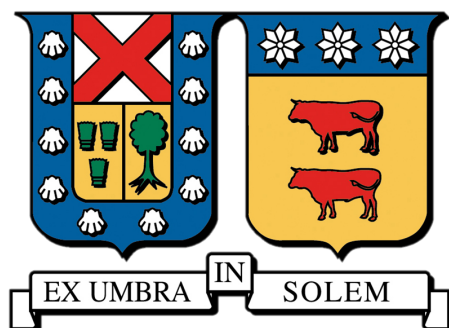




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

AUTOMATIZACIÓN DE LA METODOLOGÍA FMECA A TRAVÉS DE
LLM CON ANÁLISIS DE DISCUSIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

GABRIEL MAXIMILIANO JARA RIVERA

PROFESOR GUÍA

PABLO VIVEROS

PROFESOR CO-REFERENTE

CHRISTOPHER NIKULIN

SIMÓN GOMEZ

VALPARAÍSO CHILE, JULIO 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Automatización de la Metodología FMECA a través de LLM con Análisis de Discusiones

Nombre del candidato(a): Gabriel Maximiliano Jara Rivera

Carrera / Grado: Ingeniería Civil Industrial

Campus: Casa Central Valparaíso **Departamento:** Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Pablo Viveros Gunckel, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

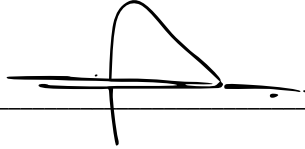
☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

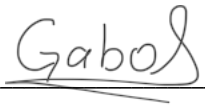
Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 22/07/2026 **Firma:** 

Estudiante o Candidato(a):

Fecha: 21/07/2026 **Firma:** 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Tabla de Contenidos

1. Problema de Investigación	6
2. Objetivos	10
2.1. Objetivo General	10
2.2. Objetivos Específicos	10
3. Marco Teórico	11
3.1. Gestión de activos en entornos industriales	11
3.2. Fundamentos del Análisis de Modos de Falla, Efectos y Criticidad (FMEA y FME- CA)	12
3.2.1. Elementos clave del análisis FMECA	13
3.3. Modelos de Lenguaje de Gran Escala (LLMs)	16
3.3.1. Aplicaciones industriales de los LLMs	17
3.3.2. Ventajas de los LLMs en entornos industriales	18
3.3.3. Desafíos Actuales	18
3.4. Reconocimiento Automático de Voz (ASR)	19
3.4.1. Método de evaluación del rendimiento y calidad	20
3.4.2. Aplicación del ASR en la industria	20
3.4.3. Desafíos y consideraciones en su uso	21
3.5. Colaboración humano-inteligencia artificial (IA) en ingeniería	22
3.5.1. Human-in-the-loop	23
3.5.2. Beneficios y Riesgos	23

3.6. Síntesis crítica del estado del arte y brecha de investigación	24
4. Metodología	26
4.1. Recolección de información (conversación técnica y uso de ASR)	27
4.1.1. Contextual / descriptiva	27
4.1.2. Funcional / analítica	28
4.1.3. Evaluativo / decisional	28
4.2. Estructuración del FMECA (utilización de modelos de lenguaje)	30
4.3. Estrategia de validación experimental	32
4.3.1. Diseño del experimento y caso de estudio	33
4.3.2. Construcción de las referencias de validación	33
4.3.3. Procedimiento experimental	34
4.3.4. Modelos utilizados	35
4.3.5. Métricas de evaluación	35
5. Análisis cuantitativo y evaluación de beneficios	36
5.1. Definición de tiempos de trabajo a partir de un caso referencial	38
5.1.1. Supuestos generales	38
5.1.2. Tiempos en el método tradicional	39
5.1.3. Tiempos en el método propuesto con asistencia de inteligencia artificial . .	41
5.1.4. Comparación de tiempos	42
5.1.5. Resumen de beneficios de tiempo	43
5.2. Asignación de costos por rol y ahorro económico	44
5.2.1. Definición de salarios promedio	44

5.2.2.	Participación del ingeniero junior	45
5.2.3.	Ahorro económico por FMECA	46
5.3.	Escalamiento a proyectos con múltiples equipos	47
5.3.1.	Ahorro total en proyectos de 50 a 100 equipos	48
5.3.2.	Escenarios conservador y optimista	49
5.4.	Punto de equilibrio y período de recuperación de la inversión	49
5.4.1.	Inversión inicial	49
5.4.2.	Punto de equilibrio (Breakeven)	50
5.4.3.	Período de recuperación de la inversión (Payback)	51
5.5.	Evaluación de la calidad estructural de la matriz generada	52
5.5.1.	Tasa de Error de Palabras	52
5.5.2.	Cobertura del contenido técnico	53
5.5.3.	Precisión del contenido generado	54
5.6.	Discusión de resultados	55
5.6.1.	Desempeño del reconocimiento automático de voz	56
5.6.2.	Calidad de la estructuración del FMECA	57
5.6.3.	Implicancias para la colaboración humano-IA	59
5.6.4.	Limitaciones y validez externa	61
5.7.	Conclusiones del análisis	63

Resumen

En esta investigación se evaluó el desempeño de un sistema que integra reconocimiento automático de voz y modelos de lenguaje de gran escala para convertir conversaciones técnicas en FMECA. El modelo se puso a prueba mediante casos controlados aplicados al sistema mecánico de un activo industrial. Para revisar la calidad de la información entregada por los modelos propuestos, los resultados obtenidos se compararon con información de referencia.

Los resultados fueron favorables y mostraron una tasa de error en las palabras procesadas por el ASR inferior al 10 %, junto con una recuperación del contenido estructurado del 93 %. Aun así, se detectaron diferencias que requieren atención, principalmente aquellas relacionadas con la organización jerárquica de la información y con la vinculación entre el equipo, el elemento mantenible, su modo de falla y la correspondiente clasificación de criticidad. Estas variaciones muestran que la herramienta puede contribuir de manera importante al análisis, pero todavía requiere la revisión de un especialista antes de considerar los resultados como definitivos.

Otro aspecto relevante de la investigación fue la reducción significativa del esfuerzo necesario para elaborar un FMECA. De acuerdo con lo observado, la metodología propuesta permite disminuir cerca del 40 % de la carga total de trabajo, especialmente en las tareas de registro y documentación de la información, las cuales suelen consumir una parte considerable del tiempo.

En términos generales, los resultados muestran que la combinación de ASR y LLM puede ser una herramienta de apoyo útil en los procesos de gestión de activos. Su principal aporte se encuentra en la agilización del trabajo, la mejora de la trazabilidad de la información y una mayor uniformidad en los procesos de documentación. Sin embargo, su función no es sustituir la experiencia de los especialistas, sino complementar su trabajo y facilitar el análisis técnico.

1. Problema de Investigación

La gestión de activos en entornos industriales se enfrenta a crecientes desafíos relacionados con la confiabilidad operacional, la eficiencia en el uso de recursos y la toma de decisiones informadas en el ámbito del mantenimiento de maquinaria y activos relevantes para el funcionamiento correcto y eficiente de los procesos involucrados. En un escenario de desarrollo industrial, donde los sistemas son cada vez más complejos y la necesidad de minimizar tiempos de inactividad es constante, resulta fundamental contar con metodologías robustas que permitan anticipar fallas, evaluar riesgos y priorizar acciones correctivas de manera estructurada y eficiente, teniendo la capacidad de crear beneficios a distintos ingenieros y profesionales de las áreas involucradas en los procesos (Márquez 2007).

En este contexto, metodologías ampliamente utilizadas como el Análisis de Modos de Falla (FMEA) y su versión más completa, el Análisis de Modos de Falla, Efectos y Criticidad (FMECA), han demostrado ser herramientas fundamentales para analizar modos de falla potenciales, estimar sus efectos sobre el sistema y definir estrategias correctivas que mitiguen consecuencias, y por ende se mejore la confiabilidad del proceso en general.

El FMECA, al incorporar la dimensión de criticidad de cada uno de los objetos a analizar, permite priorizar las fallas con base en criterios como la severidad, la probabilidad de ocurrencia y la capacidad de detección, contribuyendo así a una toma de decisiones más informada en la gestión del ciclo de vida de los activos. No obstante, la aplicación tradicional de estas metodologías depende en gran medida del juicio de expertos, discusiones colaborativas prolongadas y documentación manual, lo cual puede ralentizar los procesos, introducir sesgos y dificultar la estandarización de resultados (Price and Taylor 2002). Además, en entornos industriales con sistemas altamente com-

plejos, como instalaciones mineras o líneas de producción avanzada, donde identificar las partes involucradas en el proceso y límites de esta resulta laborioso, teniendo en cuenta los análisis de costos de ciclos de vida (LCC, por sus siglas en inglés) respectivos, la evaluación exhaustiva de todas las combinaciones de fallas posibles se torna compleja, incluso si se automatizara parcialmente la generación de reportes (Kumar et al. 2023).

En paralelo, el desarrollo de tecnologías de inteligencia artificial ha generado nuevas oportunidades para abordar estos desafíos. En particular, los modelos de lenguaje de gran escala (LLM, por sus siglas en inglés) han demostrado una notable capacidad para interpretar lenguaje natural, sintetizar información y generar contenido coherente a partir de insumos complejos como conversaciones humanas, teniendo en cuenta, a diferencia de los asistentes tradicionales de IA, el contexto de la situación planteada (Mahmood et al. 2025). Cuando estos modelos se combinan con sistemas de reconocimiento automático de voz (ASR), es posible transcribir con alta precisión discusiones técnicas y, a partir de ello, automatizar procesos de análisis estructurado (Piñones et al. 2023). Esto permitiría una reducción significativa de la carga analítica de los equipos técnicos, una mejora en la trazabilidad de los análisis realizados y una mayor estandarización de los resultados al momento de evaluar el funcionamiento de un sistema en la industria.

Además, esta integración presenta el potencial de reducir la carga de documentación de los equipos técnicos, mejorar la trazabilidad de la información y favorecer una mayor estandarización de los resultados. Sin embargo, estos beneficios dependen de que el sistema sea capaz de recuperar correctamente el conocimiento expresado por los especialistas, relacionar de manera coherente los elementos del análisis y evitar la incorporación de información no respaldada por la discusión técnica.

Sin embargo, a pesar del avance de estas tecnologías, como se menciona en el párrafo

anterior, la generación automática de una matriz FMECA no garantiza por sí misma que el conocimiento técnico contenido en una discusión haya sido capturado de manera completa y correcta. La utilidad de un sistema basado en ASR y LLM depende de su capacidad para identificar los elementos relevantes del análisis, conservar las relaciones entre equipos, componentes, funciones, modos de falla, causas y efectos, y mantener la trazabilidad entre la información generada y la conversación original.

En este marco, la problemática no se limita a determinar si es técnicamente posible automatizar la creación de matrices de análisis FMECA, sino que también requiere evaluar la calidad del conocimiento recuperado y estructurado por el sistema. Debido al carácter técnico y crítico de las decisiones asociadas al mantenimiento industrial, resulta necesario determinar el nivel de supervisión humana requerido para revisar y validar los resultados generados.

En base a lo anterior, surgen las siguientes interrogantes de investigación:

¿En qué medida un sistema integrado de reconocimiento automático de voz y modelos de lenguaje de gran escala puede recuperar y estructurar correctamente el conocimiento técnico contenido en discusiones previas a la generación de un FMECA, y qué nivel de revisión humana requiere para producir una matriz técnicamente aceptable?

¿Qué nivel de precisión puede alcanzar una transcripción automática de discusiones técnicas mediante ASR en contextos industriales?

¿En qué medida el procedimiento asistido por inteligencia artificial permite reducir el tiempo y costo de documentación respecto del método tradicional, considerando el esfuerzo requerido para validar el resultado?

En consecuencia, esta investigación busca evaluar la capacidad de un sistema integrado de reconocimiento automático de voz y modelos de lenguaje para transformar discusiones técnicas en

análisis FMECA estructurados. Para ello, se desarrolla un prototipo que actúa como instrumento de evaluación, permitiendo analizar la calidad del contenido generado, la intervención humana necesaria y los beneficios operacionales asociados a su utilización.

2. Objetivos

2.1. Objetivo General

Evaluar la capacidad de un sistema integrado de reconocimiento automático de voz y modelos de lenguaje de gran escala para recuperar y estructurar el conocimiento técnico contenido en discusiones especializadas desarrollando e implementando un prototipo basado en inteligencia artificial guiada por voz que permita automatizar la generación de reportes de gestión de activos en la industria junto al análisis FMECA (Failure Modes, Effects and Criticality Analysis) a partir del procesamiento de archivos de audio, generando un informe técnico estructurado en formato excel para apoyar la gestión de activos industriales.

2.2. Objetivos Específicos

Definir un método de transcripción de audios correspondientes a discusiones técnicas utilizando herramientas de reconocimiento automático.

Diseñar un módulo de extracción automática de información desde el texto transcrito utilizando un modelo de inteligencia artificial LLM (Large Language Model), para identificar los modos de falla, sus causas, efectos, controles existentes y niveles de criticidad.

Implementar la generación automática de la matriz de análisis FMECA (Failure Modes, Effects and Criticality Analysis), con parámetros clave extraídos y analizados por el LLM para así facilitar la priorización de fallas y apoyar decisiones técnicas de mantenimiento.

Evaluar precisión, utilidad y eficiencia del modelo y sus herramientas comparando con la aplicación manual del procedimiento FMECA (Failure Modes, Effects and Criticality Analysis)

3. Marco Teórico

3.1. Gestión de activos en entornos industriales

La gestión de activos físicos, para cualquier empresa industrial moderna, ha adquirido un rol estratégico al permitir a estas optimizar el ciclo de vida de los activos, mejorar la eficiencia operacional y reducir riesgos asociados a fallas inesperadas en los procesos. Según la norma internacional ISO 55000, la gestión de activos se define como el “conjunto de actividades coordinadas de una organización para generar valor a partir de sus activos” (International Organization for Standardization 2014). Esta área implica tomar decisiones informadas respecto al diseño, operación, mantenimiento y renovación de los activos, considerando criterios de diferentes aspectos contextuales, como son los económicos, técnicos y de riesgo.

Más específicamente, en entornos industriales como plantas mineras, termoeléctricas o plantas de manufactura avanzada, los activos físicos suelen representar un alto porcentaje del capital invertido, llegando, en el caso de la minería, a valores tan elevados como que el 88 % de la inversión en activos se compone de activos no corrientes como propiedad, plantas y equipos (Corporación Nacional del Cobre de Chile (Codelco) 2024). Las fallas en estas áreas o un mal funcionamiento pueden derivar en interrupciones de la producción, sobrecostos en mantenimientos, impactos ambientales o incluso riesgos de seguridad, pudiendo, como se indica en estudios realizados por el World Economic Forum (2020), alcanzar pérdidas de más de 50 mil millones de dólares anuales a escala global. En el caso de Chile, se ha destacado que una parte importante de las ineficiencias operacionales y pérdidas en la industria extractiva se deben a la naturaleza de las estrategias impuestas, enfocándose en el mantenimiento reactivo en vez de una planificación preventiva de las fallas (Consejo de Competencias Mineras 2022).

Por esta razón, cada vez más organizaciones han dado un vuelco hacia la toma de decisiones estratégicas más informadas sobre cómo y cuándo intervenir un activo, cuál es el impacto que tiene la falla de este en el proceso y cómo priorizar inversiones de mantenimiento o reinversión. Para este enfoque, la toma de decisiones basada en datos —donde se exige información confiable, trazable y procesable respecto al comportamiento de los activos— presenta la problemática de sistematizar la información que se genera durante los análisis y diagnósticos, además de depender en gran medida del conocimiento técnico de los operadores disponibles o bien de los registros manuales, que pueden dificultar interpretaciones.

En este escenario, automatizar los procesos de análisis estructurado, como el Análisis de Modos de Falla, Efecto y Criticidad (FMECA), aparece como una vía eficaz para estandarizar el conocimiento técnico, reducir la dependencia de expertos y mejorar la eficiencia en el proceso de toma de decisiones al momento de gestionar los activos.

3.2. Fundamentos del Análisis de Modos de Falla, Efectos y Criticidad (FMEA y FMECA)

El análisis de Modos de Falla y sus Efectos (FMEA) es una metodología estructurada que permite identificar, analizar y priorizar fallas potenciales en un sistema o proceso, con el objetivo de prevenir sus ocurrencias o bien mitigar sus efectos antes de que se generen consecuencias mayores. Desde su origen, en 1960, como parte de procedimientos de confiabilidad para misiones aeroespaciales de la NASA, FMEA se ha convertido en una herramienta fundamental en diversas industrias (Stamatis 2003).

Este análisis se estructura como una matriz donde se analizan distintos aspectos de un

sistema para poder organizar el conocimiento técnico del equipo necesario, generar consensos y establecer la mejor línea de acción y decisiones estratégicas para la gestión de los activos del proceso, en base a una evaluación sistemática de las partes.

Sin embargo, con el paso del tiempo y la creciente necesidad por mejorar la eficacia operacional de los procesos, se desarrolla una versión más completa denominada FMECA (Failure Modes, Effects and Criticality Analysis), donde se incorpora un análisis más cuantitativo sobre la criticidad de cada modo de falla, con la finalidad de poder mejorar el orden de prioridad de los riesgos que afectan al sistema.

3.2.1. Elementos clave del análisis FMECA

Los elementos clave del análisis FMECA son los siguientes (Stamatis 2003; Rekabi 2024):

- **Desglose del sistema:** El sistema se debe dividir en subsistemas, en base a las relaciones entre las diferentes variables, dependiendo el número de niveles en el objetivo final y la especificación necesaria para el análisis.
- **Descripción de componentes:** Describe qué se espera que haga un elemento del sistema y el estado o condición específica que este debe tener. Toda evaluación parte de una función bien definida; si esta no está clara, no es posible detectar fallas.
- **Modo de Falla (Failure Mode):** Es la forma específica en la que un componente del sistema o proceso puede fallar al momento del cumplimiento de su función.
- **Causa de la falla (Failure Cause):** Es el origen técnico o físico del modo de falla. Puede deberse a errores de diseño, defectos de fabricación, condiciones ambientales adversas,

mantenimiento deficiente, entre otros. Comprender las causas permite establecer controles preventivos más eficaces.

- **Controles:** Incluye todas las medidas técnicas, operativas o administrativas actualmente implementadas para prevenir o detectar la falla antes de que genere un efecto adverso.
- **Nivel de criticidad:** Se agrega el análisis cuantitativo de riesgo por cada modo de falla:
 - **Severidad:** Se indica la gravedad del efecto que traería el modo de falla correspondiente a si sistema.
 - **Ocurrencia:** Se evalúa la probabilidad o frecuencia con la que se espera que ocurra la falla basándose en datos históricos o tasas de fallas.
 - **Detección:** Se mide la capacidad que tienen los controles determinados actuales para detectar las fallas antes de que estas ocurran y que afecten al funcionamiento del proceso.

El producto entre la severidad, ocurrencia y detección determina el número de prioridad de riesgo (RPN). Este número permite ordenar los modos de falla según su riesgo relativo y priorizar acciones de mejora o bien de mitigación (Manar Rekabi, 2024). La Figura 1 representa la matriz que permite el cálculo de este número, comparando severidad con probabilidad de ocurrencia (Rekabi 2024).

		Ocurrencia			
		Poco probable	Podría ocurrir	Ocurrencia ocasional	Probable o frecuente
Severidad	Lesión menor (Crítica)	1	2	3	4
	Lesión menor o efectos en la salud (Menor)	2	4	6	8
	Lesión múltiples o efectos en la salud (Mayor)	4	6	9	12
	Fatalidades (Crítica)	6	9	12	16

Figura 1: Matriz RPN. Tomado de Rekabi, 2024

En el FMECA, además del RPN, se suele realizar un análisis de criticidad más completo, que puede incluir fórmulas basadas en confiabilidad estadística, tiempo medio entre fallas (MTBF), consecuencias operacionales o impacto económico, permitiendo un análisis más técnico y ajustado a entornos de alta exigencia.

Este análisis mejorado, FMECA, se ha utilizado desde entonces para prevenir fallos en las etapas tempranas del diseño de un sistema, así como también para evaluar los riesgos operacionales de procesos existentes. En entornos industriales, permite identificar puntos críticos que requieren refuerzo, diseñar estrategias de mantenimiento predictivo y preventivo, evaluar la confiabilidad de equipos o bien justificar inversiones en tecnología. Tiene una gran eficacia al ahorrar en pérdidas, como por ejemplo en la industria minera chilena, donde la identificación temprana de las fallas puede resultar en un ahorro en pérdidas por sobre los \$100.000 USD por hora de detención del proceso Servicio Nacional de Geología y Minería (2021).

A pesar de su valor, este enfoque presenta limitaciones importantes, al ser altamente intensivo en tiempo, requerir de expertos disponibles de diferentes áreas, contar con poca estandariza-

ción y con un grado importante de subjetividad. Incluso se ha sugerido que, aunque se automatice la generación de reportes, el volumen de combinaciones de posibles fallas en un sistema puede superar la capacidad analítica de los ingenieros, por lo que se requiere una forma de sintetizar los resultados automáticamente (Price and Taylor 2002).

Frente a esta problemática, y en un contexto industrial donde los tiempos son limitados, la posibilidad de aplicar tecnologías emergentes para automatizar procesos como el FMECA representa una ventaja competitiva. La integración de inteligencia artificial en la gestión de activos no solo tiene el potencial de mejorar la eficiencia operativa de los procesos, sino que también contribuye a la estandarización de criterios esenciales para la toma de decisiones.

3.3. Modelos de Lenguaje de Gran Escala (LLMs)

Los modelos de Lenguaje de Gran Escala (Large Language Models, LLMs) representan uno de los avances más significativos en la inteligencia artificial en la última década. Estos modelos son entrenados con grandes volúmenes de datos textuales para comprender, generar y manipular lenguaje natural con alto nivel de contexto y coherencia. La versatilidad de estos modelos es la que les permite adaptarse a una amplia gama de tareas sin necesidad de mayores ajustes, siendo esto especialmente atractivo en entornos industriales.

A diferencia de los asistentes virtuales tradicionales potenciados con inteligencia artificial, los LLMs pueden mantener el hilo de conversaciones con una extensión considerable, interpretar el lenguaje técnico y realizar tareas cognitivas complejas como la síntesis de información y la redacción de reportes técnicos. Estas capacidades han abierto nuevas posibilidades para la automatización de procesos que anteriormente requerían la atención de un experto calificado, como lo es

el análisis FMECA (Brown et al. 2020; Li et al. 2024).

3.3.1. Aplicaciones industriales de los LLMs

En el contexto de la industria, el impacto de los LLMs ha presentado un abanico de posibilidades para la generación automática de diversos dispositivos, tanto como en el ámbito de control de calidad como también de planificación de procesos o bien instrucciones de operación automatizadas. Li et al. (2024) plantean un estudio integrando LLMs con sensores industriales y sistemas de automatización, donde se observa un cierre de la brecha entre los datos generados por las máquinas y la toma de decisiones humanas, resultando en mejoras medibles de eficacia y trazabilidad de los problemas.

Por otra parte, los LLMs son utilizados actualmente para optimizar procesos industriales y mejorar el rendimiento predictivo. Estos modelos son capaces de anticipar fallas al analizar reportes históricos, estructurar datos provenientes de sensores que recopilan indicadores y generar recomendaciones en lenguaje natural que pueden ser entendidas sin necesidad de un análisis técnico de mayor magnitud. Es decir, se alivia la necesidad de contar con expertos disponibles al momento de la falla o bien al momento de predecirla (Myöhänen 2023). Además, se destaca la capacidad de los LLMs para reducir el tiempo necesario para generar y preparar los informes técnicos estandarizados requeridos para el análisis de datos.

Asimismo, se ha podido demostrar que un modelo LLM puede extraer criterios y ponderaciones desde grabaciones de audio, facilitando un análisis AHP (Analytic Hierarchy Process) automatizado Piñones et al. (2023). Lo que presenta la posibilidad de extender la metodología a análisis más robustos como FMECA, donde se requiere identificar, analizar y priorizar información crítica en base a conversaciones técnicas.

3.3.2. Ventajas de los LLMs en entornos industriales

- Comprenden y procesan lenguaje técnico sin necesidad de programación específica.
- Permiten capturar conocimiento verbal no documentado de operadores.
- Mejoran la trazabilidad y estandarización de análisis técnicos.
- Reducen los tiempos de elaboración de reportes o matrices estructuradas.

3.3.3. Desafíos Actuales

- Necesidad de validación humana en aplicaciones críticas.
- Riesgo de respuestas inexactas si el contexto no está bien definido.
- Requieren adaptar el modelo a dominios con jerga técnica especializada (Myöhänen, 2023).

Por ende, el modelo de lenguaje cumple con la función central de interpretar la transcripción de conversaciones técnicas obtenidas vía ASR, extraer de estas conversaciones los elementos necesarios para el análisis FMECA y generar automáticamente los informes técnicos estructurados. La integración de ambos sistemas, LLMs y ASR, no solo facilita el análisis, sino que reduce el tiempo y esfuerzo requerido por los expertos y encargados de la supervisión y gestión del proceso en cuestión, pudiendo estandarizar los procedimientos y dar trazabilidad a las fallas.

No obstante, la capacidad de generar contenido coherente no garantiza necesariamente su eficacia. En una metodología como FMECA, el principal desafío no consiste únicamente en identificar términos aislados, sino en establecer correctamente las relaciones entre componentes, modos de falla, causas, efectos y niveles de criticidad. Por ello, su utilización requiere criterios de valida-

ción que permitan distinguir entre información recuperada desde la conversación evitar contenido generado por los modelos.

3.4. Reconocimiento Automático de Voz (ASR)

El Reconocimiento Automático de Voz (ASR, por sus siglas en inglés) es una tecnología que permite convertir el habla humana en texto mediante algoritmos computacionales, representando un puente fundamental entre la comunicación verbal y los sistemas digitales, y a su vez, aportando trazabilidad, automatización y eficiencia en la toma de decisiones estratégicas.

Los sistemas ASR modernos funcionan a través de una arquitectura de redes neuronales profundas, entre ellas las redes de tipo transformer, capaces de procesar secuencias completas de audio y extraer patrones fonéticos, léxicos y semánticos del lenguaje. En los últimos años, los sistemas ASR han mostrado avances significativos gracias al uso de modelos de aprendizaje profundo, especialmente redes neuronales recurrentes (RNN) y redes de atención, permitiendo mejoras en precisión y adaptabilidad a distintos acentos y ruidos de fondo (Li et al. 2022).

Un sistema ASR típicamente está compuesto por tres componentes principales, un modelo acústico, el que convierte la señal de audio en representaciones fonéticas, un modelo lingüístico que predice secuencias de palabras en base a los contextos y un decodificador que transforma estas predicciones en un texto final.

Hoy en día, gracias a los avances en deep learning, especialmente con arquitecturas como RNNs, CNNs y transformers, el reconocimiento de voz ha alcanzado niveles de precisión cercanos al desempeño humano en condiciones ideales (Graves et al. 2013; Hannun et al. 2014).

Actualmente, modelos como Whisper (OpenAI), DeepSpeech (Mozilla) y Google Speech-

to-Text API son capaces de procesar múltiples idiomas, acentos y condiciones acústicas adversas, lo que los hace especialmente útiles en entornos industriales donde la grabación puede estar expuesta a ruido ambiental.

3.4.1. Método de evaluación del rendimiento y calidad

La calidad de un sistema ASR se mide comúnmente mediante la tasa de error de palabras o Word Error Rate (WER), que calcula la cantidad de inserciones, omisiones y sustituciones de palabras respecto a una transcripción de referencia. Un sistema ASR se considera de alta precisión cuando el WER está por debajo del 10 %. Modelos como Whisper han demostrado WER menores al 5 % en condiciones controladas, e incluso bajo el 10 % en contextos reales con ruido, múltiples hablantes o jerga técnica (Nikulin et al. 2025).

Además, en ambientes técnicos se pueden complementar estos indicadores con métricas de usabilidad como el tiempo de corrección humana, la aceptación de los usuarios o el grado de comprensión del contenido transcrito. Por estas razones, es fundamental seleccionar un modelo ASR robusto, capaz de adaptarse a condiciones variables y terminología técnica, especialmente si se busca integrar esta tecnología en un flujo automatizado de análisis como el FMECA.

3.4.2. Aplicación del ASR en la industria

En entornos industriales, el ASR ha sido incorporado como herramienta para documentar automáticamente conversaciones técnicas, diagnósticos en terreno y reuniones operativas. Estas verbalizaciones muchas veces contienen información crítica que no queda registrada por escrito, y cuya sistematización podría alimentar análisis estructurados como FMECA o RCA (Root Cause Analysis). A su vez, esto es particularmente útil en sectores como minería, manufactura o energía,

donde la operación en terreno limita la capacidad de registrar información de forma manual.

Nikulin et al. (2025) desarrollaron un sistema que integra ASR con modelos de lenguaje (LLMs) para automatizar análisis de causa raíz (RCA) a partir de discusiones grabadas entre operadores en una faena minera chilena. La arquitectura ASR-LLM permitió transcribir audios técnicos, extraer causas, efectos y sugerencias de acción, y generar automáticamente un reporte técnico. En el caso del análisis de falla de un molino SAG, el sistema logró reconstruir todo el proceso sin intervención manual en la codificación de datos. Los resultados mostraron que el sistema fue capaz de transcribir y estructurar correctamente más del 85 % de las entradas técnicas recogidas en entornos reales.

3.4.3. Desafíos y consideraciones en su uso

A pesar de su utilidad, el uso de ASR en entornos industriales presenta ciertos desafíos que deben ser considerados: Ruido de fondo elevado, propio de entornos como plantas o minas, vocabulario técnico especializado y siglas, que pueden no ser reconocidas por modelos genéricos, interrupciones, superposición de voces y habla espontánea, que dificultan la segmentación del audio y requerimientos de hardware y calidad de grabación, que influyen directamente en la calidad de la transcripción.

Nikulin et al. (2025) recomiendan mitigar estos factores mediante protocolos de grabación estandarizados (ubicación de micrófonos, canales separados), entrenamiento de modelos en corpus técnicos, y la incorporación de supervisión humana (human-in-the-loop) en etapas críticas del análisis.

En el contexto de esta memoria, el ASR actúa como el primer eslabón del proceso de automatización del análisis FMECA. Su función principal será la de capturar y transcribir con alta

fidelidad las discusiones técnicas realizadas por equipos de mantenimiento o ingeniería en terreno.

La integración ASR–LLM permitirá transformar conversaciones no estructuradas en matrices FMECA automatizadas y reportes PDF trazables, reduciendo la dependencia de registros manuales, mejorando la eficiencia en la toma de decisiones y asegurando la conservación del conocimiento técnico generado oralmente en entornos reales de trabajo.

En consecuencia, el desempeño del ASR no debe analizarse únicamente mediante una tasa global de error de palabras. En discusiones industriales, una transcripción puede presentar pocos errores generales y, al mismo tiempo, alterar términos técnicos críticos como nombres de componentes, condiciones de operación o modos de falla. La relevancia del ASR para esta investigación depende, por tanto, del efecto que sus errores ejercen sobre la posterior estructuración del FMECA.

3.5. Colaboración humano-inteligencia artificial (IA) en ingeniería

La colaboración entre humanos e inteligencia artificial (IA) es un área emergente que ha cobrado especial relevancia en el ámbito de la ingeniería, especialmente en la toma de decisiones técnicas, el análisis de fallas y la optimización de procesos. Lejos de buscar reemplazar la experiencia humana, estos sistemas tienen como objetivo amplificar la capacidad cognitiva de los expertos, facilitando el procesamiento de grandes volúmenes de información, la estandarización del análisis y la documentación automatizada del conocimiento técnico.

Esta colaboración humano–IA no solo permite acelerar el proceso de análisis, sino también aumentar la transparencia y trazabilidad de las decisiones. Al registrar no solo los resultados, sino también el razonamiento subyacente, se reduce la subjetividad y se facilita la revisión futura.

En estos contextos, se ha visto que los usuarios tienden a ver a la IA como un miembro

más del equipo técnico cuando su comportamiento es transparente, su lenguaje es comprensible y su aporte es justificado con evidencia técnica (Zhou et al. 2021).

3.5.1. Human-in-the-loop

La visión predominante en la literatura es la del modelo human-in-the-loop, en el cual los humanos supervisan, guían y validan las decisiones generadas por sistemas de IA. Este enfoque busca mantener el control humano en la toma de decisiones mientras se capitaliza la eficiencia de los modelos automatizados (Amershi et al. 2019).

Desde una perspectiva organizacional, se plantea que el éxito en la implementación de este tipo de colaboración requiere no solo de capacidades técnicas, sino que también de tácticas efectivas para la gestión de la transcripción, indicando que los resultados más confiables de la colaboración humano-IA surgen cuando se emplea para aumentar las capacidades humanas en lugar de sustituirlas (Tveita et al. 2025).

3.5.2. Beneficios y Riesgos

Entre los principales beneficios de esta colaboración destacan:

- Reducción del tiempo en análisis repetitivos.
- Aumento de la cobertura y profundidad del análisis técnico.
- Mejora de la trazabilidad y documentación.
- Promoción del aprendizaje organizacional.

Sin embargo, los estudios también advierten de riesgos como la delegación excesiva de

decisiones a la IA, la pérdida de juicio técnico o la confianza ciega en resultados no explicables (Holstein et al. 2020). Sin embargo, para mitigarlos, se proponen mecanismos de verificación humana, transparencia algorítmica y capacitación del personal para interactuar críticamente con los sistemas de inteligencia artificial.

3.6. Síntesis crítica del estado del arte y brecha de investigación

Los antecedentes revisados muestran una evolución desde la automatización de análisis de fallas mediante sistemas estructurados hacia el uso de modelos de lenguaje de gran escala capaces de procesar información técnica expresada en lenguaje natural. Los trabajos iniciales sobre automatización de FMEA demostraron que las herramientas computacionales pueden ampliar la cobertura del análisis y reducir el esfuerzo requerido para examinar múltiples combinaciones de fallas (Price and Taylor 2002). Sin embargo, estos enfoques dependen principalmente de información previamente estructurada y no abordan la captura directa del conocimiento generado durante discusiones entre especialistas.

La incorporación de modelos de lenguaje de gran escala ha permitido avanzar hacia el procesamiento de información técnica no estructurada. En particular, Piñones et al. (2023) demostraron que es posible extraer criterios y ponderaciones desde conversaciones para apoyar un análisis AHP, mientras que Nikulin et al. (2025) aplicaron un flujo basado en ASR y LLM para estructurar información asociada a un análisis de causa raíz en minería. Estos antecedentes son relevantes porque evidencian que las conversaciones técnicas pueden utilizarse como fuente de información para análisis de ingeniería. No obstante, los elementos que componen un FMECA presentan una estructura más exigente, dado que requieren identificar y relacionar correctamente múltiples aris-

tas como son los equipos, sus componentes, funciones, modos de falla, causas, efectos, controles y parámetros de criticidad en la misma matriz.

Es por esto que la automatización del FMECA no debería entenderse como una sustitución del juicio experto, sino como un proceso de captura y preestructuración de información que requiere mecanismos explícitos de revisión y validación.

Por otro lado, a partir de los antecedentes revisados, se observa que existen investigaciones orientadas a la automatización del FMEA a partir de información previamente estructurada, así como aplicaciones de modelos ASR–LLM para extraer criterios desde discusiones técnicas y apoyar análisis de decisión o de causa raíz. Sin embargo, dentro de la literatura analizada no se identificaron estudios que evalúen sistemáticamente la capacidad de un flujo ASR–LLM para recuperar, relacionar y estructurar desde conversaciones técnicas todos los campos que componen una matriz FMECA, incluyendo sus relaciones jerárquicas, causas, efectos y parámetros de criticidad.

En particular, permanece abierta la necesidad de determinar en qué medida los errores de transcripción y de interpretación del lenguaje afectan la calidad técnica del FMECA generado, qué configuración del modelo de lenguaje presenta un mejor desempeño y qué nivel de revisión experta se requiere antes de aceptar la matriz resultante.

Por lo tanto, la brecha de investigación abordada en esta memoria corresponde a la falta de una evaluación experimental de la calidad, confiabilidad y necesidad de supervisión humana de matrices FMECA generadas automáticamente a partir del análisis de conversaciones técnicas mediante sistemas ASR–LLM.

4. Metodología

Para poder llevar a cabo esta memoria y alcanzar los objetivos propuestos, se debe establecer un flujo claro de trabajo que permita estructurar el proceso completo, desde la conversación técnica inicial entre los especialistas hasta el ordenamiento final de la información identificada y extraída de dichas conversaciones.

Es por esto que, tal como se muestra en la Figura N°2, el flujo metodológico consta de tres pasos principales, los cuales se encuentran alineados con tres grandes aristas a tratar: la capacidad de transcripción de los sistemas de reconocimiento automático del habla (ASR), la utilidad de la inteligencia artificial, y más específicamente de los modelos de lenguaje de gran escala (LLM), en la identificación de patrones específicos y contextuales en conversaciones transcritas, y finalmente la creación y estructuración de la información necesaria para llegar al FMECA final.

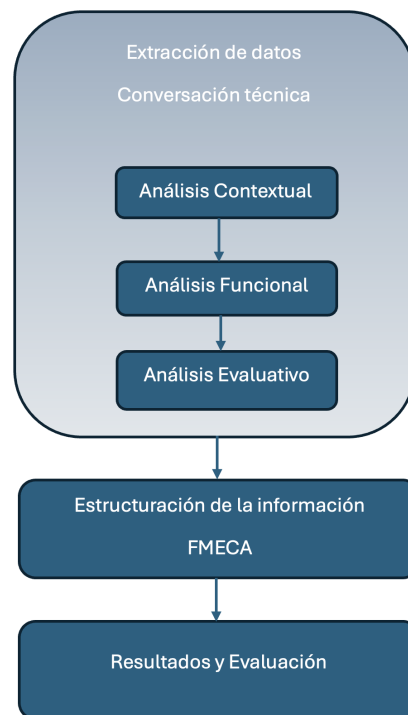


Figura 2: Flujo general de la metodología propuesta

4.1. Recolección de información (conversación técnica y uso de ASR)

Como primer paso del flujo metodológico se debe establecer un formato base para la conversación técnica que será utilizada, con el fin de agilizar el proceso y entregar a la inteligencia artificial información clara, estructurada e interpretable.

Esta conversación debe contener, como mínimo, tres aristas principales que se deben abordar durante el intercambio entre los especialistas.

4.1.1. Contextual / descriptiva

Todo FMECA parte con la necesidad de comprender el sistema y delimitar su alcance. Por ello, en esta etapa se debe generar una contextualización del proceso o sistema bajo análisis, indicando cuál es su función, cuáles son sus condiciones normales de operación y permitiendo reconocer los equipos que componen dicho sistema, así como los componentes que deben considerarse dentro de cada equipo.

Esta arista permite que el modelo comprenda el contexto semántico general antes de procesar la información asociada a las fallas.

Los participantes de la conversación deben considerar, como mínimo, las siguientes condiciones:

1. Se describe el equipo o sistema a analizar (nombre, ubicación, modelo y función principal).
2. Se delimita el alcance del equipo o sistema.
3. Se mencionan las condiciones operacionales.
4. Se pueden introducir datos históricos relevantes.

4.1.2. Funcional / analítica

En esta etapa se debe entregar un análisis en el cual se identifiquen las formas en que cada componente puede fallar y el efecto que dichas fallas generan sobre el sistema bajo análisis. El propósito es transformar el conocimiento técnico de los expertos en información estructurada asociada a modos de falla, causas, efectos y controles existentes.

Al igual que en la arista anterior, los participantes deben considerar como mínimo los siguientes aspectos:

1. Se deben identificar correctamente las funciones del sistema, de los equipos y de sus componentes.
2. Se deben describir los modos de falla.
3. Se deben analizar las causas técnicas o físicas de cada falla.
4. Se deben determinar los efectos de cada falla sobre el sistema.
5. Se deben mencionar los controles existentes.

4.1.3. Evaluativo / decisional

Una vez identificadas las fallas y sus relaciones causales, el enfoque evaluativo se centra en la identificación de la severidad, la ocurrencia y la facilidad de detección, de modo que sea posible cuantificar la criticidad de cada uno de los modos de falla identificados.

En esta arista, los participantes deben considerar como mínimo los siguientes aspectos dentro de la conversación:

1. Se deben asignar valores numéricos o términos específicos para los tres criterios de criticidad (S, O y D).
2. Se debe discutir la facilidad de detección de las fallas, lo que puede complementarse con los puntos de control mencionados anteriormente y que, a su vez, complementa la probabilidad de falla.
3. Se debe mencionar la probabilidad de ocurrencia de la falla.
4. Se debe mencionar la severidad que tendrá la ocurrencia de la falla en el sistema bajo análisis.

Cabe mencionar que la conversación no debe seguir necesariamente una estructura rígida ni secuencial, sino que puede desarrollarse de manera natural, mezclando etapas y fragmentando cada paso, tal como se ilustra en el siguiente ejemplo.

Ejemplo

1. “Se está analizando un sistema de impulsión de ácido que cuenta con una bomba de impulsión modelo XXXX”.
2. “Si el sello mecánico pierde estanqueidad, se produce una fuga de solución ácida sobre el pretil”.
3. “Cuando ocurre esta falla y hay una fuga de ácido, la severidad es alta ya que podría provocar una parada de producción. Sin embargo, debido a que existen detectores de fuga en el pretil, la probabilidad de que ocurra esta falla es baja”.

A partir de este ejemplo es posible extraer el contexto tanto de la primera como de la segunda oración, identificando inicialmente el sistema de impulsión de ácido, luego el equipo

específico, que corresponde a la bomba, y finalmente el sello mecánico como un componente o elemento mantenible del equipo.

Posteriormente, se identifica que el sello mecánico falla mediante la pérdida de estanqueidad, lo que provoca una fuga de solución ácida en el sistema. Finalmente, tal como se indica en la tercera oración, se identifican la alta severidad, la alta facilidad de detección y la baja probabilidad de ocurrencia del modo de falla mencionado, correspondiente a la pérdida de estanqueidad de la bomba.

Una vez finalizada la conversación y disponiendo del archivo de audio en formato .mp3, este será transcrito utilizando un sistema de reconocimiento automático del habla (ASR).

4.2. Estructuración del FMECA (utilización de modelos de lenguaje)

En esta etapa, el propósito es transformar el lenguaje natural utilizado en la conversación técnica en información estructurada y utilizable para la construcción del FMECA. Para ello, cada fragmento de la conversación debe contener frases con entidades clave identificables, las cuales el modelo de lenguaje (LLM) debe detectar, clasificar y relacionar, siguiendo la lógica definida en la metodología.

De esta forma, es posible generar un archivo Excel que contenga una base de datos estructurada con todos los modos de falla identificados, clasificados y ordenados, junto con su respectiva criticidad.

El proceso de estructuración se basa en la identificación de categorías y variables específicas dentro del texto transcrito. La Tabla 1 resume las principales entidades que el modelo debe detectar y el significado asociado a cada una de ellas.

Tabla 1: Entidades a identificar para la estructuración del FMECA

Categoría	Etiqueta (variable)	Qué se debe detectar
Contexto	Equipo / Componente	Nombre del activo o parte bajo análisis.
Función	Función	Qué debe hacer el componente correctamente.
Disfunción	Modo de Falla	Qué falla ocurre o puede ocurrir.
Causa	Causa de Falla	Origen o razón técnica de la falla.
Efecto	Efecto de Falla	Consecuencia visible o impacto en el sistema.
Control actual	Control Existente	Medidas, barreras o sensores que mitigan o detectan la falla.
Parámetros de criticidad	Severidad, Ocurrencia, Detección	Nivel de riesgo asociado, que puede ser estimado o discutido verbalmente.

Cabe mencionar que este modelo no solo considera cada entidad de manera individual, sino que también establece una relación jerárquica entre los distintos elementos identificados. Esta relación jerárquica permite mantener la coherencia del FMECA, asegurando que cada modo de falla se encuentre correctamente asociado a su componente, equipo y sistema correspondiente, y que los parámetros de criticidad se vinculen de manera consistente con cada modo de falla identificado.

La estructura jerárquica utilizada sigue la siguiente lógica:

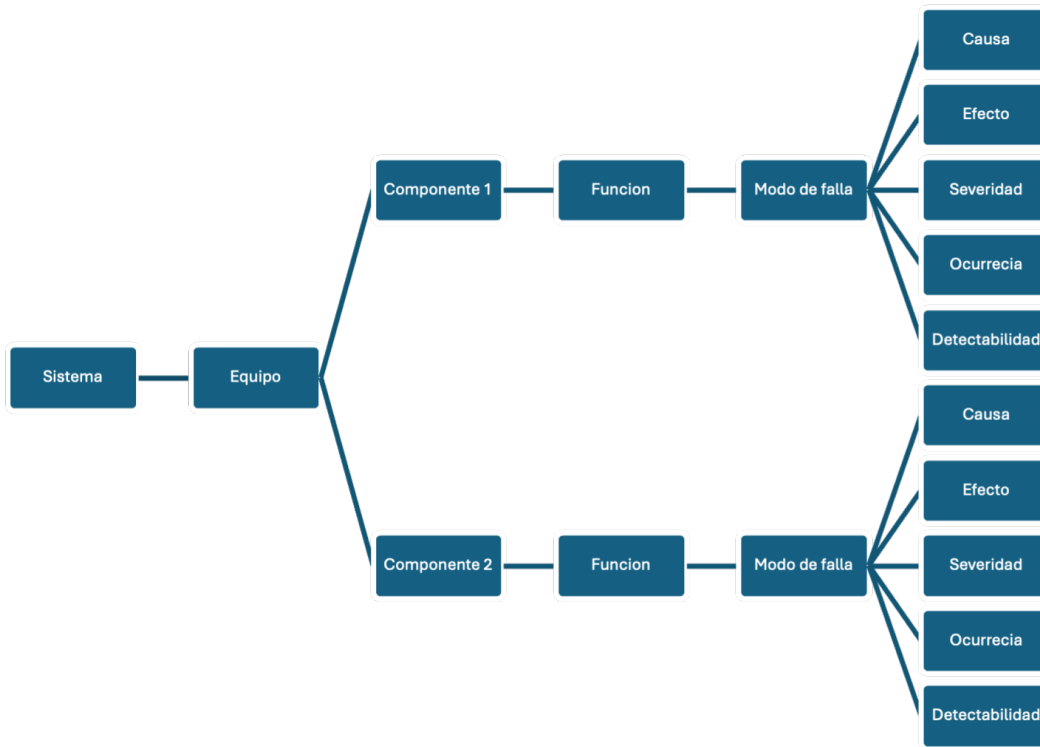


Figura 3: Estructura jerárquica FMECA

4.3. Estrategia de validación experimental

Con el propósito de evaluar la capacidad del sistema propuesto para capturar y estructurar el conocimiento técnico contenido en una discusión especializada, se definió una estrategia de validación experimental basada en la comparación del resultado automático con referencias elaboradas y revisadas manualmente.

La validación se dividió en dos niveles. El primero correspondió a la evaluación de la transcripción generada mediante reconocimiento automático de voz, utilizando como referencia una versión corregida manualmente del audio. El segundo nivel correspondió a la evaluación del FMECA generado por el modelo de lenguaje, comparándolo con una matriz de referencia revisada por especialistas.

Esta separación permitió identificar si las diferencias observadas en el resultado final se originaron durante la transcripción del audio o durante la posterior extracción y estructuración de la información mediante el modelo de lenguaje.

4.3.1. Diseño del experimento y caso de estudio

La validación se desarrolló mediante un estudio de caso aplicado al sistema mecánico de un compresor industrial. El equipo fue seleccionado por presentar un nivel de complejidad media-alta y disponer de un FMECA de referencia compuesto por los modos de falla asociados a sus distintos elementos mantenibles. Esta extensión permitió evaluar el desempeño del sistema sobre un conjunto amplio de relaciones técnicas, manteniendo un alcance abordable para la revisión manual.

La discusión técnica utilizada aborda las condiciones operacionales del equipo, sus componentes, funciones, modos de falla, causas, efectos, controles existentes y parámetros de criticidad.

El FMECA de referencia estuvo compuesto por 38 modos de falla asociados a los elementos mantenibles del equipo. Para efectos de la evaluación, se consideró como unidad de análisis cada conjunto formado por el componente, su función, el modo de falla y los campos técnicos asociados a este último.

4.3.2. Construcción de las referencias de validación

Para evaluar el desempeño del sistema se construyeron dos referencias. En primer lugar, el audio de la discusión técnica fue transcrito y posteriormente corregido de forma manual, obteniéndose una versión de referencia para identificar inserciones, omisiones y sustituciones producidas por el sistema ASR.

En segundo lugar, se utilizó un FMECA de referencia elaborado a partir de la misma discusión técnica y revisado por 2 ingenieros mecánicos. Esta matriz estableció los componentes, funciones, modos de falla, causas, efectos, controles y parámetros de criticidad considerados técnicamente válidos para el caso analizado.

4.3.3. Procedimiento experimental

El procedimiento de validación se desarrolló mediante las siguientes etapas:

1. Registro de la discusión técnica en formato de audio.
2. Procesamiento del archivo mediante el sistema de reconocimiento automático de voz para obtener una transcripción inicial.
3. Comparación de la transcripción automática con la versión corregida manualmente, identificando inserciones, omisiones, sustituciones y errores asociados a terminología técnica.
4. Procesamiento de la transcripción automática mediante el modelo de lenguaje, utilizando una instrucción estructurada para identificar las entidades y relaciones requeridas por el FMECA.
5. Generación automática de la matriz FMECA en formato estructurado.
6. Alineación de las filas y campos generados con el FMECA de referencia, utilizando como claves principales el componente y el modo de falla.
7. Clasificación de cada elemento generado según su correspondencia con la referencia.
8. Registro del tiempo requerido para la generación, revisión y corrección del FMECA.

4.3.4. Modelos utilizados

La transcripción se realizó mediante el modelo whisper-1 de OpenAI, invocado a través de la API de OpenAI desde un cuaderno ejecutado en Google Colab. Los cuatro archivos de audio fueron procesados de manera independiente, obteniéndose una transcripción para cada uno de ellos.

Posteriormente, las transcripciones fueron procesadas mediante el modelo GPT-4, también invocado mediante la API de OpenAI desde Google Colab. Este modelo identificó, relacionó y estructuró la información técnica utilizada para generar la matriz FMECA en formato Excel.

4.3.5. Métricas de evaluación

La evaluación consideró indicadores asociados a la transcripción, la recuperación del contenido técnico, la incorporación de información no respaldada y la intervención humana requerida.

Tasa de Error de Palabras (WER)

$$WER = \frac{S + D + I}{N}, \quad (1)$$

Cobertura del contenido técnico

$$Cobertura = \frac{N_{\text{elementos correctamente recuperados}}}{N_{\text{elementos de referencia}}}. \quad (2)$$

Precisión del contenido Generado

$$Precisin = \frac{N_{\text{elementos correctamente generados}}}{N_{\text{elementos totales generados}}}. \quad (3)$$

5. Análisis cuantitativo y evaluación de beneficios

La Figura 4 presenta el diagrama de flujo de la metodología utilizada para la elaboración del FMECA por equipo con sus respectivos tiempos de ejecución estimados. El proceso considera como etapas comunes a ambos escenarios el análisis previo del sistema y la sesión técnica de discusión FMECA, las cuales son necesarias para comprender el funcionamiento del equipo y definir los modos de falla. La diferencia entre el método tradicional y la propuesta se concentra exclusivamente en la etapa de documentación del FMECA en formato Excel. Mientras que en el enfoque tradicional esta tarea se realiza de forma manual, la metodología propuesta incorpora una herramienta de apoyo basada en transcripción automática y extracción estructurada de información, que permite generar directamente el FMECA en Excel. El proceso finaliza en ambos casos con la revisión y validación del FMECA por parte del especialista, quedando el archivo Excel como insumo para etapas posteriores no consideradas dentro del alcance de este estudio.

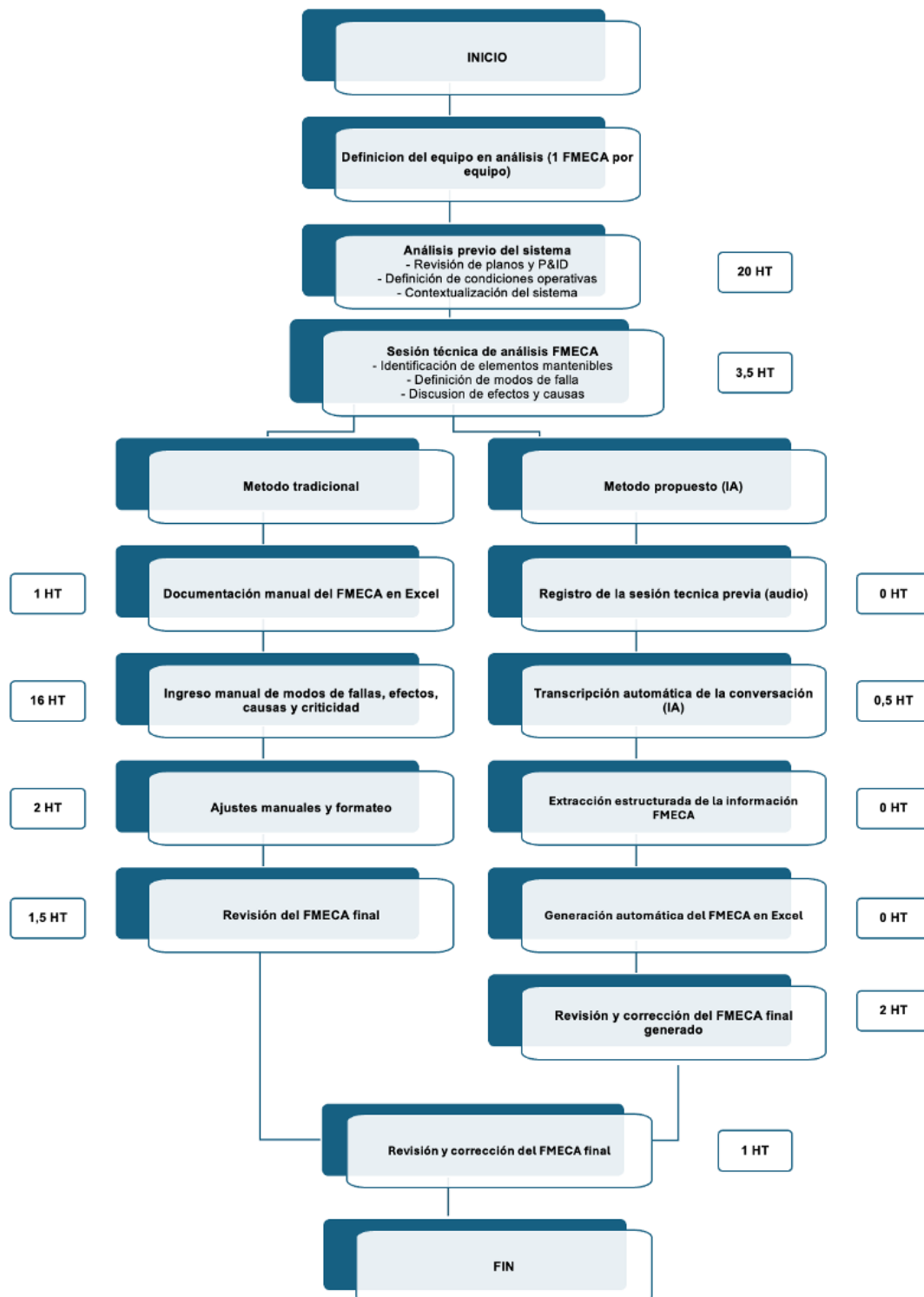


Figura 4: Metodologías FMECA tradicional v/s propuesto

5.1. Definición de tiempos de trabajo a partir de un caso referencial

Con el fin de cuantificar el impacto de la propuesta, se definió primero un caso de referencia concreto, correspondiente al FMECA del sistema mecánico de un compresor. Este equipo se clasifica como de complejidad media dentro de un sistema industrial. El FMECA considerado presenta $N_m = 38$ modos de falla, asociados a los distintos elementos mantenibles del compresor, lo que lo convierte en un ejemplo representativo de un equipo con una cantidad significativa de modos de falla, pero aún manejable desde el punto de vista del análisis unitario.

5.1.1. Supuestos generales

Para este caso se consideró un equipo de trabajo compuesto por dos especialistas: un ingeniero senior, responsable del análisis técnico y de la validación final del FMECA, y un ingeniero junior, que presta apoyo en tareas de recopilación de información y documentación.

Se trabaja con el concepto de horas de trabajo (HT) por especialista, entendidas como las horas efectivas que cada profesional dedica a las distintas actividades del proceso. De este modo, el esfuerzo total del equipo para elaborar un FMECA se obtiene como:

$$HT_{\text{equipo}} = HT_{\text{senior}} + HT_{\text{junior}}$$

En esta sección se presentan inicialmente las horas de trabajo correspondientes al ingeniero senior. Posteriormente, en el análisis económico, estas mismas magnitudes se consideran también para el ingeniero junior, ponderadas por su nivel de participación y su costo horario.

Se distinguen dos tipos de actividades. Por una parte, actividades comunes que se realizan tanto en el método tradicional como en el método asistido por inteligencia artificial, correspondientes al análisis previo del sistema y a la sesión técnica FMECA. Por otra parte, actividades de

documentación, donde se produce la principal diferencia entre ambos métodos, es decir, la documentación manual del FMECA frente a su generación automatizada en formato Excel.

5.1.2. Tiempos en el método tradicional

En el método tradicional, la secuencia de trabajo por equipo se descompone en las siguientes etapas. En primer lugar, se realiza un análisis previo del sistema, que incluye la revisión de planos y diagramas P&ID, condiciones operativas y contexto del equipo dentro del sistema. Para el compresor mecánico analizado, se estiman 20 HT del ingeniero senior:

$$HT_{\text{análisis}} = 20 \text{ HT}$$

En segundo lugar, se desarrolla una sesión técnica de análisis FMECA, correspondiente a una reunión de trabajo en la que se identifican los elementos mantenibles y se discuten los modos de falla, junto con sus efectos y causas. Para este equipo se considera una sesión de 3,5 HT:

$$HT_{\text{sesión}} = 3,5 \text{ HT}$$

Finalmente, se lleva a cabo la documentación manual del FMECA en Excel. Esta etapa incluye la preparación de la plantilla, el ingreso manual de los modos de falla, efectos, causas y criticidad, los ajustes de formato y ordenamiento de filas, y la revisión completa del FMECA con sus correcciones finales.

El punto más exigente de esta etapa corresponde al ingreso manual de la información. Aunque los modos de falla ya han sido discutidos y acordados durante la sesión técnica, en el método tradicional el especialista debe reconstruir mentalmente dicha información al momento de plasmarla en la hoja de cálculo. Para el caso del compresor, se estima que esta actividad requiere 16 HT del ingeniero senior:

$$HT_{\text{ingreso, trad}} = 16 \text{ HT}$$

Considerando adicionalmente 1 HT para la preparación de la plantilla, 2 HT para ajustes y formateo, 1,5 HT para la revisión detallada del FMECA y 1 HT para la revisión final de cierre, el tiempo total dedicado a la documentación manual es:

$$\begin{aligned} HT_{\text{doc, trad}} &= 1 + 16 + 2 + 1,5 + 1 \\ &= 21,5 \text{ HT} \end{aligned}$$

En consecuencia, el tiempo total del proceso tradicional por especialista para este FMECA resulta:

$$\begin{aligned} HT_{\text{trad}} &= HT_{\text{análisis}} + HT_{\text{sesión}} + HT_{\text{doc, trad}} \\ &= 20 + 3,5 + 21,5 \\ &= 45 \text{ HT} \end{aligned}$$

Este valor es coherente con información real proveniente de modelos de gestión de proyectos y cartas Gantt, donde el desarrollo promedio de un FMECA considera alrededor de una semana completa de trabajo desde el inicio hasta el cierre, con dedicación completa de al menos un ingeniero experto.

A partir de estos tiempos, es posible obtener un valor promedio del esfuerzo de documentación por modo de falla. Para la etapa de ingreso manual se tiene:

$$\bar{t}_{\text{modo, trad}} = \frac{HT_{\text{ingreso, trad}} \times 60}{N_m} = \frac{16 \times 60}{38} \approx 25,3 \text{ min/modos}$$

Esto indica que, en promedio, el especialista requiere del orden de 25 minutos por modo de falla para reconstruir la discusión técnica y plasmarla en la tabla FMECA, incluyendo la redacción de descripciones, causas, efectos y la asignación de índices de criticidad.

5.1.3. Tiempos en el método propuesto con asistencia de inteligencia artificial

En la metodología propuesta, el análisis técnico se conserva, es decir, se mantienen el análisis previo del sistema y la sesión FMECA. Sin embargo, la información generada en la reunión se captura mediante el registro de audio y se procesa utilizando una herramienta de inteligencia artificial que realiza la transcripción automática de la conversación, la extracción estructurada de la información relevante y la generación del FMECA directamente en formato Excel.

Desde la perspectiva del ingeniero senior, las actividades de documentación se reducen a la supervisión del correcto funcionamiento del pipeline de inteligencia artificial, con un esfuerzo aproximado de 0,5 HT, y a la revisión y corrección del FMECA generado para asegurar la consistencia técnica y de criticidad. Esta revisión contempla 2 HT de revisión detallada y 1 HT de revisión global.

El tiempo total de documentación en el escenario propuesto es:

$$\begin{aligned} HT_{\text{doc, IA}} &= 0,5 + 2 + 1 \\ &= 3,5 \text{ HT} \end{aligned}$$

Por lo tanto, el tiempo total del proceso con inteligencia artificial por especialista se calcula como:

$$\begin{aligned} HT_{\text{IA}} &= HT_{\text{análisis}} + HT_{\text{sesión}} + HT_{\text{doc, IA}} \\ &= 20 + 3,5 + 3,5 \\ &= 27 \text{ HT} \end{aligned}$$

En términos de esfuerzo promedio por modo de falla, la etapa principal de revisión detallada corresponde a:

$$\bar{t}_{\text{modo, IA}} = \frac{2 \times 60}{38} \approx 3,2 \text{ min/modo}$$

A esta cifra se suman aproximadamente 1,6 minutos adicionales por modo asociados a la revisión global. En conjunto, el orden de magnitud es de alrededor de 4,7 minutos por modo de falla para revisar y ajustar el contenido generado por la herramienta, lo que contrasta significativamente con los aproximadamente 25 minutos por modo requeridos en el método tradicional.

5.1.4. Comparación de tiempos

Con los valores definidos, el ahorro de tiempo por especialista para este FMECA de compresor se calcula como:

$$\Delta HT_{\text{total}} = HT_{\text{trad}} - HT_{\text{IA}} = 45 - 27 = 18 \text{ HT}$$

lo que equivale a una reducción relativa del:

$$\frac{\Delta HT_{\text{total}}}{HT_{\text{trad}}} = \frac{18}{45} \approx 40 \%$$

Es decir, la metodología propuesta reduce en torno a un 40 % el esfuerzo total requerido por especialista para elaborar el FMECA de un equipo de complejidad similar al compresor utilizado como referencia.

Si se analiza únicamente la etapa de documentación en Excel, el ahorro es aún más significativo:

$$\Delta HT_{\text{doc}} = HT_{\text{doc, trad}} - HT_{\text{doc, IA}} = 21,5 - 3,5 = 18 \text{ HT}$$

$$\frac{\Delta HT_{\text{doc}}}{HT_{\text{doc, trad}}} = \frac{18}{21,5} \approx 83,7 \%$$

Esto corresponde a una reducción cercana al 84 % en las horas de trabajo dedicadas a la documentación del FMECA en Excel. Esta diferencia se explica principalmente porque, en el método tradicional, el especialista debe reinterpretar y reescribir la información ya discutida durante la sesión técnica, mientras que en la metodología propuesta dicha sesión se utiliza directamente como entrada a la herramienta de inteligencia artificial, reduciendo el trabajo posterior a una tarea de revisión y ajuste.

Finalmente, dado que las horas de trabajo se definen por ingeniero, el esfuerzo total del equipo se obtiene sumando las horas del ingeniero senior y del ingeniero junior para cada escenario. Esta estructura permite, en el análisis económico posterior, ponderar el impacto de la propuesta tanto en términos de horas totales de trabajo como en términos de costo, asignando costos horarios diferenciados a cada tipo de especialista.

5.1.5. Resumen de beneficios de tiempo

El proceso tradicional de elaboración del FMECA demanda aproximadamente 45 HT por equipo. La metodología propuesta reduce este esfuerzo a 27 HT por especialista, principalmente al eliminar la necesidad de reprocesar cognitivamente la información durante la etapa de documentación. En términos absolutos, el ahorro es de 18 HT por FMECA, lo que equivale a una reducción cercana al 40 % del esfuerzo total requerido por especialista.

Si se observa específicamente la etapa de documentación del FMECA en Excel, la reducción alcanza aproximadamente un 84 % respecto al método tradicional, manteniendo intacto el contenido técnico del análisis. Este resultado respalda que la herramienta propuesta actúa sobre la parte más intensiva en tiempo del proceso, sin alterar la calidad ni la profundidad del estudio FMECA.

5.2. Asignación de costos por rol y ahorro económico

A partir de los tiempos de trabajo definidos en la sección anterior, se incorpora ahora la dimensión económica, distinguiendo entre el costo horario del ingeniero senior y el del ingeniero junior. Se definen:

- C_s : costo por hora de trabajo del ingeniero senior [USD/HT].
- C_j : costo por hora de trabajo del ingeniero junior [USD/HT].

El costo total asociado a la elaboración de un FMECA para un equipo se obtiene como:

$$C_{\text{total}} = HT_{\text{senior}} \cdot C_s + HT_{\text{junior}} \cdot C_j$$

donde HT_{senior} y HT_{junior} corresponden a las horas de trabajo efectivas de cada especialista.

5.2.1. Definición de salarios promedio

Para la estimación del valor económico de las horas de trabajo, se utilizaron referencias salariales de ingenieros mecánicos en Chile. Según datos de sueldos presentados en plataformas de empleo y encuestas salariales, el salario promedio de un ingeniero mecánico en Chile se ubica alrededor de CLP 21,600,000 a 21,800,000 anuales (aproximadamente CLP 1,800,000 mensuales) Levels.fyi (2024). Este valor corresponde a la compensación total usual para profesionales en esta especialidad.

Asimismo, rangos reportados en plataformas de salarios y empleo muestran que los ingresos pueden ir desde aproximadamente CLP 1,180,000 hasta más de CLP 3,000,000 mensuales, dependiendo de la experiencia y del sector industrial (World Salaries 2024). Para convertir estas cifras a valores por hora de trabajo, se asumió una jornada estándar de 44 horas semanales (cerca

de 176 horas mensuales), lo que da un costo promedio del orden de CLP 10,200 por hora de trabajo para un ingeniero mecánico.

Con el fin de distinguir roles en el análisis económico, se considera que el ingeniero senior tiene un costo horario superior, aproximadamente entre CLP 13,000 y CLP 15,000 por hora, reflejando su mayor experiencia y responsabilidad. En cambio, el ingeniero junior se sitúa en un rango cercano a CLP 9,000 a CLP 11,000 por hora. Estos valores sirven como base para la estimación de costos horarios que luego se expresan en dólares estadounidenses (USD) en el análisis económico.

5.2.2. Participación del ingeniero junior

Con base en la experiencia de proyectos reales, se considera que en el método tradicional el ingeniero junior participa de manera importante en la recopilación de información y en la documentación manual, dedicando aproximadamente 22,5 HT por FMECA, lo que corresponde a alrededor de un 50 % del tiempo del ingeniero senior. En el método propuesto, la participación del junior disminuye en la etapa de documentación, debido a la menor cantidad de trabajo manual, y se concentra en apoyar el análisis previo y la revisión del FMECA generado, con aproximadamente 15,5 HT por FMECA.

Así, para el caso de referencia, los tiempos quedan de la siguiente forma:

- Método tradicional: $HT_{\text{senior, trad}} = 45 \text{ HT}$, $HT_{\text{junior, trad}} = 22,5 \text{ HT}$
- Método propuesto con inteligencia artificial: $HT_{\text{senior, IA}} = 27 \text{ HT}$, $HT_{\text{junior, IA}} = 15,5 \text{ HT}$

El esfuerzo total del equipo por FMECA se calcula entonces como:

$$HT_{\text{equipo, trad}} = 45 + 22,5 = 67,5 \text{ HT}$$

$$HT_{\text{equipo, IA}} = 27 + 15,5 = 42,5 \text{ HT}$$

lo que implica un ahorro absoluto de 25 horas de trabajo por equipo. Este valor equivale a una reducción aproximada del 37 % en las horas de trabajo totales del equipo por FMECA:

$$\Delta HT_{\text{equipo}} = HT_{\text{equipo, trad}} - HT_{\text{equipo, IA}} = 25 \text{ HT}$$
$$\frac{\Delta HT_{\text{equipo}}}{HT_{\text{equipo, trad}}} = \frac{25}{67,5} \approx 37 \%$$

5.2.3. Ahorro económico por FMECA

El costo por FMECA en cada escenario se expresa como:

$$C_{\text{trad}} = HT_{\text{senior, trad}} \cdot C_s + HT_{\text{junior, trad}} \cdot C_j$$

$$C_{\text{IA}} = HT_{\text{senior, IA}} \cdot C_s + HT_{\text{junior, IA}} \cdot C_j$$

y el ahorro económico por FMECA es:

$$\Delta C_{\text{FMECA}} = C_{\text{trad}} - C_{\text{IA}}$$

Reemplazando los valores de horas se obtiene:

$$\begin{aligned} \Delta C_{\text{FMECA}} &= (45 - 27) C_s + (22,5 - 15,5) C_j \\ &= 18 C_s + 7 C_j \end{aligned}$$

Esta expresión es general y puede evaluarse con cualquier combinación de tarifas de mercado para ingenieros senior y junior.

A modo ilustrativo, si se consideran los valores definidos anteriormente expresados en dólares estadounidenses, con un costo horario de referencia de $C_s = 15.000 \text{ CLP/HT} \approx 16,7 \text{ USD/HT}$ para el ingeniero senior y $C_j = 10.000 \text{ CLP/HT} \approx 11,1 \text{ USD/HT}$ para el ingeniero junior, se obtiene:

$$\begin{aligned}C_{\text{trad}} &= 45 \cdot 16,7 + 22,5 \cdot 11,1 & C_{\text{IA}} &= 27 \cdot 16,7 + 15,5 \cdot 11,1 \\&\approx 751,5 + 249,8 & &\approx 450,9 + 172,1 \\&\approx 1.001 \text{ USD/FMECA} & &\approx 623 \text{ USD/FMECA}\end{aligned}$$

Por lo que el ahorro económico unitario se aproxima a 378 USD por de ahorro por FMECA, con una reducción relativa del costo directo de aproximadamente un 38 %

$$\Delta C_{\text{FMECA}} = 1.001 - 623 \approx 378 \text{ USD/FMECA}$$

$$\frac{\Delta C_{\text{FMECA}}}{C_{\text{trad}}} = \frac{378}{1.001} \approx 37,8 \%$$

Este ahorro se explica principalmente por la reducción de horas dedicadas a la documentación manual del FMECA en Excel, sin modificar el análisis técnico ni el contenido del estudio. Dado que los costos horarios utilizados corresponden a referencias salariales reales del mercado chileno para ingenieros mecánicos, el resultado obtenido puede considerarse representativo y conservador, y es directamente escalable a proyectos con múltiples equipos.

5.3. Escalamiento a proyectos con múltiples equipos

A partir del ahorro económico unitario por FMECA, es posible estimar el impacto de la metodología propuesta en proyectos de mayor escala, donde se requiere elaborar análisis FMECA para decenas de equipos. En la sección anterior se determinó que, para un equipo de complejidad media, la diferencia de costo entre el método tradicional y el método asistido por inteligencia artificial es del orden de 380 USD por equipo.

Sin embargo, en un proyecto real no todos los equipos presentan la misma cantidad de modos de falla ni el mismo número de elementos mantenibles, existiendo variaciones relevantes

en la complejidad del análisis. Por esta razón, se introduce un factor de complejidad relativo α_i para cada equipo i , que permite ajustar el ahorro económico unitario en función de la complejidad específica del equipo analizado.

Bajo este enfoque, si un equipo presenta aproximadamente la mitad de los modos de falla del equipo de referencia, su factor de complejidad será $\alpha_i \approx 0,5$. De forma análoga, un equipo más complejo, con por ejemplo un 50 % más de modos de falla, tendrá un factor de complejidad $\alpha_i \approx 1,5$. Como primera aproximación, el ahorro por FMECA del equipo i se escala de forma proporcional según la expresión:

$$\Delta C_{\text{FMECA},i} \approx \alpha_i \cdot \Delta C_{\text{FMECA, ref}}$$

5.3.1. Ahorro total en proyectos de 50 a 100 equipos

Para un proyecto con N_{eq} equipos, el ahorro económico total asociado a la implementación de la metodología propuesta se expresa como:

$$\Delta C_{\text{proyecto}} \approx \sum_{i=1}^{N_{\text{eq}}} \alpha_i \cdot \Delta C_{\text{FMECA, ref}}$$

Si se asume que, en promedio, los equipos de un proyecto presentan una complejidad similar, o bien que las variaciones individuales se compensan entre sí, se puede utilizar un valor promedio de $\alpha_i \approx 1$. Bajo este supuesto, el orden de magnitud del ahorro económico para proyectos de 50 y 100 equipos es el siguiente:

$$\Delta C_{\text{proyecto}}^{(50)} \approx 50 \cdot 380 \approx 19.000 \text{ USD}$$

$$\Delta C_{\text{proyecto}}^{(100)} \approx 100 \cdot 380 \approx 38.000 \text{ USD}$$

5.3.2. Escenarios conservador y optimista

Si se adapta la aproximación anterior introduciendo distintos escenarios de complejidad promedio del conjunto de equipos, se define $\bar{\alpha}$ como un valor representativo de dicha complejidad.

En particular, se consideran los siguientes casos:

Escenario conservador ($\bar{\alpha} = 0,8$), donde los equipos del proyecto presentan, en promedio, un 20 % menos de modos de falla que el equipo de referencia.

Escenario optimista ($\bar{\alpha} = 1,2$), donde los equipos son, en promedio, un 20 % más complejos que el equipo de referencia.

Para un proyecto con $N_{eq} = 100$ equipos, el ahorro económico total se ubicaría aproximadamente en el rango:

$$\Delta C_{\text{proyecto}}^{(100)} \approx 100 \cdot \Delta C_{\text{FMECA, ref}} \cdot \bar{\alpha} \Rightarrow \begin{cases} \bar{\alpha} = 0,8 : & \Delta C_{\text{proyecto}} \approx 30.400 \text{ USD} \\ \bar{\alpha} = 1,2 : & \Delta C_{\text{proyecto}} \approx 45.600 \text{ USD} \end{cases}$$

Estos resultados muestran que, incluso bajo supuestos conservadores de complejidad, la magnitud del ahorro se mantiene en el orden de decenas de miles de dólares por proyecto, lo que resulta significativo desde el punto de vista de la gestión de activos y de la evaluación económica de la solución propuesta.

5.4. Punto de equilibrio y período de recuperación de la inversión

5.4.1. Inversión inicial

Para la evaluación económica de la metodología propuesta se considera una inversión inicial asociada a la implementación de la herramienta de apoyo para la generación automática de

FMECAs. Esta inversión incluye el desarrollo y configuración del software, la integración con servicios de transcripción y procesamiento mediante APIs, la ejecución de pruebas iniciales y una capacitación básica de los usuarios finales. Como valor referencial se adopta una inversión inicial del orden de 25.000 USD.

La estimación de este monto se basa en rangos reportados en estudios de mercado y literatura especializada sobre costos de desarrollo de software y soluciones de inteligencia artificial. De acuerdo con diversas fuentes, la implementación de soluciones de IA empresarial puede oscilar entre 6.000 y 300.000 USD, dependiendo de la complejidad, grado de personalización y servicios asociados. Asimismo, estudios sobre desarrollo de software personalizado indican que proyectos pequeños a medianos, como herramientas internas de automatización, suelen presentar presupuestos desde 30.000 hasta 70.000 USD o superiores, según el alcance funcional requerido.

En este contexto, una inversión de 25.000 USD para desarrollar e implementar una herramienta que automatiza parcialmente el proceso FMECA se ubica dentro de un rango razonable para soluciones de IA dirigidas y software a medida, especialmente cuando se aprovecha infraestructura y servicios existentes. Este valor se considera un costo fijo, independiente del número de equipos analizados dentro de un proyecto específico y aplicable a escenarios donde una empresa puede ejecutar múltiples proyectos en paralelo.

5.4.2. Punto de equilibrio (Breakeven)

En este estudio, el punto de equilibrio se define como la cantidad mínima de FMECAs necesarias para que el ahorro acumulado en costos directos de elaboración del FMECA iguale la inversión inicial asociada a la implementación de la herramienta. No se consideran otros costos propios de la ejecución de los proyectos, tales como visitas a terreno, gestión de proyecto o activi-

dades administrativas, dado que dichos costos no se ven modificados por la metodología propuesta.

El punto de equilibrio se calcula como:

$$N_{BE} = \frac{25.000}{380} \approx 65,8$$

Por lo tanto, el punto de equilibrio se alcanza aproximadamente al completar 66 FMECAs. Esto implica que, en un proyecto de alrededor de 100 equipos con un horizonte de ejecución cercano a 6 meses, el costo de implementación de la metodología se recupera antes de finalizar el proyecto, generándose un margen de ahorro neto durante la etapa final del mismo.

5.4.3. Período de recuperación de la inversión (Payback)

El período de recuperación de la inversión corresponde al tiempo requerido para que los ahorros económicos atribuibles a la metodología propuesta igualen la inversión inicial asociada a su implementación. A diferencia del punto de equilibrio, definido en función del número de FMECAs elaborados, el período de recuperación incorpora explícitamente la dimensión temporal del proyecto.

Para este análisis se considera un proyecto tipo con un alcance aproximado de 100 equipos y una duración estimada de 6 meses calendario. Se utilizan los siguientes valores definidos previamente:

$$\Delta C_{\text{proyecto}} \approx 38.000 \text{ USD}, \quad I_0 = 25.000 \text{ USD}$$

El ahorro económico considerado corresponde exclusivamente a la reducción de costos en la etapa de elaboración y documentación de los FMECAs. Otros costos asociados a la ejecución del proyecto no se ven afectados por la metodología propuesta y, por lo tanto, no se incluyen en el cálculo.

Asumiendo que la generación de ahorros se distribuye de manera relativamente homogénea a lo largo del proyecto, el ahorro mensual promedio se estima como:

$$\Delta C_{\text{mensual}} = \frac{38.000}{6} \approx 6.333 \text{ USD/mes}$$

El período de recuperación se calcula entonces como:

$$T_{\text{payback}} = \frac{25.000}{6.333} \approx 3,95 \text{ meses}$$

Por lo tanto, el período de recuperación de la inversión se sitúa en torno a 4 meses. Este resultado indica que, para un proyecto de aproximadamente 100 equipos, la inversión inicial se recupera antes de finalizar el proyecto, generando ahorros económicos netos durante los últimos meses de ejecución. En proyectos posteriores, donde la herramienta ya se encuentra implementada, el ahorro económico se traduce casi íntegramente en reducción de costos operacionales, incrementando el retorno global de la inversión en el mediano y largo plazo.

5.5. Evaluación de la calidad estructural de la matriz generada

5.5.1. Tasa de Error de Palabras

La tasa de error de palabras se calculó comparando las transcripciones generadas automáticamente con los textos transcritos revisados manualmente correspondientes a los audios utilizados.

En el conjunto de transcripciones entregadas por el modelo se identificaron 76 sustituciones, 11 omisiones y 58 inserciones, sobre un total de 1.952 palabras de referencia. Por lo tanto:

$$WER = \frac{S + D + I}{N} = \frac{76 + 11 + 58}{1952} = 0,0743. \quad (4)$$

En consecuencia, la tasa de error de palabras alcanzó un valor de **7,43 %**, lo que equivale aproximadamente a 7,4 errores por cada 100 palabras de referencia.

Tabla 2: Tasa de error de palabras por audio

Audio	S	D	I	N	WER
Parte 1	21	1	16	624	6,09 %
Parte 2	18	4	10	489	6,54 %
Parte 3	18	6	20	438	10,05 %
Parte 4	19	0	12	401	7,73 %
Total	76	11	58	1952	7,43 %

5.5.2. Cobertura del contenido técnico

La cobertura del contenido técnico evalúa la proporción de información esperada en el FMECA de referencia que fue recuperada correctamente por el modelo. Para su cálculo se consideraron siete campos por registro: elementos mantenibles, modos de falla, causas de falla, efectos de falla, ocurrencias, detecciones y severidades.

Los campos correspondientes al sistema, subsistema y equipo no se incluyeron, debido a que representan información contextual común a todos los registros al estar analizando solo un equipo en particular.

El FMECA de referencia contiene 38 registros únicos de modos de falla, por lo que el número total de celdas con información esperadas es el producto entre la cantidad de registros particulares por la cantidad de columnas a evaluar:

$$N_{\text{elementos de referencia}} = 38 \cdot 7 = 266. \quad (5)$$

La comparación entre la matriz de referencia y el resultado generado permitió identificar 249 elementos correctamente recuperados. Por lo tanto, la cobertura se calcula como:

$$\begin{aligned} Cobertura &= \frac{N_{\text{elementos correctamente recuperados}}}{N_{\text{elementos de referencia}}} \\ &= \frac{249}{266} = 0,9361. \end{aligned} \quad (6)$$

Este resultado indica que el modelo recuperó correctamente el 93,6 % de la información técnica esperada en el FMECA. Las principales pérdidas de cobertura se concentraron en la asignación del elemento mantenible y en la recuperación de algunas causas de falla. No obstante, los 38 registros de referencia fueron identificados y las ponderaciones de ocurrencia, detección y severidad fueron recuperadas correctamente en su totalidad.

5.5.3. Precisión del contenido generado

La precisión del contenido generado evalúa qué proporción de la información efectivamente producida por la herramienta coincide correctamente con el FMECA de referencia. A diferencia de la cobertura, esta métrica no incorpora en su denominador aquellos campos que fueron omitidos y quedaron sin información.

De los 266 elementos técnicos posibles, cuatro causas de falla no fueron generadas. Por consiguiente, el número total de elementos efectivamente producidos por la herramienta fue:

$$N_{\text{elementos totales generados}} = 266 - 4 = 262. \quad (7)$$

De estos 262 elementos generados, 249 coincidieron correctamente con la información contenida en el FMECA de referencia. La precisión se obtiene como:

$$\begin{aligned} \textit{Precisin} &= \frac{N_{\text{elementos correctamente generados}}}{N_{\text{elementos totales generados}}} \\ &= \frac{249}{262} = 0,9504. \end{aligned} \quad (8)$$

El resultado indica que el 95,0 % de los campos técnicos efectivamente generados por la herramienta coincidió con la matriz de referencia. Las desviaciones restantes se relacionaron principalmente con la clasificación jerárquica de ciertos componentes como elementos mantenibles y, en menor medida, con la asignación de información entre los campos de modo y causa de falla.

La precisión obtenida fue superior a la cobertura, debido a que la primera considera únicamente el contenido efectivamente generado, mientras que la segunda también penaliza los campos omitidos. En conjunto, ambas métricas muestran que la herramienta produjo información mayoritariamente correcta, aunque todavía requiere una revisión técnica posterior para corregir errores de estructuración y completar los campos no recuperados.

Tabla 3: Resultados de la comparación por campo técnico

Campo evaluado	Correctos	Referencia	Exactitud
Elemento mantenible	20	20	100,0 %
Modo de falla	37	38	97,4 %
Causa de falla	33	38	86,8 %
Efecto de falla	38	38	100,0 %
Ocurrencia	38	38	100,0 %
Detección	38	38	100,0 %
Severidad	38	38	100,0 %
Total	249	266	97,7 %

5.6. Discusión de resultados

Los resultados obtenidos permiten evaluar la metodología propuesta desde dos perspectivas complementarias. Por una parte, se analiza la capacidad del sistema de reconocimiento automáti-

co de voz para conservar el contenido de las conversaciones técnicas. Por otra, se examina la capacidad del modelo de lenguaje para transformar dicha información en una matriz FMECA estructurada. Esta distinción es relevante, debido a que un error presente en el resultado final puede originarse tanto en la transcripción del audio como en la posterior interpretación, clasificación o asociación de los campos técnicos.

En términos generales, el sistema mostró un desempeño favorable bajo las condiciones controladas del experimento. Sin embargo, las diferencias observadas entre los campos evaluados evidencian que una transcripción lingüísticamente aceptable no garantiza por sí sola una estructuración técnica completamente correcta. En consecuencia, los resultados deben interpretarse considerando tanto la precisión global alcanzada como la naturaleza y el impacto de los errores particulares.

5.6.1. Desempeño del reconocimiento automático de voz

El sistema de reconocimiento automático de voz alcanzó una tasa de error de palabras de 7,43 %. Este resultado se encuentra bajo el umbral del 10 % utilizado habitualmente como referencia para identificar transcripciones de alta precisión, de acuerdo con lo señalado en la literatura revisada (Nikulin et al. 2025). Desde una perspectiva global, el desempeño obtenido indica que el modelo whisper-1 fue capaz de conservar la mayor parte del contenido verbal expresado en las cuatro conversaciones técnicas.

Este resultado es consistente con los avances reportados para sistemas ASR modernos, los cuales han demostrado una capacidad elevada para procesar conversaciones en lenguaje natural incluso ante la presencia de vocabulario especializado (Li et al. 2024). No obstante, el análisis de los errores individuales muestra que el valor global de WER no representa completamente el riesgo

asociado a una transcripción técnica.

En particular, se observaron sustituciones como "detención" por "detección", "subsistema" por "su sistema", alteraciones en nombres de componentes. Aunque cada sustitución representa solamente una palabra dentro del cálculo del WER, su efecto sobre el análisis posterior puede ser significativo. Por ejemplo, confundir la detención de un equipo con su detección modifica directamente el significado del efecto de falla y puede producir una asociación incorrecta dentro de la matriz FMECA.

Este hallazgo coincide con lo planteado por Nikulin et al. (2025), quienes señalan que el desempeño de un sistema ASR en aplicaciones industriales no debe evaluarse únicamente mediante métricas lingüísticas globales, debido a que los errores en términos técnicos pueden propagarse hacia las etapas posteriores del análisis. En este sentido, el resultado obtenido confirma que la WER constituye una medida necesaria, pero no suficiente, para evaluar el desempeño de un sistema orientado a la generación de análisis de fallas.

5.6.2. Calidad de la estructuración del FMECA

La matriz generada alcanzó una cobertura del contenido técnico de 97,7 % y una precisión de 95,0 %. Además, los 38 registros considerados en el FMECA de referencia fueron identificados por la herramienta. Estos resultados indican que el sistema fue capaz de recuperar y estructurar una proporción elevada de la información expresada en las conversaciones pero agregando elementos que no se encontraban directamente en el FMECA de referencia.

El desempeño no fue homogéneo entre los campos evaluados. Los efectos de falla y las ponderaciones de ocurrencia, detección y severidad fueron recuperados correctamente en la totalidad de los registros. Los modos de falla también presentaron un alto nivel de coincidencia con

la referencia. En cambio, las principales desviaciones se concentraron en la identificación del elemento mantenible donde se agregaron datos que no estaban en el original y, en menor medida, en las causas de falla que no se han detectado.

La menor exactitud del elemento mantenible muestra que el principal desafío del modelo no corresponde necesariamente a reconocer los términos técnicos, sino a ubicarlos dentro de la jerarquía correcta del activo. En algunos casos, el modelo clasificó cojinetes, engranajes o colectores como elementos mantenibles independientes, mientras que en el FMECA de referencia estos formaban parte de conjuntos de nivel superior, como la unidad compresora, la caja multiplicadora o el separador de aceite–aire.

Este resultado es relevante porque la estructura FMECA no se compone únicamente de entidades aisladas. Requiere conservar relaciones jerárquicas y causales entre el sistema, el equipo, los elementos mantenibles, los modos de falla, las causas y los efectos. Por lo tanto, un campo puede contener información técnicamente plausible y, aun así, ser incorrecto si se encuentra asignado a un nivel jerárquico diferente del definido en la matriz de referencia.

Los resultados amplían lo observado en estudios anteriores. Piñones et al. (2023) demostraron que es posible extraer criterios y ponderaciones desde discusiones técnicas para apoyar análisis AHP. Sin embargo, la generación de un FMECA exige conservar una estructura relacional más compleja, debido a que cada modo de falla debe vincularse simultáneamente con un elemento, una causa, un efecto y parámetros de criticidad. La principal dificultad identificada en este estudio se encuentra precisamente en esa asociación relacional.

Asimismo, el trabajo de Price y Taylor (2002) demuestra que la automatización puede ampliar la capacidad de análisis de múltiples combinaciones de fallas. No obstante, dicho enfoque parte de información previamente estructurada. En el caso desarrollado en esta memoria, la in-

formación se origina en lenguaje natural, por lo que antes de automatizar el análisis es necesario interpretar, clasificar y organizar el conocimiento expresado verbalmente. Esta diferencia introduce nuevas fuentes de error que no están presentes cuando la información de entrada ya se encuentra estructurada.

Los resultados obtenidos son también coherentes con el trabajo de Nikulin et al. (2025), donde la integración entre ASR y LLM permitió estructurar información técnica para análisis de causa raíz. Sin embargo, no corresponde establecer una comparación directa entre los porcentajes reportados en ambos estudios, debido a que las unidades de evaluación, los tipos de análisis y las condiciones experimentales son diferentes. La evidencia conjunta sí permite sostener que la combinación de reconocimiento de voz y modelos de lenguaje presenta capacidad para transformar discusiones técnicas en información estructurada, aunque la confiabilidad final depende del nivel de complejidad de las relaciones que se busca reconstruir y la precisión que se espera del modelo.

5.6.3. Implicancias para la colaboración humano-IA

Los resultados respaldan un enfoque de colaboración humano-IA en lugar de una automatización completamente autónoma. La herramienta produjo una primera versión del FMECA con alta cobertura y precisión, reduciendo la necesidad de transcribir y ordenar manualmente toda la información discutida. Sin embargo, los errores detectados muestran que todavía se requiere una revisión técnica para validar la jerarquía de los activos, la correspondencia entre causas y modos de falla y el uso correcto de la terminología.

Este comportamiento es consistente con el enfoque *human-in-the-loop* descrito por Amershi et al. (2019), en el cual la inteligencia artificial realiza tareas de procesamiento y estructuración, mientras que el especialista mantiene la responsabilidad sobre la validación y la decisión final. En

el contexto del FMECA, esta distribución resulta especialmente adecuada, debido a que la matriz puede influir posteriormente en la definición de estrategias de mantenimiento, priorización de recursos, decisiones asociadas al riesgo operacional y KPI que surgen del plan matriz de mantenimiento de SAP.

Los errores identificados también respaldan las advertencias de Holstein et al. (2020) respecto de la confianza excesiva en resultados generados por sistemas de inteligencia artificial. Una salida puede parecer coherente desde el punto de vista lingüístico y, al mismo tiempo, contener una clasificación técnica incorrecta. Por ejemplo, la asignación de un componente a un nivel jerárquico diferente puede no ser evidente para un usuario que revise superficialmente la matriz, aunque altere la lógica del análisis.

Por esta razón, la revisión humana no debe considerarse únicamente como una etapa destinada a corregir errores de redacción. Debe incluir una verificación formal de la estructura del activo, de las relaciones entre modos de falla, causas y efectos, y de la consistencia de los parámetros de criticidad. La herramienta puede reducir la carga documental y acelerar la elaboración de la matriz, pero la aceptación técnica del FMECA debe permanecer bajo la responsabilidad del especialista.

Desde esta perspectiva, la principal contribución del sistema consiste en transformar el trabajo del experto. En lugar de reconstruir y documentar manualmente toda la discusión, el especialista puede concentrarse en revisar una matriz previamente estructurada, identificar inconsistencias y validar el contenido. Esto permite desplazar el esfuerzo desde tareas repetitivas de documentación hacia actividades de mayor valor técnico.

5.6.4. Limitaciones y validez externa

Los resultados deben interpretarse considerando las condiciones bajo las cuales se desarrolló el experimento. La validación se realizó sobre un único equipo, correspondiente a un compresor industrial, y utilizando conversaciones técnicas previamente estructuradas a partir de un FMECA de referencia. Por lo tanto, el experimento corresponde a una validación controlada de la capacidad del sistema para recuperar y estructurar información verbalizada.

Las conversaciones fueron diseñadas para contener los elementos necesarios del FMECA y se desarrollaron con una secuencia relativamente clara. Además, cada audio incorpora una contextualización del sistema, subsistema y equipo. Estas condiciones favorecen la recuperación de la información y no representan completamente la dinámica de una sesión industrial espontánea, donde pueden existir interrupciones, superposición de voces, ambigüedades, cambios de tema, información incompleta y desacuerdos entre especialistas.

De igual forma, se debe tener en cuenta que existen distintos tipos de entornos industriales donde el ruido externo, la cercanía con el micrófono o bien el uso de EPP pueden llegar a afectar la efectividad de la medida entregada por el WER, sin embargo, considerando tecnologías modernas y las implementaciones del teletrabajo disminuyen la probabilidad de que la métrica se vea afectada.

Asimismo, el experimento se realizó utilizando una configuración específica de los modelos ASR y LLM. Los resultados pueden cambiar según el modelo, la versión utilizada, la estructura del prompt y la forma de dividir los audios.

La clasificación de los campos correctos se realizó mediante comparación semántica con el FMECA de referencia. Aunque se establecieron criterios para distinguir entre variaciones de redacción y errores técnicos, esta evaluación incorpora un componente de juicio. Una validación

posterior podría incluir evaluadores independientes y una medida de concordancia entre revisores, con el fin de aumentar la objetividad del proceso.

En consecuencia, los resultados demuestran la factibilidad y el potencial de la metodología bajo condiciones controladas, pero no permiten asegurar el mismo nivel de desempeño en cualquier entorno industrial.

5.7. Conclusiones del análisis

Los resultados obtenidos muestran que la metodología propuesta es capaz de transformar una discusión técnica en una matriz FMECA estructurada con un nivel elevado de recuperación y consistencia. La transcripción automática alcanzó una tasa de error de palabras de 7,43 %, mientras que la matriz generada presentó una cobertura del contenido técnico de 97,7 % y una precisión de 95,0 %. En conjunto, estas métricas indican que la mayor parte de la información contenida en las conversaciones fue correctamente identificada, estructurada y asignada a los campos correspondientes del FMECA.

La herramienta recuperó los 38 registros considerados en el caso de referencia y convirtió correctamente la totalidad de las ponderaciones cualitativas de ocurrencia, detección y severidad a sus respectivos valores numéricos. Asimismo, los efectos de falla fueron recuperados correctamente en todos los registros y los modos de falla presentaron un alto nivel de coincidencia con la matriz original. Las principales desviaciones se concentraron en la identificación del nivel jerárquico de algunos elementos mantenibles y en la separación entre los campos de modo y causa de falla. También se observaron errores de transcripción técnicamente relevantes, como la sustitución de términos de significado similar en el lenguaje general, pero diferente dentro del contexto del mantenimiento.

Por lo tanto, los resultados no permiten afirmar que la generación del FMECA sea completamente autónoma ni que el contenido técnico permanezca inalterado. Sin embargo, sí demuestran que la herramienta puede producir una primera versión ampliamente completa y mayoritariamente correcta, reduciendo el trabajo posterior del especialista a la revisión, corrección y validación de un número acotado de campos. En este sentido, la metodología debe entenderse como una herramienta

de apoyo a la toma de decisiones y no como un reemplazo del juicio experto.

Desde el punto de vista operacional y económico, el análisis desarrollado demuestra que la metodología propuesta permite obtener beneficios económicos cuantificables y relevantes, manteniendo inalterado el contenido técnico del análisis FMECA. El período de recuperación calculado es consistente con el punto de equilibrio definido previamente y se basa en un escenario conservador, considerando únicamente los ahorros directamente atribuibles a la etapa de elaboración y documentación del FMECA.

Este enfoque evita sobreestimar los beneficios económicos de la solución, al excluir deliberadamente aquellos costos del proyecto que no se ven afectados por la metodología propuesta. De este modo, el análisis proporciona una evaluación realista, transparente y defendible desde el punto de vista de la ingeniería económica.

En síntesis, la principal contribución de la metodología propuesta no se limita a la reducción de tiempos y costos. Su valor radica en que permite capturar y estructurar de manera sistemática una proporción elevada del conocimiento técnico expresado durante las sesiones FMECA. No obstante, la presencia de omisiones, errores terminológicos y diferencias de clasificación confirma la necesidad de mantener una etapa formal de validación humana. Bajo esta condición, la herramienta presenta un potencial significativo para mejorar la eficiencia, trazabilidad y estandarización del proceso de elaboración de matrices FMECA.

Referencias

Adevs (2024). Custom software development costs.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Cakmak, M.,

Kamar, E., and Horvitz, E. (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam,

P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901.

Consejo de Competencias Mineras (2022). Reporte fuerza laboral de la gran minería chilena.

<https://www.ccmineria.cl>.

Corporación Nacional del Cobre de Chile (Codelco) (2024). Análisis razonado de los estados

financieros consolidados al 31 de marzo de 2024. <https://www.codelco.com>. Informe corporativo.

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neu-

ral networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6645–6649.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S.,

Sengupta, S., Coates, A., and Ng, A. Y. (2014). Deep speech: Scaling up end-to-end speech recognition. <https://arxiv.org/abs/1412.5567>. arXiv preprint arXiv:1412.5567.

Holstein, K., Wortman Vaughan, J., Daumke, P., Dudley, J. J., Gray, W. D., and Shah, H. (2020).

The ai in the loop: User-centered design patterns for human-in-the-loop machine learning systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM.

International Organization for Standardization (2014). Iso 55000: Asset management – overview, principles and terminology. Standard No. ISO 55000:2014.

Kumar, U., Karim, R., Galar, D., and Kour, R., editors (2023). *Proceedings of the International Congress and Workshop on Industrial AI and eMaintenance 2023*. Luleå University of Technology, Luleå, Sweden.

Levels.fyi (2024). Mechanical engineer salary in chile.

Li, J., Deng, L., Gong, Y., and Haeb-Umbach, R. (2022). Recent advances in deep learning for speech research at microsoft. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:471–488.

Li, Y., Zhao, H., Jiang, H., Pan, Y., Liu, Z., Wu, Z., Shu, P., Tian, J., Yang, T., Xu, S., Lyu, Y., Blenk, P., Pence, J., Rupram, J., Banu, E., Liu, N., Wang, L., Song, W., Zhai, X., Song, K., Zhu, D., Li, B., Wang, X., and Liu, T. (2024). Large language models for manufacturing. <https://arxiv.org/abs/2410.21418>. Preprint.

Mahmood, A., Wang, J., Yao, B., Wang, D., and Huang, C.-M. (2025). User interaction patterns and breakdowns in conversing with llm-powered voice assistants. *International Journal of Human-Computer Studies*, 195:103406. 10.1016/j.ijhcs.2024.103406.

Myöhänen, J. (2023). Improving industrial performance with language models: A review of pre-

dictive maintenance and process optimization. <https://doi.org/10.48550/arXiv.2310.1234>. White paper.

Márquez, A. C. (2007). *The maintenance management framework: Models and methods for complex systems maintenance*. Springer Science & Business Media, Berlin, Germany.

Nikulin, C., Piñones, E., Carreño, C., Viveros, P., and Kristjanpoller, F. (2025). Integrating automatic speech recognition and large language models for root cause analysis in mining projects. In *ESREL 2025 - European Safety and Reliability Conference*. Manuscrito inédito.

Piñones, E., Nikulin, C., and Zhou, Z. (2023). Human-ai collaboration in engineering design decisions: Automating ahp through llm analysis of team discussions. *Proceedings of the International Conference on Engineering Design*.

Portal Innova (2023). Entre 6.000 y 300.000 usd cuesta implementar inteligencia artificial en una empresa.

Price, C. J. and Taylor, N. S. (2002). Automated multiple failure fmea. *Reliability Engineering & System Safety*, 76(1):1–10.

Rekabi, M. (2024). Failure mode, effects and criticality analysis (fmeca).

Servicio Nacional de Geología y Minería (2021). Informe anual de accidentes en la minería chilena. <https://www.sernageomin.cl>.

Stamatis, D. H. (2003). *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality Press, Milwaukee, WI, 2nd edition.

- Tveita, L. J., Kotsialos, A., and Vassilakopoulou, P. (2025). Human-ai collaboration for case management: Exemplars, required capabilities and transition management. *Procedia Computer Science*, 256:166–173.
- World Economic Forum (2020). Shaping the future of advanced manufacturing and production. <https://www.weforum.org>. Recuperado de sitio web del World Economic Forum.
- World Salaries (2024). Average mechanical engineer salary in chile.
- Zhou, X., Zhang, J., and Chen, J. (2021). Human-centered ai in engineering workflows: A qualitative study. *IEEE Transactions on Engineering Management*, 68(6):1662–1673.