

Análisis de Caso GenAI Fallido



PRÁCTICA PROFESIONALIZANTE I

JUL
17

Fecha de Entrega: Viernes 29 de Agosto, 2025 - 21:00hs

Mauricio Molina

45.699.198

TSCDIA

B - Viernes

Repositorio



Dianela Accietto

A) INVESTIGACIÓN PROFUNDA DEL CASO

Causas raíz del fracaso:

El principal problema fue un entrenamiento insuficiente del modelo, ya que la base de conocimiento utilizada no contemplaba terminología médica específica ni situaciones complejas frecuentes en la clínica. A esto se sumó una mala integración con los sistemas internos (HIS y gestión de turnos), lo que generaba información desactualizada y respuestas inconsistentes. Otro factor clave fue la falta de pruebas reales antes del lanzamiento, lo que impidió detectar con anticipación errores críticos en el funcionamiento del chatbot. Finalmente, las decisiones apresuradas por presión de tiempo llevaron a implementar un sistema que no estaba listo para producción.

Stakeholders involucrados:

Entre los actores más importantes se encontraba la dirección de la clínica, interesada en reducir tiempos de espera y costos operativos; el equipo técnico, con limitaciones en experiencia en IA aplicada a la salud; los pacientes, usuarios finales en su mayoría adultos mayores que requerían una interfaz clara y confiable; y el personal médico, que tuvo poca participación en la validación de las respuestas, lo cual resultó un error crítico.

Decisiones clave tomadas:

Un punto determinante fue la elección de un modelo genérico (GPT-3.5) sin un entrenamiento profundo en aspectos médicos, lo que limitó la precisión de las respuestas. La falta de supervisión médica en la construcción y validación de la base de conocimiento agravó este problema, permitiendo que circulara información incorrecta o incompleta. Por último, la decisión de lanzar el chatbot sin resolver las alertas detectadas en la fase beta marcó un punto de inflexión que aceleró el fracaso del proyecto.

B) ¿QUÉ PUDO PREVENIRSE?

Problemas prevenibles:

Varios de los inconvenientes del proyecto pudieron haberse evitado con una mejor planificación. La falta de pruebas exhaustivas antes del lanzamiento fue determinante, ya que en la beta ya se habían detectado fallas de precisión. También se podría haber resuelto el problema de la información médica incorrecta, validando previamente las respuestas con personal calificado. Otro aspecto prevenible fue la sincronización deficiente de los turnos, que generaba desconfianza en los pacientes.

Red flags ignoradas:

Las pruebas iniciales con 50 pacientes mostraron baja precisión en las respuestas y problemas de compatibilidad con el sistema hospitalario, pero estos indicadores no fueron atendidos. Asimismo, la latencia alta y las caídas del sistema ya eran visibles antes del lanzamiento oficial.

Limitaciones técnicas prevenibles:

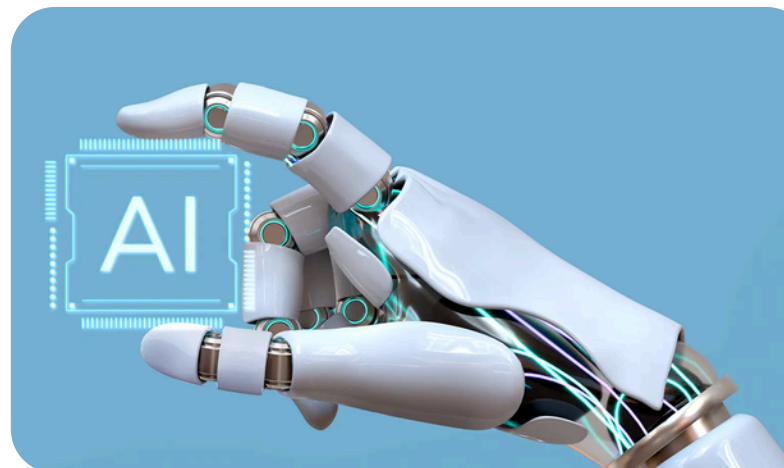
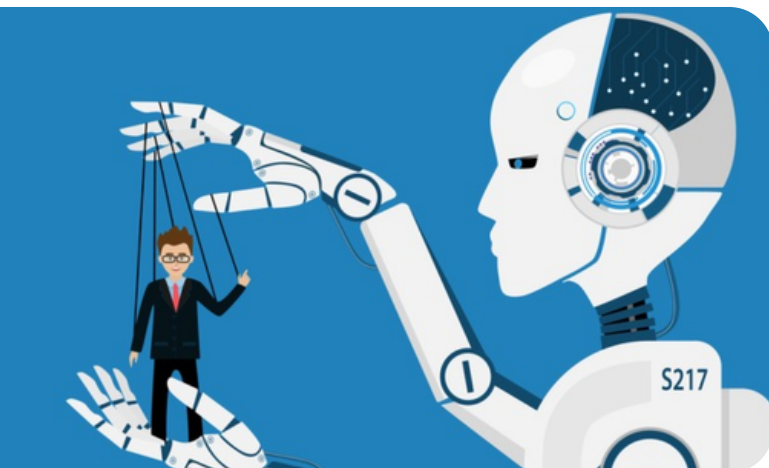
Al revisar el análisis técnico, se observa que varias de las fallas identificadas podían mitigarse con mayor profundidad en el diseño y la implementación. El entrenamiento insuficiente del modelo, limitado a una base de 500 preguntas frecuentes, era claramente inadecuado para el dominio médico y podría haberse reforzado con un set de datos más amplio y específico. La falta de integración profunda con el sistema hospitalario impidió acceder a historiales médicos y disponibilidad real de turnos, algo que pudo haberse trabajado con una integración gradual y pruebas de compatibilidad. Asimismo, la ausencia de un mecanismo de escalamiento a humanos era un riesgo previsible que podría haberse incorporado desde el inicio como parte de la lógica del sistema. Finalmente, aunque el procesamiento de lenguaje natural limitado y la falta de memoria de contexto respondían en parte a restricciones del modelo, una mejor estrategia de entrenamiento y diseño de prompts hubiera reducido en gran medida los errores de comprensión.

C) ¿QUÉ NO PUDO PREVENIRSE?

Algunos aspectos del fracaso estuvieron directamente vinculados a limitaciones propias de la tecnología GenAI utilizada. El modelo seleccionado, GPT-3.5 con un fine-tuning básico, no estaba diseñado para manejar con precisión terminología médica ni casos complejos, lo que derivó en respuestas incompletas o incorrectas. Esta carencia no podía resolverse totalmente dentro de las capacidades del modelo en ese momento, ya que se trataba de un sistema generalista y no de un modelo entrenado específicamente en salud.

Otra limitación fue la ausencia de memoria de contexto, que hacía que cada interacción se tratara de forma independiente, impidiendo al chatbot sostener conversaciones largas o reconocer situaciones de urgencia.

Si bien existió una inversión importante, ciertos resultados negativos fueron consecuencia de estas limitaciones sistémicas propias de la herramienta y no de la gestión del proyecto. En definitiva, aunque varios errores pudieron prevenirse con mejor planificación, otros estaban ligados al alcance real que la GenAI podía ofrecer en ese momento para un entorno tan sensible como el de la salud.



D) CÓMO TRABAJARÍAS EN EL FUTURO

Para implementar un sistema similar en el futuro, sería necesario un proceso de viabilidad más riguroso, evaluando desde el inicio riesgos médicos, legales y técnicos.

El proyecto debería avanzar de forma escalonada, comenzando con testing interno controlado, luego un piloto con un grupo reducido de pacientes y finalmente el lanzamiento general.

Los criterios de éxito/fallo deberían estar claramente definidos, por ejemplo: nivel mínimo de automatización alcanzado, tiempo máximo de respuesta y ausencia de errores en información médica crítica. También resultaría fundamental contar con un plan de contingencia, que asegure la intervención de un humano cuando el bot no pueda resolver la consulta. A nivel de gestión, la clínica debería trabajar con una metodología que priorice el feedback temprano y la revisión continua, para reducir el riesgo de repetir fallas críticas.

E) LECCIONES APRENDIDAS PERSONALES

Este caso deja como principal enseñanza la importancia de validar y supervisar con expertos médicos toda la información que maneje un sistema de IA en salud. También evidencia que los proyectos tecnológicos en entornos sensibles requieren testing profundo y sostenido, ya que errores aparentemente menores pueden derivar en consecuencias graves, tanto para los pacientes como para la institución.

Otro aprendizaje clave es que ignorar señales de alerta en fases tempranas puede escalar rápidamente a una crisis mayor. En lo personal, refuerza la idea de que trabajar con IA no es solo un desafío técnico, sino también ético y profesional: se deben asumir responsabilidades por el impacto de cada decisión. Para el futuro, es fundamental desarrollar competencias en gestión de riesgos, trabajo interdisciplinario y ética en el uso de inteligencia artificial.



F) RECOMENDACIONES ESTRATÉGICAS

Para la organización afectada: implementar un modelo híbrido, donde el bot solo gestione consultas simples y todo lo demás se derive a personal humano.

Para profesionales del sector: capacitarse en el uso responsable de IA y comprender los riesgos de aplicar tecnología genérica en contextos críticos como la salud.

Para proveedores de tecnología GenAI: desarrollar soluciones entrenadas específicamente en el dominio médico, con mecanismos de verificación y actualización de información en tiempo real.

Para reguladores del sector salud: establecer protocolos de seguridad, privacidad y calidad en el uso de IA, asegurando que las respuestas estén validadas y que los datos sensibles estén protegidos.

Como parte del seguimiento, se recomienda definir métricas concretas como: tasa de automatización alcanzada, precisión de respuestas médicas, tiempos de espera promedio, satisfacción del paciente y número de errores críticos detectados. Además, implementar indicadores de alerta temprana (como caídas del sistema o latencia elevada) y protocolos claros de escalamiento ayudaría a evitar que los problemas se repitan.

