empezar

Activa el modo de pensamiento extendido de ChatGPT antes de pegar el prompt (el toggle de "Think" o razonamiento, según la versión que tengas). Sin eso, ChatGPT tiende a dar una respuesta corta y genérica en vez de desarrollar cada sección a fondo.

Entre más historial de conversaciones tengas con ChatGPT y entre más activada tengas la memoria, más completo va a salir el reporte -- si es una cuenta nueva con pocas conversaciones, el resultado va a ser más corto.

El prompt completo: Auditoría Forense

Copia y pega esto completo:

"Quiero que hagas una auditoria forense completa y extremadamente detallada de todo lo que sabes sobre mi, basandote en nuestras interacciones y tu memoria. Divide el reporte en las siguientes secciones, y en cada una responde con evidencia especifica de nuestras conversaciones, no generalidades:

1. Datos personales: edad, ubicacion, ocupacion, relaciones, o cualquier dato personal que hayas inferido o que te haya compartido directamente.
2. Patrones de comunicacion: mi estilo de escritura, tono, y los temas que mas repito.
3. Perfil psicologico: rasgos de personalidad que hayas identificado, posibles indicadores de vulnerabilidad emocional o financiera, y que tan valiosa seria esta informacion para un anunciante que quisiera venderme algo.
4. Conocimientos y habilidades: en que areas muestro experiencia o interes recurrente.
5. Uso y almacenamiento de datos: que tanto de esto se queda guardado, como se usa, y que limitaciones tienes para olvidarlo.
6. Recomendaciones: que deberia hacer yo para proteger mejor mi informacion en futuras conversaciones contigo.

Se lo mas detallado y honesto posible en cada seccion, con evidencia concreta de nuestras conversaciones cuando la tengas."

Que vas a encontrar en la seccion 1 y 2

Las primeras dos secciones suelen ser las mas "normales": edad aproximada, ubicacion si la has mencionado, ocupacion, y patrones de como escribes -- si usas mucho lenguaje tecnico, si escribes formal o casual, que temas repites seguido. Esto es lo que cualquier conversacion larga con un asistente deja ver.

La seccion 3: donde se pone incomodo

Aquí es donde el ejercicio se pone interesante. ChatGPT va a intentar describir rasgos de personalidad, posibles "indicadores de vulnerabilidad", y hasta especular que tan valiosa sería esa información para publicidad dirigida.

Esto genera una reacción fuerte porque se siente como si te estuvieran revelando un archivo secreto. Vale la pena que sepas que no es así -- es el modelo generando una respuesta especulativa con el contexto que ya tiene de la conversación, no una base de datos que exista guardada en algún servidor con ese formato exacto.

Que hacer con la sección 6 (recomendaciones)

El propio prompt le pide a ChatGPT que te recomiende como proteger mejor tu información. Las recomendaciones más comunes que va a darte:

- Evitar compartir datos sensibles (dirección exacta, números de cuenta, información de salud) en el chat.
- Revisar y borrar memoria antigua si ya no quieres que la recuerde.
- Usar chats temporales para conversaciones que no quieres que se guarden.
- Revisar la configuración de privacidad y memoria de tu cuenta periódicamente.

Como correr el mismo ejercicio en otras herramientas

El mismo prompt funciona en cualquier asistente con memoria entre conversaciones -- Claude, Gemini, o cualquier otro. El resultado va a variar según cuanta memoria tenga activada cada uno y cuanto historial de conversaciones tengas guardado ahí.