

Análisis de Caso GenAI Fallido

Práctica Profesionalizante I
Instituto Superior Politécnico Córdoba

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

Información del Estudiante

NOMBRE Y APELLIDO
Juan Gabriel Cabrera

DNI
39448561

CARRERA
Tec. Sup. Cs de Datos e IA

COMISIÓN
A

REPOSITORIO

DOCENTE
Ing. Accietto Daniela

Caso: ChatBot de Atención al Cliente "MediBot"

Implementación fallida de IA conversacional para clínica médica

Resumen Ejecutivo del Problema

La Clínica MediCare implementó un chatbot de IA generativa llamado "**MediBot**" para automatizar la atención inicial de pacientes y gestión de turnos. A pesar de una inversión de **\$85,000 USD** y 8 meses de desarrollo, el sistema no logró resolver los pain points identificados, generando más problemas operacionales y afectando negativamente la experiencia del paciente.

23%

TASA DE RESOLUCIÓN EXITOSA

\$85k

INVERSIÓN TOTAL

47%

AUMENTO EN QUEJAS

8

MESES DE DESARROLLO

1. Contexto del Cliente

Cliente: Clínica MediCare - Centro médico privado con 150 empleados

Ubicación: Córdoba, Argentina

Sector: Salud y medicina privada

Situación Inicial:

- ❑ Recibían 800+ llamadas diarias para consultas y turnos
- 🕒 Tiempo de espera promedio: 12 minutos
- ❑ Solo 3 operadoras para atención telefónica
- ❑ 85% de consultas eran repetitivas (horarios, ubicación, preparación estudios)
- ❑ Alta rotación de personal de call center por estrés

2. Objetivos vs Resultados

OBJETIVO	PLANTEADO	META	ESPERADA	RESULTADO	OBTENIDO	ESTADO
Reducir tiempo de espera		< 3 minutos		15 minutos		✗ Empeoró
Automatizar consultas básicas		70% automatización		23% automatización		✗ No cumplido
Reducir costos operativos		40% reducción		15% aumento		✗ Contraproducente
Mejorar satisfacción paciente		8.5/10 NPS		4.2/10 NPS		✗ Crítico

3. Análisis Técnico

Tecnología Utilizada:

- Modelo: GPT-3.5 Turbo con fine-tuning básico
- Plataforma: Chatbot personalizado en React
- Integración: API REST con sistema de gestión HIS
- Base de conocimiento: 500 preguntas FAQ básicas

Limitaciones Técnicas Identificadas:

- ⚠

Entrenamiento Insuficiente: La base de conocimiento no incluía terminología médica específica ni casos edge complejos
- ⚠

Falta de Integración Profunda: El bot no tenía acceso a historiales médicos ni disponibilidad real de turnos
- ⚠

Sin Escalamiento Inteligente: No detectaba cuándo derivar a un humano, generando frustración
- ⚠

Procesamiento de Lenguaje Natural Limitado: No comprendía jerga local cordobesa ni variaciones en expresiones
- ⚠

Sin Memoria de Contexto: Cada consulta era independiente, perdiendo el hilo de conversaciones complejas

4. Timeline de Implementación

ENERO 2024

Kick-off del Proyecto

Reunión inicial con stakeholders. Definición de requirements básicos.

FEB - MAR 2024

Desarrollo MVP

Creación del chatbot básico con 200 preguntas FAQ. Primeras pruebas internas.

ABRIL 2024

Integración con HIS

Conexión básica con el sistema hospitalario. Problemas de compatibilidad.

MAYO 2024

Pruebas Beta

Testing con 50 pacientes. Primeras señales de alerta por baja precisión.

JUNIO 2024

Lanzamiento Oficial

Go-live completo. Inmediata sobrecarga del sistema y quejas de usuarios.

JULIO - AGOSTO 2024

Crisis y Desactivación Parcial

Escalamiento de problemas. Decisión de reducir funcionalidades del bot.

✕ 5. Problemas Críticos Identificados

PROBLEMA CRÍTICO: El bot daba información médica incorrecta o desactualizada, generando riesgo para los pacientes y responsabilidad legal para la clínica.

Categorías de Problemas:

Problemas de Seguridad y Compliance:

- ⚠ Violación de privacidad: El bot almacenaba datos sensibles sin encriptación adecuada
- ⚠ Información médica incorrecta: 18% de las respuestas sobre preparación de estudios eran erróneas
- ⚠ Sin validación profesional: Respuestas no revisadas por personal médico calificado

Problemas Técnicos:

- ⚠ Latencia alta: 8-12 segundos de respuesta promedio
- ⚠ Caídas del sistema: 15 interrupciones en el primer mes
- ⚠ Sincronización deficiente: Información de turnos desactualizada en tiempo real

Problemas de Experiencia Usuario:

- ⚠ Interfaz confusa para adultos mayores (65% de la base de pacientes)
- ⚠ No reconocía urgencias médicas ni las derivaba correctamente
- ⚠ Respuestas robóticas que generaban frustración en situaciones delicadas

6. Impacto en el Cliente

-47%

SATISFACCIÓN DEL CLIENTE

+15%

COSTOS OPERATIVOS

+25%

TIEMPO DE ESPERA

-12%

NUEVOS PACIENTES

Consecuencias Específicas:

- Pérdida de Confianza:** 340 pacientes cambiaron de prestador médico
- Sobrecarga de Staff:** Las operadoras humanas recibieron 60% más llamadas de escalamiento
- Reputación Online:** Calificación en Google bajó de 4.3 a 2.8 estrellas
- Costos Adicionales:** \$23,000 USD extra en soporte técnico y recursos humanos
- Riesgo Legal:** 3 reclamos por información médica incorrecta

7. Consignas de Trabajo - Análisis Profundo

Instrucciones: A continuación se presentan las consignas que debes desarrollar en profundidad para completar tu análisis del caso GenAI fallido. Cada punto requiere investigación, análisis crítico y propuestas fundamentadas.

A) Investigación Profunda del Caso

Realiza una investigación exhaustiva sobre el caso MediBot. Analiza en detalle las causas raíz del fracaso, investigando aspectos técnicos, organizacionales, de gestión de proyecto y factores humanos que contribuyeron al resultado negativo. Incluye análisis de stakeholders, decisiones clave tomadas y puntos de inflexión.

B) ¿Qué Pudo Prevenirse?

Identifica qué problemas eran prevenibles con mejor planificación, testing, o metodología. Analiza las red flags que fueron ignoradas, los riesgos que no se evaluaron correctamente y las mejores prácticas que no se aplicaron. Propón un plan de prevención de riesgos específico para este tipo de proyectos.

C) ¿Qué NO Pudo Prevenirse?

Examina qué aspectos del fracaso eran inherentes a las limitaciones tecnológicas actuales de GenAI, restricciones presupuestarias realistas, o factores externos fuera del control del equipo. Distingue entre problemas evitables y limitaciones sistémicas o tecnológicas genuinas.

D) Cómo Trabajarías en el Futuro

Desarrolla una metodología mejorada para implementaciones futuras de GenAI en el sector salud. Incluye: framework de evaluación de viabilidad, proceso de testing escalonado, criterios de éxito/fallo, estrategias de gestión del cambio y plan de contingencia. Considera avances tecnológicos esperados.

E) Lecciones Aprendidas Personales

Reflexiona sobre las lecciones más importantes que extraes de este caso. ¿Cómo cambiaría tu enfoque profesional? ¿Qué principios adoptarías para futuros proyectos de IA? ¿Qué competencias consideras esenciales desarrollar? Incluye reflexiones sobre ética en IA y responsabilidad profesional.

F) Recomendaciones Estratégicas

Formula recomendaciones específicas y accionables para: la organización afectada, profesionales del sector, proveedores de tecnología GenAI, y reguladores del sector salud. Incluye métricas de seguimiento, indicadores de alerta temprana y protocolos de escalamiento.

Entrega: Tu análisis debe ser exhaustivo, bien documentado y demostrar pensamiento crítico. Utiliza fuentes confiables, cita ejemplos reales similares cuando sea posible, y asegúrate de que tus recomendaciones sean prácticas y implementables. El trabajo será evaluado en base a la profundidad del análisis, calidad de las propuestas y madurez profesional demostrada.

Este template te proporciona el espacio para desarrollar tu análisis con profundidad académica y visión crítica.

Las consignas presentadas arriba con este template te ayudarán a desarrollar tu caso en un aprendizaje valioso!

RESPUESTAS EVIDENCIA N°1

A-

Los stakeholders son, Dirección/Gerencia, porque definió y financió el proyecto y tomó la decisión de lanzamiento pese a resultados, impactando costos y quejas; Equipo médico, porque debía garantizar la calidad del contenido clínico y su ausencia de revisión derivó en errores de preparación con riesgo legal y pérdida de confianza; Sistemas (TI) y proveedores tecnológicos, porque su integración con agenda/HIS fue superficial, generando incompatibilidades, desincronización, latencias de 8–12 s y unas 15 caídas, sin tolerancia a fallos, lo que afectó la continuidad del servicio; Operadoras del call center, porque recibieron 60% más derivaciones, se saturaron y aumentó el estrés, contradiciendo el objetivo de alivianar la carga; y Pacientes, porque son el destinatario final: sufrieron una interfaz poco accesible y lenguaje no adaptado, lo que produjo frustración, migración de personas, con impacto económico y en la relación a largo plazo.

Para detectar dónde se traba el sistema conviene mirar el recorrido típico. Todo empieza cuando la persona busca información o un turno. En ese primer contacto, el bot debía reconocer si se trataba de un trámite administrativo, una consulta general o una preparación clínica, y orientar de inmediato. En la práctica, no conservaba contexto y, ante frases cotidianas o preguntas encadenadas, respondía a medias o fuera de tema, empujando a la gente nuevamente a la cola o a repetir el pedido.

En el triage y la seguridad, la debilidad fue mayor: no detectaba posibles urgencias ni derivaba a tiempo con una operadora. Sumando demoras por respuesta, cualquier caso algo ambiguo se volvía un callejón sin salida.

En información clínica, las respuestas sobre preparaciones fallaban seguido: plantillas genéricas, sin revisión médica ni fuentes citadas, que dieron pie a reclamos. En turnos, la integración con la agenda era superficial: el bot mostraba disponibilidad inexistente y el “turno confirmado” se caía al concretarlo, forzando nuevos contactos y recargando el call center. Sin memoria de conversación ni reglas de derivación, no cerraba bien los casos ni dejaba claro el siguiente paso. El canal, que debía simplificar, terminó sumando fricción.

Puntos de inflexión y decisiones

- **Mayo 2024 (Beta):** ya había señales de **baja precisión**.
- **Junio (go-live total):** sobrecarga y quejas inmediatas.
- **Julio–agosto: desactivación parcial** tras acumular incidentes.

B-

Mucho de lo ocurrido era evitable con contenido clínico curado y aprobado por médicos, la tasa de errores en preparaciones no habría existido. Con integración real a la agenda no se habrían prometido turnos inexistentes. Con memoria básica, reglas de que cuando dudo, paso a humano se habría reducido la frustración. Y con metas técnicas previas (respuesta rápida, estabilidad) más pruebas de carga, las caídas y demoras se habrían visto a tiempo. Además, se ignoraron alertas: incompatibilidades en abril y baja precisión en mayo; pese a eso, se lanzó en junio. También se subestimaron riesgos clínicos, operativos y reputaciones. En conclusión, faltó gobernanza clínica, integración de verdad, cortes claros para derivar, ensayos serios antes de abrir y diseño accesible para mayores.

Un plan preventivo resumido definir con precisión qué responde y qué no responde el sistema; todo lo clínico sale de fichas internas con responsable médico y versión vigente; agenda conectada al mismo motor que usa el personal; pruebas por etapas y derivación sin fricción; reglas de desempeño y privacidad por diseño.

C-

Hay límites reales del contexto y de la tecnología. Los modelos de lenguaje siempre dejan un margen de error por la ambigüedad del habla humana; la memoria es limitada y conversaciones largas pueden perder hilo; la latencia y las caídas dependen también de redes y proveedores; con USD 85 000 y 8 meses no es simple lograr, de una, integración profunda, orquestación avanzada y validación clínica completa; los sistemas hospitalarios no siempre ofrecen APIs completas/estables; y con una base con alta proporción de adultos mayores, siempre habrá quien prefiera trato humano. Estos factores se mitigan, pero no se eliminan.

Evitables: contenido sin revisión médica, corpus pobre, go-live pese a alertas, sin derivación, sin objetivos técnicos.

Sistémicos: error residual del lenguaje, memoria limitada, variaciones de red/proveedor, techo de presupuesto/tiempo, límites del HIS, brecha digital.

D-

Primero, evaluaría la viabilidad antes de empezar ¿qué problema concreto queremos resolver y con qué impacto? Definiría con claridad qué hará y qué no hará la IA, qué datos necesita y quién es responsable del contenido clínico. Verificaría que haya consentimiento y privacidad adecuados, y que el sistema de la clínica (HIS) esté listo para integrarse. También miraría costos, tiempos, regulación y riesgos. Contaría con avances esperados (mejores modelos, respuestas más “ancladas” en fuentes, integraciones más simples), pero sin confiar ciegamente: todo debe validarse.

Luego, aplicaría un testing escalonado. Etapa 1: pruebas internas con preguntas reales (incluyendo lenguaje local), sin usuarios. Etapa 2: “modo sombra”, donde la IA responde en paralelo pero la decisión final sigue siendo humana. Etapa 3: lanzamiento chico y por horarios, y recién después escalar. En todo momento, si hay baja confianza o un tema sensible, la IA deriva a una persona y deja un breve resumen para no repetir.

Los criterios de éxito/fallo serían simples y seguros: cero errores clínicos graves, confirmaciones reales de turnos, respuestas estables y claras, y posibilidad visible de hablar con una persona. Si aparecen errores clínicos, turnos prometidos que no existen, caídas reiteradas o quejas sostenida, se vuelve a la etapa anterior hasta corregir.

Por último, gestionaría el cambio y el plan de contingencia. Capacitar al personal, explicar a los pacientes qué puede y qué no puede hacer la herramienta, y abrir un canal para recibir mejoras. Mantendría un ciclo de actualización del contenido con responsables médicos. Si algo falla, modo “degradado” (solo información segura) y el resto lo toma un humano; si el problema es serio, rollback y análisis del incidente antes de reactivar. Así, aprovechamos la IA y los futuros avances, pero con pasos cortos, control y foco en la seguridad de las personas.

E-

La IA es un medio, no un fin. Antes de usarla, hay que entender el problema y evaluar soluciones más simples. Si se usa, que sea donde suma sin comprometer a las personas ni a la organización. Mi enfoque sería empezar

probando por etapas, tener plan B, escalar sólo si funciona en la realidad y decir “todavía no” cuando el contexto no está listo.

Me basaría en algunos principios como por ejemplo, persona primero, seguridad por diseño, privacidad, transparencia y trazabilidad, equidad, robustez y responsabilidad.

Competencias clave: planteo de problemas, calidad de datos, evaluación, gestión de riesgos y normativa, integración con sistemas, diseño de experiencias, gestión del cambio y comunicación clara. Técnicamente, conocer alternativas para elegir lo adecuado, no lo “de moda

F-

Algunas recomendaciones serían que la clínica decide avance con prudencia. Empieza con un alcance acotado: el sistema solo brinda información segura y ofrece turnos que estén realmente confirmados en el sistema de la institución. Todo lo clínico nace de fichas escritas y validadas por médicos, y si en medio de una conversación el sistema duda o detecta algo sensible, deriva de inmediato a una persona, dejando un breve resumen para no hacer repetir al paciente. El diseño está pensado para personas mayores: texto claro y grande, pasos simples y un botón visible de “hablar con alguien”. Detrás de escena, los datos se protegen con cifrado y accesos mínimos, se define con precisión qué hace y qué no hace el sistema, se asegura una integración firme con el HIS, y se mantiene un “modo degradado” para momentos de caída o lentitud, donde el sistema solo entrega información segura y el resto lo toma un humano.

En paralelo, los profesionales crean y mantienen vivas esas fichas: escriben en lenguaje simple, cada una con su responsable clínico, y las actualizan de forma periódica. Cuando hay que pasar la conversación a una persona, se incluye un miniresumen automático con lo que ya se habló. Además, el equipo observa dónde la gente se confunde y ajusta textos, pasos y mensajes para facilitar el recorrido del paciente.

Del lado de los proveedores de IA, la solución incorpora respuestas con fuentes citables, un indicador de confianza para decidir cuándo derivar, y bloqueos explícitos para temas clínicos que no estén habilitados. Los conectores con el HIS son robustos, hay trazabilidad de fallas y un “botón de apagado” rápido para cortar el servicio ante incidentes. Y, como marco de seguridad, los reguladores exigen validación médica previa de los contenidos, un botón humano obligatorio, privacidad por diseño en todo el ciclo de datos y una notificación ágil de incidentes con revisión antes de reactivar el sistema.