Análisis de Caso GenAI Fallido

Institución: ISPC
Carrera: Ciencia de Datos e IA
Cohorte: 2024
Asignatura: Practica Profesionalizante I
Docente: Accietto Daniela
Estudiante: Lorenzati Valentino
Repositorio: Link

A) Investigación Profunda del Caso "MediBot"

El caso "MediBot" de la Clínica MediCare representa un ejemplo de una implementación fallida de tecnología GenAI en un sector tan crítico como la salud. La inversión de $85,000 USD y 8 meses de desarrollo concluyo en una tasa de resolución exitosa del 23%, un aumento del 47% en las quejas de pacientes y un impacto negativo significativo en la experiencia del paciente y la reputación de la clínica.

Causas del Fracaso:

1. **Factores de gestión del proyecto:**

◦ **Planificación Deficiente**: El proyecto inició con una definición de requerimientos básicos; pero no en la complejidad del dominio de salud. Esto se evidencia en la base de conocimiento de 500 FAQ básicas la cual era insuficiente, NO incluía terminología médica específica ni casos complejos. La Guía PMBOK enfatiza la importancia de establecer el alcance y detallar los objetivos durante la fase de planificación. Aquí, el alcance fue muy superficial, ignorando las particularidades críticas del sector.

◦ **Gestión inadecuada de los stakeholders:**
- **Pacientes**: No se consideraron las necesidades de los adultos mayores, para quienes la interfaz resultó confusa (65% de la base de pacientes).
- **Personal Médico**: Las respuestas del bot no fueron validadas por personal médico, lo que llevó a información incorrecta y riesgo legal.
- **Operadores de Call Center**: El bot generó más problemas operacionales y sobrecargó al personal humano, que recibió un 60% más de llamadas. Esto indica que no se logró uno de los objetivos clave (reducir el estrés del personal y los costos operativos)

◦ **Ausencia de una metodología**: No se evaluaron correctamente los riesgos inherentes a la IA en un contexto médico (información incorrecta, seguridad, privacidad). El PMBOK destaca la gestión de las restricciones de calidad y riesgo como responsabilidades clave del director de proyecto.

2. **Factores Técnicos:**

◦ **Entrenamiento del modelo**: El uso de GPT-3.5 Turbo con "fine-tuning básico" y una base de conocimiento limitada no fue una buena opción para la complejidad de las consultas médicas. El bot no comprendía jerga local ni variaciones de expresiones, ni casos complejos.

◦ **Falta de integración profunda y contexto**: La ausencia de acceso a historiales médicos o disponibilidad real de turnos limito mucho la capacidad del bot para ofrecer respuestas aceptables, un fallo en la unicidad y el valor del

servicio prometido. La falta de memoria de contexto hizo que cada consulta fuera independiente, perdiendo el hilo de conversaciones complejas.

◦ **Problemas de rendimiento**: El bot mostró alta latencia (8-12 segundos) y caídas del sistema (15 interrupciones en el primer mes). Esto empeoró el tiempo de espera, un objetivo crítico que se quería mejorar.

◦ **Deficiencias en la detección de urgencias**: El bot no reconocía urgencias médicas ni derivaba correctamente a un humano.

◦ **Vulnerabilidades de seguridad y compliance**: El bot almacenaba datos sensibles sin encriptación adecuada y daba información médica incorrecta o desactualizada, lo que constituye una violación de la privacidad y un riesgo legal.

3. **Factores Humanos y de experiencia de Usuario:**

◦ **Respuestas inapropiadas y robóticas**: La falta de empatía del bot en situaciones delicadas y su incapacidad para entender el lenguaje natural, generaron enojo y frustración.

◦ **Información incorrecta y riesgosa**: El 18% de las respuestas sobre preparación de estudios no eran correctas, lo que generó riesgo para los pacientes y llevó a 3 reclamos legales.

◦ **Pérdida de confianza**: La combinación de todos estos problemas llevó a la pérdida de 340 pacientes y una caída en la reputación online

Decisiones clave y puntos de inflexión:

• **Definición de requisitos básicos en enero 2024**: Fue el primer error, al subestimar la complejidad del dominio médico.

• **Ignorar las señales de baja precisión en mayo 2024 (Beta Testing)**: Un punto crítico, el cual pudo haber sido, revisado o detener el proyecto. Kerzner destaca la **i**dentificación temprana de problemas para acciones correctivas.

• **Lanzamiento oficial en junio 2024**: La decisión de lanzar el Bot (go-live) a pesar de las alertas previas, resultó en una sobrecarga inmediata y muchas quejas.

• **Decisión de reducir funcionalidades en julio-agosto 2024**: Una medida desesperada ante una crisis, se mostró el fracaso del enfoque inicial.

En resumen, el fracaso del MediBot se atribuye a una mala planificación, una gestión de riesgos nula, una validación técnica mínima. Subestimar la complejidad del dominio de salud.

No se tomó en cuenta NINGUNO de Los principios de la Guía PMBK sobre la gestión de procesos y las restricciones de un proyecto, junto con los factores críticos de éxito de Kerzner

---

B) ¿Qué Pudo Prevenirse?

Problemas prevenibles:

1. **Planificación y requisitos:**

◦ **Definición de alcance incompleta**: La poca base de conocimiento, la falta de terminología médica específica y el no saber manejar los casos complejos pudieron prevenirse con una fase de planificación mejor, involucrando a médicos y especialistas en salud desde el principio.

◦ **Subestimación de recursos y tiempo**: La inversión de $85,000 USD y 8 meses no fueron suficiente para un proyecto tan complejo. Una mejor inversión y asignación de recursos para el desarrollo, entrenamiento, y validación profesional debería haber sido prioritario.

2. **Testing:**

◦ **Ignorar señales de alerta**: Las "primeras señales de alerta por baja precisión" en las pruebas beta fueron una bandera roja clara a la cual no le prestaron atención. Se debió haber estudiado esta misma

◦ **Ausencia de pruebas exhaustivas**:
- **Precisión de contenido médico**: El 18% de respuestas erróneas era totalmente prevenible con un proceso más seguro de **v**alidación profesional.
- **Usabilidad y accesibilidad**: La interfaz es confusa para gente mayor, esto pudo haberse prevenido con pruebas concretas para este rango de edad.
- **Rendimiento y estabilidad**: La alta latencia y caídas del sistema pudieron detectarse y arreglarse con pruebas de estrés y rendimiento antes del lanzamiento.
- **Derivación**: La incapacidad de reconocer urgencias debía ser una funcionalidad crítica sin margen de error

3. **Metodología y enfoque:**

◦ **Falta de un enfoque**: El método Lean Startup propone la experimentación sobre la planificación elaborada y retroalimentación del cliente sobre intuición. Un desarrollo iterativo centrado en seguridad, con ciclos rápidos, habría permitido corregir el rumbo mucho antes.

◦ **Gestión de riesgos inexistente o ineficaz**: Los riesgos legales, de reputación, operativos y de seguridad se debieron identificarse, y planear estrategias de corrección desde el inicio.

Plan de prevención de riesgos específico para proyectos en salud:

1. **Fase de inicio:**
   - **Comité especializado:** Establecer un comité de especialistas (médicos, especialistas en TI/IA, legal) para controlar la viabilidad y los requisitos.
   - **Viabilidad ética y legal**: Realizar una auditoría inicial de cumplimiento (ej. HIPAA, GDPR) y un análisis de riesgos éticos específicos de la IA.
   - **Impacto de bajo riesgo**: Arrancar con GenAI en áreas donde un error tiene un impacto mínimo (ej., FAQs generales no médicas, trámites administrativos internos) y escalar con el tiempo.
   - **Requisitos detallados**: Elaborar un documento de requisitos detallado que incluya terminología médica específica, jerga local, casos límite, expectativas de privacidad y protocolos de seguridad.
   - **Presupuesto y cronograma realista**: Asignar tiempo y recursos significativos para entrenamiento del modelo, integración segura, testing (incluyendo validación humana continua) y cumplimiento normativo.

2. **Fase de ejecución:**
   - **Desarrollo de MVPs**: Adoptar un enfoque ágil/Lean. El "Producto Mínimo Viable" debe ser Mínimo en funcionalidades, pero Máximo en seguridad.
   - **Integración de "Human-in-the-Loop" (HITL)**: Diseñar el sistema para que las respuestas del GenAI en áreas críticas siempre sean revisadas o validadas por personal humano calificado antes de llegar al paciente.

3. **Fase de pruebas:**
   - **Pruebas de precisión especializada**: Desarrollar un banco de pruebas específico con preguntas y escenarios médicos. Cada respuesta del bot debe ser validada por especialistas.
   - **Auditorías de seguridad y privacidad**: Pruebas de penetración, análisis de vulnerabilidades y verificación del cumplimiento de normativas de datos sensibles.
   - **Pruebas de rendimiento y escalabilidad**: Sobrecargas simuladas y reales para garantizar la estabilidad del sistema y baja latencia.

4. **Preparación para el lanzamiento:**
   - **Programa de capacitación integral**: Para el personal que interactúe con el bot y los equipos de soporte.
   - **Comunicación transparente con pacientes**: Informar claramente sobre las capacidades y limitaciones del bot.

C) ¿Qué NO Pudo Prevenirse?

Limitaciones de la tecnología GenAI (GPT-3.5 turbo en 2024):

◦ **Tendencia a "alucinar" y precisión**: Los modelos de lenguaje como GPT-3.5 Turbo, incluso con fine-tuning, tienden a generar información plausible pero incorrecta ("alucinaciones"). El lograr una precisión del 100% en información médica sin validación humana constante es una limitación que no se puede "eliminar" por completo solo con planificación.

◦ **Comprensión de la jerga**: La dificultad para comprender "jerga local cordobesa" puede ser una limitación del NLP del modelo base.

◦ **Memoria de contexto limitada**: Aunque se pueden implementar estrategias para mejorar la memoria de contexto, para GPT-3.5 Turbo, mantener un "hilo de conversaciones complejas" en un dominio técnico como el médico es bastante difícil.

D) Cómo Trabajarías en el Futuro (Metodología Mejorada para GenAI en Salud)

Para futuras implementaciones de GenAI en el sector salud, se adoptaría una metodología mejorada que combine los principios de la gestión de proyectos (PMBOK, Kerzner) con el enfoque iterativo y de aprendizaje de Lean Startup, priorizando la seguridad, la ética y la experiencia del paciente.

1. Framework de evaluación de viabilidad (Pre-proyecto - "Go/No-Go Decision"):

• **Alineación estratégica (Kerzner, PMBOK)**:

◦ **Problema real?**:¿Qué "pain points" de la clínica se buscan resolver? ¿Tiene el bot, la capacidad de resolverlos de manera segura y eficiente?

◦ **Análisis costo-beneficio**: Además de la inversión inicial, ¿cuáles son los costos operativos a largo plazo (mantenimiento, supervisión humana, licenciamiento)?

◦ **Impacto Ético y Legal**: Evaluación temprana del cumplimiento normativo (HIPAA, etc.) y los riesgos éticos (sesgos, privacidad, responsabilidad)

2. Proceso de testing:

• Fase 1: MVP de aprendizaje:

◦ **Alcance ultra-limitado**: Desarrollar un chatbot para un conjunto de prueba y de bajo riesgo de consultas internas

◦ **Validación humana:** Cada respuesta generada por la IA debe ser validada por especialistas antes de ser considerada "aprobada".

◦ **Métricas**: Tasa de precisión interna, tiempo de respuesta del modelo, etc

• Fase 2: Expansión:

◦ **Aumento progresivo**: Solo después de superar las fases anteriores con métricas claras.

◦ **Implementación de "Human-in-the-Loop" (HITL**: Donde los casos complejos, las urgencias o las consultas sensibles sean derivadas a un operador humano. El bot debe ser un asistente, no un sustituto

◦ **Pruebas de resistencia**: Asegurar que el sistema pueda manejar picos de demanda sin degradar el rendimiento o la seguridad.

3. Criterios de Éxito/Fallo (SMART):

• **Éxito**:
◦ 0% de información médica incorrecta/engañosa.
◦ Tasa de resolución exitosa
◦ Reseñas positivas de pacientes
◦ Reducción del tiempo de espera
◦ Cumplimiento total de regulaciones de privacidad
◦ Reducción de la sobrecarga del personal humano

• **Fallo**:
◦ Cualquier incidente de información médica incorrecta con riesgo para el paciente.
◦ Violación de datos de privacidad.
◦ Reseñas negativas de pacientes
◦ Aumento de quejas o escalamientos humanos relacionados con la IA.
◦ Inestabilidad del sistema o fallos de rendimiento prolongados.

E) Lecciones Aprendidas Personales

1. **Enfoque absoluto en seguridad y eficiencia**

En el sector salud, hay un riesgo crítico para la vida y la salud de las personas, con implicaciones legales y éticas. La búsqueda de automatización y reducción de costos no debe interferir jamás con la seguridad del paciente ni la precisión de la información médica.

2. **La importancia de los expertos**

Siempre integraría una validación de un experto. Esto implica diseñar sistemas donde los especialistas de cualquiera que sea el dominio estén sumamente involucrados en la revisión de respuestas y la toma de decisiones. El rol de la IA sería aumentar las capacidades humanas, no reemplazarlas.

3. **Gestión detallada:**

Siempre priorizar una gestión de riesgos exhaustiva y continua, identificando activamente todos los riesgos más posibles (técnicos, legales, éticos, etc.) y desarrollando planes de mitigación y corrección. La gestión de stakeholders sería una prioridad, asegurando que todas los involucrados sean escuchados y sus necesidades tomadas en cuenta en el diseño.

F) Recomendaciones Estratégicas

• Invertir en una oficina de gestión de proyectos (PMO) o, mínimamente, en un director de proyecto con experiencia en gestión de riesgos y proyectos de tecnología enfocadas en salud, para liderar futuras iniciativas.

• Participar activamente y liderar la definición de requisitos y validación de contenido en cualquier proyecto de IA. No darle toda la responsabilidad exclusivamente a los equipos de TI.

• Fomentar la alfabetización digital y en IA para comprender las capacidades y limitaciones de estas tecnologías, así como sus implicaciones éticas.

• Desarrollar y comercializar "modelos especializados y pre-entrenados" para el sector salud que incorporen terminología médica.
• Proporcionar herramientas robustas y personalizables que permitan a los clientes integrar fácilmente la supervisión humana y la validación en el flujo de trabajo de la IA.

• Ser transparentes y proactivos sobre las limitaciones inherentes de los modelos GenAI para evitar expectativas poco realistas en sectores críticos. Incluir guías claras de uso responsable.