Título: CHECKLIST DE AUDITORÍA: ROBUSTEZ Y GOBERNANZA DE SISTEMAS IA

Subtítulo: Protocolo de 25 puntos críticos para el paso a producción de sistemas generativos.

Versión: 1.2 (Febrero 2026)

Autor: Antonio Gutiérrez – jefedeproyectos.com

INTRODUCCIÓN Y METODOLOGÍA

Esta auditoría está diseñada para detectar vulnerabilidades en tres ejes: **Operatividad (Ops)**, **Fidelidad Semántica** y **Soberanía de Infraestructura**. No es una revisión estética; es un control de seguridad sistémica.

EL CHECKLIST TÉCNICO

SECCIÓN 1: GOBERNANZA Y PROMPT OPS (LÓGICA)

- [ ] C1. Versionado de Artefactos: ¿Están todos los prompts versionados en Git como código y no "incrustados" en el backend?
- [ ] C2. Test de Regresión Semántica: ¿Existe un pipeline de CI/CD que valide que un cambio en el prompt no degrada respuestas anteriores?
- [ ] C3. Esquemas de Salida (JSON Schema): ¿Se utiliza Pydantic o similares para forzar que el modelo responda con una estructura de datos válida?
- [ ] C4. Golden Dataset: ¿Contamos con un conjunto de 50-100 ejemplos de "respuestas perfectas" para comparar con cada nueva build?
- [ ] C5. Auditoría de Variables de Contexto: ¿Se limpian y sanitizan las entradas de usuario antes de inyectarlas en el prompt (prevención de Prompt Injection)?

SECCIÓN 2: INFRAESTRUCTURA E INFERENCIA (SISTEMAS)

- [ ] S1. Estrategia de Inferencia Local/Cloud: ¿Se ha evaluado el uso de SLMs locales (Phi-4, Llama 3) para reducir el TCO y aumentar la privacidad?
- [ ] S2. Gestión de Latencia (TTFT): ¿El tiempo hasta el primer token es inferior a 2 segundos en el 95% de las peticiones?
- [ ] S3. Cuantización Optimizada: En caso de usar modelos locales, ¿se ha validado que el formato (GGUF/EXL2) no afecta la perplejidad crítica de la tarea?
- [ ] S4. Circuit Breaker: ¿Existe un mecanismo de fallback automático que active un flujo determinista si la API externa falla?
- [ ] S5. Rate Limiting: ¿Está protegida la infraestructura contra ráfagas de peticiones que puedan disparar los costes de tokens?

SECCIÓN 3: GESTIÓN DEL CONTEXTO Y RAG (DATOS)

- [ ] D1. Densidad de Tokens: ¿Se aplica poda de contexto (Pruning) para evitar el fenómeno "Lost in the Middle"?
- [ ] D2. Re-ranking Semántico: ¿Se utiliza un re-ranker (ej. Cohere/BGE) tras la búsqueda vectorial para asegurar que solo lo más relevante entra en la ventana de contexto?
- [ ] D3. Trazabilidad de Fuentes: ¿El sistema cita explícitamente el origen de cada dato extraído para evitar alucinaciones?
- [ ] D4. Ventana de Memoria Dinámica: ¿Se gestiona el historial de conversación mediante resúmenes automáticos para no saturar la ventana de tokens?
- [ ] D5. Desacoplamiento de Datos: ¿Están los datos sensibles anonimizados antes de ser procesados por modelos de terceros?

MÉTRICAS DE ÉXITO Y OBSERVABILIDAD

SECCIÓN 4: EVALUACIÓN Y CALIDAD (OUTPUT)

- [ ] E1. Monitorización LLM-as-a-judge: ¿Tenemos un modelo superior evaluando la calidad semántica de las respuestas del modelo de producción?
- [ ] E2. Detección de Alucinaciones: ¿Existen checks de "Grounding" (verificación contra la base de conocimientos) antes de mostrar la respuesta?
- [ ] E3. Análisis de Deriva (Drift): ¿Se revisa mensualmente si el modelo está cambiando su tono o precisión debido a actualizaciones del proveedor?
- [ ] E4. Feedback Loop: ¿Hay un mecanismo para que los usuarios marquen respuestas erróneas y estas alimenten el Golden Dataset?
- [ ] E5. Coste por Tarea: ¿Sabemos exactamente cuánto cuesta cada ejecución de extremo a extremo (Token In + Token Out + Cómputo)?

JEFEDEPROYECTOS.COM

CONCLUSIÓN DE AUDITORÍA

Interpretación de Resultados:

- 20-25 Puntos: Sistema de grado industrial. Listo para escala masiva.
- 15-19 Puntos: Riesgo moderado. Priorizar la implementación de Prompt Ops y Seguridad.
- Menos de 15 Puntos: Prototipo frágil. Riesgo alto de alucinaciones y costes descontrolados. No apto para producción crítica.

Nota: " *La robustez en IA no es el resultado de un modelo potente, sino de una arquitectura que asume que el modelo va a fallar. Auditar es el primer paso para dejar de jugar con la IA y empezar a construir ingeniería con ella.*"

Más recursos en: 🌐 https://jefedeproyectos.com