



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

**Ecosistema de monitorización, predicción y
visualización del consumo energético en edificios
inteligentes**

TESIS

Para optar el Grado Académico de Doctor en
Ingeniería de Sistemas e Informática

AUTOR

Jorge Luis VELOZ ZAMBRANO

ASESOR

Dr. Yván Jesús TÚPAC VALDIVIA

Lima, Perú

2026



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

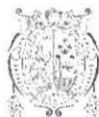
Referencia bibliográfica

Veloz, J. (2026). *Ecosistema de Monitorización, Predicción y Visualización del Consumo Energético en Edificios Inteligentes*. [Tesis de doctorado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Jorge Luis Veloz Zambrano
Tipo de documento de identidad	Cédula de Identidad
Número de documento de identidad	EC/ 1307540722
URL de ORCID	https://orcid.org/0000-0002-9001-4478
Datos de Asesor	
Nombres y apellidos	Dr. Yván Jesús Túpac Valdivia
Tipo de documento de identidad	DNI
Número de documento de identidad	29600586
URL de ORCID	https://orcid.org/0000-0002-7279-3105
Datos del jurado	
Presidente del Jurado	
Nombres y apellidos	Dr. Ciro Rodríguez Rodríguez
Tipo de documento de identidad	DNI
Número de documento de identidad	06020241
Miembro del Jurado 1	
Nombres y apellidos	Dr. Pedro Martin Lezama Gonzales
Tipo de documento de identidad	DNI
Número de documento de identidad	09656793
Miembro del Jurado 2	
Nombres y apellidos	Dr. Ivan Carlo Petrlik Azabache
Tipo de documento de identidad	DNI
Número de documento de identidad	10140461
Miembro del Jurado 3	
Nombres y apellidos	Dr. Yván Jesús Túpac Valdivia
Tipo de documento de identidad	DNI
Número de documento de identidad	29600586
Datos de Investigación	
Línea de Investigación	C.0.3.3. Desarrollo de modelos y aplicación de las tecnologías de información y comunicaciones.

Grupo de Investigación	No Aplica
Agencia de Financiamiento	Sin financiamiento
Ubicación Geográfica de la Investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.056445 Longitud: -77.085994
Año o rango de años en que se realizó la investigación	2023-2025
URL de disciplinas OCDE	Ingeniería de Sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04 Ciencias de la computación https://purl.org/pe-repo/ocde/ford#1.02.01



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

**ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE
DOCTOR EN INGENIERÍA DE SISTEMAS E INFORMÁTICA**

A los ocho (8) días del mes de mayo de 2026, siendo las 11:00 am., se reunieron en el Auditorio Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. **Ciro Rodriguez Rodriguez** (Presidente)
Dr. **Ivan Carlo Petrlik Azabache** (Miembro)
Dr. **Pedro Martin Lezama Gonzales** (Miembro)
Dr. **Yván Jesús Túpac Valdivia** (Miembro Asesor)

Se inició la Sustentación invitando al candidato a Doctor **JORGE LUIS VELOZ ZAMBRANO**, para que realice la exposición oral de la tesis para optar el Grado Académico de Doctor en Ingeniería de Sistemas e Informática, siendo la Tesis intitulada:

**“ECOSISTEMA DE MONITORIZACIÓN, PREDICCIÓN Y VISUALIZACIÓN DEL CONSUMO
ENERGÉTICO EN EDIFICIOS INTELIGENTES”**

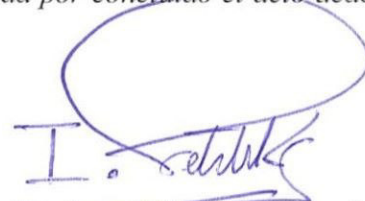
Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

Dieciséis (16) *Muy Bueno*

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga al Magíster **JORGE LUIS VELOZ ZAMBRANO** el Grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Siendo las *12:15* horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.


Dr. *Ciro Rodriguez Rodriguez*
(Presidente)


Dr. *Ivan Carlo Petrlik Azabache*
(Miembro)


Dr. *Pedro Martin Lezama Gonzales*
(Miembro)


Dr. *Yván Jesús Túpac Valdivia*
(Miembro Asesor)

Anexo 1

CERTIFICADO DE SIMILITUD

Yo, Yván Jesús Túpac Valdivia, en mi condición de asesor acreditado mediante Dictamen de la UPG N° 000297-2024-UPG-VDIP-FISI/UNMSM, de la tesis, cuyo título es **“ECOSISTEMA DE MONITORIZACIÓN, PREDICCIÓN Y VISUALIZACIÓN DEL CONSUMO ENERGÉTICO EN EDIFICIOS INTELIGENTES”**, presentado por el egresado **JORGE LUIS VELOZ ZAMBRANO**, para optar el grado Académico de Doctor en Ingeniería de Sistemas e Informática, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante la plataforma PIDS, el documento evaluado cuenta con el **8%** de similitud, nivel PERMITIDO para continuar con los trámites correspondientes y para su **publicación en el repositorio institucional**.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado Académico de Doctor en Ingeniería de Sistemas e Informática.

Lima, 15 de junio de 2026



Firma del Asesor

Nombres y Apellidos del asesor: Yván Jesús Túpac Valdivia

DNI: 29600586

* Según corresponda: bachiller, magíster, doctor; título profesional o título de segunda especialidad

La vigencia del presente certificado corresponde a un período de noventa (90) días desde la fecha de su emisión; siempre que no haya observaciones en el texto durante el proceso de revisión por el jurado informante o surja alguna contingencia.

DEDICATORIA

Para mi Haydita y mi Valita, mi mayor inspiración y la fuerza que impulsa cada uno de mis pasos.

A mi compañera de vida, Andreita, por su apoyo incondicional en cada proceso que hemos vivido.

A mis hermanos, compañeros de travesuras, tristezas y alegrías, y a mis sobrinos, por su jovial manera de ver la vida.

Y, a mi mamá, mi ñaña y mi papá, cuya ausencia duele, pero cuya memoria me impulsa a honrar sus nombres con cada logro.

Gracias, papá, por tu bendición y por enseñarme, incluso en tus últimos días, que el cáncer pudo vencer tu cuerpo, pero nunca tu fe.

AGRADECIMIENTO

Mi agradecimiento se dirige a las personas e instituciones que han contribuido en la culminación de esta investigación:

*A mi **Director de Tesis, Dr. Yván Jesús Túpac Valdivia**, por su valiosa mentoría, su visión crítica y su guía experta que transformó un problema práctico en una contribución académica formal.*

*A la **Universidad Nacional Mayor de San Marcos (UNMSM)**, por ser la institución que me brindó la oportunidad para desarrollar esta investigación.*

*A la **Universidad Técnica de Manabí (UTM)**, por proporcionar el entorno institucional y los recursos de infraestructura necesarios que permitieron la Demostración y Evaluación in situ del artefacto en un entorno real.*

*A **mi equipo de trabajo**, quienes con su apoyo y colaboración aportaron al desarrollo del trabajo de investigación.*

Gracias...

ÍNDICE GENERAL

CAPÍTULO 1. INTRODUCCIÓN	1
1.1. Situación Problemática	2
1.2. Formulación del Problema.....	5
1.2.1. Problema General.....	5
1.2.2. Problemas Específicos	6
1.3. Justificación	6
1.3.1. Teórica	6
1.3.2. Práctica.....	7
1.4. Objetivos.....	10
1.4.1. Objetivo General.....	10
1.4.2. Objetivos Específicos.....	10
1.5. Hipótesis	10
1.5.1. Hipótesis General.....	11
1.5.2. Hipótesis específicas:.....	11
CAPÍTULO 2. MARCO TEÓRICO.....	12
2.1. Antecedentes de la Investigación.....	12
2.1.1. Fallas de precisión en la data IoT.	13
2.1.2. Sistemas de gestión energética en edificios inteligentes.....	15
2.1.3. Característica de los sistemas de gestión	15
2.1.4. Sistemas de monitoreo en tiempo real	16

2.1.5. Modelos de predicción de consumo de energía	17
2.1.6. Modelos predictivos sin variables exógenas clave	18
2.1.7. Modelos predictivos para gestionar consumo energético	19
2.1.8. Usabilidad/latencia en asistentes conversacionales genéricos	20
2.1.9. Integración de avatares en la gestión Smart Building.....	21
2.1.10. Integración de avatares con algoritmos de aprendizaje.....	22
2.2. Bases Teóricas	23
2.2.1. Relacionadas con la Gestión eficiente de edificios.....	23
2.2.2. Relacionados con el desarrollo del Avatar.....	29
2.2.3. Usabilidad y Experiencia de Usuario.....	35
CAPÍTULO 3. METODOLOGÍA.....	38
3.1. Tipo de investigación.....	39
3.2. Población y Muestra	39
3.2.1. Unidad de análisis	39
3.2.2. Población de estudio	39
3.2.3. Tamaño de muestra	39
3.2.4. Selección de muestra.....	39
3.3. Técnicas de recolección de datos.....	40
3.3.1. Módulo IoT y Calidad de Datos (Validación de HE1)	40
3.3.2. Módulo Predictivo y Avatar Conversacional (Validación HE2 y HE3).....	40
3.3.3. Ecosistema Integrado y Validación de Impacto (Validación HE4 y HG) ...	40
3.4. Técnicas de análisis de datos	41

3.4.1. Análisis de Calidad y Superioridad (Validación de HE1 y HE2).....	41
3.4.2. Análisis de Rendimiento y Experiencia (Validación de HE3 y HE4)	41
3.4.3. Análisis de Impacto Global (Validación de HG).....	42
3.5. Método Ciencia del Diseño en la Investigación	42
3.6. Matriz de consistencia	48
CAPÍTULO 4. CONSTRUCCIÓN DE LA SOLUCIÓN (FASE 3)	53
4.1. Diseño del Módulo IoT.....	53
4.1.1. Diseño del Hardware.....	54
4.1.2. Creación del Software.....	58
4.1.3. Protocolo de comunicación.....	59
4.2. Selección del Modelo de Predicción.....	59
4.2.1. Proceso Extracción, Transformación y Carga (ETL)	61
4.2.2. Entrenamiento del Modelo de Predicción ARIMA	63
4.2.3. Entrenamiento del Modelo de Predicción SARIMAX	68
4.3. Diseño y Desarrollo del Módulo HCI Conversacional.....	74
4.3.1. Creación del Avatar	74
4.3.2. Personalización del Avatar	76
4.3.3. Incorporación de Voz al Avatar	81
4.3.4. Ejecución de tarea con base en el mensaje del usuario.....	86
4.3.5. Extracción de parámetros de la consulta del usuario.	89
4.3.6. Recordar parámetros de mensajes anteriores.	91
4.3.7. Ejecución de consulta SQL en la base de datos	92

4.3.8. Respuestas finales del asistente	95
CAPÍTULO 5. DEMOSTRACIÓN (FASE 4)	97
5.1. Integración y Operatividad del Software con el módulo IoT	97
5.2. Precisión del Modelo de Predicción	99
5.3. Funcionalidad del Asistente Conversacional.....	102
5.3.1. Detección de intención.....	102
5.3.2. Transcripción de Voz a Texto (ASR)	103
5.3.3. Síntesis de Texto a Voz (TTS).....	104
5.4. Aplicación de encuestas de Usabilidad y Calidad Hedónica.....	104
5.4.1. Encuestas de Usabilidad	105
5.4.2. Calidad Hedónica (UEQ).....	105
5.5. Demostración de la Interfaz de Monitoreo del Rendimiento.....	108
5.5.1. Cuantificación de la Ganancia de Eficiencia Temporal.....	109
5.5.2. Respecto a la reducción del consumo de recursos.	109
CAPÍTULO 6. RESULTADOS (EVALUACIÓN -FASE 5).....	112
6.1. Rigor Técnico y Rendimiento Computacional Óptimo	112
6.1.1. Rigor Técnico del Módulo IoT (HE1)	112
6.1.2. Eficacia del Módulo Predictivo (HE2)	112
6.1.3. Viabilidad del Ecosistema (HE3 y HE4)	113
6.1.4. Rendimiento y Eficiencia Computacional (HE4):	113
6.2. Evaluación de la Usabilidad Superior y Experiencia de Usuario	114
6.2.1. Evaluación de la Usabilidad Subjetiva (SUS)	114

6.2.2. Evaluación de la Calidad Hedónica (UEQ)	114
6.3. Evaluación de Impacto del Artefacto	115
6.3.1. Optimización en la Gestión Energética.....	115
6.3.2. Contribución al Cuerpo de Conocimiento	117
CAPÍTULO 7. COMUNICACIÓN - FASE 6	118
7.1. Contribución al Cuerpo de Conocimiento (Impacto Teórico).....	118
7.1.1. Modelo de Diseño Híbrido y Sinérgico	118
7.1.2. Principios de Diseño para la Eficiencia Computacional.....	118
7.1.3. Evidencia de Usabilidad para el Cambio Conductual.....	119
7.2. Transferencia a la Comunidad Profesional (Impacto Práctico).....	119
CAPÍTULO 8. CONCLUSIONES, RECOMENDACIONES Y LIMITACIONES DEL ESTUDIO.....	122
8.1. Conclusiones.....	122
8.2. Recomendaciones	123
8.3. Limitaciones	123
REFERENCIAS BIBLIOGRÁFICAS.....	125

LISTA DE FIGURAS

<i>Figura 1. Tendencia de edificaciones en Ecuador</i>	4
<i>Figura 2. Características claves en la gestión eficiente de edificios</i>	16
<i>Figura 3 Adaptación del Modelo del proceso de investigación en diseño (Design Science Research Process Model)</i>	38
<i>Figura 4 Esquema de módulos a desarrollar</i>	45
<i>Figura 5 Implementación de un algoritmo de predicción de consumo energético</i>	46
<i>Figura 6 Esquema propuesto del Módulo IoT</i>	55
<i>Figura 7 Esquema del circuito de la placa Arduino</i>	56
<i>Figura 8 Esquema 3D del circuito de la placa Arduino</i>	56
<i>Figura 9 Placa física fabricada de forma artesanal</i>	57
<i>Figura 10 Programación en el IDE Arduino</i>	57
<i>Figura 11 Data luego de aplicar el proceso de ETL</i>	62
<i>Figura 12 Entrenamiento del Modelo ARIMA</i>	63
<i>Figura 13 Descomposición de la serie temporal</i>	64
<i>Figura 14 Prueba de Dickey Fuller para obtener p-valor en ARIMA</i>	64
<i>Figura 15 Función de autocorrelación ACF para determinar valor de q con ARIMA</i>	65
<i>Figura 16 Función de autocorrelación parcial PACF para determinar valor de p con ARIMA</i>	65
<i>Figura 17 Obtención del mejor modelo para ARIMA</i>	66
<i>Figura 18 Verificación de existencia de autocorrelación de los residuos</i>	66
<i>Figura 19 Validación de distribución normal de los residuos</i>	67
<i>Figura 20 Resultado de la predicción ARIMA</i>	67
<i>Figura 21 Pasos para entrenamiento del Modelo SARIMAX</i>	68

<i>Figura 22 Data divida para entrenamiento</i>	69
<i>Figura 23 Selección de dato para entrenamiento SARIMAX</i>	70
<i>Figura 24 Mejor modelo encontrado para SARIMAX</i>	70
<i>Figura 25 Entrenamiento del modelo SARIMAX</i>	71
<i>Figura 26 Prueba de Ljung-Box SARIMAX</i>	71
<i>Figura 27 Prueba de Shapiro-Wilk SARIMAX</i>	71
<i>Figura 28 Respuesta del LLM con formato definido</i>	72
<i>Figura 29 Dataframe para la respuesta del LLM</i>	72
<i>Figura 30 Dataframe de variables exógenas en fechas de predicción</i>	73
<i>Figura 31 Gráfico de los datos de consumo predicho</i>	74
<i>Figura 32 Procedimiento de diseño propuesto para la comunicación avatar - usuario</i>	75
<i>Figura 33 Herramientas utilizadas</i>	76
<i>Figura 34 Avatar v1.0, creado utilizando Mixamo y Unity</i>	77
<i>Figura 35 Proceso de Conversión de texto a voz</i>	83
<i>Figura 36 Evaluación de clasificadores</i>	85
<i>Figura 37 Extracción de entidades</i>	85
<i>Figura 38 Ejemplo de llamadas a funciones de GPT</i>	86
<i>Figura 39 Detección de intención con toolcalls de GPT</i>	87
<i>Figura 40 Llamadas a funciones de Qwen2.5</i>	87
<i>Figura 41 Declaración del modelo de detección de intención</i>	88
<i>Figura 42 Resultados de detección de la intención del usuario con el modelo BERT</i>	88
<i>Figura 43 Resultados de detección de la intención con el LLM de Mistral</i>	89
<i>Figura 44 Construcción del modelo en la API de asistentes de OpenAI</i>	90

<i>Figura 45 Formato para obtener parámetros por expresiones regulares</i>	91
<i>Figura 46 Resultado del proceso de extracción de entidades</i>	91
<i>Figura 47 Resultado del proceso de recordar</i>	92
<i>Figura 48 Respuesta con reemplazo de los nombres por los identificadores por Mistral</i>	93
<i>Figura 49 Resultado y tiempo de respuesta de Codellama</i>	93
<i>Figura 50 Resultado y tiempo de respuesta de Mistral</i>	94
<i>Figura 51 Respuesta del LLM sobre el consumo energético</i>	95
<i>Figura 52 Respuesta del LLM sobre la presentación del asistente</i>	95
<i>Figura 53 Respuesta del LLM sobre errores de usuario</i>	95
<i>Figura 54 Respuesta del LLM sobre respuestas a preguntas generales</i>	95
<i>Figura 55 Puesta en marcha de sensores in situ</i>	97
<i>Figura 56 Prototipo de medición manual con pantalla incorporada</i>	98
<i>Figura 57 Prueba de funcionalidad de sensores</i>	98
<i>Figura 58 Dashboard del consumo energético del edificio a manera general</i>	99
<i>Figura 59 Dashboard de datos como humedad, temperatura</i>	99
<i>Figura 60 Comparación de Error Absoluto Medio (MAE)</i>	101
<i>Figura 61 Benchmarking</i>	107
<i>Figura 62 Dashboard Rendimiento de CPU</i>	108
<i>Figura 63 Dashboard Rendimiento de GPU</i>	108
<i>Figura 64 Arquitectura del Ecosistema Integrado</i>	115
<i>Figura 65 Interfaz final del ecosistema</i>	116
<i>Figura 66 Publicación realizada en revista Encriptar</i>	120
<i>Figura 67 Difusión con la comunidad académica en el evento INNOVATECH 2025, Portoviejo, Ecuador</i>	121

LISTA DE TABLAS

<i>Tabla 1 Solución propuesta para la investigación</i>	<i>42</i>
<i>Tabla 2 Consistencia entre problema – Objetivos y Resultado Esperados</i>	<i>48</i>
<i>Tabla 3 Análisis comparativo de modelos de predicción analizados</i>	<i>60</i>
<i>Tabla 4 Cuadro Comparativo entre LLaMA y GPT</i>	<i>78</i>
<i>Tabla 5 Comparativa de Modelos de transcripción de voz a texto</i>	<i>81</i>
<i>Tabla 6 Comparativa de Modelos de transcripción de texto a voz</i>	<i>83</i>
<i>Tabla 7 Tabla comparativa de métricas aplicadas</i>	<i>100</i>
<i>Tabla 8 Validación estadística de algoritmos de predicción</i>	<i>102</i>
<i>Tabla 9 Comparación Técnicas de Detección de Intención</i>	<i>102</i>
<i>Tabla 10 Comparativa Modelos de transcripción de Voz a Texto</i>	<i>103</i>
<i>Tabla 11 Comparativa Modelos de transcripción de Texto a Voz</i>	<i>104</i>
<i>Tabla 12 Calidad Hedónica</i>	<i>106</i>
<i>Tabla 13 Intervalo de confianza de la Calidad Hedónica por ítem</i>	<i>107</i>
<i>Tabla 14 Comparativa métricas de rendimiento</i>	<i>109</i>
<i>Tabla 15 Optimización en la CPU (VITS)</i>	<i>110</i>
<i>Tabla 16 Optimización en la GPU (XTTS)</i>	<i>111</i>
<i>Tabla 17 Métricas SARIMAX vs ARIMA</i>	<i>113</i>

RESUMEN

La investigación abordó la ineficiencia en la gestión energética de edificios inteligentes, identificando la brecha crítica entre la precisión de la monitorización técnica y la falta de compromiso conductual proactivo del usuario final, justificando así la necesidad de un sistema que integre funcionalidad y engagement. El Método se basó rigurosamente en la Ciencia del Diseño (DSR), enfocándose en la construcción y evaluación formal de un Ecosistema Inteligente híbrido que fusiona la monitorización IoT en tiempo real, la predicción basada en IA (SARIMAX) y una interfaz de avatar conversacional. Los Resultados validaron la Hipótesis General en tres pilares: Rigor Analítico, donde el modelo SARIMAX demostró un error de predicción MAE significativamente menor que ARIMA ($<38.7\%$) al integrar variables exógenas, con residuos que satisficieron la condición de ruido blanco ($p\text{-value} \approx 0.99$); Rendimiento Computacional Óptimo, alcanzando una latencia de interacción inferior a 10 segundos mediante la optimización de recursos (INT8/FP16) de los modelos de IA local, lo cual es crítico para la viabilidad en tiempo real; y Usabilidad Superior, evidenciada por un puntaje SUS promedio de 98.9 ("Mejor Imaginable") y una alta Calidad Hedónica ($HQ=0.889$), validando el engagement necesario para la adopción proactiva del usuario. Se concluye que la arquitectura propuesta no solo cumple con los requisitos de eficiencia computacional y usabilidad, sino que establece un modelo de diseño híbrido que produce una mejora significativa en la gestión energética del edificio, contribuyendo al conocimiento sobre la integración de la IA conversacional en infraestructuras críticas.

Palabras claves: *Ciencia del Diseño (DSR), Gestión Energética Inteligente, SARIMAX, Inteligencia Artificial Conversacional, Usabilidad.*

ABSTRACT

The research addressed inefficiency in the energy management of smart buildings, identifying the critical gap between the accuracy of technical monitoring and the lack of proactive behavioral engagement from end users, thus justifying the need for a system that integrates functionality and engagement. The Method was rigorously based on Design Science Research (DSR), focusing on the construction and formal evaluation of a hybrid Intelligent Ecosystem that combines real-time IoT monitoring, AI-based prediction (SARIMAX), and a conversational avatar interface. The Results validated the General Hypothesis in three pillars: Analytical Rigor, where the SARIMAX model demonstrated a significantly lower prediction error (MAE) than ARIMA (38.7%) by integrating exogenous variables, with residuals that met the white noise condition (p-value 0.99); Optimal Computational Performance, achieving interaction latency below 10 seconds through resource optimization (INT8/FP16) of local AI models, which is critical for real-time viability; and Superior Usability, evidenced by an average SUS score of 98.9 (“Best Imaginable”) and high Hedonic Quality (HQ=0.889), validating the engagement needed for proactive user adoption. It is concluded that the proposed architecture not only meets the requirements of computational efficiency and usability but also establishes a hybrid design model that produces a significant improvement in building energy management, contributing to the body of knowledge on the integration of conversational AI in critical infrastructure.

Keyword: *Design Science (DSR), Intelligent Energy Management, SARIMAX, Conversational Artificial Intelligence, Usability.*

CAPÍTULO 1. INTRODUCCIÓN

Este capítulo establece los fundamentos epistemológicos y la metodología de la tesis doctoral, cuyo objetivo es desarrollar un artefacto tecnológico validado para mejorar la gestión energética en edificios inteligentes. Este enfoque se alinea con la metodología de Ciencia del Diseño para la Investigación (DSR), que requiere una definición precisa de una brecha en el conocimiento práctico y la propuesta de una solución innovadora.

El rápido crecimiento de los dispositivos IoT y la incorporación de la Inteligencia Artificial en el entorno construido, aunque prometen avances significativos, han generado una paradoja: una mayor conectividad no garantiza automáticamente una mayor eficiencia o facilidad de uso. Persisten problemas críticos en la precisión de la recopilación de datos (IoT), limitaciones en el rigor estadístico de las predicciones, y una alta latencia en las interfaces humano-máquina que impiden una gestión energética proactiva y centrada en el usuario.

Este capítulo se estructura de la siguiente manera: Primero, se expone la situación problemática mediante un análisis del contexto global y regional (Ecuador) del consumo energético y la expansión de los edificios inteligentes. La investigación se justifica al confrontar el avance tecnológico con las deficiencias persistentes en la confiabilidad predictiva y la usabilidad de las interfaces conversacionales. Luego, se desglosa la problemática general en una serie de Problemas Específicos (PEs) interconectados, que demuestran la necesidad de una arquitectura sinérgica que aborde tanto la validez técnica (PE2) como la viabilidad computacional (PE4) y la accesibilidad (PE3). Posteriormente, se presenta la relevancia teórica, práctica y técnica del proyecto, destacando su alineación con los Objetivos de Desarrollo Sostenible (ODS) y la contribución de la investigación a la solución de la crisis de latencia y usabilidad en sistemas híbridos.

Basado en la problemática, se presenta el Objetivo General como la meta final del artefacto: “Diseñar y validar un Ecosistema Inteligente que, mediante la integración sinérgica de monitorización IoT en tiempo real, predicción de consumo basada en IA (SARIMAX) y una interfaz de avatar conversacional, demuestre una mejora significativa en la gestión energética del edificio, garantizando una usabilidad

superior y un rendimiento computacional óptimo”, seguido de los Objetivos Específicos, que trazan las etapas de construcción modular (IoT, IA, Avatar) y la evaluación de rendimiento (Optimización).

Por consiguiente, se expone la Hipótesis General, que es la proposición de valor del artefacto (mejora significativa en eficiencia y usabilidad), y las Hipótesis Específicas (HE1-HE4), que son los criterios medibles utilizados en la Fase de Evaluación para confirmar el éxito de cada componente de diseño.

Con este antecedente, este primer capítulo sienta las bases para el desarrollo del marco teórico y la arquitectura de diseño, asegurando la coherencia vertical entre el problema identificado y la solución propuesta.

1.1. Situación Problemática

A nivel mundial, los edificios inteligentes han tenido un aumento significativo en los últimos años, con una tendencia proyectada de un total de 115 millones de edificaciones para el 2026 en comparación con los 45 millones existentes en el 2022 (Juniper Research, 2025). Una tendencia similar se observa al analizar el total de dispositivos IoT conectados en edificios inteligentes pasando de 1.500 millones en el 2022 a un aproximado de 14700 millones para el 2030 (Memoori, 2025).

En este sentido, se puede definir como edificio inteligente a aquella infraestructura que utiliza conectividad para optimizar los recursos instalados y que a su vez proporcionan un ambiente seguro y cómodo para sus ocupantes (Jørgensen & Ma, 2025; Sleem & Elhenawy, 2023). Estos edificios inteligentes interconectan diferentes dispositivos IoT como son los equipos de acceso, seguridad, sensores de temperatura, cámaras, entre los más representativos (Xie & Leung, 2021). No obstante, el consumo energético generado por este tipo de edificaciones supera el 30% en comparación con las residencias que no cuentan con dispositivos IoT.

Las innovaciones en las tecnologías de la información y la comunicación (TIC) han propiciado una transformación digital, resultando en una sociedad profundamente interconectada con la tecnología (Weiss, 2019). Estas tecnologías están transformando profundamente diversos aspectos de la sociedad, impulsando cambios significativos en nuestra forma de vivir, trabajar, comunicarnos e interactuar.

En este sentido, tecnologías como la inteligencia artificial (IA), el machine learning (ML), la robótica y las interfaces hombre-máquina están transformando el mercado laboral, las industrias y la productividad (Schaffers et al., 2020). La integración de la IA con los sistemas IoT permite habilitar funciones avanzadas e innovadoras, esto conlleva a que el mercado de soluciones de IA en edificios inteligentes crezca a una tasa anual compuesta (CAGR) del 25,5 % con una tendencia a alcanzar los 6500 millones de dólares en 2028 (Inteligencia artificial del mercado de las cosas, 2022).

Uno de los aspectos por considerar en la gestión y conservación de energía en edificios es la predicción del consumo energético, que permita reducir de manera significativa el consumo energético (Memoori, 2025). En este sentido, los algoritmos de ML y la IA optimizan el consumo de energía en edificios inteligentes al permitir el análisis predictivo, los controles automatizados, la detección de anomalías y el aprendizaje adaptativo, lo que resulta en mejoras significativas en eficiencia, sostenibilidad y ahorro de costos operativos (Anan et al., 2024; Khanwalkar & Ravikanti, 2025). No obstante, a pesar de la disponibilidad de tecnologías avanzadas de monitoreo y predicción, persiste una limitación crítica en la capacidad de los edificios inteligentes para llevar a cabo una gestión energética proactiva, precisa y centrada en el usuario (Rodríguez-Gracia et al., 2023).

Los agentes conversacionales (AC) frecuentemente enfrentan dificultades para ofrecer respuestas precisas debido a limitaciones en la información disponible o en sus capacidades, lo que a menudo resulta en respuestas repetitivas como "No lo sé" (Lee & Lee, 2021). Esta situación puede inducir frustración y desilusión entre los usuarios (Brendel et al., 2020; Heuer et al., 2023). Los errores en la comprensión y el reconocimiento del habla son comunes, lo que lleva a respuestas incorrectas o irrelevantes (Choi et al., 2024). La falta de fluidez en el diálogo y la incapacidad de los AC para gestionar conversaciones de manera natural pueden interrumpir la interacción, reducir la satisfacción del usuario y resultar en una experiencia negativa (Jentzsch et al., 2019).

En América Latina, la rápida urbanización, el crecimiento económico y las mejoras en la calidad de vida están impulsando un aumento significativo del consumo de energía (Viera-Romero, 2025). Ecuador es uno de los países que ha experimentado

un aumento significativo en el consumo de energía durante las últimas décadas, lo que ejerce una mayor presión sobre la infraestructura existente y aumenta el riesgo de escasez, esto debido a que el sistema energético depende en gran medida de la generación hidroeléctrica, que representó el 67,4 % de la matriz energética nacional en 2024 (Pinzón, 2018). En este contexto, Vidal et al indican que la demanda de energía puede disminuir entre un 0,5% y un 21,9% con la instalación de controles inteligentes (2025).

El número de edificaciones (Figura 1) en Ecuador ha pasado de 30.896 en 2023 a 32.253 en 2024 lo que representa un crecimiento aproximado del 4,4%, aunque no se especifica si estas construcciones son o no inteligentes (INEC, 2025). Este crecimiento resalta la necesidad de incorporar tecnologías avanzadas de gestión energética para optimizar los recursos en la infraestructura emergente, pues si bien las tecnologías de edificios inteligentes ofrecen notables beneficios ambientales y de eficiencia, persisten en todas las regiones elevados costos iniciales, complejidad técnica y desafíos de interoperabilidad (Montúfar, 2025).

Figura 1. *Tendencia de edificaciones en Ecuador*



Fuente: www.ecuadorencifras.gob.ec

La diversidad de tipos de construcción y los patrones de uso cambiantes en hogares, oficinas y escuelas representan un reto significativo para la infraestructura de ciudades inteligentes. Los sensores inteligentes enfrentan varios desafíos, como largos procedimientos de calibración, toma de decisiones preprogramadas fijas y la necesidad de grandes conjuntos de datos etiquetados. Estos retos logran un equilibrio entre la precisión del sistema, su capacidad de ampliación y su adaptabilidad, aspectos fundamentales para la infraestructura IoT en la construcción inteligente (Oliveira et al., 2024).

En este contexto, la adopción de tecnologías IoT en el ámbito de la construcción todavía se encuentra en una fase inicial, lo que subraya la importancia de llevar a cabo investigaciones en los campos relevantes, especialmente desde una perspectiva de aplicación práctica (Herath & Mittal, 2022). Además, estas tecnologías aún no abordan completamente los desafíos de escalabilidad, adaptabilidad y seguridad, lo que requiere una mayor integración y mejoras en la interoperabilidad entre tecnologías existentes y futuras (Mazhar et al., 2022).

1.2. Formulación del Problema

Con base en los antecedentes y la problemática analizada se evidencian limitaciones respecto a la gestión eficiente del consumo energético en edificios inteligentes, se formulan el problema general y los problemas específicos de esta investigación.

1.2.1. Problema General

El incremento constante en el consumo energético de los edificios inteligentes, impulsado por el uso intensivo de dispositivos eléctricos plantea un desafío crítico en términos de sostenibilidad, eficiencia y costos. A pesar de los avances tecnológicos, muchos de estos edificios carecen de herramientas integrales que permitan monitorizar en tiempo real, predecir patrones de consumo y proporcionar recomendaciones personalizadas a los usuarios. Esta carencia limita la capacidad de optimizar el uso de la energía, aumenta los gastos operativos y reduce el impacto positivo esperado en la gestión sostenible de los recursos.

1.2.2. Problemas Específicos

PE1. Ausencia de un ecosistema que permita monitorear, predecir y visualizar en tiempo real de manera precisa y continua parámetros ambientales (temperatura, humedad) dentro de edificios inteligentes.

PE2. Limitación en la confiabilidad y efectividad predictiva de los métodos actuales de pronóstico del consumo energético, debido a la insuficiente especificación estadística de los modelos utilizados y la omisión de variables exógenas relevantes lo que impide la obtención de un modelo robusto cuyos residuos se comporten como ruido blanco, comprometiendo la validez de las predicciones a corto plazo.

PE3. Existe una limitación técnica y de rendimiento en los diseños actuales de asistentes conversacionales que impide la creación de un avatar con interacción fluida en tiempo real y alta fiabilidad en la detección de comandos por la latencia y la baja calidad del lenguaje natural.

PE4. La integración de modelos de IA local complejos (LLMs, TTS/ASR) en el prototipo híbrido IoT impone una carga computacional excesiva y latencia. Esto compromete el rendimiento óptimo y la escalabilidad del sistema, haciendo indispensable la optimización para asegurar una gestión energética eficaz.

1.3. Justificación

La justificación de esta investigación se fundamenta en la urgencia de resolver las brechas de eficiencia y usabilidad identificadas en la gestión energética de los edificios inteligentes. Este apartado establece la relevancia multi-dimensional del diseño y la validación del ecosistema propuesto, abordando tanto su trascendencia en el ámbito académico como su impacto concreto en el sector. La necesidad de esta investigación se articula en dos dimensiones clave: teórica y práctica.

1.3.1. Teórica

El Objetivo de Desarrollo Sostenible (ODS) número 9 subraya la importancia de desarrollar infraestructuras resilientes, sostenibles e innovadoras, y los edificios inteligentes se alinean con estos principios al incorporar tecnologías avanzadas que optimizan la eficiencia energética y reducen su impacto ambiental (Montúfar, 2025).

El modelado de información de construcción (BIM), como parte de la era de la industria 4.0, se ha integrado en los procesos de diseño, construcción y operación de edificios, promoviendo la sostenibilidad. Sin embargo, la implementación de aplicaciones IoT en estos edificios aún enfrenta desafíos relacionados con la interoperabilidad y la estandarización (Romero Macías & Loo Navia, 2025; Veloz Zambrano et al., 2025).

Además, la electrificación del sector de la construcción, que aspira a alcanzar una electricidad completamente limpia para el año 2035, requiere la integración de tecnologías innovadoras que reduzcan el consumo energético en los edificios. La implementación del Internet de las Cosas (IoT) ha demostrado ser una solución eficaz para mejorar la gestión energética; sin embargo, aún se carece de soluciones confiables, escalables y económicas para el diagnóstico y mantenimiento basados en IoT (Sun et al., 2022). Los edificios inteligentes, al optimizar la gestión de energía, impactan directamente en la sostenibilidad y la seguridad, pero también presentan desafíos en términos de ciberseguridad. Los sistemas automatizados pueden ser vulnerables a ataques que comprometan la red de IoT, lo que subraya la necesidad de implementar soluciones de seguridad avanzadas basadas en aprendizaje automático para proteger tanto los recursos de la red como la privacidad de los datos (Gajewski et al., 2022).

Por lo expuesto, se demuestra cómo la construcción del artefacto contribuye al avance del conocimiento en la convergencia de la Industria 4.0 (IoT, IA, Machine Learning) y la Sostenibilidad (ODS 9, electrificación 100% limpia), proporcionando soluciones a los desafíos persistentes de interoperabilidad, estandarización y la falta de arquitecturas confiables y escalables para el diagnóstico energético.

1.3.2. Práctica

La investigación propuesta tiene como objetivo aumentar la eficiencia del consumo de energía en edificios inteligentes, lo que resulta en un impacto económico considerable al disminuir los costos operativos de edificios tanto públicos (como hospitales, universidades, instituciones gubernamentales, etc.) como privados (como hoteles, edificios residenciales, etc.). Además, facilitará la optimización del uso de recursos mediante el monitoreo constante y la personalización del consumo,

promoviendo una mayor conciencia energética entre los usuarios.

En el sector de la construcción, la investigación ofrece una perspectiva crucial para desarrollar edificios más sostenibles, integrando tecnologías IoT y aprendizaje automático para mejorar la eficiencia energética y reducir los riesgos asociados al uso de recursos. La implementación de estos sistemas permitirá a las empresas del sector crear infraestructuras más resilientes, seguras y eficientes, alineadas con los objetivos globales de sostenibilidad y disminución de la huella de carbono. Desde un enfoque técnico, los sistemas de gestión energética en edificios inteligentes generan datos complejos sobre el consumo, los patrones de uso y las variaciones horarias, lo que dificulta su interpretación para los usuarios no especializados.

En este contexto, el avatar conversacional actúa como un traductor visual e interactivo, capaz de presentar estos datos en un lenguaje accesible y comprensible. Además, incluye elementos de motivación y retroalimentación positiva, promoviendo la adopción de comportamientos sostenibles y fortaleciendo el compromiso de los ocupantes con el ahorro energético. Más allá de su función informativa, el avatar puede integrarse con bases de datos y sensores IoT, lo que le permite generar respuestas inmediatas y personalizadas. Asimismo, mediante técnicas de aprendizaje automático, puede identificar patrones individuales de consumo y anticipar necesidades específicas, optimizando la eficiencia del sistema.

A diferencia de los paneles técnicos convencionales, el avatar puede diseñarse con criterios de accesibilidad cognitiva y visual, convirtiéndose en un asistente inclusivo, adaptable y orientado a mejorar la experiencia de los usuarios en la gestión energética. Las interfaces conversacionales, especialmente aquellas basadas en modelos de lenguaje avanzados permiten a los operadores de edificios interactuar con sistemas complejos de gestión energética de manera más intuitiva. En lugar de navegar por múltiples paneles de control, los operadores pueden utilizar el lenguaje natural para acceder a la información y controlar los sistemas, lo que agiliza las operaciones y reduce la curva de aprendizaje para los usuarios (Zhang & Chen, 2025).

Estas interfaces pueden proporcionar información en tiempo real y consejos personalizados a los ocupantes sobre su consumo energético, ayudándoles a tomar decisiones informadas para ahorrar energía. Al concienciar a los usuarios sobre las

consecuencias de sus acciones, los asistentes conversacionales fomentan comportamientos de ahorro energético (Bavaresco et al., 2019; Lagarde et al., 2024; Zanjani et al., 2015).

Los asistentes conversacionales pueden servir de puente entre los usuarios y los sistemas automatizados de los edificios, permitiendo a los ocupantes controlar la iluminación, la temperatura y otros dispositivos que consumen energía mediante sencillos comandos de voz o texto. Esta integración mejora la interacción del usuario y permite una gestión energética más dinámica y eficaz (Bavaresco et al., 2019). Al aprovechar dispositivos familiares como teléfonos inteligentes, tabletas y asistentes de voz personales, estas interfaces facilitan que los ocupantes participen en iniciativas de ahorro de energía, informen sobre problemas o soliciten cambios, fomentando así una cultura de concienciación energética y una gestión proactiva (Bavaresco et al., 2019; Zanjani et al., 2015).

Las arquitecturas híbridas de sistemas inteligentes, que incorporan IoT, pronósticos y aprendizaje automático conversacional, están diseñadas para ser funcionales; sin embargo, enfrentan un desafío inherente al intentar asegurar la eficiencia computacional y la baja latencia necesarias para operar en tiempo real en entornos con recursos limitados (Carmean et al., 2023; He et al., 2025). La integración sinérgica de modelos de IA complejos a nivel local, como LLMs y TTS/ASR, esenciales para la interfaz conversacional, genera una carga de procesamiento (CPU/GPU) y una latencia que comprometen la viabilidad operativa, la escalabilidad y la sostenibilidad económica del prototipo (Carmean et al., 2023). Por consiguiente, el desafío radica en identificar y validar estrategias de optimización de hardware y software que permitan reducir significativamente el consumo de recursos y la latencia del prototipo completo, garantizando una experiencia de usuario fluida y la operación continua del sistema de gestión energética.

En este sentido, se justifica el valor funcional y económico del ecosistema, abordando la reducción de costos operativos en edificios, el fomento de la conciencia energética mediante la personalización del consumo y, crucialmente, la solución a la complejidad técnica de los datos. Esta solución se materializa a través del avatar conversacional, que actúa como un traductor accesible e inclusivo, superando las limitaciones de los paneles técnicos convencionales

1.4. Objetivos

El presente estudio enmarcado en la metodología de Ciencia del Diseño para la Investigación (DSR), cuyo propósito principal es la creación y validación de un artefacto tecnológico que aborde una brecha en el rendimiento y la usabilidad en la gestión energética, presenta un objetivo general como la meta final en la construcción del artefacto, y los objetivos específicos que detallan las etapas necesarias para alcanzar dicha meta y validar las hipótesis asociadas.

1.4.1. Objetivo General

Diseñar y validar un Ecosistema Inteligente que, mediante la integración sinérgica de monitorización IoT en tiempo real, predicción de consumo basada en IA (SARIMAX) y una interfaz de avatar conversacional, demuestre una mejora significativa en la gestión energética del edificio, garantizando una usabilidad superior y un rendimiento computacional óptimo.

1.4.2. Objetivos Específicos

- OE1: Desarrollar el módulo IoT para la captura, procesamiento y almacenamiento de parámetros ambientales (temperatura, humedad) y del consumo eléctrico (KW y amperaje) en el edificio.
- OE2: Implementar un modelo predictivo del consumo energético utilizando los datos históricos y variables exógenas para estimar la demanda futura.
- OE3: Desarrollar un avatar conversacional que permita la interacción en tiempo real garantizando precisión y comunicación fluida.
- OE4: Optimizar el cálculo computacional para el funcionamiento del prototipo integral.

1.5. Hipótesis

Las hipótesis de este estudio representan las proposiciones formales y comprobables de la solución de diseño sugerida. Dentro del ámbito de la Ciencia del Diseño (DSR), estas hipótesis definen los criterios de éxito del artefacto creado y su

impacto previsto. La Hipótesis General expresa el valor superior del ecosistema integrado, mientras que las Hipótesis Específicas (HE1 a HE4) se enfocan en validar el rendimiento técnico, el rigor estadístico y la viabilidad computacional de cada módulo del artefacto, garantizando que la arquitectura final cumpla con los estándares de precisión y eficiencia exigidos.

1.5.1. Hipótesis General

La arquitectura tecnológica propuesta, resultante de la integración sinérgica de la monitorización IoT en tiempo real, la predicción de consumo basada en IA (SARIMAX) y la interfaz de avatar conversacional, demuestra una mejora significativa en la gestión energética del edificio y una usabilidad superior en la interacción comparada con métodos tradicionales, operando bajo condiciones de rendimiento computacional óptimas (demostrado por la baja latencia y el uso eficiente de recursos CPU/GPU).

1.5.2. Hipótesis específicas:

- HE1: El módulo IoT proporciona datos ambientales y de consumo energético con precisión y continuidad.
- HE2: El modelo SARIMAX ofrece una predicción del consumo de energía más precisa que ARIMA al incluir variables externas, logrando un error de predicción (evaluado mediante RMSE o MAPE) considerablemente menor y mostrando residuos que cumplen con la condición de ruido blanco.
- HE3: El diseño técnico del avatar conversacional satisface los requisitos de un sistema en tiempo real, asegurando una precisión en la detección de intenciones superior al 95% y tiempos de respuesta menores a 10 segundos en la cadena de interacción.
- HE4: La implementación de técnicas de optimización resultará en una mejora notable de la eficiencia computacional del sistema, evidenciada por una reducción medible en el consumo de recursos (CPU y GPU) y la latencia.

CAPÍTULO 2. MARCO TEÓRICO

El desarrollo de edificios inteligentes se ha consolidado como una de las estrategias más relevantes para enfrentar los desafíos de eficiencia energética y sostenibilidad en el siglo XXI. Estos espacios, equipados con tecnologías de monitoreo, automatización e inteligencia artificial, buscan optimizar el uso de los recursos energéticos sin comprometer el confort ni la funcionalidad de los usuarios. No obstante, la complejidad de los sistemas que integran y la creciente demanda de energía plantean la necesidad de contar con herramientas que permitan no solo medir y gestionar el consumo en tiempo real, sino también anticipar su comportamiento futuro y orientar decisiones más sostenibles.

En este contexto, el marco teórico se establece como la base conceptual que sustenta el diseño del ecosistema propuesto para la supervisión, predicción y visualización del consumo energético en edificios inteligentes. A lo largo de este capítulo, se examinan los principales antecedentes y fundamentos teóricos relacionados con los edificios inteligentes y la eficiencia energética, el monitoreo en tiempo real y la arquitectura IoT, la predicción de series temporales, la interacción conversacional, así como la usabilidad e integración de sistemas.

De igual manera, se revisan investigaciones y desarrollos previos que evidencian los avances y limitaciones actuales en la gestión energética de edificios inteligentes. Este análisis permitirá fundamentar la pertinencia de integrar tecnologías emergentes en un ecosistema único que combine la recolección de datos, la predicción y la interacción con el usuario, con el fin de alcanzar una optimización tangible en la gestión energética.

2.1. Antecedentes de la Investigación

Este apartado, tiene como propósito fundamental contextualizar el estudio actual mediante la revisión y el análisis crítico de investigaciones previas que abordan temáticas relacionadas, esta revisión exhaustiva no solo permite establecer el estado del arte del problema de investigación, sino que también facilita la identificación de los vacíos de conocimiento que nuestra investigación se propone complementar. Al examinar los enfoques metodológicos, los principales hallazgos y las conclusiones de los trabajos académicos y científicos más relevantes, se sienta una base teórica y

empírica sólida, justificando la necesidad y la pertinencia de este estudio

2.1.1. Fallas de precisión en la data IoT.

La recopilación de datos sobre el consumo de energía a partir de dispositivos IoT puede enfrentar diversas imprecisiones. Para asegurar que los registros de consumo energético en los sistemas IoT sean precisos y confiables, es fundamental abordar cuidadosamente varios problemas críticos, entre ellos:

Un diseño deficiente de los sensores puede provocar errores en la captura de datos. Por ejemplo, un sistema destinado a medir niveles de nutrientes en el suelo, como nitrógeno, fósforo y potasio, podría presentar discrepancias en la precisión, especialmente en ciertos elementos (Hartono et al., 2024). Es esencial que los dispositivos IoT utilicen la energía de manera eficiente. Un marco de detección de consumo de energía ultra bajo, como el de los medidores de gas inteligente, es necesario para mejorar la precisión y la fiabilidad de las lecturas (Z. Wang et al., 2022). La elección de rutas eficientes para la transmisión de datos es crucial. Un protocolo de enrutamiento robusto basado en clústeres puede aumentar la eficiencia energética de la red y prolongar la vida útil de los dispositivos (Shafiq et al., 2021). La falta de seguridad adecuada puede resultar en alteraciones de los datos, comprometiendo la precisión de la información recibida de los dispositivos IoT (Murthy et al., 2024).

Por otro lado, Lagsaiar et al (2021) desarrolla una tesis sobre un sistema de construcción inteligente que se centra en la supervisión del consumo energético y las condiciones de confort en los edificios. El estudio investiga cómo la comodidad en las aulas influye en el aprendizaje de los estudiantes. La metodología se basa en una arquitectura de software IoT para la supervisión de edificios y la gestión de viviendas sociales. Se emplean dispositivos para medir la temperatura, la humedad y el CO2 en las aulas, evaluando cómo estos parámetros afectan la percepción de confort y aprendizaje de los estudiantes. Este trabajo demuestra cómo las soluciones basadas en IoT pueden mejorar la eficiencia en el uso de recursos.

Uno de los trabajos seleccionados, es el presentado por Konstantakopoulos et al (2022) en el que se realizó un estudio experimental de cinco meses llamado 'SmartSDH' (llevado a cabo en el espacio de oficina de 27 empleados) donde los usuarios informaron sus preferencias para el brillo de la iluminación superior, se

evaluó la viabilidad del mecanismo de VCG en el contexto del control de la construcción Smart y el efecto de SmartSDH utilizando métricas como la satisfacción del nivel de luz, la satisfacción de los incentivos y el consumo de energía. Después de las ocho semanas iniciales, se logró reducir el consumo de energía en un 35,22% durante un período de cinco meses, sin incentivar directamente el menor consumo de energía, lo que proporciona evidencia de que gran parte del consumo de energía actual puede ser un desperdicio.

Balaji et al (2023) crearon un modelo de predicción de energía basado en IoT (IoT-EP) que pueda estimar de manera confiable el consumo de energía de edificios inteligentes. Para esto, se realiza un caso de prueba del mundo real sobre predicciones de energía en una cuadrícula de electricidad local para probar la practicidad del enfoque. El modelo propuesto (IoT-EP) seleccionó las características significativas como neuronas de entrada, los datos predecibles se seleccionan como nodos de salida y un perceptrón de múltiples capas se construye junto con las características del algoritmo de red neuronal de convolución (CNN). El análisis del modelo IoT-EP propuesto tiene una mayor precisión del 90%, la correlación del 89% y la varianza del 16% en menos tiempo de entrenamiento de 29.2 s, y con una mayor velocidad de predicción de 396 (observación/seg). En comparación con los modelos existentes, los resultados mostraron que el modelo propuesto (IoT-EP) supera con un nivel satisfactorio de precisión en la predicción del consumo de energía en edificios inteligentes.

Kharbouch et al (2022) en su trabajo, introducen un marco de hardware en el circuito (HIL) para la implementación y la evaluación de los enfoques de control predictivo en edificios inteligentes. Se combinaron varias plataformas de IoT y Big Data junto con algoritmos de aprendizaje automático y programas de control predictivo del modelo basados en MATLAB para habilitar las simulaciones HIL. El rendimiento del control sobre el marco HIL se ha evaluado y comparado con una estrategia de encendido/apagado. Los resultados muestran la utilidad del enfoque propuesto y su efectividad para reducir el consumo de energía en aproximadamente un 16%.

2.1.2. Sistemas de gestión energética en edificios inteligentes

Los sistemas de gestión energética para edificios inteligentes (BEMS) emplean tecnologías y estrategias de control avanzadas con el fin de optimizar el uso de energía, reducir los costos operativos y promover la sostenibilidad ambiental (Aliero et al., 2022; Zhang, et al., 2024). Los edificios inteligentes integran controles eficientes y automatizados, sensores en red y software de análisis de datos para monitorear tanto los datos ambientales como los hábitos de consumo energético de los ocupantes, mejorando así el funcionamiento y el rendimiento energético de los edificios (Aliero et al., 2022).

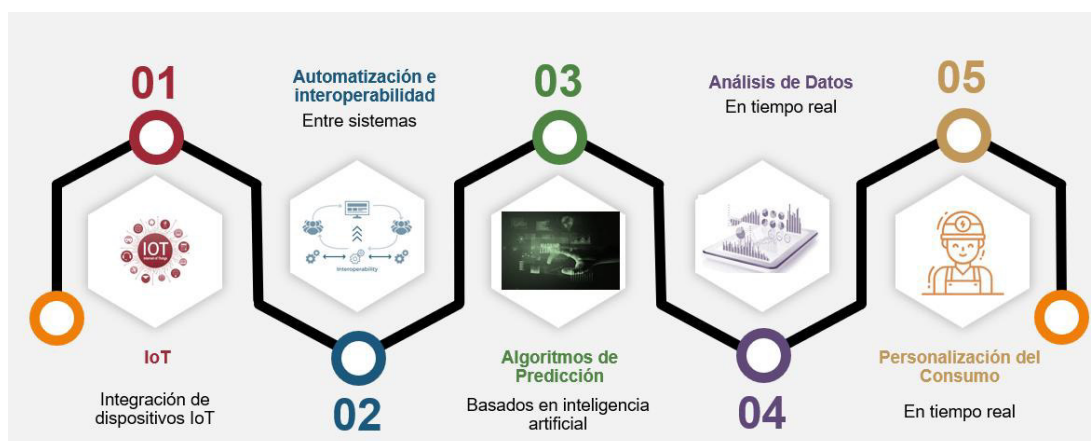
Los sistemas de gestión inteligente de energía pueden incrementar la eficiencia del consumo energético en los edificios al programar el funcionamiento de los electrodomésticos en función de la información disponible sobre energía y costos, lo que reduce los costos de uso de energía (Tito et al., 2021). Los sistemas de gestión de energía de edificios (BEMS) ofrecen una flexibilidad prometedora para la gestión de la demanda y la respuesta a la demanda (DSM&DR), interactuando con redes inteligentes para mejorar la eficiencia energética mientras se garantiza un nivel mínimo de confort para los usuarios del edificio (Hurtado et al., 2015). Los sistemas de gestión de energía (EMS) en edificios inteligentes optimizan el rendimiento de las actividades energéticas, especialmente la energía eléctrica, recuperando y explotando datos de los sistemas de gestión de edificios (BMS) y utilizando técnicas innovadoras para la generación de perfiles de carga (Mustapha et al., 2025).

2.1.3. Característica de los sistemas de gestión

Para gestionar de manera anticipada y eficiente los recursos en edificios inteligentes, es esencial considerar ciertas características clave, tales como: a) la integración de dispositivos IoT, b) la automatización e interoperabilidad, c) el uso de algoritmos de predicción basados en inteligencia artificial, d) el análisis de datos en tiempo real, y f) la personalización del consumo de energía.

En este estudio se abordan las características mencionadas en la Figura 2. Sin embargo, aunque no forman parte de estudio se destacan la sostenibilidad y reducción de la huella de carbono, la seguridad y protección de datos, así como la escalabilidad y flexibilidad de los sistemas, entre otros aspectos.

Figura 2. Características claves en la gestión eficiente de edificios



Fuente: Elaboración propia

2.1.4. Sistemas de monitoreo en tiempo real

Los sistemas de monitorización energética en tiempo real están diseñados para rastrear y visualizar el consumo energético en tiempo real, proporcionando a los usuarios información inmediata y datos históricos para su análisis. Estos sistemas están siendo cada vez más utilizados en diversos sectores debido a su capacidad para mejorar la eficiencia energética, reducir costos y fomentar prácticas sostenibles.

Los sistemas de monitoreo en tiempo real pueden disminuir significativamente los gastos de energía al identificar ineficiencias y permitir a los usuarios ajustar sus patrones de consumo. Al ofrecer información detallada sobre el uso de energía, estos sistemas reportan una reducción en los costos de energía de más del 25% durante el primer año, ayudando a los usuarios a identificar electrodomésticos que consumen mucha energía y el uso innecesario de energía, lo que lleva a un consumo de energía más eficiente (Devine & Joseph, 2012; Govindarajan et al., 2020).

El monitoreo en tiempo real promueve prácticas sostenibles al reducir el desperdicio de energía y las emisiones de carbono asociadas. Esto contribuye a un entorno más ecológico y ayuda a combatir el cambio climático (Neeraj & Gupta, 2024). Las empresas e industrias emplean la monitorización en tiempo real para optimizar el consumo energético, mejorar la eficiencia operativa y reducir costos. Estos sistemas también pueden integrarse con la infraestructura existente para una implementación sin problemas (New et al., 2019).

Por otro lado, las universidades y colegios están adoptando sistemas de monitoreo en tiempo real para rastrear el uso de energía en los edificios del campus, lo que les permite desarrollar estrategias para ahorrar energía y reducir los costos operativos (Dahlan et al., 2016; Deepika et al., 2021). La integración de los sistemas de información de gestión de la energía (SIGED) con los sistemas de gestión de instalaciones puede mejorar la eficiencia de los intermediarios en la disminución del desperdicio de energía y agua en la gestión de las instalaciones. Se discuten los desafíos y métodos para garantizar la usabilidad de los futuros sistemas inteligentes de salas de control de energía, destacando la importancia de la colaboración entre los investigadores de interacción humano-computadora, los investigadores de informática energética y las partes interesadas del sector energético. El conocimiento del diseño para las aplicaciones de los usuarios del mercado energético local se deriva, ejemplifica y evalúa para asegurar la participación de los usuarios a largo plazo, lo cual es crucial para mantener los efectos a largo plazo.

2.1.5. Modelos de predicción de consumo de energía

La predicción del consumo energético es crucial para optimizar el uso de energía, reducir costos y minimizar el impacto ambiental. Se han empleado diversos algoritmos de aprendizaje automático para mejorar la precisión y la fiabilidad de estas predicciones. Los algoritmos de aprendizaje automático pueden optimizar el consumo de energía en edificios inteligentes utilizando datos históricos y tomando decisiones óptimas basadas en esos datos (Divina et al., 2019).

Al integrar métodos avanzados como el aprendizaje automático y el aprendizaje profundo, los sistemas inteligentes de gestión de energía pueden tomar decisiones óptimas para gestionar el consumo de energía y reducir significativamente su uso (Talib & Croock, 2023).

Los modelos de aprendizaje automático se han empleado para prever la demanda y el consumo de energía en los sistemas energéticos de los edificios, impulsando el desarrollo de métodos y tecnologías en este ámbito (Ardabili et al., 2020). Además, se han aplicado técnicas de aprendizaje automático para estimar el consumo de energía en los hogares y mejorar la eficiencia energética (Kaur & Bala, 2019). A través de estas técnicas, es posible anticipar el consumo de energía en

edificios inteligentes e identificar estrategias adecuadas para optimizar la energía (Divina et al., 2019). Estos algoritmos ofrecen un enfoque prometedor para optimizar el consumo energético en edificios inteligentes y apoyar los esfuerzos de eficiencia energética.

Existen varios algoritmos comunes para predecir el consumo energético en diferentes sectores. La regresión lineal (LR) se utiliza principalmente en el sector industrial para prever el consumo de energía, mostrando eficiencia en algunos casos, aunque no siempre supera a algoritmos más avanzados (Bahij et al., 2020; Husin et al., 2023). Por otro lado, la máquina de vectores de soporte (SVM) se aplica en diversos sectores, incluidos hogares y edificios inteligentes, y es eficaz para trabajar con datos de series temporales no lineales y no estacionarias (Bahij et al., 2020; Liu et al., 2019; Moulla et al., 2025). Las redes neuronales artificiales (RNA), ampliamente utilizadas en la predicción del consumo energético, son especialmente eficaces para modelar patrones complejos y se comparan frecuentemente con otros algoritmos para evaluar su rendimiento (Bahij et al., 2020; Moulla et al., 2025; Tudora & Tîrziu, 2024). El bosque aleatorio (RF) destaca por su precisión en la predicción, mostrando los menores errores absolutos medios (MAE) y errores cuadráticos medios (MSE), y es muy eficaz en la predicción del consumo energético tanto en hogares como en edificios inteligentes (Husin et al., 2023; Moulla et al., 2025; Wu & Chu, 2021). La memoria a largo corto plazo (MLCP), especialmente eficaz para manejar datos de series temporales con tendencias e irregularidades, ha demostrado un rendimiento superior en las predicciones de consumo energético previas a la COVID-19 (Shakila, 2024). Finalmente, el impulso de gradiente extremo (XGBoost), reconocido por su alto rendimiento en tareas de predicción, se utiliza en diversos escenarios de predicción del consumo energético, destacándose en comparación con otros métodos de conjunto (Tao et al., 2021).

2.1.6. Modelos predictivos sin variables exógenas clave

Los modelos predictivos son herramientas esenciales en diversos campos, pero su eficacia puede verse significativamente comprometida si carecen de una validación estadística rigurosa o si no incluyen variables exógenas clave. El modelo ARIMA suele ser muy eficaz para las predicciones a corto plazo debido a su capacidad para modelar tendencias lineales en datos de series temporales (Lee et al., 2020). Sin

embargo, no es capaz de predecir con exactitud series temporales no lineales (Benitez et al., 2024; Y.-S. Lee & Tong, 2011).

En la investigación realizada por Wang (2022), se presentan dos modelos de predicción: el modelo autorregresivo integrado de media móvil (ARIMA) y el modelo de memoria a largo plazo (LSTM). Se destaca que el objetivo de estos modelos es mejorar la precisión en la predicción del tráfico móvil de las estaciones base. Una serie temporal espaciotemporal se describe como un registro cronológico del movimiento de una comunidad, caracterizado por su complejidad e inestabilidad. En relación con el modelo ARIMA, primero se verificó la estacionariedad de la serie temporal mediante el método ADF, y se determinaron los parámetros del modelo (p , d , q) utilizando ACF y PACF. Tras construir el modelo, se evaluó con AIC y BIC. Posteriormente, se desarrollaron modelos LSTM con tres capas ocultas, eligiendo el modelo óptimo al probar diversas combinaciones de parámetros. El rendimiento de la predicción fue evaluado a través de RMES, MRE y R (coeficiente de correlación). Al comparar con ARIMA, los resultados experimentales indicaron que LSTM ofrece una mayor precisión en relación con el conjunto de datos de tráfico de estaciones base.

Por su parte, Kumar et al (2024) utilizaron diferentes modelos basados en datos del Internet de las Cosas (IoT) para la estimación de la temperatura interior. Los datos para este trabajo se recopilaron a partir de sensores de temperatura instalados en edificios inteligentes. Para la predicción, utilizaron modelos como ARIMA, SARIMAX y LSTM, concluyeron que aplicar un solo modelo con los datos recopilados puede no ser eficiente para una predicción precisa.

2.1.7. Modelos predictivos para gestionar consumo energético

Los algoritmos de aprendizaje automático (ML) tienen el potencial de optimizar de manera significativa el consumo energético en diversos contextos mediante el uso de técnicas basadas en datos para predecir, gestionar y reducir el consumo energético (Gonzalez-Briones et al., 2019). El propósito de su uso es anticipar el consumo energético futuro basándose en datos históricos, lo que facilita la planificación y optimización de la distribución de energía, reduciendo costos y minimizando el impacto ambiental (Khanwalkar & Ravikanti, 2025; Moulla et al., 2025; Vonitsanos et al., 2024). Los modelos de ML son capaces de detectar

irregularidades en el uso de energía, lo que permite realizar mantenimiento preventivo y detectar fallos, contribuyendo a evitar el desperdicio de energía y mejorando la eficiencia general (Khanwalkar & Ravikanti, 2025). Estos algoritmos prevén patrones de consumo energético, identifican anomalías y automatizan los sistemas de control para optimizar el uso de energía en edificios inteligentes. Esto incluye la integración de fuentes de energía renovable y la gestión eficiente de los sistemas de almacenamiento de energía (Ngouessy Ep Ovono & Otunniyi, 2024). Además, los algoritmos avanzados para la optimización de hiperparámetros dentro de los marcos de AutoML pueden disminuir el consumo de energía al optimizar el proceso de desarrollo del modelo, reduciendo así la huella de carbono de las aplicaciones de ML.

2.1.8. Usabilidad/latencia en asistentes conversacionales genéricos

Los asistentes conversacionales genéricos enfrentan varios problemas de usabilidad y latencia que afectan la experiencia del usuario. En lo relacionado con la usabilidad, se destacan la eficiencia y satisfacción, flexibilidad y personalización, la interacción humana – AI y en temas de latencia están aquellos relacionados con el procesamiento de lenguaje natural y conectividad, así como también la precisión y el tiempo de respuesta en tiempo real.

El estudio sobre la eficiencia y satisfacción de asistentes conversacionales como "Max" de la UTPL mostró métricas bajas, lo que indica la necesidad de revisar la estrategia de interacción (Maza-Jimenez & Torres-Carrión, 2020). En la propuesta, se destaca que asistentes como Siri y Bixby presentan diferencias significativas en términos de velocidad y satisfacción del usuario, con Siri superando a Bixby en interacciones complejas y prevención de errores (Khandaker et al., 2025).

Los asistentes comerciales como Google Assistant, Siri y Alexa enfrentan desafíos en la gestión del diálogo, personalización y manejo de errores, especialmente en entornos de coordinación de cuidados (Zubatiy et al., 2023). La colaboración entre humanos y AI en la evaluación de la usabilidad ha demostrado que las sugerencias automáticas pueden mejorar la eficiencia y confianza de los evaluadores, aunque la intervención humana sigue siendo crucial (Kuang et al., 2024).

La latencia en los sistemas de conversación puede afectar negativamente tanto el rendimiento del usuario como la disponibilidad del sistema. Para reducir la latencia

percibida por el usuario, se han propuesto soluciones como el uso de rellenos conversacionales y gestuales (Kum & Lee, 2022). Disminuir el tamaño de los modelos de asistentes digitales puede contribuir a mejorar la latencia y la conectividad, permitiendo que los asistentes operen directamente en los dispositivos de los usuarios (Gerhard-Young et al., 2022). La latencia en la generación de respuestas representa un problema considerable. Investigaciones han indicado que una latencia que supera los 4 segundos deteriora la calidad de la experiencia del usuario (Maslych et al., 2025).

2.1.9. Integración de avatares en la gestión Smart Building

Diversos estudios han demostrado que los avatares constituyen una estrategia de comunicación más eficaz que los métodos tradicionales, como la voz o el texto, para fomentar prácticas proambientales entre los ocupantes de edificios. Esto evidencia su capacidad para influir positivamente en el comportamiento de los usuarios hacia un uso más responsable de la energía (Khashe et al., 2017). Simultáneamente, los sistemas de gestión energética que emplean el Internet de las Cosas (IoT) han demostrado su habilidad para ajustar automáticamente los dispositivos según la presencia de personas, lo que no solo optimiza la eficiencia energética, sino que también ofrece a los usuarios un mayor control sobre su consumo (Natarajan et al., 2024).

Los avatares cibernéticos, además de ser mecanismos interactivos, pueden desempeñar un papel educativo al incrementar la conciencia sobre la eficiencia energética y fomentar actitudes positivas hacia el uso sostenible de los recursos. Incluso se ha planteado su utilidad en la comunicación y en la mejora de la eficiencia de la gestión pública relacionada con políticas energéticas (Shimpo et al., 2025).

Asimismo, las aplicaciones basadas en tecnologías de la información y la comunicación (ICT) permiten transformar los datos generados por medidores inteligentes en información comprensible y valiosa para los ocupantes. Estas aplicaciones ofrecen recomendaciones personalizadas, incentivos y recompensas que promueven un uso más eficiente de los dispositivos eléctricos (Doukas et al., 2018). En conjunto, los sistemas adaptados tanto al usuario como al contexto permiten una optimización más precisa del consumo, ajustándose a las necesidades y comportamientos específicos de cada individuo, al mismo tiempo que cumplen una

función educativa al fomentar prácticas sostenibles y provocar cambios de comportamiento (Natarajan et al., 2024; Shimp et al., 2025).

2.1.10. Integración de avatares con algoritmos de aprendizaje

La integración de avatares con algoritmos de aprendizaje automático (ML) para optimizar el consumo energético presenta diversas aplicaciones, tales como actuar como gemelos digitales, proporcionar monitoreo y control en tiempo real de los sistemas energéticos, y mejorar la eficiencia operativa mediante la simulación y predicción de patrones de consumo energético, entre otros (Agostinelli et al., 2021).

En el ámbito de los edificios inteligentes, los avatares pueden emplear algoritmos de aprendizaje automático para supervisar y gestionar dinámicamente el uso de energía. Esto incluye la recopilación de datos en tiempo real, la identificación de anomalías y el mantenimiento predictivo, lo que puede mejorar significativamente la eficiencia energética y reducir los costos (Alanne & Sierla, 2022).

Además, tienen la capacidad de anticipar la demanda energética futura y los patrones de consumo. Esta habilidad predictiva permite una planificación y optimización más efectivas de los recursos energéticos, minimizando el desperdicio y estabilizando el suministro (Tan et al., 2024). Al analizar los patrones de consumo, los avatares pueden optimizar la gestión de la carga en entornos residenciales, comerciales e industriales, lo que incluye desplazar el consumo de energía a horas de menor demanda e integrar de manera más eficiente las fuentes de energía renovable (Chorozoglou et al., 2023).

Los avatares pueden proporcionar información práctica a los operadores, identificando oportunidades de optimización y aconsejando sobre medidas de ahorro energético (S. S. Kumar et al., 2022). Además, simplifican la interacción con modelos complejos de aprendizaje automático, facilitando que personas sin experiencia técnica aprovechen sistemas avanzados de gestión energética. Esto incluye la automatización de las operaciones de aprendizaje automático y la provisión de interfaces intuitivas para las partes interesadas en el sector energético (Chorozoglou et al., 2023).

En este contexto, la combinación de avatares y algoritmos de aprendizaje automático permitirá una gestión más eficiente del consumo energético,

personalizando la experiencia del usuario y ofreciendo recomendaciones adaptadas, lo que puede aumentar la satisfacción de los ocupantes y su compromiso con prácticas sostenibles, además de ajustarse a cambios en el comportamiento de los ocupantes y en las condiciones del entorno, mejorando continuamente su eficacia.

2.2. Bases Teóricas

2.2.1. Relacionadas con la Gestión eficiente de edificios

Smart Building

Los edificios inteligentes representan un avance significativo en el entorno construido, al aprovechar tecnologías de vanguardia para aumentar la eficiencia energética y el confort de los ocupantesla eficiencia operativa. Los edificios inteligentes son estructuras que incorporan tecnologías avanzadas y sistemas automatizados para gestionar diversas actividades del edificio, tales como calefacción, ventilación, aire acondicionado, iluminación y seguridad (Sardjono & Perdana, 2024). Estos edificios emplean dispositivos del Internet de las Cosas (IoT), sensores, actuadores y microchips para recopilar datos y optimizar las operaciones del edificio en tiempo real (Elhabyb et al., 2024; Khaoula et al., 2022; Whig et al., 2024)

Entre las ventajas de los edificios inteligentes se destaca que utilizan un 30% menos de energía en comparación con los edificios convencionales, gracias a la integración de sistemas que colaboran para mejorar la eficiencia (Winiarski, 2007). Además, la gestión automatizada de la energía, que incluye el control de HVAC y la iluminación, reduce significativamente el consumo energético (Sivasankari & Rathika, 2025). La automatización y el uso eficiente de los recursos disminuyen los costos operativos y de mantenimiento (Halkier & NeoCortec, 2024).

La implementación de tecnologías inteligentes puede transformar edificios existentes en estructuras más eficientes, reduciendo los gastos generales (Kishore et al., 2020; Winiarski, 2007). Los edificios inteligentes contribuyen a los objetivos de sostenibilidad al reducir las emisiones de CO₂ y optimizar el uso de recursos (Sardjono & Perdana, 2024; Sivasankari & Rathika, 2025). La integración de energías renovables y sistemas de gestión de energía avanzados promueve prácticas ecológicas (Muchafangeyi & Channi, 2024).

Eficiencia Energética en Edificios

La eficiencia energética en edificios se refiere a la mejora y optimización del uso de energía eléctrica en los mismos, con el objetivo de reducir la demanda energética y minimizar el consumo de electricidad (Kwon et al., 2022; Zanjani et al., 2015). Considerando que los edificios contribuyen significativamente al incremento del consumo global de electricidad y a las emisiones de CO₂, mejorar la eficiencia energética se considera un paso crucial para promover la sostenibilidad energética del planeta (González-Torres et al., 2022). Para lograr este objetivo, es fundamental gestionar la energía eléctrica en los edificios y obtener información detallada sobre el consumo energético, lo cual permite proponer soluciones que optimicen la eficiencia y caracterizar los edificios mediante indicadores que orienten las acciones de gestión energética (Saleem et al., 2023).

Middleware

El middleware desempeña un papel crucial en la interconexión de los Sistemas de Gestión de Edificios (BMS), al facilitar la comunicación fluida entre diferentes sistemas y tecnologías dentro de un entorno de edificio inteligente (Saleem et al., 2023). El middleware actúa como una capa intermedia que permite la interoperabilidad entre sistemas diversos, mejorando así la eficiencia y funcionalidad general de las aplicaciones de edificios inteligentes (Ahmed et al., 2022).

Los avances recientes en los sistemas de gestión de edificios han comenzado a incorporar tecnologías basadas en estándares abiertos sobre IP, dejando atrás los protocolos propietarios y el hardware especializado. Este cambio mejora la interoperabilidad entre software y hardware, pero también presenta desafíos, como la complejidad en la configuración de redes y la dependencia del middleware para funciones básicas, además de problemas de seguridad (Kaplan & Haenlein, 2019). El middleware ayuda a superar estos desafíos al actuar como un enlace que conecta diferentes sistemas dentro del edificio, permitiendo una comunicación y un intercambio de datos eficientes (Saleem et al., 2023). La integración de diversos sistemas de gestión de edificios es esencial para los edificios inteligentes modernos, que a menudo requieren conexiones con múltiples sistemas de información como BMS, Sistemas de Gestión Energética (EMS), dispositivos de Internet de las Cosas

(IoT), entre otros. El middleware facilita la creación de una arquitectura de sistemas con APIs bien definidas, lo cual es crucial para el desarrollo de aplicaciones de edificios inteligentes que sean reutilizables, modulares y escalables (Chamari et al., 2023). Esta arquitectura se alinea con principios como microservicios, diseño API-first, componentes en la nube y principios headless, y ha sido implementada como prueba de concepto en aplicaciones de edificios inteligentes, demostrando una mayor reutilización y modularidad.

El middleware desempeña un papel fundamental en la integración de diversas tecnologías y fuentes de datos en un edificio, lo cual es esencial para los Sistemas de Gestión de Edificios basados en IoT. Estos sistemas recopilan y procesan grandes volúmenes de datos para optimizar el consumo energético (Nakhuva & Champaneria, 2015). La integración facilitada por el middleware puede resultar en ahorros energéticos significativos, como se ha demostrado en casos específicos donde se integraron sistemas de aire acondicionado, iluminación y monitoreo energético, logrando una notable reducción en el consumo de energía (Mataloto, 2019).

En este aspecto, se considera al middleware esencial para conectar los Sistemas de Gestión de Edificios, permitiendo que sistemas diversos operen conjuntamente de manera eficiente y segura. Esta conectividad mejora la gestión energética, incrementa el confort de los ocupantes y optimiza la eficiencia operativa en edificios inteligentes. El middleware asegura que los distintos sistemas dentro de un edificio puedan compartir y procesar datos de manera eficaz, estableciendo una base para innovaciones y desarrollos modernos en edificios inteligentes.

Por otro lado, Silva et al. (2018) presentan un proyecto que utiliza el paradigma de computación de agente móvil, empleando JavaScript para permitir la ejecución y movilidad de agentes a través de redes locales e Internet. Este enfoque busca resolver los problemas asociados con el middleware de agentes móviles basados en Java, que requieren la instalación de la Máquina Virtual Java en los dispositivos, lo que complica su uso en sistemas operativos como macOS, iOS y otros no compatibles con Java. Este middleware funciona de manera uniforme en todos los sistemas operativos, requiriendo solo la instalación de node.js.

Internet de las Cosas (IoT)

El Internet de las Cosas (IoT) es una tecnología clave que impulsa la cuarta Revolución Industrial y el concepto de Industria 4.0, permitiendo la detección y recolección de grandes cantidades de datos para su posterior almacenamiento y análisis. Generalmente, estos datos se procesan en servidores centralizados en la nube para garantizar un alto rendimiento.

La implementación del Internet de las Cosas (IoT) en la gestión de edificios inteligentes se ha revelado como una solución prometedora para optimizar la eficiencia energética, el confort de los ocupantes y la eficacia operativa. Los edificios inteligentes se benefician de la integración de diversos sistemas a través de sensores que facilitan operaciones inteligentes y permiten el monitoreo en tiempo real y la toma de decisiones basadas en datos (Umaeswari et al., 2024). El IoT posibilita un control y monitoreo remoto eficiente al conectar diferentes partes de la infraestructura del edificio, lo que facilita la gestión de aspectos como la temperatura, la seguridad y el mantenimiento (Shah et al., 2022).

De esta manera, se pueden automatizar tareas de control climático interno o realizar mantenimientos rutinarios de manera más sencilla mediante dispositivos móviles o computadoras. En cuanto a la eficiencia energética, el IoT tiene el potencial de disminuir el consumo de energía hasta en un 30% y los costos operativos en un 20% al integrar tecnologías como la inteligencia artificial y la computación en la nube para el control energético a múltiples niveles dentro del edificio (Pan et al., 2015). Esta eficiencia energética se alcanza mediante la implementación de sensores inteligentes y la integración de energías renovables (Metallidou et al., 2020). No obstante, la adopción del IoT en la gestión de edificios inteligentes enfrenta barreras significativas, como los altos costos iniciales, que suelen representar alrededor del 15% del presupuesto de un proyecto, preocupaciones sobre la seguridad de los datos y la complejidad en la integración de sistemas (Umaeswari et al., 2024)..

Protocolo MQTT

El protocolo Message Queuing Telemetry Transport (MQTT) juega un papel crucial en la gestión de edificios inteligentes, gracias a su capacidad para facilitar la comunicación entre dispositivos en entornos de Internet de las Cosas (IoT). MQTT es

un protocolo de mensajería de publicación-suscripción que permite la transmisión de datos de manera eficiente y mínima en términos de ancho de banda, lo que lo hace ideal para aplicaciones donde los recursos de red son limitados (Yassein et al., 2017).

La adopción del protocolo MQTT en la gestión de edificios inteligentes facilita la supervisión en tiempo real y un alto nivel de control remoto, lo cual es esencial para mantener operaciones eficientes y sostenibles (S et al., 2024). Sin embargo, una preocupación significativa con respecto a MQTT es su falta de características de seguridad integradas, lo que lo hace vulnerable a ciberataques como el "man-in-the-middle", desbordamiento de búfer y ataques de denegación de servicio (DDoS) (Alotaibi et al., 2024; Mosaiyebzadeh et al., 2021).

Para abordar estos problemas de seguridad en MQTT, se han propuesto diversas soluciones. Por ejemplo, la implementación de un esquema de autenticación ligero que utiliza operaciones criptográficas ligeras puede contribuir a asegurar la comunicación entre dispositivos en redes inteligentes sin incurrir en altos costos computacionales (S et al., 2024). Además, herramientas como el Asistente de Seguridad MQTT (MQTTSA) ayudan a identificar configuraciones incorrectas en las implementaciones de MQTT y ofrecen recomendaciones para mitigar estas vulnerabilidades (Palmieri et al., 2019).

En cuanto a la detección de intrusiones, se han desarrollado sistemas que emplean técnicas de aprendizaje profundo para identificar patrones de comportamiento anómalos y mitigar las amenazas potenciales en las redes MQTT (Khan et al., 2021; Mosaiyebzadeh et al., 2021). Silva et al. (2017) proponen un servicio de notificaciones Push a través del protocolo MQTT. Estos sistemas pueden monitorear de manera distribuida el tráfico MQTT y clasificar comportamientos sospechosos, permitiendo así una respuesta más rápida a los ataques (Alotaibi et al., 2024).

El uso adecuado y seguro de MQTT en la gestión de edificios inteligentes no solo mejora la eficiencia operativa, sino que también es crucial para proteger los sistemas de control contra posibles amenazas cibernéticas. Implementar medidas de seguridad robustas, como el uso de redes de aprendizaje automático distribuidas y esquemas de autenticación ligeros, es fundamental para la seguridad de las aplicaciones que utilizan MQTT en entornos IoT.

Inteligencia Artificial (IA)

Kaplan & Haenlein (2019) definen a la IA como la capacidad de un sistema para interpretar datos externos, aprender de ellos y alcanzar objetivos a través de una adaptación flexible, este concepto es difuso y plagado de interrogantes teniendo en cuenta que la IA está inspirada en los humanos y humanizada. También existen conceptos afines como el Internet de las cosas y el Big Data, permitiendo una visión más matizada.

La inteligencia artificial tiene la capacidad de procesar en tiempo real los datos relacionados con el consumo energético, permitiendo realizar predicciones precisas sobre las necesidades futuras de energía, lo cual optimiza su uso (Chávez Bravo, 2021). A través del aprendizaje automático, la IA analiza los patrones de consumo de energía y anticipa la demanda futura. Esta habilidad predictiva posibilita la adaptación del suministro energético conforme a la demanda prevista, lo que reduce el desperdicio y mejora la eficiencia (Herath & Mittal, 2022).

Además, la IA desempeña un papel crucial en la automatización de sistemas de climatización, iluminación y otros dispositivos eléctricos en edificios (Escobar & Villazón, 2018). Mediante la integración de sensores inteligentes, la IA ajusta automáticamente estos sistemas para optimizar el consumo energético según la ocupación y las condiciones ambientales (Minoli et al., 2017). Asimismo, gestiona de manera eficiente el uso de fuentes de energía renovables. Por ejemplo, puede prever la generación de energía solar o eólica y ajustar el consumo energético del edificio para maximizar el uso de estas fuentes sostenibles (Mustapha et al., 2025).

A través del análisis de datos, la IA puede anticipar y alertar sobre posibles fallos en los sistemas energéticos antes de que ocurran, lo que permite realizar un mantenimiento preventivo, evitando interrupciones y costos adicionales (Aziz et al., 2020). Además, posee la capacidad de identificar patrones de uso ineficientes y detectar irregularidades en el consumo energético, lo que permite a los administradores de edificios tomar decisiones informadas para optimizar la eficiencia (Rani et al., 2023).

De este modo, la incorporación de la inteligencia artificial en la gestión energética de edificios inteligentes no solo ayuda a disminuir costos y mejorar la

sostenibilidad, sino que también incrementa el confort y la productividad de los ocupantes al ofrecer un entorno controlado de manera eficiente. Esta integración se convierte en un componente esencial ante la creciente demanda de energía y la necesidad de desarrollos sostenibles en el sector de la construcción.

2.2.2. Relacionados con el desarrollo del Avatar

Avatar

Los avatares y los asesores virtuales son formas de tecnología que se utilizan para diversos fines, como entretenimiento, educación y orientación. Así, los avatares pueden tener influencias psicológicas en los usuarios y pueden diseñarse para aumentar la motivación (Ashrafi et al., 2024). El uso de avatares diferentes de la apariencia real del usuario, pueden impactar positivamente la experiencia del usuario en términos de interacción y percepción (Cheymol et al., 2023). La relación entre los usuarios y sus avatares puede verse influenciada por el contexto en el que se utilizan (Bengtsson, 2019). Los avatares pueden usarse como guías, maestros, acompañantes o mentores en experiencias inmersivas, y su diseño debe considerar factores como la credibilidad y la alianza de trabajo (Bartlett et al., 2022).

La implementación de avatares en la gestión del consumo energético en edificios inteligentes está recibiendo una atención creciente. Estos avatares pueden funcionar como interfaces interactivas que se comunican con los ocupantes del edificio para influir en sus hábitos hacia prácticas de ahorro de energía. En este contexto, los avatares resultan ser más efectivos que la voz o el texto para fomentar el cumplimiento de las solicitudes ambientales entre los ocupantes de los edificios.

Esto sugiere que los avatares pueden desempeñar un papel crucial en el fomento de comportamientos de ahorro energético (Khashe et al., 2017). Estos avatares podrían asumir el rol de entrenadores o guías virtuales que asisten a los usuarios en la navegación de los sistemas del edificio o en la gestión más eficiente de su consumo de energía (Hadjidemetriou et al., 2023; Khashe et al., 2017). Esto indica un potencial en los edificios, donde los avatares podrían interactuar de manera continua con los usuarios para optimizar el uso de energía o informar sobre problemas de mantenimiento.

Además, la personalización y customización de los avatares pueden ayudar a fomentar un sentido de pertenencia y una interacción personalizada entre los usuarios. Este nivel de personalización podría traducirse en una mejor cooperación de los ocupantes con las prácticas de gestión edilicia (Hadjidemetriou et al., 2023). Adicionalmente, en sistemas de realidad aumentada (AR) y realidad virtual (VR), los avatares pueden potenciar las tareas colaborativas al aportar representaciones realistas en sistemas de telepresencia.

Estos sistemas, que permiten a los usuarios interactuar con ubicaciones remotas como si estuvieran presentes físicamente, podrían ser particularmente beneficiosos para la gestión de instalaciones complejas (Pakanen et al., 2022). Esta tecnología podría aplicarse en contextos de gestión remota, donde los administradores supervisen múltiples edificios desde un único lugar. El efecto Proteus, que describe cómo el comportamiento y la actitud de los usuarios se adaptan a su representación digital, destaca otro posible papel de los avatares en la gestión de edificios (Grech et al., 2023). Por ejemplo, la adopción de avatares que promuevan comportamientos sostenibles y eficiencia energética podría motivar a los usuarios a adoptar estas actitudes en su vida cotidiana (Kocur et al., 2021).

Por consiguiente, al integrar avatares en la gestión de edificios, las instalaciones pueden experimentar un mayor compromiso de los usuarios, interacciones personalizadas y mejores capacidades de gestión a distancia. Estas herramientas digitales no solo pueden optimizar las operaciones, sino también fomentar comportamientos ecológicos y promover prácticas de gestión colaborativa entre los ocupantes del edificio. Su diseño y el contexto en el que se emplean son fundamentales para definir la experiencia del usuario y su relación con el avatar.

BERT: Clasificadores de intención

El uso de BERT (Bidirectional Encoder Representations from Transformers) para el clasificador de intención representa un avance en el campo del procesamiento del lenguaje natural (NLP) se basa en la arquitectura Transformer (letra 'T' en el famoso 'GPT') debido a su capacidad para mejorar significativamente las capacidades de los modelos de lenguaje. Este modelo se utiliza ampliamente en diversas tareas de procesamiento de lenguaje natural, tales como la segmentación de oraciones, el

análisis de sentimientos, el reconocimiento de entidades nombradas y la detección de noticias falsas, entre otras (Gardazi et al., 2025).

Es importante destacar que BERT, al ser un modelo preentrenado y altamente adaptable, permite a los investigadores y profesionales ajustar su funcionamiento para aplicaciones específicas, lo que puede resultar en mejoras significativas en comparación con las técnicas más tradicionales. Esto fomenta una comprensión amplia y sofisticada de la tecnología para interpretar el lenguaje natural y aplicarse en diversas tareas de procesamiento de lenguaje natural (Gupta, 2024).. Dentro de las características principales de BERT, podemos mencionar que:

- Utiliza una arquitectura de transformadores bidireccionales, lo que significa que considera el contexto de las palabras tanto a la izquierda como a la derecha simultáneamente. Esto permite una comprensión más profunda de las relaciones entre palabras en una oración (Tanaka et al., 2020).
- Se pre-entrena en grandes corpus de texto, como BookCorpus y Wikipedia, utilizando tareas como la predicción de palabras enmascaradas y la predicción de la siguiente oración. Este preentrenamiento permite que BERT capture representaciones contextuales ricas (Yu et al., 2019).
- Ha demostrado un rendimiento sobresaliente en diversas tareas de PNL, incluyendo clasificación de texto, detección de plagio, análisis de sentimientos y respuesta a preguntas (Özçift et al., 2021).
- Una de las fortalezas de BERT es su capacidad para ser ajustada (fine-tuned) en tareas específicas con relativamente pocos datos adicionales. Este proceso ajusta los parámetros del modelo para optimizar su rendimiento en tareas específicas (Yu et al., 2019).

Tecnologías de conversión de voz a texto (STT)

Las tecnologías de conversión de voz a texto han avanzado significativamente gracias a los desarrollos en inteligencia artificial (IA) y aprendizaje automático (ML). Son varias las tecnologías involucradas en estos sistemas:

El reconocimiento del habla y procesamiento del lenguaje natural (NLP),

permite a las máquinas interpretar y transcribir palabras habladas a texto. Esta tecnología se utiliza ampliamente en asistentes virtuales y dispositivos inteligentes. La implementación de modelos avanzados de procesamiento del lenguaje natural (NLP) mejora la precisión del reconocimiento al analizar patrones lingüísticos (Eziamaka et al., 2024; Vashisht et al., 2021).

La traducción automática neuronal (NMT) se emplea para convertir un idioma a otro, utilizando redes neuronales profundas que primero codifican y luego decodifican los términos del idioma de manera eficiente. Dos redes neuronales recurrentes colaboran para llevar a cabo esta tarea, facilitando la transformación del texto desde un idioma fuente hacia un idioma objetivo (Vashisht et al., 2021).

Las tecnologías de clasificación de voz son esenciales para el desarrollo de aplicaciones que responden a comandos vocales, ya que permiten la identificación e interpretación de sonidos en señales acústicas naturales. Entre las aplicaciones prácticas de estas tecnologías se incluyen los sistemas de reconocimiento de idiomas, la biometría de voz y el reconocimiento de hablantes (Patra et al., 2023). Para mejorar el reconocimiento de voz, se emplean modelos de aprendizaje profundo que incorporan diversas tecnologías, como sensores de biosignales para interfaces de habla silenciosa y reconocimiento visual del habla, entre otros métodos. Estas técnicas contribuyen a aumentar la precisión y reducir los errores en la transcripción de voz (Lee et al., 2021).

Se han desarrollado infraestructuras de IoT y nubes inteligentes que integran métodos de NLP y reconocimiento de voz para proporcionar una interfaz más fluida y adaptable en dispositivos inteligentes. Estos sistemas permiten transformar frases en comandos directos y se ajustan a diferentes entornos, mejorando la experiencia del usuario con dispositivos habilitados para voz (Iliev & Ilieva, 2022). Existen soluciones de software basadas en IA que facilitan la comunicación efectiva de personas con discapacidades mediante tecnologías de reconocimiento de voz, incluso en ambientes ruidosos o con múltiples fuentes sonoras.

Estas soluciones integran modelos de atención y descomposición de fuentes de sonido, optimizados para su uso en dispositivos de IoT (Tambe et al., 2023). Todos estos avances en tecnologías de voz a texto no solo mejoran la interacción entre humanos y computadoras, sino que también fomentan la accesibilidad y la inclusión

digital, brindando mayor independencia a usuarios con diversas necesidades (Eziamaka et al., 2024). Entre estas herramientas destaca Whisper AI, desarrollada por OpenAI, un servicio de reconocimiento automático de voz (ASR) de última generación que busca transcribir la entrada de audio con alta precisión.

En el 2023 se ha destacado por superar a muchos competidores del sector. Sin embargo, no está exento de desafíos, incluidas alucinaciones ocasionales en las transcripciones, donde el sistema genera frases u oraciones que no están presentes en el audio original (Dani et al., 2023). Whisper AI se puede integrar con interfaces gráficas de usuario (GUI) y bibliotecas de comandos para permitir la interacción manos libres con la tecnología (Fernandez et al., 2024). Esto es especialmente beneficioso para personas con discapacidad, personas mayores o quienes realizan múltiples tareas (Koenecke et al., 2024). La integración de Whisper AI con las funciones de IA conversacional mejora la interacción persona-computadora, haciéndola más intuitiva y accesible. Esta combinación permite una comunicación más atractiva y eficiente, mejorando la atención al cliente y la asistencia.

Tecnologías de conversión de texto a voz (TTS)

Las tecnologías de conversión de texto a voz (TTS) han avanzado significativamente gracias a los desarrollos recientes en inteligencia artificial y redes neuronales. Estas tecnologías han mejorado la calidad de la síntesis de voz y se utilizan ampliamente en aplicaciones como la interacción humano-computadora, los asistentes virtuales y la accesibilidad para personas con discapacidades auditivas o del habla (Yao et al., 2024).

Una de las tecnologías más innovadoras es el sistema SynthRhythm-TTS (SR-TTS), que utiliza una estructura basada en Transformer para mejorar no solo la calidad fonológica y la naturalidad del habla, sino también la eficiencia del proceso de generación de voz. Este sistema es adecuado para aplicaciones que requieren un habla de alta calidad, como asistentes inteligentes y podcasts sintetizados por voz (Yao et al., 2024).

Otra propuesta relevante es el modelo SpeechX, que combina la modelización de lenguaje codificado neural con el aprendizaje multitarea para tareas de conversión de texto a voz sin preajustes. SpeechX es capaz de manejar señales limpias y ruidosas,

demostrando su eficacia en tareas de mejora y transformación del habla, lo que lo hace valioso para aplicaciones en entornos acústicos adversos (X. Wang et al., 2024).

Además, ZMM-TTS busca ampliar los límites de la síntesis de voz multilingüe y multispeaker al utilizar representaciones de voz discretas autodidactas. Este modelo demuestra capacidad de generalización sin preajustes para hablantes y lenguajes no vistos, siendo una promesa para lenguajes de bajo recurso, además de ofrecer naturalidad y similitud en el habla producida (Gong et al., 2024)

Estas tecnologías no solo ofrecen ventajas significativas en términos de calidad del habla y flexibilidad, sino que también presentan desafíos en cuanto a la velocidad de inferencia y la naturalidad del habla sintetizada. La ventriloquía algorítmica, un término para describir cómo estas tecnologías manipulan las voces de artistas de doblaje profesionales, también plantea cuestiones culturales y filosóficas sobre la relación entre lo humano y lo virtual (Ramati, 2024)

Además, se utilizan ampliamente en áreas como la accesibilidad para personas con discapacidades visuales, asistentes virtuales (como Siri y Alexa), aprendizaje de idiomas, automatización de servicios al cliente y contenido multimedia. Los desarrollos recientes en el campo de la síntesis de voz incluyen la capacidad de "zero-shot" y funciones interlinguales, que permiten a los sistemas TTS generar voces para múltiples idiomas y hablantes sin requerir datos de entrenamiento específicos para cada idioma o nuevo hablante (Gong et al., 2024; Ramati, 2024; Yao et al., 2024).

Dentro de estas tecnologías, el modelo VITS (Inferencia Variacional Texto a Voz) se presenta como un sistema de texto a voz de una sola etapa, mejorado y adaptado para diversas aplicaciones, tales como la síntesis de voz y la conversión de voz. Este modelo integra el modelado autorregresivo, BERT, HuBERT y el aprendizaje autosupervisado para producir un habla de alta fidelidad que imita con precisión el habla humana (Guo et al., 2023).

El modelo VITS es capaz de generar un habla de alta calidad que reproduce con exactitud la entonación, el ritmo y la prosodia humanos (Chan et al., 2024). Mejoras como ABS-VITS integran BERT para la extracción de características y la predicción aleatoria de la duración para diversos ritmos lingüísticos (Zhao & Yang, 2023). Los modelos TTS de extremo a extremo han transformado el campo de la

síntesis de voz al generar formas de onda directamente a partir del texto, simplificando la arquitectura y mejorando la naturalidad e inteligibilidad del habla sintetizada.

Este modelo mejora la expresividad prosódica mediante la integración profunda de las propiedades jerárquicas del habla con autocodificadores variacionales jerárquicos. Modela variables latentes multiescala en varios niveles lingüísticos (marco, fonema, subpalabra, palabra y oración) para generar un habla más natural y diversa (Shang et al., 2023). En este contexto, se destaca que las tecnologías TTS continúan evolucionando para ofrecer experiencias de usuario más naturales y accesibles, abriendo nuevas posibilidades para la accesibilidad y la interacción hombre-máquina en un mundo cada vez más digitalizado. No obstante, estos avances también subrayan la necesidad de abordar desafíos técnicos y éticos para garantizar que estas tecnologías sean inclusivas y equitativas.

Llama 2 y Llama 3

Llama 2 y Llama 3 son series de modelos de lenguaje grande creados como parte de la familia LLaMa, destinados a tareas de procesamiento de lenguaje natural (NLP) y aplicaciones más especializadas. Entre sus características y comparaciones principales: Llama 2 ha mostrado una notable eficacia en la clasificación de texto, especialmente al compararse con modelos como GPT-3.

Por ejemplo, Llama 2 13B y 70B superaron a GPT-3 en eficiencia y efectividad en tareas de clasificación de texto, logrando altas puntuaciones F1 con datos completos (Kumar et al., 2024). Tiene variantes adaptadas para aplicaciones específicas, como TC-Llama 2, que se centra en la comercialización de tecnología (Yeom et al., 2024), y ha sido evaluado para tareas médicas (Pandya et al., 2024).

2.2.3. Usabilidad y Experiencia de Usuario

La evaluación de la usabilidad y la experiencia de usuario (UX) es una tarea compleja que abarca varias metodologías y enfoques. El enfoque del diseño centrado en el usuario, combinado con evaluaciones heurísticas, permite identificar problemas de la UI y UX, lo que es esencial para mejorar la satisfacción del cliente en aplicaciones móviles de banca y otros sistemas (Fatkhuri et al., 2024; Sudirjo et al., 2024). El uso de cuestionarios validados, como el Sistema de Escala de Usabilidad (System Usability

Scale) y el Cuestionario de Experiencia de Usuario (UEQ), son utilizados para evaluar la usabilidad de aplicaciones y dispositivos. Estas herramientas ayudan a refinar y optimizar dichos sistemas (Moulaei et al., 2024).

System Usability Scale (SUS)

La Escala de Usabilidad del Sistema (SUS) es una herramienta ampliamente utilizada para evaluar la usabilidad de diversos sistemas, especialmente en el desarrollo de software y la investigación de la experiencia de usuario. Sirve para evaluar la usabilidad percibida de un sistema de forma rápida y eficaz mediante un cuestionario estructurado (Gronier & Baudet, 2021; Zwakman et al., 2021).

El SUS consta de 10 afirmaciones valoradas en una escala de Likert de 5 puntos, que van desde "Totalmente en desacuerdo" hasta "Totalmente de acuerdo". Estas preguntas abordan múltiples aspectos de la usabilidad, como la facilidad de uso y la satisfacción del usuario. Las respuestas se convierten en una puntuación compuesta, que va de 0 a 100, indicando la usabilidad percibida del sistema, donde una puntuación superior a 80.3 equivale a una usabilidad excelente, entre 68 a 80.3 valoración Buena, 68 aceptable, menor de 68 pobre (Sevilla-Gonzalez et al., 2020; Zwakman et al., 2021).

SUS es elogiado por su simplicidad, facilidad de administración y capacidad para ofrecer información rápida sobre la usabilidad. Facilita la comparación entre sistemas, lo que contribuye a la identificación de problemas de usabilidad (Sevilla-Gonzalez et al., 2020; Vlachogianni & Tselios, 2021). Aunque fue desarrollado inicialmente para software, el SUS ha sido adaptado a diversos campos, como las interfaces web, móviles y de voz (Deshmukh & Chalmeta, 2024).

A pesar de su robustez, existen algunas limitaciones. Los críticos señalan que SUS podría no captar completamente las sutilezas de usabilidad, especialmente en entornos sin interfaz gráfica como las interfaces de usuario de voz. Esta limitación ha llevado al desarrollo de nuevas escalas adaptadas a estos contextos (Ensink et al., 2022; Zwakman et al., 2021).

User Experience Questionnaire (UEQ)

El Cuestionario de Experiencia del Usuario (UEQ) es una herramienta

fundamental para evaluar la calidad de la experiencia del usuario en una variedad de productos, centrándose especialmente en interfaces digitales como sitios web y aplicaciones móviles. Evalúa la experiencia del usuario a través de seis escalas diferentes: Atractivo, Eficiencia, Fiabilidad, Estimulación, Novedad y Claridad (Saleh et al., 2022; Takano et al., 2023).

El Cuestionario de Experiencia de Usuario (UEQ) se integra en las pruebas de usabilidad para obtener datos cuantitativos sobre las experiencias de los usuarios, identificando áreas que requieren mejoras (Kushendriawan et al., 2021; Sabukunze & Arakaza, 2021). Además, se emplea tanto en contextos académicos como industriales para evaluar la satisfacción y el compromiso del usuario en diversos ámbitos, como aplicaciones móviles, sitios web y plataformas de aprendizaje electrónico (Cuadros et al., 2021; Saleh et al., 2022).

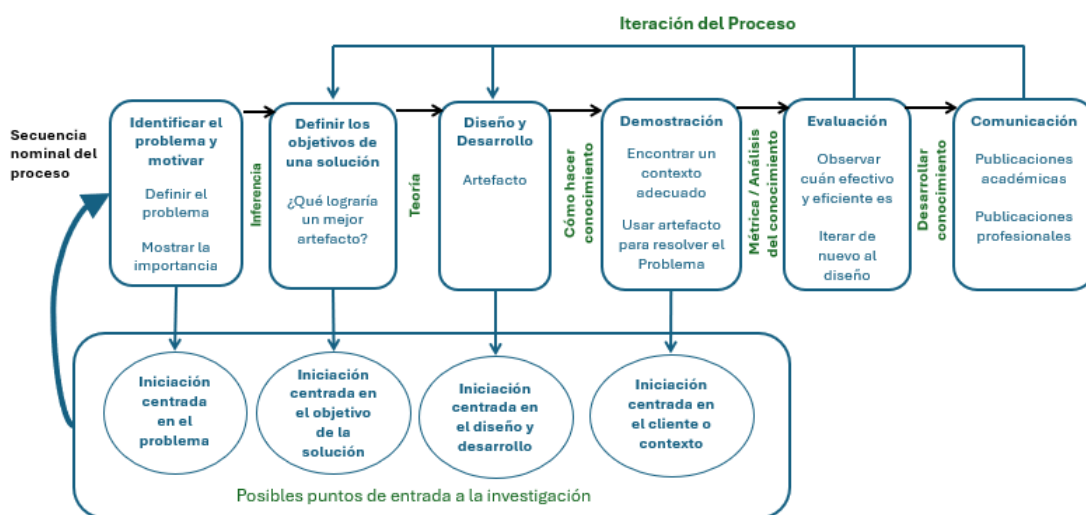
Asimismo, los diseñadores y desarrolladores de productos utilizan los conocimientos derivados del UEQ para tomar decisiones informadas durante los procesos de diseño iterativo, mejorando y refinando el diseño de productos basándose en la retroalimentación de los usuarios (Kushendriawan et al., 2021; Takano et al., 2023). Por esta razón, el UEQ es valorado por su facilidad de uso y su capacidad para proporcionar información inmediata sobre las experiencias del usuario, orientando las mejoras en el diseño y la funcionalidad del producto.

CAPÍTULO 3. METODOLOGÍA

La metodología de investigación constituye la guía fundamental para garantizar la coherencia y validez de un estudio. El presente estudio adopta un paradigma de Investigación en Ciencia del Diseño (DSR). El objetivo central de la DSR es generar conocimiento a través de la creación y evaluación de artefactos innovadores que resuelven problemas prácticos del mundo real. Por ello, el artefacto central de esta tesis doctoral es el "Ecosistema Inteligente de Gestión Energética".

La Ciencia del Diseño se centra en la construcción y validación de soluciones tecnológicas que abordan problemas complejos de la realidad, combinando el rigor científico con la relevancia práctica. En este sentido, la metodología de Investigación en Ciencia del Diseño se estructura en seis fases cíclicas (Figura 3) que orientan el proceso de investigación a desarrollar. Cada una de estas fases permite garantizar que el “Ecosistema Inteligente de Gestión Energética” no solo sea funcional desde el punto de vista técnico, sino también pertinente y útil en el contexto de la gestión eficiente de los edificios inteligentes.

Figura 3 Adaptación del Modelo del proceso de investigación en diseño (*Design Science Research Process Model*)



Fuente: Peffers et al. (2007).

3.1. Tipo de investigación

La presente investigación es de carácter aplicado y adopta un enfoque mixto, combinando métodos cuantitativos y cualitativos, utilizando como metodología la ciencia del diseño (Design Science Research). Este enfoque se considera un marco amplio que permite describir, desde una perspectiva práctica, el proceso de desarrollo y evaluación de una solución.

3.2. Población y Muestra

La selección de la población y muestra de este estudio se fundamenta en los principios de la Ciencia del Diseño para la Investigación (DSR), la cual prioriza la validación rigurosa del artefacto en un contexto de aplicación real y pertinente, por encima de la generalización estadística de tipo probabilístico. En consecuencia, se ha adoptado una estrategia de muestreo no probabilístico, enfocada en el estudio de caso, para asegurar la viabilidad y la profundidad de la evaluación:

3.2.1. Unidad de análisis

Se define por los edificios que disponen de equipos IoT para el monitoreo de energía en una universidad pública de la provincia de Manabí en Ecuador.

3.2.2. Población de estudio

Corresponde a las universidades públicas de la Provincia de Manabí que cumplen con los requisitos de infraestructura inteligente para la medición y control de energía.

3.2.3. Tamaño de muestra

Para el cálculo del tamaño de la muestra, se utiliza un muestreo no probabilístico en el que se selecciona a la Universidad Técnica de Manabí por disponer de varias edificaciones que cuentan con equipos IoT instalados en los últimos años y por la facilidad de acceso a las instalaciones.

3.2.4. Selección de muestra

Se utiliza un Muestreo Intencional Estratégico (Purposeful Sampling),

focalizando la investigación en una edificación específica de la Universidad Técnica de Manabí (UTM). Esta selección es clave para la Fase 5 de Evaluación, donde se elegirá específicamente a los usuarios finales directos del sistema (administradores, autoridades y personal técnico) cuya relevancia práctica es indispensable para validar la usabilidad y la eficacia del paradigma de interacción conversacional propuesto.

3.3. Técnicas de recolección de datos

Las técnicas de recolección se dividen según la naturaleza del dato: Observación Técnica (IoT, Rendimiento), Pruebas de Conjunto de Datos (IA) y Demostración Experimental (Usabilidad, Impacto). Así:

3.3.1. Módulo IoT y Calidad de Datos (Validación de HE1)

- Muestreo continuo de datos (temperatura, humedad, consumo eléctrico - kW y amperaje) a intervalos fijos durante un periodo de prueba.
- Medición simultánea con un instrumento calibrado para obtener valores de referencia, esenciales para cuantificar el error de precisión.

3.3.2. Módulo Predictivo y Avatar Conversacional (Validación HE2 y HE3)

- Aplicación del LLM a un conjunto de datos de prueba etiquetado (test set) con un mínimo de N interacciones para la detección de intención.
- Uso de herramientas de profiling para registrar el tiempo de procesamiento (timestamp) y el consumo de recursos CPU/GPU durante la ejecución de cada componente de la cadena conversacional (ASR, LLM, TTS).
- Obtención de la serie temporal de los residuos del modelo SARIMAX (diferencia entre pronóstico y valor real) para el diagnóstico estadístico.

3.3.3. Ecosistema Integrado y Validación de Impacto (Validación HE4 y HG)

- Monitoreo del rendimiento computacional (uso de CPU/GPU y latencia final) del sistema integral tras la aplicación de técnicas de optimización.
- Recolección del consumo energético global del edificio durante un periodo

de intervención con el ecosistema activo, en contraste con un periodo de referencia (baseline).

- Aplicación del Cuestionario SUS y los ítems de "Atractivo" y "Novedad" del UEQ a los usuarios finales directos del sistema.

3.4. Técnicas de análisis de datos

Las técnicas de análisis se enfocan en contrastar los datos recolectados contra los umbrales de las hipótesis.

3.4.1. *Análisis de Calidad y Superioridad (Validación de HE1 y HE2)*

- Cálculo del Error Absoluto Medio (MAE) entre los datos del sensor IoT y las mediciones del instrumento de referencia.
- Verificación del porcentaje de datos nulos o perdidos en la serie de tiempo.
- Cálculo del Error Cuadrático Medio (RMSE) y Error Porcentual Absoluto Medio (MAPE) para ambos modelos (ARIMA y SARIMAX). La superioridad se demuestra si el RMSE del SARIMAX es significativamente menor.
- Test de Ljung-Box aplicado a los residuos del modelo SARIMAX. El criterio de aceptación es que el p-valor sea mayor que 0.05 ($p > 0.05$) para confirmar la condición de ruido blanco.

3.4.2. *Análisis de Rendimiento y Experiencia (Validación de HE3 y HE4)*

- Cálculo de la Precisión (Accuracy) de detección de intención, verificando que el umbral sea $\geq 95\%$.
- Cálculo del tiempo de respuesta total y del Percentil 95 (P95) de la latencia de la cadena conversacional para confirmar el cumplimiento del requisito de < 10 segundos.
- Comparación de las métricas de consumo de CPU/GPU y latencia obtenidas en el baseline frente a las métricas post-optimización para cuantificar la

reducción porcentual.

3.4.3. *Análisis de Impacto Global (Validación de HG)*

- Análisis de Eficacia (Impacto Energético): Aplicación de una Prueba t de Student (o ANOVA) sobre los datos de consumo energético global para demostrar que la reducción porcentual del consumo durante la intervención del ecosistema es estadísticamente significativa.
- Análisis de Usabilidad (Factor Humano): Cálculo de la Puntuación Única del SUS (Escala 0-100) y su comparación con el umbral de aceptabilidad ($SUS > 68$) para demostrar la usabilidad superior. Las subescalas de UEQ (Atractivo y Novedad) se analizarán para sustentar la naturaleza superior de la experiencia con el avatar.

3.5. Método Ciencia del Diseño en la Investigación

Una característica distintiva de la investigación en ciencias del diseño es que el artefacto o solución constituye la principal contribución de esta investigación al conocimiento. Por consiguiente, en la **Tabla 1** se pueden observar las actividades a realizar y los outputs esperados para cada fase.

Tabla 1 *Solución propuesta para la investigación*

Fase	Actividades	Output
Fase 1: Identificación del Problema y Motivación	<i>Definir el problema</i> (Identificar las brechas de conocimiento) Justificar la <i>relevancia</i> de la investigación.	Estado del arte
Fase 2: Definición de	Conceptualización del "Ecosistema Inteligente"	Objetivos reformulados

Objetivos de la Solución	como la solución.	
	Definición de requisitos funcionales para el artefacto	Especificación de requisitos técnicos de la solución
Fase 3: Diseño y Desarrollo del Artefacto	Desarrollo Módulo IoT	Prototipo funcional e integrado del ecosistema
	Desarrollo Módulo Predictivo	(Avatar conversacional funcional con capacidad de diálogo natural e informativo, datos visuales y comprensibles sobre consumo energético, evidencia empírica de mejora en la eficiencia energética).
	Desarrollo Módulo Conversacional	
Fase 4: Demostración	Uso de la solución en el edificio mediante un estudio experimental controlado.	Evidencia del rendimiento técnico y usabilidad en el contexto real de gestión energética.
Fase 5: Evaluación	Evaluación Predictiva (Ljung-Box, RMSE).	Resultados analizados (confirmación de Ruido Blanco y Superioridad).
	Evaluación Técnica (Rendimiento CPU/GPU, Latencia). Evaluación de Impacto (Estudio Experimental A/B y Cuestionario SUS).	Validación de la Hipótesis General (Reducción de consumo y Usabilidad Superior).
Fase 6: Comunicación	Redacción del documento de tesis doctoral y presentación	Tesis doctoral y publicación de artículo científico con el Aporte de Diseño (Design

de resultados.	Rules) para sistemas inteligentes sostenibles.
----------------	---

A continuación, se describe cada una de las fases del proceso metodológico utilizando Ciencia del Diseño (DSR).

Fase 1: Identificación del Problema y Motivación

La base de esta investigación se establece mediante el análisis riguroso del estado del arte (knowledge base). Esta fase se fundamenta en la revisión sistemática de literatura y desarrollos previos en IoT, predicción de series temporales mediante IA, y modelos de interacción humano-computadora. El conocimiento adquirido sobre las limitaciones de las soluciones actuales (ej., la insuficiencia de modelos predictivos sin variables exógenas y la latencia en interfaces conversacionales) funge como el insumo principal para definir la brecha y el diseño, garantizando el rigor metodológico.

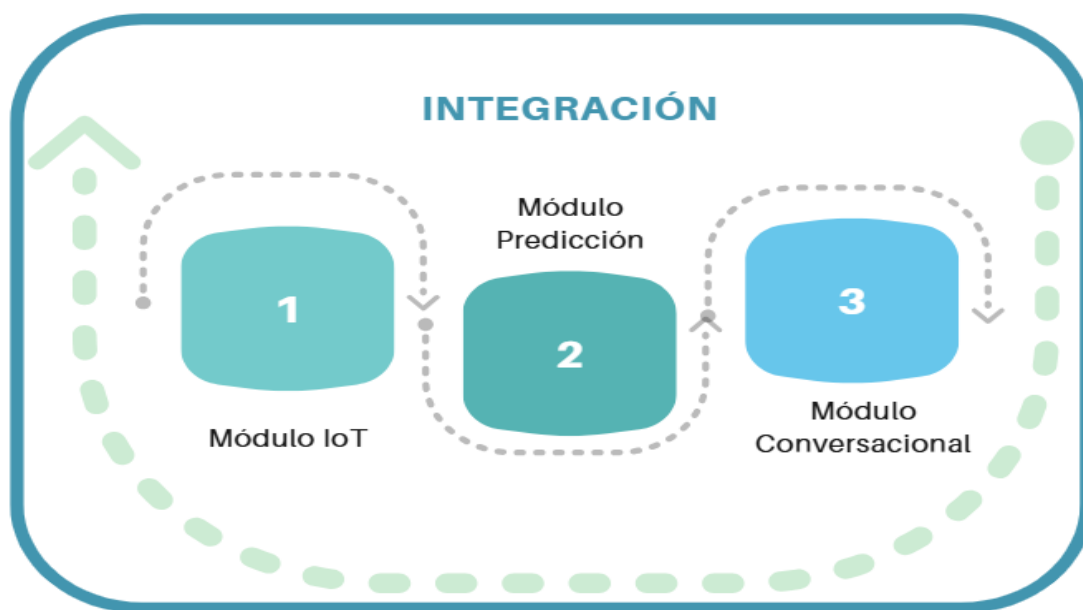
Fase 2: Definición de Objetivos de la Solución

Con base en el problema y la brecha de diseño identificados en la Fase 1, se establecen los Objetivos Generales y Específicos del artefacto. Esta fase traduce los requerimientos funcionales y no funcionales en metas medibles que guiarán la construcción. La conceptualización del Ecosistema Inteligente como la solución se formaliza aquí, asegurando que cada objetivo específico (OE1-OE4) sea una acción directa para resolver un problema específico (PE).

Fase 3: Diseño y Desarrollo del Artefacto

Esta fase representa la construcción tecnológica e iterativa del artefacto (Figura 4) y la materialización de los tres primeros objetivos específicos (OE1, OE2 y OE3), además de la integración inicial del OE4 (Optimización).

Figura 4 Esquema de módulos a desarrollar



Diseño de la Arquitectura.

El componente inicial se centra en la definición de una arquitectura de software y hardware híbrida que asegure la integración sinérgica y el rendimiento en tiempo real. Se establecen los protocolos (ej., MQTT) y los middlewares para la conexión técnica entre los dispositivos IoT y los módulos de IA. Crucialmente, se planifica la estrategia de gestión de datos y se definen los criterios de precisión y calidad (validación de HE1) necesarios para el input del sistema de predicción.

Desarrollo del Módulo IoT

Implica la identificación, adquisición e instalación de sensores y dispositivos IoT en ubicaciones estratégicas del edificio para la captura de datos de consumo (kW) y ambientales (temperatura). Se desarrolla la plataforma de back-end para la gestión de la conectividad segura y la transferencia de datos en tiempo real, culminando con el desarrollo de interfaces visuales básicas para el análisis y pre-procesamiento de datos históricos y en tiempo real.

Desarrollo del Módulo Predictivo

El proceso se inicia con el preprocesamiento riguroso de los datos, incluyendo normalización y partición. La selección del modelo se fundamenta en la literatura para

evaluar alternativas de aprendizaje automático (como SARIMAX frente a ARIMA) para la predicción de patrones. Se procede al entrenamiento y ajuste de hiperparámetros del modelo SARIMAX y se define la estructura para integrar las variables exógenas (Figura 5).

Figura 5 Implementación de un algoritmo de predicción de consumo energético



Desarrollo del Módulo Conversacional

Esta fase se enfoca en la ingeniería de la interfaz de usuario (UI) y la experiencia de usuario (UX) del avatar. Se realiza el análisis de requisitos para la interacción (NLU) y se desarrolla el diseño físico y de personalidad del avatar. Se implementa el stack de Machine Learning (ASR, NLU, TTS), integrando el algoritmo SARIMAX para la generación de respuestas personalizadas y predictivas. Se realizan pruebas funcionales para asegurar la claridad de las recomendaciones y la capacidad de diálogo.

Fase 4: Demostración

Esta fase se concentra en la ejecución del prototipo integral y optimizado en el entorno real del edificio (Facultad de Ciencias Humanísticas, UTM). Se lleva a cabo un estudio experimental controlado y se procede a la recolección de datos empíricos. El objetivo es obtener la evidencia directa y los datos de rendimiento para demostrar que el artefacto funciona en el contexto para el cual fue diseñado. Esto incluye la medición continua de la latencia y el consumo de recursos (logging) después de la optimización (OE4) y la obtención de los datos de consumo energético durante el

periodo de intervención.

Fase 5: Evaluación

Este proceso se centra en el análisis riguroso de los datos recolectados en la Fase 4 para validar o refutar las hipótesis (HE1-HE4 y HG).

Evaluación de Rigor y Superioridad (HE2): Análisis de los errores de predicción RMSE/MAPE y la aplicación del Test de Ljung-Box sobre los residuos para validar la validez estadística del modelo final (condición de ruido blanco).

Evaluación Técnica (HE3 y HE4): Análisis comparativo de las métricas de latencia y uso de recursos CPU/GPU antes y después de la optimización, para confirmar la viabilidad en tiempo real latencia < 10 s y la mejora significativa de la eficiencia computacional.

Evaluación de Impacto (HG):

- **Eficiencia:** Contraste estadístico de los datos del consumo energético del edificio (periodo de intervención vs. baseline) para demostrar la reducción medible del consumo.
- **Usabilidad:** Análisis de las puntuaciones del cuestionario SUS para validar la usabilidad superior del avatar.

Fase 6: Comunicación

Los hallazgos de la investigación se consolidan en el documento de tesis doctoral. Esta fase se enfoca en la presentación formal del Problema, la descripción del Artefacto, el rigor de su Diseño y la Efectividad demostrada por la Evaluación. El aporte central a la comunidad científica y profesional es la generación de Reglas de Diseño (Design Rules) generalizables, derivadas de la validación exitosa de las hipótesis, que orientan la construcción de futuros sistemas inteligentes sostenibles.

3.6. Matriz de consistencia

Tabla 2 *Consistencia entre problema – Objetivos y Resultado Esperados*

Problema Específico	Objetivo Específico	Hipótesis Específica	Técnicas de Recolección de Datos	Técnica de Análisis	de Output DSR (Artefacto)
PE1: Ausencia de sistemas de monitorización en tiempo real que permitan medir de manera precisa y continua parámetros ambientales (temperatura, humedad) dentro de los edificios inteligentes.	OE1: Desplegar el módulo IoT para la captura, procesamiento y almacenamiento en la nube de datos ambientales (temperatura, humedad) y de consumo eléctrico.	HE1: El módulo IoT proporciona datos de consumo energético con precisión y continuidad.	Recolección en continuo de datos (temperatura, humedad, consumo) a intervalos fijos durante un periodo de prueba.	Logs del sistema (continuidad); Análisis de QoS (integridad y uptime).	Software de captura y monitoreo de datos desde sensores (IoT)
PE2: Limitación en la	OE2: Implementar un	HE2: El modelo	Obtención de la	Estadísticos	Modelo

confiabilidad y efectividad predictiva de los métodos actuales de pronóstico del consumo energético, debido a la insuficiente especificación estadística de los modelos utilizados y la omisión de variables exógenas relevantes lo que impide la obtención de un modelo robusto cuyos residuos se comporten como ruido blanco, comprometiendo la validez de las predicciones a corto plazo.	y modelo de predicción del consumo energético utilizando los datos históricos y variables exógenas para estimar la demanda futura.	SARIMAX predice el consumo energético de manera superior a ARIMA al considerar variables exógenas, logrando un error de predicción (medido por RMSE o MAPE) significativamente menor y presentando residuos que satisfacen la condición de ruido blanco.	serie temporal de los residuos tras el entrenamiento y el testing.	Ljung-Box y Test de Shapiro-Wilk; Análisis de Residuos (Normalidad e Independencia).	SARIMAX implementado y validado
---	--	--	--	--	---------------------------------

PE3: Limitación técnica en el desarrollo de un avatar conversacional para sistemas de gestión que ofrezca interacción en tiempo real.	OE3: Desarrollar un avatar conversacional que permita la interacción con el sistema en tiempo real, garantizando precisión y comunicación fluida.	HE3: El módulo conversacional cumple con los requisitos de un sistema en tiempo real, garantizando una precisión en la detección de intención superior al 95% y tiempos de respuesta inferiores a 10 segundos.	Registro automático de interacciones con un corpus de prueba (test set).	Cálculo de la Matriz de Confusión (Precisión, Recall y F1-Score); Tiempos de respuesta en la cadena de interacción	Diseño y desarrollo del Módulo HCI conversacional.
PE4: La integración de modelos de IA local complejos en el prototipo híbrido impone una carga computacional excesiva y latencia, poniendo en riesgo la operación en tiempo real, la	OE4: Optimizar el cálculo computacional para el funcionamiento del prototipo integral.	HE4: La implementación de técnicas de optimización resultará en una mejora significativa de la eficiencia computacional del sistema, demostrada por	Pruebas de Carga y Estrés (Benchmark): Registro continuo del consumo de recursos (CPU y GPU) y la	Análisis Comparativo de Rendimiento (Benchmarking): Cálculo de la tasa de optimización computacional antes y después de	Interfaz que permite medir el rendimiento del funcionamiento del artefacto en función al consumo de

escalabilidad y la sostenibilidad del sistema.	una reducción medible del consumo de recursos (CPU y GPU) y la latencia.	latencia.	la optimización.	CPU y GPU.
Problema General: La integración de modelos de IA local complejos (LLMs, TTS/ASR) en el prototipo híbrido IoT impone una carga computacional excesiva y latencia. Esto compromete el rendimiento óptimo y la escalabilidad del sistema, haciendo indispensable la optimización para asegurar una gestión energética eficaz.	Objetivo General: Diseñar y validar un Ecosistema Inteligente que, mediante la integración sinérgica de monitorización IoT en tiempo real, predicción de consumo basada en IA (SARIMAX) y una interfaz de avatar conversacional, demuestre una mejora significativa en la gestión energética del edificio, garantizando	Hipótesis General: La arquitectura tecnológica propuesta, resultante de la integración sinérgica de la monitorización IoT en tiempo real, la predicción de consumo basada en IA (SARIMAX) y la interfaz de avatar conversacional, demuestra una mejora significativa en la gestión energética del edificio y una usabilidad superior en	Encuestas de usabilidad y satisfacción. Análisis comparativo de métricas energéticas y rendimientos históricos.	Prototipo funcional e integrado del ecosistema.

una usabilidad superior la interacción
y un rendimiento comparada con métodos
computacional óptimo. tradicionales, operando
bajo condiciones de
rendimiento
computacional óptimas
(demostrado por la baja
latencia y el uso
eficiente de recursos
CPU/GPU).

CAPÍTULO 4. CONSTRUCCIÓN DE LA SOLUCIÓN (FASE 3)

Este capítulo expone el procedimiento de creación física y formal de los artefactos tecnológicos (outpus) definidos en la matriz, demostrando que los prototipos funcionales han sido contruidos e integrados como principal aporte de la presente investigación doctoral. En consonancia con el Objetivo General: Construir un ecosistema inteligente validado, basado en la integración sinérgica de monitorización IoT en un edificio, predicción de consumo basada en IA y una interfaz de avatar conversacional, el enfoque se adhiere al marco de la Ciencia del Diseño (Design Science Research - DSR), cuyo propósito no es solo describir o explicar un fenómeno, sino construir una solución innovadora para un problema relevante del mundo real y evaluar su eficacia.

Este capítulo describe las decisiones de ingeniería, las tecnologías específicas seleccionadas (protocolos IoT, bases de datos de series temporales, Machine Learning, tecnologías para crear avatares conversacionales, el cálculo computacional y la integración arquitectónica de estos módulos. La implementación se estructura siguiendo las fases de Diseño y Desarrollo (Fase 3) del modelo DSR, donde el objetivo es construir: 1) Software de captura y monitoreo de datos (IoT). 2) El Modelo SARIMAX implementado. 3) El Avatar conversacional desarrollado. 4) Interfaz de medición de rendimiento (CPU y GPU).

La correcta ejecución de este procedimiento de construcción del artefacto sienta las bases técnicas para comprobar si la arquitectura tecnológica propuesta, resultado de la integración sinérgica, demuestra una mejora significativa en la gestión energética del edificio y una usabilidad superior comparada con métodos tradicionales declarada en la Hipótesis General.

4.1. Diseño del Módulo IoT

En esta etapa, se diseña una infraestructura de Internet de las Cosas (IoT) para la captura de datos en tiempo real, lo que permite la recolección de información, incluyendo la medición directa del consumo de energía a través de sensores instalados y la monitorización de estos datos en tiempo real mediante una interfaz web.

Este proceso implica la identificación y adquisición de dispositivos IoT

adecuados para medir y registrar el consumo de energía en edificios inteligentes, como los sensores de energía. Se procedió a la instalación y configuración de los dispositivos IoT en puntos estratégicos dentro del edificio para capturar datos precisos y representativos. Una vez definido el hardware, se procede a crear una plataforma informática que permita la integración y gestión de los dispositivos IoT. Se definieron aspectos relacionados con la conectividad y comunicación para proporcionar una conexión segura y confiable entre los dispositivos IoT y la plataforma centralizada para la transferencia de datos en tiempo real, además de establecer protocolos para la recopilación de datos en tiempo real o de manera periódica.

En este sentido, se describen cada uno de los procedimientos realizados en la creación del módulo IoT: Diseño del Hardware, Diseño del Software y Protocolos de comunicación entre hardware y software. Así:

4.1.1. Diseño del Hardware

Los sensores instalados en diferentes puntos del edificio recopilan datos de consumo de energía y temperatura, que se envían al microcontrolador Arduino. La programación del microcontrolador Arduino incluye la configuración de los sensores, la lectura de datos, y la transmisión de estos datos a la base de datos. Todos los componentes se integran en una placa base, y se aseguran de que las conexiones eléctricas sean correctas. Los elementos integrados incluyen entre otros:

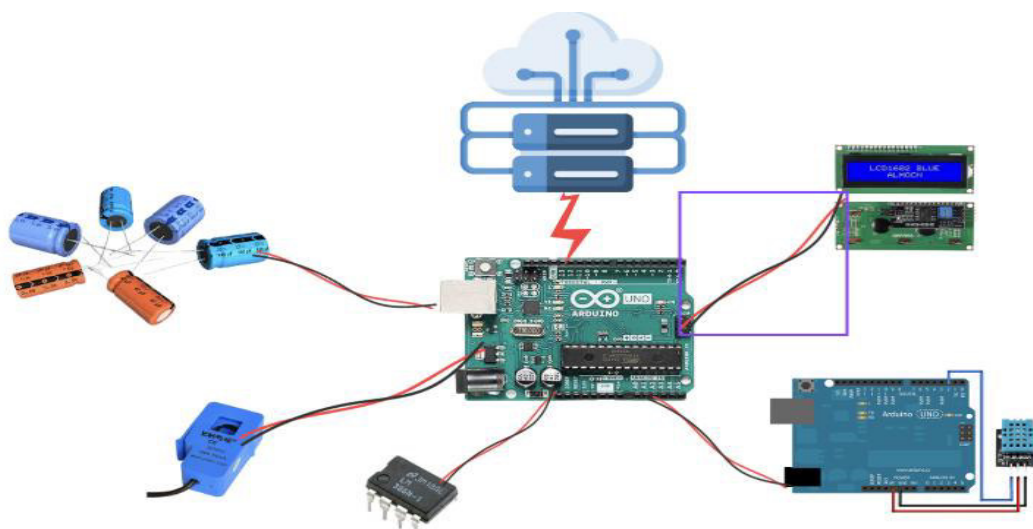
El sensor SCT-013 es un sensor de corriente no invasivo que mide la intensidad de la corriente eléctrica que circula por un cable. Este sensor funciona en modo "pinza amperimétrica", lo que significa que puede medir la corriente sin necesidad de interrumpir el circuito eléctrico. Se utiliza principalmente en aplicaciones de monitoreo de consumo eléctrico debido a su capacidad para proporcionar mediciones precisas y su facilidad de uso.

El sensor de temperatura DHT11 es un módulo que permite medir la temperatura en grados Centígrados o Fahrenheit y la humedad relativa del ambiente. Tiene un rango de operación entre 0°C y 50°C. Para la humedad, tiene un rango de operación de 20% y el 90% de humedad relativa. Los componentes necesarios para el monitoreo eléctrico son:

- Placa Arduino Uno: Microcontrolador utilizado para procesar los datos recibidos del sensor.
- Placa Arduino Shield Ethernet: Permite la conexión a internet para la transmisión de datos.
- Sensor SCT-013: Sensor de corriente no invasivo.
- Resistor de carga: Utilizado para convertir la corriente a un voltaje medible por el Arduino.
- Amplificador LM358N: Amplificador operacional para mejorar la señal.
- Condensadores: Utilizados para estabilizar la señal.
- Pantalla LCD 16x2 con módulo I2C: Para mostrar las lecturas en tiempo real.

La Figura 6 muestra el esquema del sistema integrado, el cual incluye: componentes de hardware utilizado, protocolo de comunicación, software diseñado e integración del sistema de monitorización.

Figura 6 *Esquema propuesto del Módulo IoT*



El desarrollo del prototipo consistió en ensamblar todos los componentes según el esquema mostrado en la Figura 7 y el esquema 3D del circuito Arduino se visualiza en la Figura 8, además de programar el sistema para que funcione correctamente. Para el funcionamiento del sensor de temperatura DHT11, se conecta a un

microcontrolador, como es la placa Arduino. La conexión se realiza utilizando tres pines: uno para la alimentación (VCC), otro para la señal (DATA) y el tercero para la tierra (GND).

Figura 7 Esquema del circuito de la placa Arduino

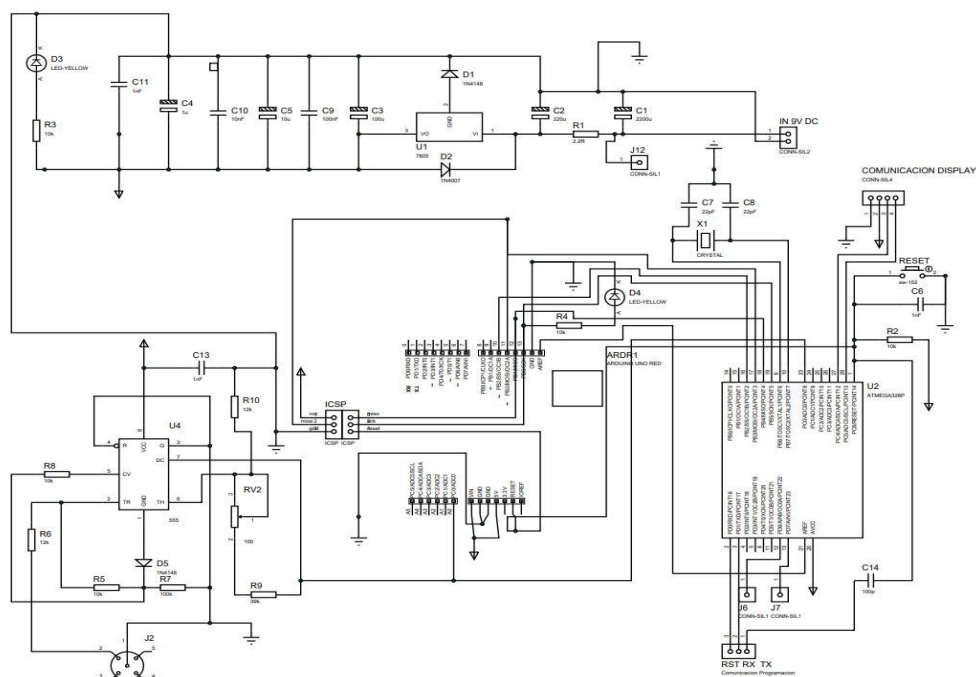
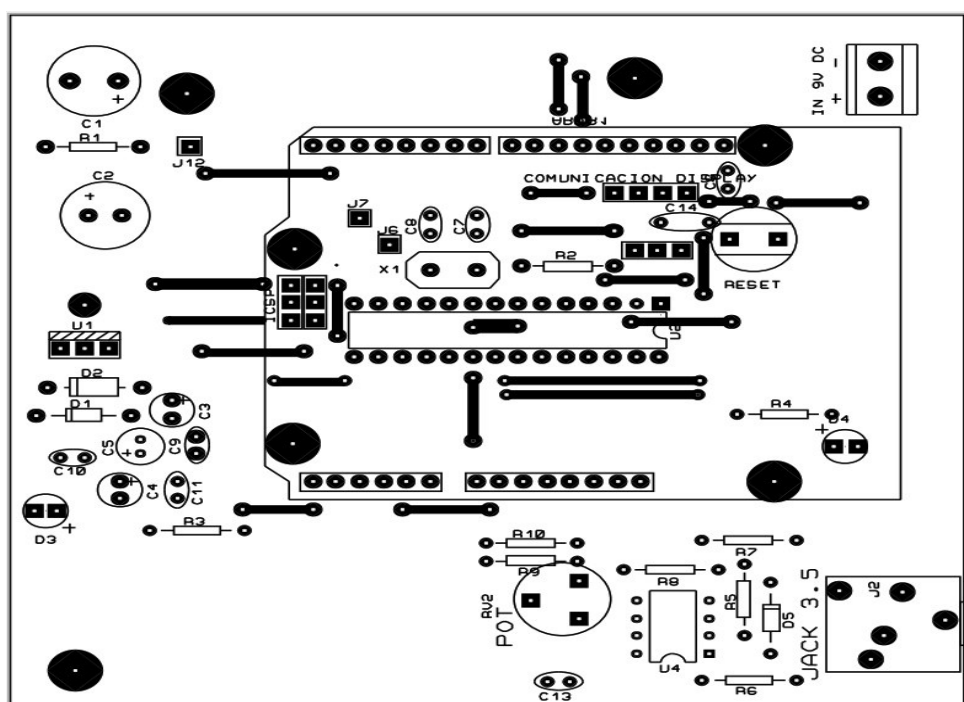


Figura 8 Esquema 3D del circuito de la placa Arduino



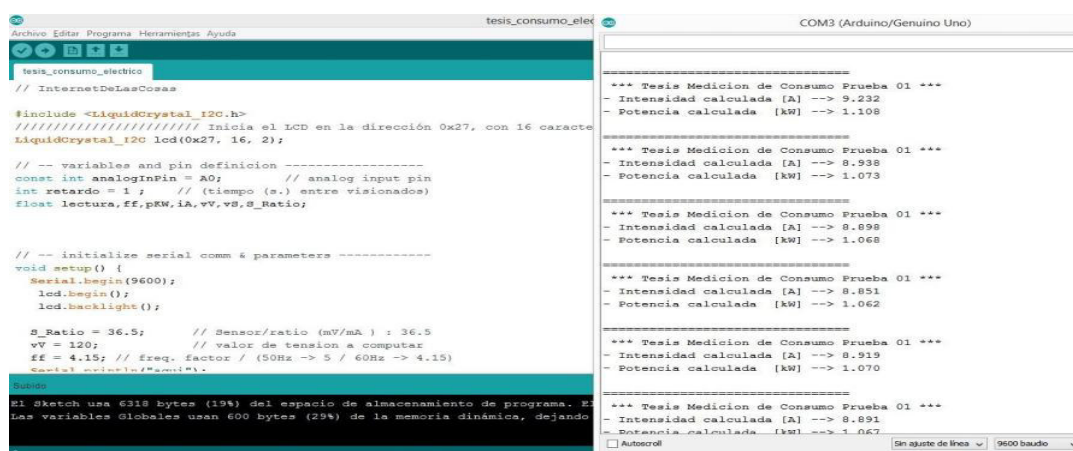
Se utiliza el entorno de desarrollo integrado (IDE) de Arduino para escribir y cargar el código en la placa Arduino. En el código, se incluye la biblioteca DHT que facilita la comunicación con el sensor y la lectura de datos de temperatura y humedad. Una vez cargado el código, el microcontrolador comienza a recibir datos del sensor de temperatura. El sensor envía pulsos eléctricos que el microcontrolador convierte en datos digitales representativos de la temperatura y la humedad. Todos los componentes se integran en una placa base (Figura 9), asegurando que las conexiones eléctricas sean correctas.

Figura 9 Placa física fabricada de forma artesanal



Se utiliza el entorno de desarrollo integrado (IDE) de Arduino para escribir y cargar el código en la placa Arduino (Figura 10).

Figura 10 Programación en el IDE Arduino



En el código, se incluye la biblioteca DHT que facilita la comunicación con el

sensor y la lectura de datos de temperatura y humedad. Una vez cargado el código, el microcontrolador empieza a recibir datos del sensor de temperatura. El sensor envía pulsos eléctricos que el microcontrolador convierte en datos digitales representativos de la temperatura y la humedad.

4.1.2. Creación del Software

El sistema de monitoreo web fue desarrollado utilizando herramientas avanzadas de desarrollo web, con un enfoque particular en JavaScript debido a su versatilidad, facilidad de uso y adaptabilidad. Este lenguaje de alto nivel facilitó la creación eficiente de los módulos necesarios para el proyecto. Un aspecto crucial del diseño del sistema fue la creación de una interfaz de usuario amigable, intuitiva y fácil de usar. Se consideró esencial la adaptación del usuario final para asegurar que cualquier persona pudiera interactuar con el sistema sin dificultad, optimizando así la experiencia del usuario.

La arquitectura empleada para el desarrollo del sistema fue el Modelo Vista Controlador (MVC), que divide el sistema web en dos partes principales: el Back-end y el Front-end. Esta separación facilita el desarrollo, mantenimiento y escalabilidad del sistema. El desarrollo del Back-end permite gestionar la lógica del manejo de los datos, utilizando servidores, bases de datos y APIs para manejar las solicitudes del cliente, procesar los datos del prototipo y almacenarlos en una base de datos. De esta manera, el Back-end recibe los datos del prototipo de monitoreo de energía, los procesa y los guarda en una base de datos PostgreSQL, incluyendo mediciones de consumo de energía en tiempo real.

Por otro lado, el Front-end permite presentar los datos procesados al usuario de manera intuitiva y visualmente atractiva. Las interfaces de usuario se desarrollan utilizando HTML, CSS y JavaScript, integrándolas con las APIs del Back-end para obtener y visualizar datos en tiempo real. Se emplean gráficos y tablas con el uso de librerías de gráficos como Chart.js para mostrar el consumo de energía, permitiendo a los usuarios monitorear y analizar los datos de manera eficiente. Con ello, se logra que la información se transmita a un servidor donde se almacena en una base de datos y se presenta a través de una interfaz web que permite la visualización gráfica y el análisis en tiempo real.

4.1.3. Protocolo de comunicación

Durante el transcurso de la implementación, se realizaron varios cambios significativos para mejorar el rendimiento y la fiabilidad del sistema. Así, inicialmente se planteó utilizar el protocolo TCP/IP, sin embargo, debido a la gran cantidad de paquetes transmitidos, se optó por cambiar al protocolo MQTT (Message Queuing Telemetry Transport), que es más eficiente para el manejo de grandes volúmenes de datos en tiempo real.

Los firewalls de la institución bloqueaban ocasionalmente las IP de salida de los sensores, por lo cual se tuvieron que reprogramar las placas de Arduino para que la asignación de IP sea dinámica y se realice a través de un dominio, permitiendo una mayor flexibilidad y evitando bloqueos por parte del firewall. Además, debido a las dificultades encontradas como bloqueos ocasionales de salida web, se optó por levantar un servidor local que realiza lecturas paralelas de los sensores. Este servidor actúa como respaldo en caso de que el acceso web sea bloqueado, asegurando que la recolección de datos no se interrumpa.

4.2. Selección del Modelo de Predicción

Este componente se centró en realizar una comparación eficaz, se estudiaron diversos algoritmos de aprendizaje automático que han sido empleados con éxito en la predicción del consumo energético. Entre los modelos evaluados (

Tabla 3), se incluyeron: Máquinas de Soporte Vectorial (SVM / SVR), Redes Neuronales Artificiales (ANN), Memoria a Corto y Largo Plazo (LSTM), Media Móvil Autorregresiva Integrada (ARIMA), Media Móvil Integrada Autorregresiva Estacional con Variables Exógenas (SARIMAX) y Modelos Híbridos (SARIMAX - LSTM).

Se establecieron criterios clave para la comparación de los modelos (Tabla 4): capacidad para manejar relaciones lineales y no lineales en los datos de consumo energético, eficiencia en el manejo de datos estacionales y factores exógenos, como la temperatura ambiente u otros parámetros que puedan influir en el consumo energético y requerimientos computacionales y disponibilidad de datos: algunos algoritmos requieren grandes cantidades de datos y recursos computacionales, lo que puede ser

una limitación dependiendo del contexto del proyecto.

Con base en la revisión y comparación de algoritmos se seleccionó el modelo **SARIMAX** para la predicción del consumo energético en edificios inteligentes. Este modelo fue elegido debido a su capacidad para capturar las tendencias lineales y estacionales presentes en los datos de consumo, así como su efectividad al trabajar con series temporales con patrones estacionales. A pesar de sus limitaciones con relaciones no lineales y factores exógenos complejos, su desempeño en situaciones con datos limitados y fuertes patrones estacionales lo convierte en la opción más adecuada para esta investigación.

En el proceso de selección también se incluyeron pruebas preliminares utilizando otros modelos como **LSTM** y **ANN**, que, aunque mostraron una excelente capacidad para capturar relaciones complejas y no lineales, requerían un mayor volumen de datos y recursos computacionales significativos. Estos factores fueron considerados como limitantes para la implementación en el contexto de la investigación, donde la disponibilidad de datos y los recursos computacionales eran factores para tomar en cuenta.

Tabla 3 *Análisis comparativo de modelos de predicción analizados*

Modelo	Fortalezas	Debilidades	Mejores Casos de Uso
SARIMAX	Capta bien las tendencias lineales y la estacionalidad	Dificultad con relaciones no lineales y factores exógenos complejos	Datos limitados, fuertes patrones estacionales
SVM	Bueno para datos no lineales con variables de entrada limitadas	La precisión puede verse limitada por la falta de datos completos	Relaciones no lineales, datos limitados

ANN	Excelente para capturar relaciones no lineales complejas	Requiere grandes conjuntos de datos y recursos computacionales	Datos complejos y no lineales
LSTM	Eficaz para series temporales con dependencias a largo plazo	Computacionalmente intensivo, requiere grandes conjuntos de datos	Previsión de carga máxima, dependencias a largo plazo
Modelos Híbridos	Combina las fortalezas de los modelos lineales y no lineales, con precisión elevada	Mayor coste computacional, complejidad en el entrenamiento y ajuste	Escenarios que requieren alta precisión y robustez

4.2.1. *Proceso Extracción, Transformación y Carga (ETL)*

Para que un sistema pueda predecir valores futuros, se emplea forecasting. El Forecasting es un proceso crucial que se utiliza en diversos campos para predecir eventos o condiciones futuras basándose en datos históricos, tendencias y patrones. Desempeña un papel fundamental en la toma de decisiones, la planificación y la evaluación de nuevas ideas. La precisión de los pronósticos depende de la calidad de los datos, los métodos utilizados y las suposiciones hechas sobre el futuro.

Entonces, para poder realizar la predicción se necesita de la data de consumo energético, esta consulta se la realiza a la base de datos con la información que menciona el usuario, de donde se obtienen. Para lograr conseguir los mejores resultados en predicción de datos de consumo, se realizaron pruebas con los modelos de Forecasting ARIMA y SARIMAX con el histórico de los datos de consumo energético agrupados por días, los cuales se obtuvieron de la base de datos.

Toda esta data que se usará para entrenar el modelo de forecasting, pasa por un proceso crítico en la integración de datos denominado ETL (Extraer, Transformar, Cargar). Este proceso implica extraer datos de diversas fuentes, transformarlos a un formato adecuado y cargarlos en un almacén o repositorio de datos (Prema & Pethalakshmi, 2013; Qin et al., 2012; Younes et al., 2022), garantizando que los datos estén limpios, estandarizados y listos para el análisis, lo cual es esencial para realizar forecasting precisos.

La información de consumo energético se obtiene de la base de datos a partir de los parámetros de la consulta del usuario, esta data pasa por un proceso de limpieza en donde se selecciona solo la data real sin variaciones (a partir del 16 de mayo del 2025), además se rellenan las fechas faltantes con valores de consumo promedio de la columna, además de reducir la dimensionalidad al seleccionar la columna a evaluar en el forecasting. Esta información se carga en un dataframe de pandas (Figura 11) para ser operada por los algoritmos de los modelos de forecasting en Python.

Figura 11 Data luego de aplicar el proceso de ETL

	total_kilovatio	feriado	evento_especial	total_temperatura
2025-05-16	48359.54	1.0	0.0	19.24
2025-05-17	41744.29	0.0	1.0	22.64
2025-05-18	41501.16	1.0	0.0	28.36
2025-05-19	55133.92	1.0	1.0	23.01
2025-05-20	58050.27	0.0	1.0	28.65
...	...	...	...	...
2025-08-14	47113.76	1.0	0.0	18.82
2025-08-15	43246.61	1.0	1.0	21.86
2025-08-16	46535.96	0.0	0.0	24.03
2025-08-17	44679.08	0.0	1.0	25.97
2025-08-18	55793.37	0.0	1.0	29.94

95 rows × 4 columns

4.2.2. Entrenamiento del Modelo de Predicción ARIMA

El primer modelo escogido para realizar pruebas es ARIMA, un modelo estadístico ampliamente utilizado para analizar y pronosticar datos de series temporales univariadas. ARIMA captura las relaciones lineales dentro de los datos mediante la combinación de tres componentes: autorregresivo (AR), integrado (I) y media móvil (MA).

ARIMA, sigue una serie de pasos para lograr una correcta construcción y que permita obtener valores de predicción correctos. En este orden, los pasos (Figura 12) a seguir son: 1. Se determina el orden de los componentes AR, I, MA; 2. Se estiman los parámetros del modelo; y, por último, 3. Se comprueba la adecuación del modelo mediante pruebas diagnósticas.

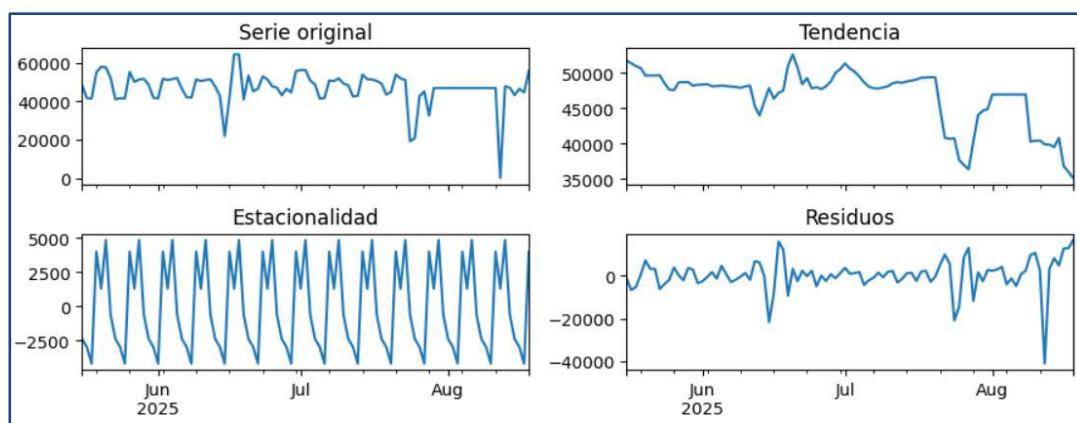
Figura 12 Entrenamiento del Modelo ARIMA

Pasos para el entrenamiento del Modelo ARIMA



1 Determinación del orden de los componentes

Lo primero que se hace a los datos para tener una visión general de la información con la que estamos a punto de entrenar el modelo, es la descomposición de series temporales (Figura 13), que se trata de descomponerla en sus componentes fundamentales: tendencia, estacionalidad y residuos. Para ello, se debe determinar los parámetros p , d y q los cuales representan AR, I y MA respectivamente.

Figura 13 *Descomposición de la serie temporal*

Un modelo ARIMA requiere que la serie sea estacionaria, esto sucede cuando su media, varianza y auto covarianza son invariantes en el tiempo. Para determinar el parámetro d se evalúa observando si existe estacionariedad en la descomposición de la serie temporal presentada anteriormente o haciendo un Test de Dickey Fuller, la cual es una prueba estadística que verifica si la raíz unitaria es una serie de tiempo.

Al hacer la prueba de Dickey Fuller se obtiene un valor de $1.17e-10$, lo que cumple con la hipótesis de $p\text{-value} > 0.05$ y por lo tanto la serie es estacionaria (Figura 14). Si en caso la serie no fuese estacionaria habría que diferenciarla tantas veces sea necesario hasta que logre serlo.

Figura 14 *Prueba de Dickey Fuller para obtener p -valor en ARIMA*

```
from statsmodels.tsa.stattools import adfuller
adf_result = adfuller(df_full['total_kilovatio'])
print(f'ADF Statistic: {adf_result[0]}, p-value: {adf_result[1]}') #d=1
```

ADF Statistic: -7.324030771854943, p-value: 1.173772885992285e-10

2 Estimación de los parámetros del modelo

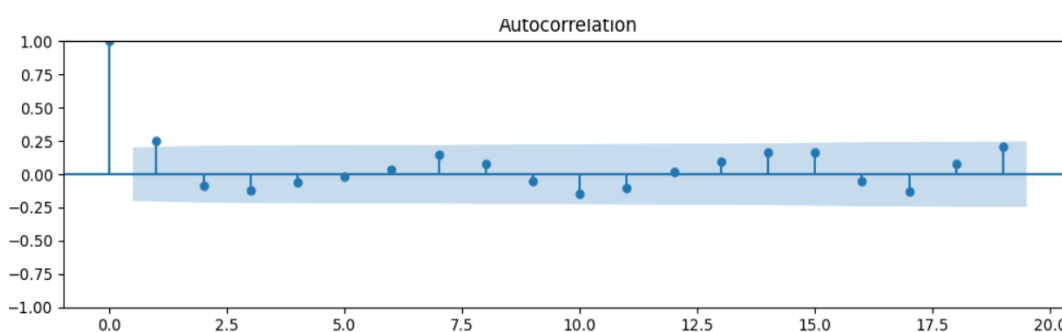
Para determinar los valores que tendrán p y q , se evalúa la data con dos funciones: la función de autocorrelacion (ACF) para determinar el valor de q , y la función de autocorrelacion parcial (PACF) para determinar p .

El modelo ACF mide la correlación entre observaciones con diferentes desfases temporales, identifica componentes de media móvil (MA) mostrando como

valores pasados influyen en valores actuales. Por otro lado, PACF mide la correlación entre observaciones con diferentes rezagos, pero tras eliminar los efectos de rezagos más cortos, sirve para determinar AR, ya que muestra la relación directa entre una observación y sus valores rezagados.

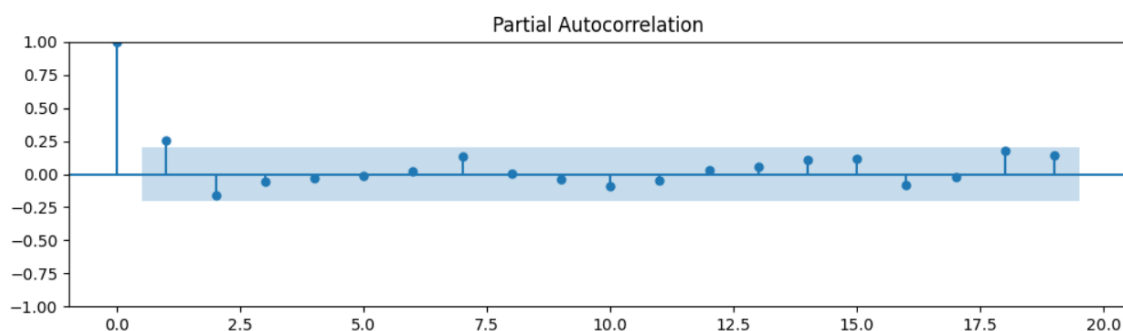
Revisando la gráfica de ACF en la Figura 15 se puede determinar que el posible valor para q podría ser 1 ya que el primer rezago tiene un pico que supera de manera significativa las bandas de confianza.

Figura 15 *Función de autocorrelación ACF para determinar valor de q con ARIMA*



En la Figura 16 se puede observar que los primeros dos rezagos tienen picos que superan o están cerca de las bandas de confianza, por lo que los posibles valores de p están entre 1 o 2.

Figura 16 *Función de autocorrelación parcial PACF para determinar valor de p con ARIMA*



Una vez se tienen los posibles valores de cada componente, podemos realizar una estimación de los posibles modelos, que siguiendo la forma del modelo ARIMA se construye como (p,d,q) . Se tienen 2 posibles combinaciones: Modelo1 = $(1,1,1)$ y Modelo2 = $(2,1,1)$.

Para determinar el mejor modelo a utilizar, se entrenan todos estos posibles modelos con la función ARIMA de python y verificar los valores de AIC y BIC que tienen cada uno. AIC permite seleccionar el modelo que minimice la pérdida de información, BIC pretende seleccionar el modelo con la mayor probabilidad posterior dados los datos. En la Figura 17 se aprecia que el modelo 1 es el que menor AIC y BIC representa, por lo tanto, será el modelo ideal.

Figura 17 Obtención del mejor modelo para ARIMA

	Model	Order	AIC	BIC
0	Model1	(1, 1, 1)	1967.485791	1975.115676
1	Model2	(2, 1, 1)	1968.903729	1979.076908

Modelo óptimo según AIC: Model1 (1, 1, 1)

3 Comprobar adecuación del modelo

Procedemos a validar que los residuos sean Ruido Blanco, es decir que tengan una distribución normal y no exista autocorrelación entre ellos. Para ello, la prueba de Ljung-Box nos permite evaluar si existe o no autocorrelación entre los residuos, debido a su gran desempeño en determinar la existencia de autocorrelaciones significativas en los residuos de un modelo de series de tiempo, lo que ayuda en la validación y selección del modelo.

En la Figura 18 se puede observar que el p-value arrojado por la prueba de Ljung-Box es de 0.83, y dado que la hipótesis dice que si p-value es mayor a 0.05 entonces los residuos son ruido blanco y por lo tanto no existe correlación.

Figura 18 Verificación de existencia de autocorrelación de los residuos

	lb_stat	lb_pvalue
20	14.001592	0.830415

Para validar si los residuos se distribuyen de manera normal, se aplica el test de normalidad de Shapiro-Wilk el cual es el mejor para comprobar esa distribución. Como se puede apreciar en la Figura 19, el valor p es menor que un nivel de significancia de 0.05, lo que se podría concluir en que la muestra no sigue una

distribución normal.

Figura 19 Validación de distribución normal de los residuos

Estadístico de Shapiro-Wilk: 0.7285

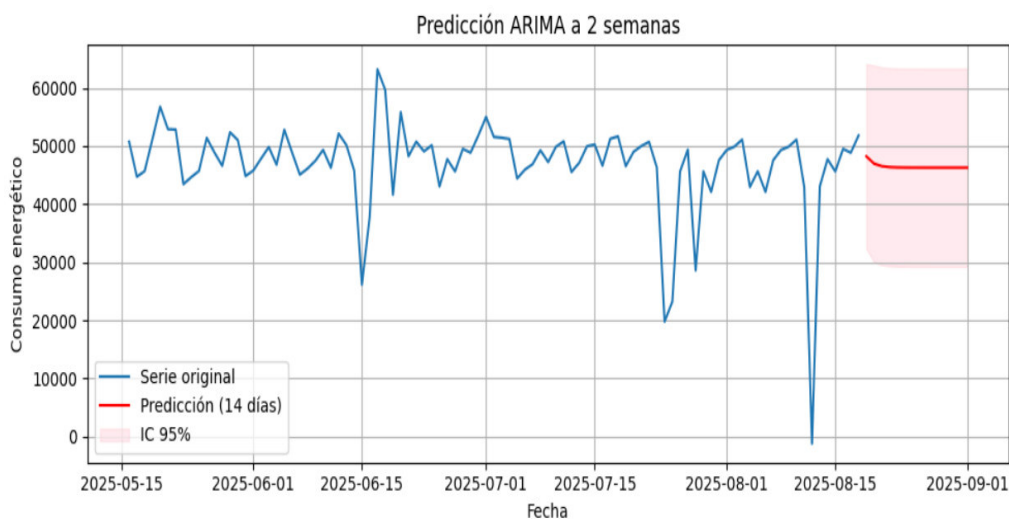
Valor-p: 0.0000

Se rechaza H_0 : los residuos NO parecen seguir una distribución normal.

Una vez realizado todos los pasos, y comprobar que todos los parámetros elegidos son correctos, se procede a realizar la predicción con la data original del consumo energético, pasándole el conjunto de componentes (p, d, q) encontrados en los pasos anteriores, entrenando el modelo y prediciendo para los siguientes 14 días. En la Figura 20, se observa que los valores de predicción a partir del 25 de agosto comienzan a ser casi similares, no hay mucha variación, y esto se puede deber a que no se están considerando posibles eventos que puedan ocurrir en esas fechas.

Es importante destacar que la razón del por qué se desea predecir a 14 días si se requiere que el sistema realice la predicción a una semana, es debido a los datos incompletos reales de la semana en la que el usuario se encuentra. Las predicciones necesitan de datos reales, y si nos encontramos, por ejemplo, en un miércoles, no tenemos los datos completos de la semana, por lo que tenemos que entrenar con datos reales del domingo de la semana anterior, para predecir los datos de toda la semana actual desde el lunes hasta el domingo de la siguiente semana. De esta forma, solo seleccionamos la predicción de la última semana.

Figura 20 Resultado de la predicción ARIMA

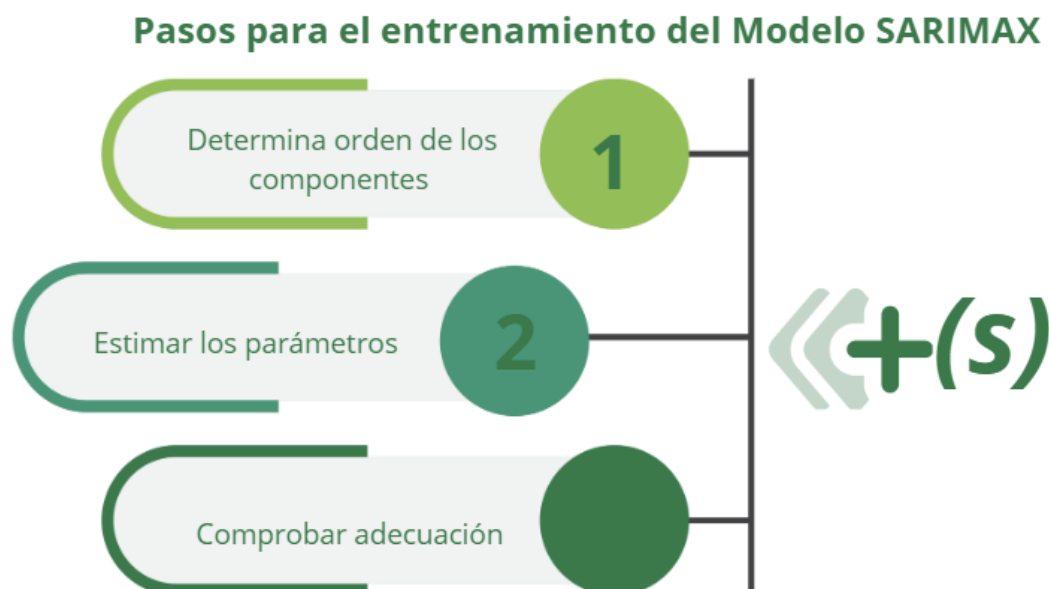


4.2.3. Entrenamiento del Modelo de Predicción SARIMAX

Luego de ver que la predicción del modelo de ARIMA no daba resultados tan precisos, y tomando en cuenta cualquier novedad que pueda ocurrir en el tiempo, decidimos probar con el modelo SARIMAX. Este modelo extiende ARIMA al incorporar variables estacionales (S) y exógenas (X). Al incluir variables estacionales, este agrega componentes autorregresivos estacionales (P), de diferenciación estacional (D), de promedio móvil estacional (Q) y la estacionalidad (s), además de las variables exógenas que mejora la precisión del modelo.

Para obtener un resultado preciso se necesitó de una gran cantidad de datos con todas las variaciones de las variables exógenas de modo que el algoritmo pueda aprender de todas las posibles situaciones que han ocurrido y estar preparado para predecir de manera más acertada el consumo de la siguiente semana. Los pasos (Figura 21) que seguir para la construcción de SARIMAX son los mismos que los de ARIMA, pero añadiendo los componentes estacionales y las variables exógenas.

Figura 21 Pasos para entrenamiento del Modelo SARIMAX

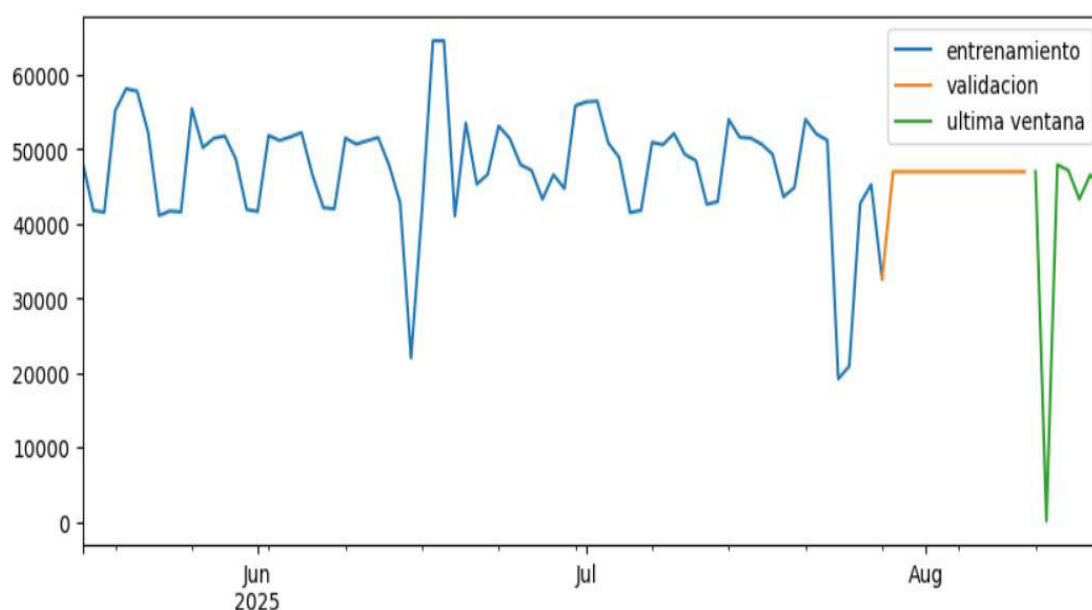


Ya que se hace uso de la misma data de consumo para entrenar el modelo SARIMAX, los componentes extraídos de la descomposición de la serie temporal son los mismos, al igual que los parámetros p , d , q obtenidos, solo queda obtener los componentes estacionales. Como en la anterior prueba el valor de p fue de 1, el valor

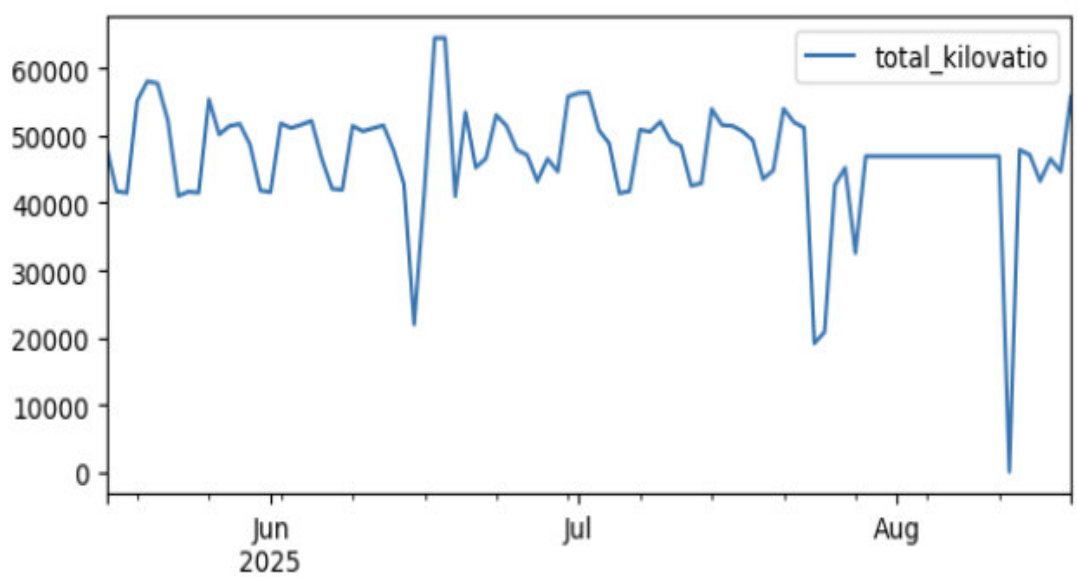
de d será de 1 también y s será de 14 ya que se desea predecir para 14 días. El resto de los valores para p y q podrían estar entre 0, 1 o 2. Con todos estos posibles valores se pueden obtener algunos posibles modelos para entrenar SARIMAX.

Antes de continuar con el segundo paso de la construcción del modelo SARIMAX, es necesario realizar unas divisiones en la data de consumo. Se debe dividir la data en tres partes: entrenamiento, evaluación y última ventana. Las dos primeras que sirven para buscar el mejor modelo para SARIMAX y la última ventana que en este caso siempre corresponderá a la última semana completa y es la que se usa al momento de realizar la predicción, la sección de evaluación corresponderá, para este caso, a las 2 semanas anteriores a la última ventana y el resto corresponderá a la data de *entrenamiento* (Figura 22).

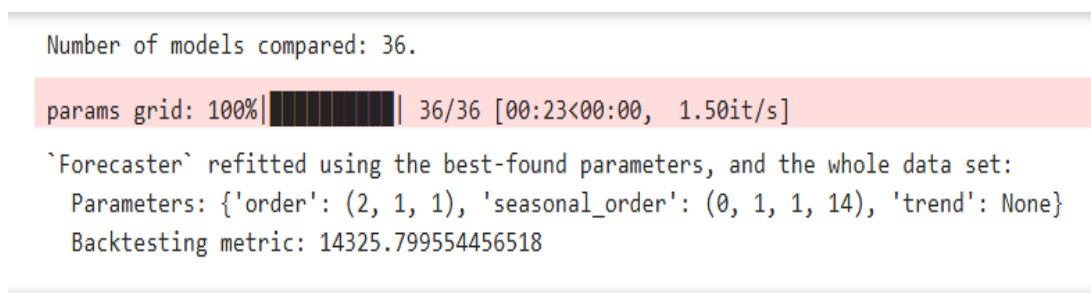
Figura 22 Data dividida para entrenamiento



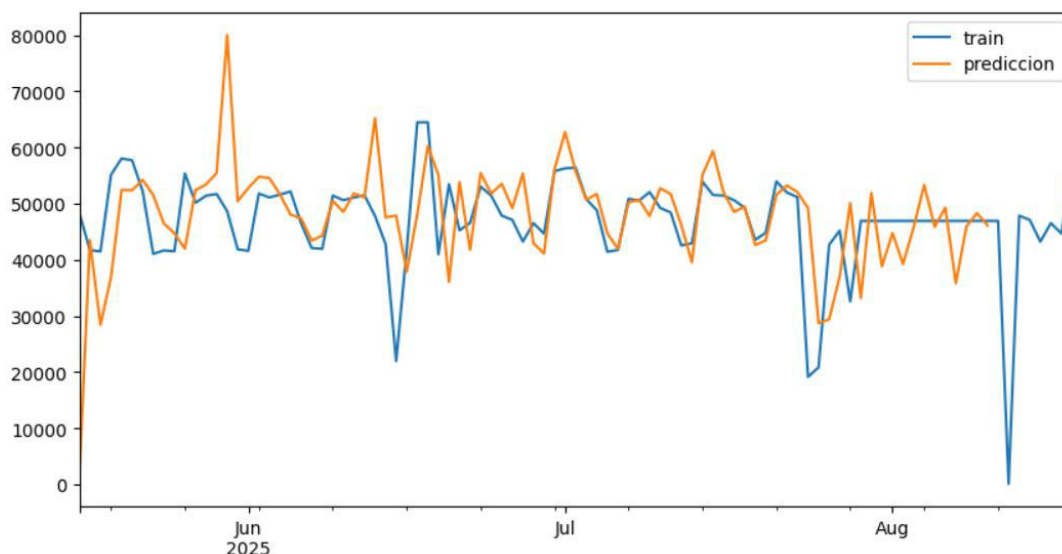
Una vez se tiene la data de entrenamiento y de evaluación, se selecciona solo la columna de `total_kilovatio` (Figura 23) para estimar el mejor modelo para SARIMAX. Se realiza un entrenamiento por cada combinación de cada posible valor obtenido en los componentes ARIMA y los componentes estacionales, para encontrar el que menor AIC y BIC tiene.

Figura 23 Selección de dato para entrenamiento SARIMAX

Este proceso realiza una validación cruzada de la data de entrenamiento y la data de evaluación combinando todos los componentes encontrados. Como se aprecia en la Figura 24 los mejores componentes ARIMA fueron (2,1,1) y componentes estacionales (0,1,1,14).

Figura 24 Mejor modelo encontrado para SARIMAX

Con estos parámetros se realiza un entrenamiento del modelo SARIMAX usando la data de entrenamiento y validación (Figura 25). El siguiente paso es la validación en la que se analizan los residuos. Se realiza la prueba de Ljung-Box para evaluar si existe o no autocorrelación en los residuos.

Figura 25 *Entrenamiento del modelo SARIMAX*

Como se puede apreciar en la Figura 26 el p -value es de 0.99 cumpliendo la hipótesis al ser mayor que 0.05, por lo tanto, no existe autocorrelación en los residuos.

Figura 26 *Prueba de Ljung-Box SARIMAX*

```
lb_stat  lb_pvalue
20  7.270523  0.995699
```

Para verificar la distribución normal de los residuos, al igual que en ARIMA, se aplica el test de Shapiro-Wilk. Según la Figura 27, el p -value obtenido con esta prueba es menor a 0.05 con lo que se concluye que la muestra no sigue una distribución normal.

Figura 27 *Prueba de Shapiro-Wilk SARIMAX*

```
Estadístico de Shapiro-Wilk: 0.8690
```

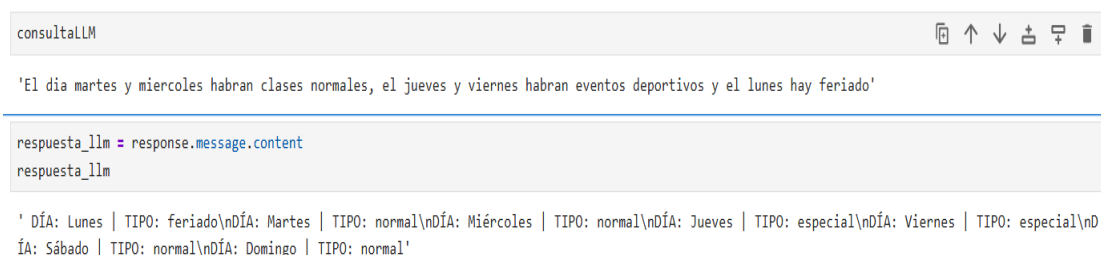
```
Valor-p: 0.0000
```

```
Se rechaza H0: los residuos NO parecen seguir una distribución normal.
```

Una vez ya tenemos construido el modelo SARIMAX, haremos uso de las variables exógenas que tenemos en los datos, en este caso son `evento_especial`, `feriado` y `temperatura`. Pero, además, debemos tener los datos exógenos de la semana a predecir.

Como se necesita predecir las dos semanas siguientes, se requiere disponer de las variables exógenas de los próximos 14 días. El sistema debe permitir que el usuario diga por voz las instrucciones necesarias para construir la data necesaria de las variables exógenas para los últimos 7 días de la predicción. Es por esto, que se le da la tarea al LLM para que genere una respuesta en base a lo que el usuario diga en texto, especificando en el *prompt* (Figura 28) que la respuesta se entregue en un formato definido.

Figura 28 Respuesta del LLM con formato definido



```

consultaLLM

'El día martes y miercoles habran clases normales, el jueves y viernes habran eventos deportivos y el lunes hay feriado'

respuesta_llm = response.message.content
respuesta_llm

'DÍA: Lunes | TIPO: feriado\nDÍA: Martes | TIPO: normal\nDÍA: Miércoles | TIPO: normal\nDÍA: Jueves | TIPO: especial\nDÍA: Viernes | TIPO: especial\nDÍA: Sábado | TIPO: normal\nDÍA: Domingo | TIPO: normal'

```

Así, la información se entrega en un formato definido, del cual se extraen los valores con una combinación de patrones definido en expresiones regulares para construir, a partir de esto, se crea un dataframe (Figura 29) con toda la información de las variables exógenas con el formato necesario para entrenar el modelo.

Figura 29 Dataframe para la respuesta del LLM

	feriado	evento_especial	total_temperatura
2025-08-25	1	0	28.007459
2025-08-26	0	0	20.999810
2025-08-27	0	0	20.204014
2025-08-28	0	1	20.207757
2025-08-29	0	1	21.914883
2025-08-30	0	0	22.287282
2025-08-31	0	0	26.676602

Se construyó en la primera interacción (semana previa) un dataframe con valores aleatorios para guardarlo en la base de datos *Figura 30*, con el propósito de que, en las siguientes interacciones, al pasar las semanas, se recupere como primera

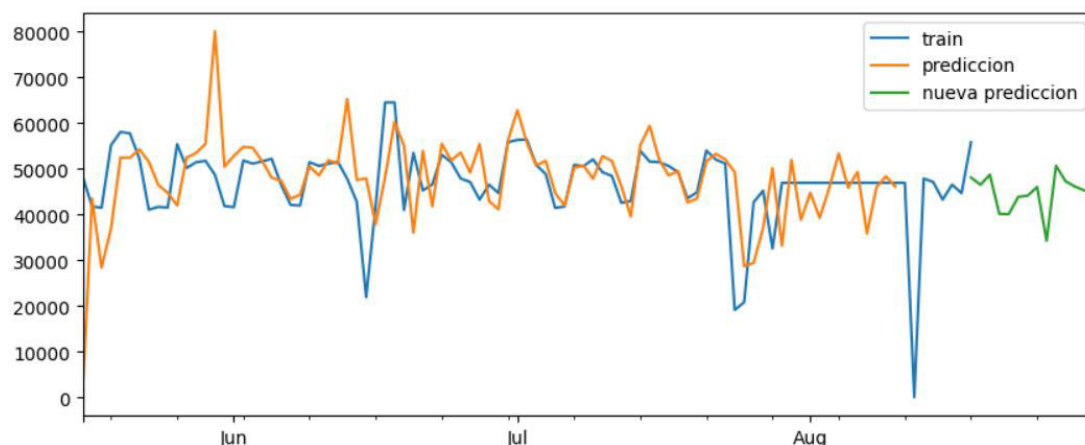
semana la última predicción (realizada de la semana anterior) y el dataframe construido a partir de la consulta del usuario actual.

Figura 30 *Dataframe de variables exógenas en fechas de predicción*

	feriado	evento_especial	total_temperatura
2025-08-18	0	1	22.403109
2025-08-19	1	1	25.789727
2025-08-20	0	0	28.335207
2025-08-21	0	1	23.930406
2025-08-22	1	1	24.756880
2025-08-23	1	0	21.696005
2025-08-24	1	0	27.750855
2025-08-25	1	0	28.007459
2025-08-26	0	0	20.999810
2025-08-27	0	0	20.204014
2025-08-28	0	1	20.207757
2025-08-29	0	1	21.914883
2025-08-30	0	0	22.287282
2025-08-31	0	0	26.676602

Por último, se le pasa al algoritmo entrenado el dataframe de las variables exógenas de ambas semanas unidas, los valores de consumo de la última ventana y las variables exógenas de la última ventana se observan en la Figura 31.

Con estos datos, se obtiene el modelo SARIMAX como el ideal en este tipo de casos, es decir, cuando se requiere predecir el consumo energético de días posteriores, por considerar posibles eventos que ocurrirán en esas fechas, ajustándose al contexto y por lo tanto se obtienen resultados más precisos.

Figura 31 *Gráfico de los datos de consumo predicho*

4.3. Diseño y Desarrollo del Módulo HCI Conversacional

En este apartado se incluye el procedimiento desarrollado para la creación del avatar conversacional que actúe como asistente energético. Esta solución de inteligencia artificial conversacional diseñada para facilitar la gestión y el análisis del consumo energético en los edificios de la Universidad Técnica de Manabí funciona mediante interacción por voz y texto; su finalidad es apoyar la toma de decisiones energéticas, adaptándose a nuevas incorporaciones de edificios, pisos y ambientes a través de actualizaciones periódicas de la base de datos logrando optimizar los recursos de procesamiento.

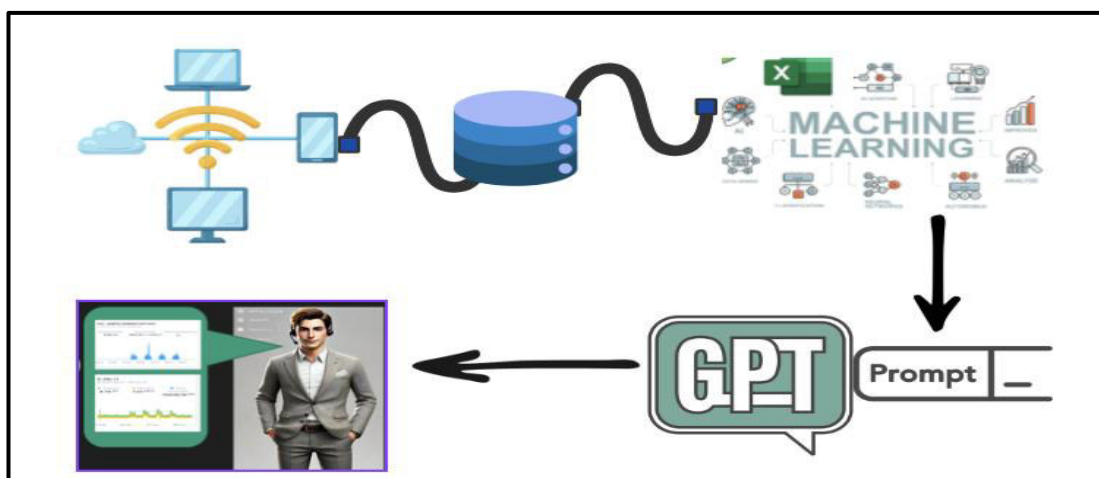
4.3.1. Creación del Avatar

Se procede a diseñar una interfaz amigable que permita la interacción entre el usuario y el asistente. La interfaz incluye un ícono para iniciar la conversación y un área de visualización del texto del asistente. En una primera fase, se utilizó la API ChatGPT de OpenAI para acelerar la creación de un prototipo funcional, sobre esta plataforma se implementó lo siguiente: El asistente genera cada respuesta en base a un prompt de personalidad y al historial de la conversación. Se ejecutan funciones programadas específicas que se invocan selectivamente según la intención detectada en el mensaje del usuario, retornando una retroalimentación acorde al resultado. Los textos de archivos y bases de datos se convierten en vectores para que el asistente pueda interpretarlos y ofrecer respuestas más personalizadas (Figura 32).

El historial de conversaciones se realiza configurando el sistema para

almacenar el historial de conversaciones en una variable global y enviarlo a la API para mantener la coherencia en la interacción. Para ello, se aprovecha la herramienta `'function_call'` de GPT para definir y utilizar funciones en formato JSON, permitiendo al asistente ejecutar acciones específicas basadas en la entrada del usuario. Se integraron las API nativas del navegador para reconocimiento y síntesis de voz, de modo que el audio captado por el micrófono se transforme en texto y el texto generado por el asistente se reproduzca como sonido.

Figura 32 *Procedimiento de diseño propuesto para la comunicación avatar - usuario*



Una vez validados los resultados con datos simulados, el sistema se conectó a una API real de consulta de consumo energético de edificios, permitiendo así trabajar con información veraz y actualizada. Cuando el proyecto empezó a mostrar resultados prometedores, surgió un obstáculo: la dependencia de la conexión a Internet para comunicarse con el asistente y con las bibliotecas de voz del navegador. Para eliminar ese cuello de botella, se optó por trasladar todos los procesos al ámbito local, priorizando la búsqueda de algoritmos y modelos LLM que pudieran ejecutarse sin conexión externa.

Durante la exploración de soluciones locales se detectó que ninguna incorporaba las herramientas integradas de ChatGPT. Inicialmente se consideraron modelos ligeros y extensibles capaces de invocar funciones personalizadas descartando opciones como Deepseek, que carecían de esa capacidad. Se realizaron pruebas con Llama 3.2 mediante Ollama, pero sus respuestas resultaban vagas y repetitivas, sin adaptarse a la retroalimentación de las funciones. A continuación, se

evaluó Mistral y Owen: de ambos, Mistral ofreció el mejor equilibrio entre precisión en las respuestas y eficiencia de ejecución (Para el diseño y modelado del avatar 3D que represente de manera efectiva al asistente y que el diseño sea visualmente atractivo y funcional, se procede a implementar animaciones que permiten al avatar expresar emociones y gestos e integrar estas animaciones con un motor de síntesis de voz para que el avatar pueda hablar de manera natural (Figura 33).

Para el diseño y modelado del avatar 3D que represente de manera efectiva al asistente y que el diseño sea visualmente atractivo y funcional, se procede a implementar animaciones que permiten al avatar expresar emociones y gestos e integrar estas animaciones con un motor de síntesis de voz para que el avatar pueda hablar de manera natural.

Figura 33 *Herramientas utilizadas*



4.3.2. Personalización del Avatar

Para lograr la creación e integración del Avatar se utiliza la plataforma web Mixamo. Esta aplicación ha resultado ser una solución innovadora y efectiva para mejorar la interacción entre los usuarios y el sistema, debido a que ofrece una amplia gama de avatares personalizables que son altamente accesibles y fáciles de integrar en diversas plataformas de desarrollo, tales como Unity y otras plataformas web. A continuación, se describen algunas de sus características y bondades:

Se personaliza el avatar (La integración de los avatares de Mixamo en Unity es sencilla, gracias a las herramientas intuitivas y a la compatibilidad que ofrece la

plataforma. En Unity, se desarrolló una aplicación o ejecutable WebGL que genera un archivo HTML. Este archivo HTML es el principal, aunque se generan otros archivos complementarios, y puede ser fácilmente insertado en cualquier página web para su visualización y uso.

Figura 34) lo que facilita la adaptación a las necesidades específicas del proyecto, ya que los avatares pueden expresar una amplia variedad de emociones y gestos, mejorando la comunicación y la experiencia del usuario.

La integración de los avatares de Mixamo en Unity es sencilla, gracias a las herramientas intuitivas y a la compatibilidad que ofrece la plataforma. En Unity, se desarrolló una aplicación o ejecutable WebGL que genera un archivo HTML. Este archivo HTML es el principal, aunque se generan otros archivos complementarios, y puede ser fácilmente insertado en cualquier página web para su visualización y uso.

Figura 34 *Avatar v1.0, creado utilizando Mixamo y Unity*



Unity facilita la creación de aplicaciones interactivas y la ejecución de avatares en tiempo real. La generación de un ejecutable WebGL en Unity simplifica el proceso de llevar los avatares a una plataforma web. La aplicación permite a los usuarios interactuar con los avatares directamente desde el navegador, sin necesidad de instalar software adicional, aumentando así la accesibilidad y la facilidad de uso.

La implementación de avatares de Mixamo ha demostrado una mejora significativa en la interacción entre los usuarios y el sistema, proporcionando una experiencia más dinámica y atractiva. Por tanto, se logran avatares humanizados que permiten una comunicación más natural y eficiente, lo que es especialmente beneficioso en entornos donde la interacción usuario-sistema es crucial. Mixamo y Unity son compatibles con una variedad de plataformas y dispositivos, lo que permite que los avatares sean utilizados en diferentes contextos y aplicaciones. La flexibilidad de integración asegura que los avatares pueden ser adaptados y mejorados continuamente, manteniendo la relevancia y efectividad en su uso.

Por lo expuesto, se destaca el uso de Mixamo y Unity para la creación e integración de avatares ha resultado en una solución eficiente y altamente efectiva para mejorar la interacción en plataformas web. La facilidad de uso, la personalización, y la accesibilidad proporcionada por estas herramientas han permitido desarrollar un sistema interactivo y atractivo, beneficiando tanto a los desarrolladores como a los usuarios finales.

En un principio, se utilizaron avatares independientes donde la interacción se realizaba mediante la conversión de voz a texto y viceversa, utilizando herramientas como ElevenLabs. Sin embargo, este proceso de conversión resultaba en tiempos de espera elevados, afectando la fluidez de la interacción usuario-sistema.

Después de algunos meses, con el lanzamiento del modelo conversacional ChatGPT, se realizaron pruebas comparativas. Los resultados mostraron que los tiempos de espera se redujeron significativamente en comparación con el método anterior. Esto permitió integrar el avatar ya diseñado con una mejora notable en la interacción conversacional. Se utilizó la API de ChatGPT-4 para gestionar las conversaciones, optimizando la respuesta y la interacción en tiempo real. En la Tabla 4, se puede observar una comparativa entre las herramientas LLaMA y GPT-4.

Tabla 4 *Cuadro Comparativo entre LLaMA y GPT*

Característica	LLaMA	GPT
Desarrollador	Meta	OpenAI
Fecha de Lanzamiento	Febrero 2023	Junio 2020 (GPT-3), Marzo 2023 (GPT-4)
Arquitectura	Basado en transformadores	Basado en transformadores
Modelos Disponibles	LLaMA-7B, LLaMA-13B, LLaMA-30B, LLaMA-65B	GPT-3 (175B), GPT-3.5, GPT-4
Escalabilidad	Escala desde 7B hasta 65B parámetros	Escala hasta 175B parámetros
Tamaño del Conjunto de Tokens	Aproximadamente 1.4 trillones de tokens	Aproximadamente 300 billones de tokens
Dataset de Entrenamiento	Variado, enfocado en documentos en múltiples dominios	Diverso, incluye libros, artículos y contenido web
Tareas Destacadas	Análisis de texto, generación de texto, traducción automática	Redacción de texto, generación de código, respuestas a preguntas, traducción automática
Rendimiento	Buen rendimiento en tareas de procesamiento de lenguaje	Excelente rendimiento en múltiples tareas de NLP

	natural	
Implementación	Enfocado en eficiencia y ajuste fino para tareas específicas	Generalizado, con enfoque en rendimiento y versatilidad
Personalización	Altamente ajustable para tareas específicas	Ofrece capacidades de ajuste fino y API para personalización
Acceso y Disponibilidad	Licencias para uso académico y comercial	API comercial disponible, acceso mediante suscripción
Comunicación y Soporte	Soporte mediante comunidad y foros de investigación	Soporte mediante documentación extensa, foros y comunidad activa
Dificultades de Instalación	Requiere hardware de alta gama, configuración de dependencias, optimización de modelos	Relativamente sencillo con acceso a API, pero requiere suscripción para acceso completo
Evaluación de Desempeño	Bueno en tareas específicas de NLP y series temporales	Excelente en tareas generales de NLP y aplicaciones amplias
Usos Comunes	Aplicaciones de investigación, sistemas de recomendación, predicciones específicas	Chatbots, asistentes virtuales, generación de contenido, herramientas de desarrollo de código
Limitaciones	Complejidad en la configuración, alto consumo de recursos, necesidad de	Costo de acceso a API, posibles sesgos en los datos, limitaciones de contexto debido a la ventana

Actualmente, el sistema Web diseñado permite realizar consultas a archivos CSV que se obtienen a partir de consultas SQL cargadas al sistema. Este método proporciona los datos necesarios, pero requiere una actualización manual periódica.

4.3.3. Incorporación de Voz al Avatar

Una de las características claves de un avatar es que sea capaz de comunicarse a través de diferentes modalidades que incluyen voz, gestos y expresiones faciales para poder replicar la comunicación de manera más real. Es primordial que exista una sincronización precisa entre las señales verbales y no verbales para evitar resultados desfavorables y mejorar la interacción con el usuario. Por ello, la latencia de respuesta debe ser mínima y constante para imitar el tiempo de respuesta humano, las transiciones entre el silencio y el habla deben ser suaves para evitar saltos visuales que puedan parecer poco naturales. Por ello, se describe a continuación el procedimiento de validación de los diferentes modelos de transcripción de voz a texto y viceversa para identificar el que permita una comunicación fluida más real.

Modelo de transcripción de voz a texto

Para el experimento de transcripción de voz a texto, se ha utilizado Vosk y Whisper. Ambas herramientas hacen el mismo proceso, el modelo lee el archivo de voz enviado por el usuario desde el frontend hacia el backend (python). En este proceso se observa que la implementación y uso no eran complejos, los resultados de tiempos y calidad de la respuesta demuestran el correcto funcionamiento (Tabla 5).

Tabla 5 *Comparativa de Modelos de transcripción de voz a texto*

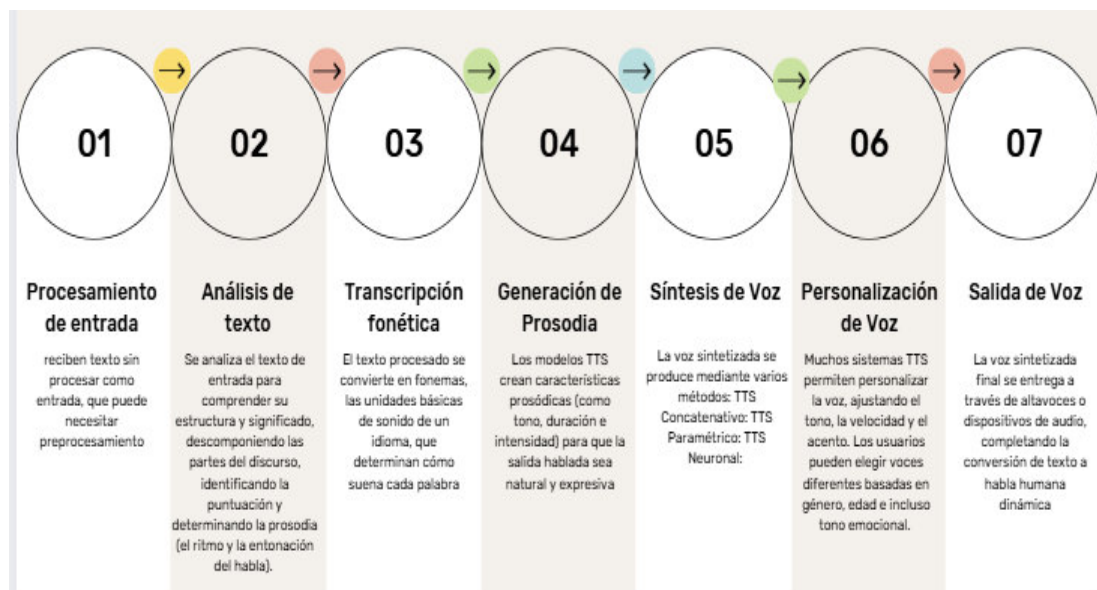
Modelo	Uso	Tiempo	Calidad del texto transcrito
---------------	------------	---------------	-------------------------------------

VOSK	CPU	6.81 s	Texto generado contiene faltas ortográficas y falta de signos, no genera una frase gramatical correcta.
WHISPER	CPU	25.70 s	Texto generado contiene los signos de punto, comas, tildes, etc., siendo una frase más real al lenguaje natural.
WHISPER	GPU	5.56 s	Texto generado contiene los signos de punto, comas, tildes, etc., siendo una frase más real al lenguaje natural.

Con base en los resultados, se decide utilizar Whisper debido a su aceleración por GPU y su capacidad de generar transcripciones rápidas, precisas y con puntuación adecuada garantizando la mejor experiencia de usuario y la máxima fiabilidad en la interpretación de comandos y consultas.

Modelo de transcripción de texto a voz

Mediante el uso de diversos algoritmos y modelos, se logra transformar texto escrito en palabras habladas, el proceso de conversión se observa en la Figura 35. De manera similar, se lleva a cabo el experimento de conversión de voz a texto utilizando TTS de Coqui, el cual emplea dos esquemas de licenciamiento distintos según el tipo de modelo. Los modelos "VITS" distribuidos por Coqui (por ejemplo, aquellos entrenados sobre LJSpeech o VCTK) están publicados bajo la Apache License, Versión 2.0, una licencia permisiva que permite su uso, modificación y redistribución, incluso con fines comerciales.

Figura 35 *Proceso de Conversión de texto a voz*

Por otro lado, el modelo "XTTS" se publica bajo la Coqui Public Model License (CPML), que permite el uso libre no comercial de los pesos y de sus salidas, pero requiere obtener una licencia comercial separada para su uso con fines lucrativos. Los modelos se descargan ejecutando la siguiente línea de código en Python:

```
#Descarga del modelo Vits
```

```
tts = TTS(
    model_name="tts_models/es/css10/vits",
    progress_bar=False
).to(device)
```

```
#Descarga del modelo XTTS
```

```
tts = TTS(
    model_name="tts_models/multilingual/multi-dataset/xtts_v2",
    progress_bar=False
).to(device)
```

Ambas herramientas ejecutan el mismo procedimiento. El modelo recibe el texto y genera el archivo de voz que será enviado al usuario. Ambos modelos son funcionales y se obtienen los resultados en términos de tiempos y calidad de la respuesta (Tabla 6).

Tabla 6 *Comparativa de Modelos de transcripción de texto a voz*

Modelo	Uso	Tiempo	Tiempo optimizado	Mejora aproximada
VITS	CPU	9.01 s	3-4 s	-55%
VITS	GPU	8.97 s	2-4 s	-67%
XTTS	GPU	35.39 s	<10 s	-72%

Basado en los resultados, se decide emplear XTTS, ya que, a pesar del tiempo de respuesta, su aceleración mediante GPU y su capacidad para generar un resultado de voz muy natural garantizan la mejor experiencia de usuario. Aunque VITS puede ejecutarse rápidamente en CPU aplicando cuantización e hilos, el uso de GPU apenas aporta beneficios debido al overhead, lo cual puede mejorarse cuantizando a INT8, activando multithreading y agrupando fragmentos (batching) o manteniendo la ejecución en CPU.

En este mismo contexto, aunque XTTS requiere una GPU optimizada (precisión reducida y vocoder ligero) para alcanzar latencias aceptables, es posible exportarlo a FP16 y desplegarlo con TensorRT u ONNX Runtime. Además, se puede cambiar a un vocoder ultrarrápido (HiFi-GAN pruned o MelGAN optimizado) y reducir la frecuencia de muestreo (de 24 kHz a 16–22 kHz). Con estas estrategias, el modelo VITS se reduciría a aproximadamente 3 segundos y XTTS a menos de 10 segundos por frase, haciendo viable la síntesis casi en tiempo real.

Con las mejoras implementadas, el sistema era funcional; sin embargo, persistían errores, como la falta de optimización en la orquestación de funciones de Mistral. Para resolverlo, se inicia un segundo enfoque de trabajo: dotar al asistente de

una capa de ML específica para cada tarea. Para ello, se utiliza BERT debido a su precisión en la identificación del proceso que el usuario desea ejecutar (Figura 7).

Figura 36 *Evaluación de clasificadores*

```
Tiempo de ejecución de la intención (BERT): 3.23 segundos
Resultado de la intención: solicita_datos_consumo
```

Para la extracción de entidades se exploran tres enfoques: a) Embeddings de la estructura JSON de edificios + búsqueda por similitud de sentencias. b) Modelo SpaCy entrenado con ejemplos etiquetados. Y, c) Coincidencia difusa de cadenas (fuzzy string matching) aplicada a la ficha de edificios, pisos y ambientes. Este último resultó ser el más robusto para extraer parámetros de la consulta.

Figura 37 *Extracción de entidades*

```
Tiempo de ejecución extracción de entidades (Fuzzy Lookup): 0.01 segundos
Resultado de la extracción: [{'edificio': {'id': 10, 'nombre': 'HUMANÍSTICA'}, 'piso': {'id': 18, 'nombre': 'PLANTA BAJA'}, 'ambiente': {'id': 174, 'nombre': 'Centro d
e datos 1'}}]
```

Durante el proceso surgió la necesidad de que el asistente recordara consultas previas de consumo energético para no volver a solicitar edificios, pisos o ambientes ya mencionados. Para ello, se implementó un módulo de historial que almacena solo esas interacciones y construye, a partir de ellas, un prompt consolidado que el LLM trata como una nueva petición completa.

También en su primer momento, para obtener los datos de la API de SmartBuilding, se pasaban los parámetros de la extracción de entidad y la API estáticamente ejecutaba una consulta SQL con los parámetros que llegan. En esta situación, se optimiza el proceso de la consulta SQL y se decide que luego de la extracción de entidad, el LLM Mistral cree la consulta SQL a partir de un prompt refinado con los datos de la extracción, esta funcionalidad brinda más libertad a que el usuario pregunte diferentes necesidades de consumo energético y ya no se utiliza la API, sino que se conecta directamente a la base de datos de SmartBuilding.

Se realizan veinte pruebas en Jupyter usando un prompt que permitía al LLM generar consultas SQL. Todas las consultas se ejecutaron en la base de datos, y ocho de ellas devolvieron resultados nulos, principalmente por rangos de fechas fuera de los

datos disponibles. De esos ocho casos, en tres la columna *total_kilovatio_edificio* también quedó en blanco, debido a un error en la generación de la consulta por parte del LLM.

4.3.4. Ejecución de tarea con base en el mensaje del usuario

Las intenciones que un usuario desea comunicar en un mensaje pueden ser muy variadas, por lo que es crucial para los asistentes virtuales y chatbots comprender, responder y delimitar estas intenciones para llevar a cabo diferentes operaciones que cumplan con las necesidades del usuario.

En la primera versión del sistema, cuando se hacía uso del modelo GPT desde de la API de OpenAI para poder ejecutar alguna de las tareas que debía hacer el asistente, se usaban las llamadas a funciones. En la Figura 38 se muestra una de las funciones definidas para el asistente con GPT. Se puede apreciar la estructura de estas funciones en donde se definía el nombre de la función, una descripción sobre en qué momento debe ejecutarse la función y los parámetros que debe devolver, y enviando una respuesta negativa o positiva de acuerdo con la forma como se ejecutó la tarea.

Figura 38 Ejemplo de llamadas a funciones de GPT

```
{
  "name": "predecir_consumo_energetico",
  "description": "Predice el consumo energético futuro a partir de los datos
proporcionados por el usuario",
  "strict": true,
  "parameters": {
    "type": "object",
    "properties": {
      "fecha_inicio": {
        "type": "string",
        "description": "Fecha de inicio de la predicción en formato AAAA-MM-DD"
      },
      "fecha_fin": {
        "type": "string",
        "description": "Fecha de fin de la predicción en formato AAAA-MM-DD"
      },
      "datos_historicos": {
        "type": "array",
        "description": "Serie temporal de datos históricos de consumo energético en
kilovatios hora (kWh)",
        "items": {
          "type": "object",
          "properties": {
            "fecha": {
```

Además, como se puede visualizar en la Figura 39, lo que el modelo de GPT hacía es que leía la consulta del usuario e inteligentemente interpretaba la intención.

También se puede ver que el modelo no dio resultados erróneos, sin embargo, el tiempo de respuesta era muy lento, en gran parte porque depende de recibir respuestas de un servidor externo y existe el retraso de conexiones.

Figura 39 *Detección de intención con toolcalls de GPT*

```
Tiempo de ejecución de la intención: 16.52 segundos
Consulta: Dame el consumo energético del edificio de humanística, piso planta baja, ambiente centro de da
Resultado de la intención: solicita_datos_consumo
-----
Tiempo de ejecución de la intención: 8.18 segundos
Consulta: Cómo estás?
Resultado de la intención: pregunta_respuesta_general
-----
Tiempo de ejecución de la intención: 7.65 segundos
Consulta:Cuál será la predicción para la siguiente semana?
Resultado de la intención: solicita_prediccion
-----
Tiempo de ejecución de la intención: 4.74 segundos
Consulta: Quisiera saber que consumo tendra el edificio de humanistica para la siguiente semana
Resultado de la intención: solicita_prediccion
-----
Tiempo de ejecución de la intención: 5.29 segundos
Consulta: Cual será el consumo para la siguiente semana
Resultado de la intención: solicita_prediccion
-----
Tiempo de ejecución de la intención: 5.76 segundos
Consulta: Cómo te encuentras el día de hoy?
Resultado de la intención: pregunta_respuesta_general
-----
Tiempo de ejecución de la intención: 7.28 segundos
Consulta: Considerando que la siguiente semana tendremos feriado el lunes, cuál sería la predicción?
Resultado de la intención: solicita_prediccion
-----
```

Cuando se migró a modelos locales, esta opción ya no era factible, por lo que se profundizó en los LLMs locales disponibles, la mayoría de estos no cuentan con llamadas a funciones y algunos de estos responden de manera ineficiente o son repetitivos en sus respuestas, por ejemplo, el LLM de Qwen2.5 en su modelo 14b. En la Figura 40 se puede ver que, de las 3 consultas, Qwen solo respondió correctamente una de ellas, el resto de las consultas no se ejecutaron, debido a que ocurrió una pérdida de conexión con la API local de Ollama. Esta desconexión se debe a los largos tiempos de consulta.

Figura 40 *Llamadas a funciones de Qwen2.5*

```
Tiempo de ejecución de la intención: 25.26 segundos
Consulta: Dame el consumo energético del edificio de humanística, piso planta baja, ambiente centro de datos, del 1 al 5 de junio del 2025.
Resultado de la intención: solicita_prediccion
-----
Tiempo de ejecución de la intención: 9.52 segundos
Consulta: Cómo estás?
Resultado de la intención: solicita_prediccion
-----
Tiempo de ejecución de la intención: 48.56 segundos
Consulta:Cuál será la predicción para la siguiente semana?
Resultado de la intención: solicita_prediccion
-----
```

Al enfrentar estos inconvenientes, se optó por emplear modelos basados en aprendizaje profundo, los cuales ofrecen un rendimiento superior en la clasificación de intenciones. Estos modelos se ajustan de manera precisa a conjuntos de datos específicos para alcanzar resultados de vanguardia. El modelo que se usó fue BERT entrenado en la detección de intención, proporcionando también una lista de categorías de entre las que pueda escoger según la consulta del usuario, además de pasarle un pequeño prompt (llamado hipótesis en el modelo) para que sepa qué hacer.

Figura 41 *Declaración del modelo de detección de intención*

```
model_name = "Recognai/bert-base-spanish-wwm-cased-xnli"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)
classifier = pipeline("zero-shot-classification", model=model, tokenizer=tokenizer)

resultado = classifier(consulta, etiquetas, hypothesis_template="Que accion desea realizar el usuario con esta consulta: {}.")
```

Para probar con este modelo se le paso una lista de posibles consultas del usuario con variaciones en el texto, pero manteniendo la intención del mensaje, pero en el 43% de las consultas el modelo falla al intentar predecir la intención.

Figura 42 *Resultados de detección de la intención del usuario con el modelo BERT*

```
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 1.33 segundos
Consulta: Dame el consumo energético del edificio de humanística, piso planta baja, ambiente centro de
Resultado de la intención: solicita_datos_consumo
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.69 segundos
Consulta: Cómo estás?
Resultado de la intención: pregunta_respuesta_general
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.71 segundos
Consulta:Cuál será la predicción para la siguiente semana?
Resultado de la intención: solicita_prediccion
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.71 segundos
Consulta: Quisiera saber que consumo tendra el edificio de humanistica para la siguiente semana
Resultado de la intención: solicita_datos_consumo
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.71 segundos
Consulta: Cual será el consumo para la siguiente semana
Resultado de la intención: solicita_datos_consumo
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.71 segundos
Consulta: Cómo te encuentras el día de hoy?
Resultado de la intención: pregunta_respuesta_general
-----
Device set to use cpu
Tiempo de ejecución de la intención (BERT): 0.71 segundos
Consulta: Considerando que la siguiente semana tendremos feriado el lunes, cuál sería la predicción?
Resultado de la intención: pregunta_respuesta_general
-----
```

Al ver que este modelo en varias ocasiones se confundía se decidió por hacer uso del LLM de Mistral con un prompt fuerte y acorde a lo que se necesitaba.

Pasándole la misma lista de posibles consultas del usuario, la respuesta del LLM arrojó mejores resultados disminuyendo considerablemente el margen de error.

Figura 43 *Resultados de detección de la intención con el LLM de Mistral*

```

Consulta: Dame el consumo energético del edificio de humanística, piso planta baja, ambiente centro de datos, d
Intencion: solicita_datos_consumo
Tiempo: 0.72 segundos
-----
Consulta: Cómo estás?
Intencion: pregunta_respuesta_general
Tiempo: 0.67 segundos
-----
Consulta:Cuál será la predicción para la siguiente semana?
Intencion: solicita_prediccion
Tiempo: 0.58 segundos
-----
Consulta: Quisiera saber que consumo tendra el edificio de humanistica para la siguiente semana
Intencion: solicita_prediccion
Tiempo: 0.62 segundos
-----
Consulta: Cual será el consumo para la siguiente semana
Intencion: solicita_prediccion
Tiempo: 0.58 segundos
-----
Consulta: Cómo te encuentras el día de hoy?
Intencion: pregunta_respuesta_general
Tiempo: 0.71 segundos
-----
Consulta: Considerando que la siguiente semana tendremos feriado el lunes, cuál sería la predicción?
Intencion: solicita_prediccion
Tiempo: 0.63 segundos

```

4.3.5. *Extracción de parámetros de la consulta del usuario.*

Para poder realizar una búsqueda de la información del consumo energético en un ambiente en específico, se necesita de ciertos parámetros como nombre del edificio, nombre del piso, nombre del ambiente, la fecha de inicio y la fecha de fin de consulta. Esto se vuelve una tarea desafiante al momento de extraer estos parámetros a partir del texto de consulta del usuario, que requiere de identificar las palabras claves en cada parte de la oración.

En la versión inicial del asistente, al utilizar la API de OpenAI, la extracción de parámetros se realizaba mediante un toolcall que especificaba el momento adecuado para ejecutar la función, como cuando se deseaba consultar el consumo de energía. En este proceso, se definían los parámetros derivados de los identificadores del edificio, el piso, el ambiente mencionado y las fechas. Además del toolcall, se empleaba la técnica de embedding, que se utiliza para mapear datos de alta dimensión en espacios vectoriales de menor dimensión, con el fin de vectorizar el contenido del archivo json de los datos del edificio. Esta vectorización se almacenaba en bases de datos vectoriales, que son comúnmente utilizadas para guardar y recuperar eficientemente

incrustaciones de alta dimensión.

Figura 44 Construcción del modelo en la API de asistentes de OpenAI



El proceso implementado por la API de OpenAI representa un procedimiento implícito de extracción de entidades, en el cual las operaciones se realizan de manera subyacente. En el ámbito del Procesamiento de Lenguaje Natural, la extracción de entidades es una tarea que implica la identificación y categorización de entidades dentro del texto. Este proceso es fundamental para diversas aplicaciones, tales como la recuperación de información, la traducción automática y la respuesta a preguntas. Para obtener un resultado comparable con los modelos de IA locales, es crucial comprender cada una de las técnicas de extracción de entidades.

Para la extracción de entidades de un texto, se adoptó un enfoque basado en diccionarios, que consiste en comparar subcadenas de texto libre con un diccionario de entidades. Una de las técnicas empleadas es el uso de expresiones regulares dentro de este enfoque. Antes de proceder a la extracción de entidades mediante expresiones regulares, el texto de consulta refinado, obtenido del paso anterior, es estructurado por el LLM de Mistral. Este proceso de estructuración permite formatear el texto de manera que cada parámetro de la consulta pueda ser separado por comas y otros símbolos, facilitando así la extracción de las entidades.

Figura 45 Formato para obtener parámetros por expresiones regulares

```
self.regexpr = [
    re.compile(r"edificio\s+(de\s+)*(?P<edificio>[\wáéíóúñ ]+)", re.IGNORECASE),
    re.compile(r"piso\s+(?P<piso>[\wáéíóúñ\d ]+)", re.IGNORECASE),
    re.compile(r"ambiente[s]?s+(?P<ambiente>[\wáéíóúñ\-\d ]+)", re.IGNORECASE),
    re.compile(r"inicio:?\s+(?P<fechainicio>[\wáéíóúñ\-\d ]+)", re.IGNORECASE),
    re.compile(r"fin:?\s+(?P<fechafin>[\wáéíóúñ\-\d ]+)", re.IGNORECASE)
]
```

Mediante la aplicación del método de lógica difusa, específicamente desde el enfoque de diccionarios, se pretende identificar coincidencias aproximadas entre los parámetros derivados de la consulta y los parámetros reales almacenados en la base de datos que contiene información sobre edificios, pisos y ambientes. Este proceso es viable gracias a la combinación de la lógica difusa con medidas de similitud semántica, lo cual optimiza la recuperación en tareas de extracción de entidades, abordando situaciones en las que cadenas sintácticamente distintas comparten el mismo significado. El resultado de este procedimiento (Figura 46) es un diccionario en Python que incluye los nombres e identificadores del ambiente, piso y edificio mencionados en la consulta, junto con las fechas de inicio y fin correspondientes.

Figura 46 Resultado del proceso de extracción de entidades

```
Resultado de la extracción: {'edificio': {'id': 10, 'nombre': 'HUMANÍSTICA'}, 'piso': {'id': 18, 'nombre': 'PLANTA BAJA'}, 'ambiente': {'id': 174, 'nombre': 'Centro de datos 1'}, 'fecha_inicio': '2025-06-01', 'fecha_fin': '2025-06-05'}
```

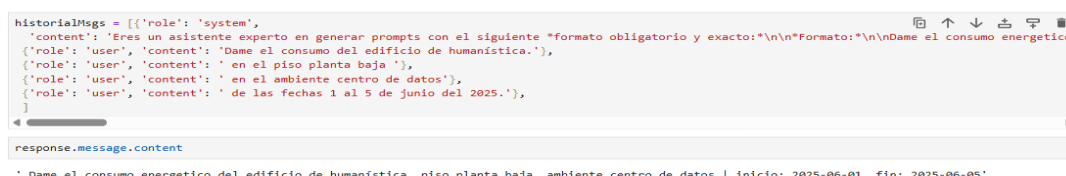
4.3.6. Recordar parámetros de mensajes anteriores.

Uno de los desafíos que surgían al extraer los parámetros clave para avanzar al siguiente proceso de consulta de consumo en la base de datos era que, cuando el usuario no mencionaba ciertos parámetros, estos se transferían como nulos a la fase siguiente, lo que provocaba errores. Era imperativo que el asistente energético tuviera la capacidad de prevenir la omisión de estos parámetros relevantes dentro de la consulta del usuario. Esto no se lograba simplemente rellenándolos con valores por defecto; el asistente debía solicitar al usuario que completara su solicitud en una interacción posterior y, además, debía poseer la capacidad de recordar parámetros de consultas previas.

En consonancia con esta idea, se programó al asistente una funcionalidad que le permite informar al usuario sobre las omisiones de los parámetros cuando la función

de extracción de entidades genera un error debido a valores nulos o inexistentes en los datos reales de los edificios. Mediante un prompt, la consulta incompleta del usuario y el comentario del error detectado, el asistente puede guiar al usuario para completar la solicitud (Figura 47). Además, para mantener la coherencia entre los mensajes antiguos y la nueva solicitud, cada mensaje del usuario se almacena en la base de datos utilizando una columna denominada categoría, donde se guarda el valor de ‘consumo_energético’.

Figura 47 *Resultado del proceso de recordar*



```
historialMsgs = [{"role": "system",
  "content": "Eres un asistente experto en generar prompts con el siguiente *formato obligatorio y exacto:*\n\n*Formato:*\n\nDame el consumo energetico
{'role': 'user', 'content': 'Dame el consumo del edificio de humanistica.'},
{'role': 'user', 'content': ' en el piso planta baja '},
{'role': 'user', 'content': ' en el ambiente centro de datos'},
{'role': 'user', 'content': ' de las fechas 1 al 5 de junio del 2025.'},
}]

response.message.content

' Dame el consumo energetico del edificio de humanistica, piso planta baja, ambiente centro de datos | inicio: 2025-06-01, fin: 2025-06-05'
```

De este modo, únicamente los mensajes del usuario que tienen la intención de realizar consultas sobre consumo son almacenados y recuperados al interactuar con el LLM, el cual cuenta con un prompt específico para este fin. Esto permite al asistente generar una nueva consulta a partir de mensajes previos, completándolas con nuevos parámetros e incluso sustituyendo parámetros antiguos si existen nuevos con valores diferentes.

4.3.7. *Ejecución de consulta SQL en la base de datos*

La construcción de una cadena SQL con una sintaxis funcional es fundamental para obtener los datos correctos y conforme a lo solicitado por el usuario mediante comandos de voz. Para lograr este objetivo, se consideró desde el inicio utilizar un modelo de lenguaje grande (LLM) que construyera dinámicamente la consulta, con el propósito de proporcionar flexibilidad en la obtención de la información. Inicialmente, se evaluó la opción de seleccionar un LLM especializado en la generación de código que, al mismo tiempo, no requiriera un procesamiento intensivo. En este contexto, se optó por Codellama, que aprovecha la inteligencia artificial para facilitar el desarrollo de software y es capaz de generar, depurar e incluso traducir código entre diferentes lenguajes de programación.

Dado que Codellama es ideal para construir código a partir de descripciones en lenguaje natural, se utilizó para realizar las primeras pruebas de construcción de la

cadena SQL. La intención era emplear este LLM para generar el código a partir de la consulta del usuario; sin embargo, previamente era necesario sustituir los nombres de edificios, pisos y ambientes por los identificadores obtenidos mediante búsqueda difusa. Esta tarea fue llevada a cabo por el LLM de Mistral, al proporcionarle como entrada el texto de la consulta refinada y el diccionario completo del edificio obtenido en el proceso anterior, permitiendo que el LLM tradujera inteligentemente estos segmentos de texto.

Ahora la cadena de texto resultante (Figura 48) podía ser utilizada para generar el código SQL junto con un prompt que contiene el ejemplo de la sintaxis con la que ejecutará el código, y demás restricciones para que solo responda con el código sin texto adicional.

Figura 48 *Respuesta con reemplazo de los nombres por los identificadores por Mistral*

```
Paso #1.3: Reemplazo de valor por identificadores por el LLM.
Tiempo de la extracción: 5.11 segundos
Respuesta de traducción: "Dame el consumo energetico del edificio 10, piso 18, ambiente 174 | inicio: 2025-06-01, fin: 2025-06-05"
```

Como se puede observar en la Figura 49 el LLM de Codellama no dio tan buenos resultados, ignorando las restricciones propuestas sobre añadir texto adicional, además del tiempo de respuesta largo, en este caso 42 segundos. Como este LLM no era factible para lo que se necesitaba, se probó con el LLM de Mistral siguiendo el mismo procedimiento.

Figura 49 *Resultado y tiempo de respuesta de Codellama*

```
WITH agrupados AS (
  SELECT DATE(fecha_creacion) AS fecha,
         SUM(amperio) AS total_amperio,
         SUM(kilovatio) AS total_kilovatio
  FROM monitoreo.vmostrardatoselectricidad
  WHERE idempresa = 2
         AND idedificacion = 10
         AND idpiso = 18
         AND idambiente = 174
         AND DATE(fecha_creacion) >= '2025-06-01'
         AND DATE(fecha_creacion) <= '2025-06-05'
  GROUP BY DATE(fecha_creacion)
)
SELECT g.fecha::TEXT,
       g.total_amperio,
       g.total_kilovatio,
       (SELECT SUM(B.kilovatio)
        FROM monitoreo.vmostrardatoselectricidad AS B
        WHERE B.idedificacion = 10
              AND DATE(B.fecha_creacion) = g.fecha) AS total_kilovatio_edificio
FROM agrupados g
ORDER BY g.fecha ASC;
--
```

En este SQL, hemos reemplazado los valores de los identificadores y las fechas en la plantilla de consulta con los valores proporcionados por el usuario. También hemos eliminado la cláusula 'AND idambiente = 174' ya que no se especificó en la petición del usuario.

Es importante destacar que este SQL solo devuelve los datos de la tabla 'monitoreo.vmostrardatoselectricidad' y no incluye ninguna otra información, como el consumo energético total del edificio o el consumo por habitaciones. Si se requiere esta información, se deberá agregar a la consulta utilizando una subconsulta 'JOIN'.

```
print(f"Tiempo de ejecución la generacion del SQL: {t_fin - t_ini:.2f} segundos")
```

Tiempo de ejecución la generacion del SQL: 42.68 segundos

En la Figura 50 se puede observar la diferencia del tiempo, incluso la forma en cómo responde, entendió las restricciones correctamente y generó la sentencia en la mitad de tiempo que Codellama. Esta diferencia de ambos LLM se debe a que, si bien tanto Mistral como Codellama presentan fortalezas en la generación de código, Mistral parece tener ventajas en términos de eficiencia, reducción de errores y robustez.

A pesar de las excelentes respuestas que arrojaba Mistral en la generación de código, el tiempo de espera para construirla aún seguía siendo bastante lento, por lo que se decidió omitir todo el proceso tanto de reemplazo de valores por identificadores y generación de SQL con el LLM, para consultar directamente a una vista en la base de datos usando los parámetros obtenidos del proceso de búsqueda difusa. Otra razón por la que se reemplazó fue porque tampoco se le estaba dando la libertad al LLM para que construya consultas dinámicas, ya que se le especificaba en el prompt una sintaxis fija y se le permitía solo reemplazar los valores de los parámetros.

Figura 50 Resultado y tiempo de respuesta de Mistral

```
WITH agrupados AS (
  SELECT DATE(fecha_creacion) AS fecha,
         SUM(amperio) AS total_amperio,
         SUM(kilovatio) AS total_kilovatio
  FROM monitoreo.vmostrardatoselectricidad
  WHERE idempresa = 2
        AND idedificacion = 10
        AND idpiso = 18
        AND idambiente = 174
        AND DATE(fecha_creacion) >= '2025-06-01'
        AND DATE(fecha_creacion) <= '2025-06-05'
  GROUP BY DATE(fecha_creacion)
)
SELECT g.fecha::TEXT,
       g.total_amperio,
       g.total_kilovatio,
       (SELECT SUM(B.kilovatio)
        FROM monitoreo.vmostrardatoselectricidad AS B
        WHERE B.idedificacion = 10
              AND DATE(B.fecha_creacion) = g.fecha) AS total_kilovatio_edificio
FROM agrupados g
ORDER BY g.fecha ASC;
```

```
: print(f"Tiempo de ejecución la generacion del SQL: {t_fin - t_ini:.2f} segundos")
```

Tiempo de ejecución la generacion del SQL: 20.20 segundos

4.3.8. *Respuestas finales del asistente*

El asistente debe proporcionar un informe coherente mediante un reporte de voz sobre la operación realizada, informando al usuario sobre los logros obtenidos o cualquier error ocurrido durante el proceso. Para cada proceso, ya sea consulta de consumo, predicción, respuestas a preguntas generales, bienvenida al usuario, o errores encontrados al consultar, se han diseñado diferentes prompts para que el asistente pueda responder adecuadamente a cada situación.

Para informar sobre el consumo de energía del edificio en los rangos de fechas especificados, una vez obtenida toda la información, esta se organiza en forma de tabla y se utiliza como contexto, instruyendo al asistente a informar al usuario sobre el consumo total consultado y el promedio diario, así como sobre picos o anomalías, identificar los ambientes con consumos inusuales y sugerir recomendaciones para optimizar el consumo. Este mismo prompt se emplea para los datos de predicción, proporcionando esta información como contexto.

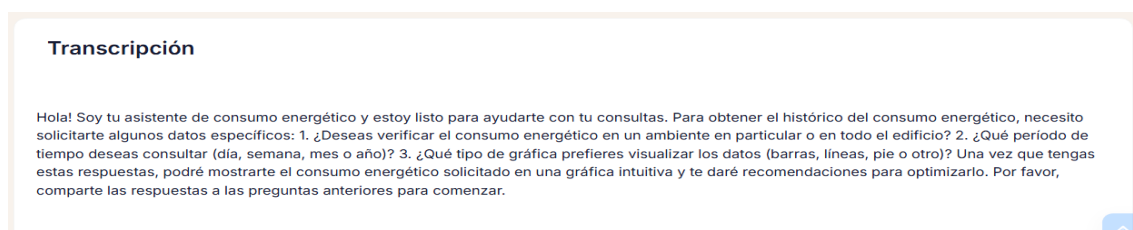
Figura 51 *Respuesta del LLM sobre el consumo energético*

Transcripción

Analizando los datos proporcionados para el rango de fechas del 1 al 7 de septiembre de 2025, el consumo energético total es de 289426.5202960836 kWh. Examinando los patrones observados en los datos, no hay anomalías o picos significativamente elevados durante el rango especificado. Sin embargo, se han detectado dos días con un consumo ligeramente superior al promedio: el 3 de septiembre y el 6 de septiembre. A continuación, ofrezco algunas recomendaciones prácticas para optimizar el consumo energético basándome en los patrones observados: 1. Auditoría energética del edificio: Realizar una auditoría completa de la infraestructura y los sistemas del edificio con el fin de identificar posibles fuentes de pérdidas energéticas y ofrecer soluciones para mejorar la eficiencia energética. 2. Ajuste de horarios de uso: Encourage the use of energy-intensive appliances during off-peak hours to reduce strain on the grid and take advantage of lower electricity prices. 3. Instalación de sensores y monitores: Install energy monitoring devices that can track and analyze real-time energy consumption data, providing insights into energy usage patterns and enabling proactive management of energy consumption. 4. Mantenimiento preventivo de los equipos: Prioritize regular maintenance for HVAC systems, lighting, and other major energy consumers to ensure they are operating at maximum efficiency and minimize energy waste. 5. Utilización de energías renovables:

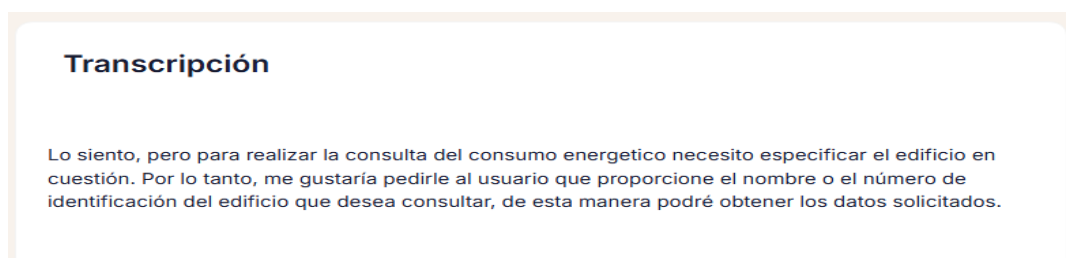
En relación con la bienvenida al usuario, este mensaje se activa cuando el usuario hace clic en el botón de información situado debajo del asistente. Se le indica en el prompt que describa brevemente el propósito del sistema, las acciones que el usuario puede realizar en el mismo, y que oriente al usuario para su correcta utilización.

Figura 52 *Respuesta del LLM sobre la presentación del asistente*



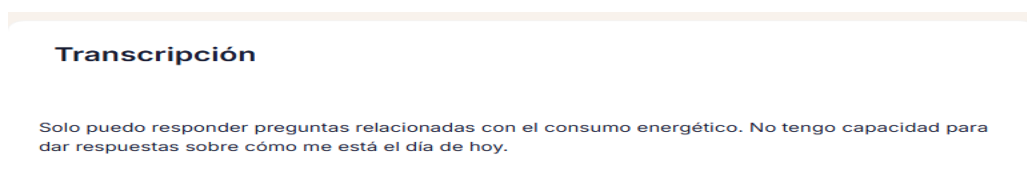
El asistente también tiene la capacidad de notificar al usuario en caso de que se produzca un error durante procesos como la detección de entidades. En tal caso, se incorpora tanto la consulta inicial como el mensaje de error, especificando el problema identificado y su ubicación.

Figura 53 *Respuesta del LLM sobre errores de usuario*



Finalmente, para responder a las preguntas generales formuladas por el usuario, el asistente dispone de un breve prompt en el que se especifica su rol y se le indica cómo debe responder a las consultas del usuario, estableciendo una restricción para que solo pueda formular preguntas relacionadas con el tema del consumo energético (Figura 54).

Figura 54 *Respuesta del LLM sobre respuestas a preguntas generales*



Las respuestas finales de este tipo otorgan al asistente una cualidad más humana, incrementando su capacidad de interacción con el usuario. Esto le confiere una personalidad que trasciende la de un mero sistema de consulta, emulando de manera más precisa el comportamiento de un asistente humano.

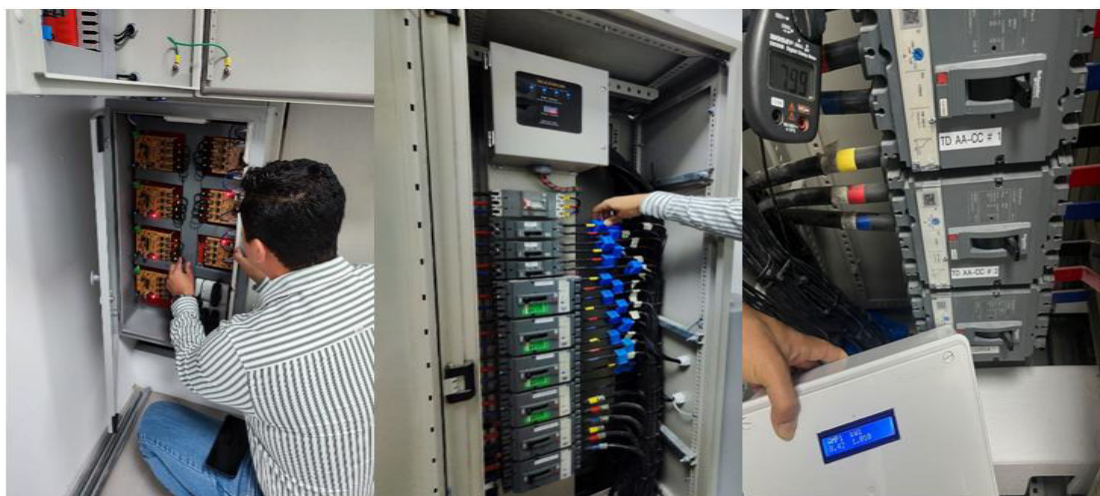
CAPÍTULO 5. DEMOSTRACIÓN (FASE 4)

En este apartado se describe la fase 4 de la metodología Ciencia del Diseño (Design Science Research, DSR), orientada a validar la operatividad y funcionalidad integral del ecosistema en su entorno real (el edificio), confirmando que la integración sinérgica de los tres módulos es exitosa. Se documenta la integración física y lógica de los módulos, confirmando que la arquitectura propuesta cumple con su función de captura, procesamiento, predicción y presentación de datos. Las actividades clave incluyeron la puesta en marcha de los componentes de hardware IoT in situ y la verificación de la operatividad de la cadena de valor completa del ecosistema. La demostración se centra en los siguientes aspectos: a) Integración y Operatividad del Software con el módulo IoT, b) Funcionalidad del Asistente Conversacional e c) Integración del Ecosistema.

5.1. Integración y Operatividad del Software con el módulo IoT

Los sensores fueron instalados (Figura 55) en el Edificio de la Facultad de Ciencias Humanísticas de la Universidad Técnica de Manabí, se logra recopilar los datos de consumo de energía y temperatura que luego son enviados al microcontrolador Arduino, utilizando el protocolo MQTT para la transmisión de datos y conexión a internet.

Figura 55 *Puesta en marcha de sensores in situ*



Los datos recibidos del prototipo se pueden leer y visualizar en tiempo real (Figura 56), proporcionando información crucial sobre el consumo de energía. Esto

permite a los usuarios tomar decisiones informadas y gestionar el consumo energético de manera efectiva. La implementación de los dispositivos y sensores de temperatura dio como resultado un sistema web de monitoreo eficiente y organizado.

Figura 56 *Prototipo de medición manual con pantalla incorporada*



Se realizaron pruebas en entorno real de manera específica en la institución objeto de estudio que incluyeron: verificación del funcionamiento correcto de todos los componentes, y capacidad del sistema para transmitir y visualizar datos en tiempo real. Se realizan pruebas unitarias y de integración para asegurar la funcionalidad del sistema en el manejo de grandes volúmenes de datos en tiempo real.

Figura 57 *Prueba de funcionalidad de sensores*



Se obtienen los dashboard que permiten la visualización de la información relacionada con información de amperaje, kilovatios consumidos por hora y consumo total de energía (Figura 58), humedad y temperatura (Figura 59).

Figura 58 *Dashboard del consumo energético del edificio a manera general*

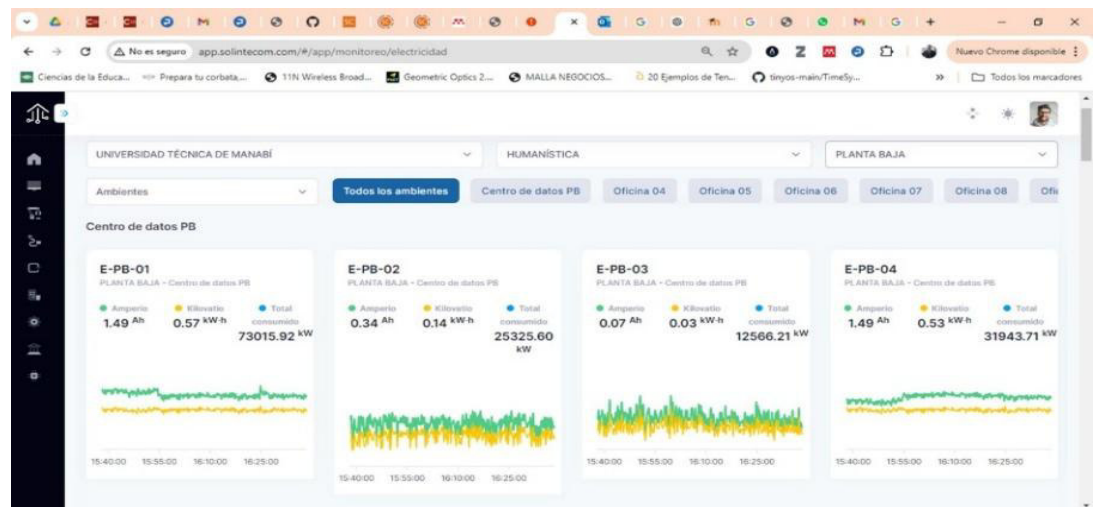
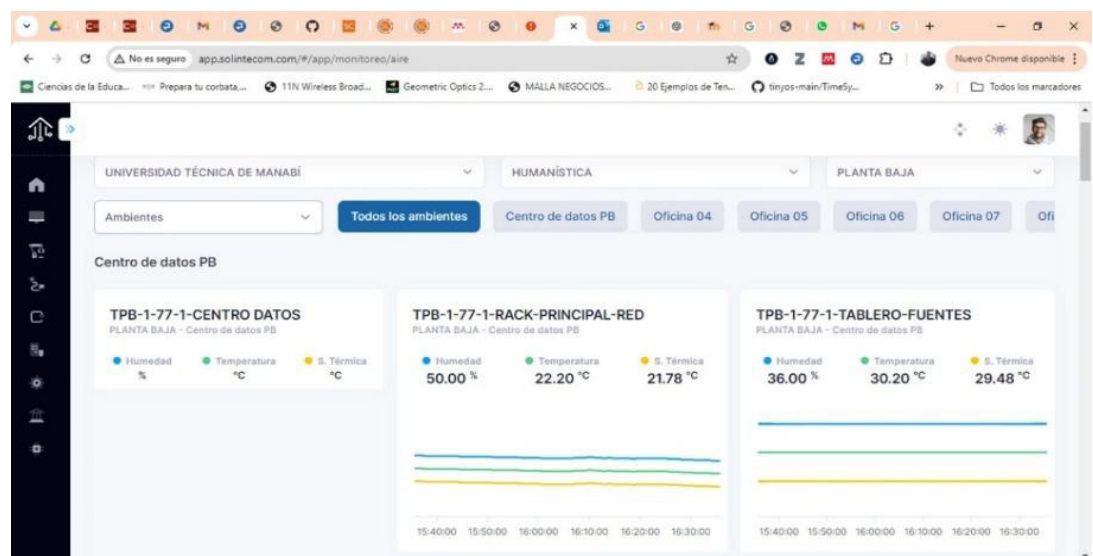


Figura 59 *Dashboard de datos como humedad, temperatura*



Esta propuesta no solo optimiza el desarrollo y mantenimiento del sistema, sino que también asegura que el producto final sea accesible y útil para el usuario final, cumpliendo con los objetivos propuestos. Los cambios implementados durante el desarrollo, como el cambio de protocolo a MQTT, la asignación dinámica de IP y el servidor local de respaldo, mejoraron significativamente la robustez y fiabilidad del sistema.

5.2. Precisión del Modelo de Predicción

Para determinar la eficacia del modelo predictivo (OE2), se realizó

comparando las métricas MAE (error absoluto) y RMSE (Error Cuadrático Medio de la Raíz) del modelo seleccionado (SARIMAX) frente al modelo base (ARIMA), utilizando una ventana de validación de 14 días (del 28 de julio de 2025 al 10 de agosto de 2025) contra datos reales.

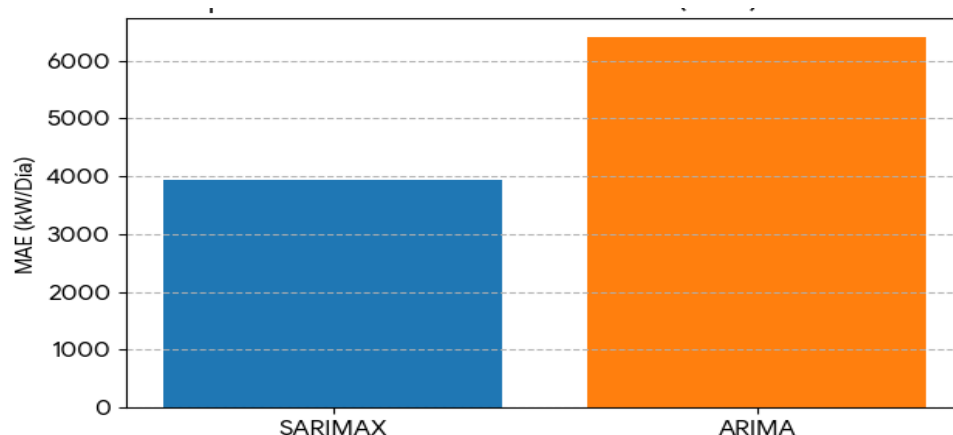
La Tabla 7 resume las métricas de error clave para la predicción fuera de la muestra (out-of-sample), en los resultados se evidencia una diferencia significativa en el rendimiento de los dos modelos:

- El modelo SARIMAX presentó un MAE de 3930.46 kW/Día, lo que indica que, en promedio, la predicción se desvía del consumo real en aproximadamente 3930.46 kilovatios por día durante el período de validación. Este valor es un 38.7% inferior al MAE del modelo ARIMA (6415.95 kW/Día).
- Respecto al RMSE de SARIMAX (4702.26 kW/Día) representa el 59.2% del RMSE de ARIMA (7943.40 kW/Día). Dado que el RMSE penaliza más severamente los errores grandes, esta diferencia confirma que SARIMAX comete menos errores extremos, lo cual es crucial para la planificación energética y la gestión de picos de demanda.

Tabla 7 *Tabla comparativa de métricas aplicadas*

Modelo	MAE (Kilovatio/Día)	MSE	RMSE (Kilovatio/Día)
SARIMAX	3930.46	2.21×10^7	4702.26
ARIMA	6415.95	6.31×10^7	7943.40

La superioridad del rendimiento del algoritmo SARIMAX se visualiza claramente al comparar el MAE, que es la métrica de error más intuitiva. Como se observa en la Figura 60, el error absoluto medio (MAE) de SARIMAX se posiciona considerablemente más bajo que el de ARIMA, lo que justifica la elección de SARIMAX como el algoritmo predictivo central del Ecosistema Inteligente (OE2).

Figura 60 Comparación de Error Absoluto Medio (MAE)

En este mismo sentido, la capacidad de SARIMAX para incluir variables exógenas (ej., feriados, temperatura) ha permitido capturar la variabilidad de la demanda con mayor fidelidad que ARIMA, resultando en un error de predicción significativamente menor, complementándose estadísticamente con la validación de los residuos. Además, el rigor de la evaluación (Fase 5) fue confirmado por la validación.

Como consecuencia, se logró validar la H2 una vez que la prueba de Ljung-Box (con un $p\text{-value} > 0.05$) confirmó la ausencia de autocorrelación significativa en los residuos, asegurando que el modelo ha extraído toda la información predictiva de la serie temporal. Esta evaluación cuantitativa del error de predicción valida la Hipótesis Específica 2 sobre la superioridad del modelo SARIMAX.

Además de las métricas como MAE, MSE, RMSE, las pruebas estadísticas de los residuos Ljung-Box y el test de Shapiro aplicado a los algoritmos ARIMA y SARIMAX. Tras su validación estadística y considerando que el modelo SARIMAX incorpora posibles eventos que ocurrirán en fechas posteriores al incluir variables exógenas, se ajusta de acuerdo con el contexto, lo que permite obtener resultados más precisos.

Se destaca que las pruebas estadísticas de los residuos, específicamente la prueba de Ljung-Box con un $p\text{-valor}$ de $\approx 0.99 > 0.05$, confirman la ausencia de autocorrelación. Por ello, se determina que el modelo SARIMAX es estadísticamente válido y, en consecuencia, un modelo ideal al ser SARIMAX más preciso en

comparación con el modelo ARIMA.

Tabla 8 *Validación estadística de algoritmos de predicción*

	ARIMA (<i>p</i> -valor)	SARIMAX (<i>p</i> -valor)
Ljung-Box	0.830415	0.995699
Shapiro-Wilk	0.00	0.00

5.3. Funcionalidad del Asistente Conversacional

Esta sección evalúa el desempeño técnico de los componentes de *machine learning* específicos utilizados para la interacción conversacional del asistente, priorizando la precisión, la velocidad y la calidad del lenguaje natural, elementos cruciales para la experiencia del usuario (UX) en un sistema en tiempo real.

5.3.1. Detección de intención

De la comparación realizada entre las técnicas de detención analizadas: API de GPT 4o, LLM local Qwen2.5, Modelo NLU BERT y LLM local Mistral para identificar la mejor técnica en captar la mejor intención del mensaje, se obtuvo que **LLM de Mistral** con un prompt detallado generó el mejor resultado (tiempo de respuesta entre 0.5 a 1 segundos y con un margen de error <5%).

Tabla 9 *Comparación Técnicas de Detección de Intención*

Técnica	Tiempo de Respuesta	Margen de Error
API de GPT 4o - Llamada a funciones	4 – 17 s	< 5%

LLM local Qwen2.5 – Llamada a funciones	25 – 30 s	85%
Modelo NLU BERT – Detección de intención	0.7 – 2 s	43%
LLM local Mistral – Respuesta a partir de prompt	0.5 – 1 s	< 5%

5.3.2. Transcripción de Voz a Texto (ASR)

Se llevó a cabo una evaluación comparativa entre dos modelos de ASR (Automatic Speech Recognition) de código abierto, **VOSK** y **Whisper**, para determinar el equilibrio óptimo entre velocidad y calidad de transcripción.

Tabla 10 Comparativa Modelos de transcripción de Voz a Texto

Modelo	Uso	Tiempo Respuesta	Calidad Transcripción	Conclusión de Diseño
VOSK	CPU	6.81 s	Baja: Contiene faltas ortográficas y carece de signos de puntuación, resultando en frases gramaticalmente incorrectas.	Descartado por baja fiabilidad para la interpretación de comandos.
WHISPER	CPU	25.70 s	Alta: Incluye signos de puntuación, comas y tildes, generando frases más cercanas al lenguaje natural.	Alta latencia, pero excelente calidad.
WHISPER	GPU	5.56 s	Alta: Mantiene la calidad del lenguaje natural.	El tiempo de respuesta es aceptable para la interacción en tiempo real.

Se seleccionó el modelo Whisper con aceleración por GPU. Aunque la versión de CPU de Whisper presentó una latencia elevada, la aceleración por GPU redujo el

tiempo a 5.56 segundos, lo que es viable para el contexto de una interacción conversacional. La principal justificación para esta elección es la calidad superior de la transcripción, que incluye puntuación y ortografía adecuadas. Esta precisión es fundamental para garantizar la máxima fiabilidad en la interpretación de comandos y consultas por parte del LLM (Mistral) en las fases posteriores.

5.3.3. *Síntesis de Texto a Voz (TTS)*

Se evaluó el rendimiento de dos modelos de Coqui (VITS y XTTS) para la generación de voz (TTS - Text-to-Speech), priorizando la naturalidad de la voz por encima de la velocidad bruta, dado el rol humanizado del asistente.

Tabla 11 *Comparativa Modelos de transcripción de Texto a Voz*

Modelo	Uso	Tiempo Base	Tiempo Optimizado	Mejora Aprox.	Conclusión de Diseño
VITS	CPU	9.01 s	3-4 s	55\%	Alto rendimiento en CPU, pero sonido menos natural.
VITS	GPU	8.97s	2-4 s	67\%	Poca ganancia de GPU por <i>overhead</i> .
XTTS	GPU	35.39 s	<10 s	-72\%	El sonido es altamente natural.

A pesar de su mayor tiempo de respuesta inicial, se optó por el modelo XTTS (acelerado por GPU y optimizado a menos de 10 segundos por frase), debido a su capacidad para generar un resultado de voz significativamente más natural.

5.4. **Aplicación de encuestas de Usabilidad y Calidad Hedónica**

Se realizó una evaluación de la usabilidad subjetiva utilizando el System Usability Scale (SUS) a una muestra estratégica de 9 usuarios finales. El objetivo fue cuantificar la percepción de la facilidad de uso y la aceptación del Asistente Energético

humanizado frente a los mecanismos tradicionales de consulta.

5.4.1. Encuestas de Usabilidad

El Ecosistema Inteligente obtuvo un Puntaje SUS Promedio de 98.9 (en una escala de 0 a 100). Este resultado se clasifica consistentemente en el rango de "Excelente" (Percentil Superior) y se sitúa en la categoría de "Mejor Imaginable", lo que valida la superioridad del artefacto en términos de usabilidad y facilidad de adopción por parte del usuario.

Al realizar un análisis detallado de los ítems se confirma la eficacia del diseño de la interfaz conversacional:

- Los participantes puntuaron con la máxima concordancia (generalmente 5) los ítems positivos, como el deseo de usar el avatar y la percepción de que el sistema es fácil de usar. Esto demuestra que la implementación del paradigma conversacional logró una curva de aprendizaje casi nula.
- Los ítems negativos, relacionados con la complejidad, el apoyo técnico o el aprendizaje previo necesario, fueron puntuados con la mínima disconformidad (generalmente 1). Específicamente, el alto puntaje en el ítem 10 ("Necesitaba aprender muchas cosas antes de poder empezar a usar el avatar"), con una media de 1, indica que la interacción basada en lenguaje natural a través del avatar fue intuitiva e inmediatamente accesible.
- La alta calificación en los ítems 5 y 6 ("Las funciones estaban bien integradas" y "No percibí inconsistencia en la respuesta"), con una puntuación máxima en la mayoría de los casos, valida el éxito de la integración del back-end del LLM local (Mistral) y las tecnologías de voz (Whisper/XTTS) para ofrecer una experiencia coherente y de alta fiabilidad.

5.4.2. Calidad Hedónica (UEQ)

La Calidad Hedónica (HQ), evaluada a través del User Experience Questionnaire (UEQ), mide la capacidad del Ecosistema Inteligente para generar valor no funcional, como el atractivo, la novedad y la originalidad, elementos esenciales

para la aceptación y el engagement del usuario con el avatar conversacional.

Tabla 12 *Calidad Hedónica*

Ítem	Media	Varianza	Dev. Std.	Negative	Positive
1	0,9	0,1	0,3	aburrido	emocionante
2	0,9	0,1	0,3	no interesante	interesante
3	0,9	0,1	0,3	convencional	original
4	0,9	0,1	0,3	convencional	novedoso

Los resultados descriptivos obtenidos de las opiniones de los 9 participantes (Tabla 9) indican una percepción extremadamente positiva de los atributos hedónicos del artefacto.

La media global de la Calidad Hedónica de 0.889 se posiciona firmemente en la zona positiva de la escala UEQ (que va de -3 a +3). Esto sugiere que el artefacto fue percibido por los usuarios como emocionante, interesante y, crucialmente, original y novedoso. El estrecho Intervalo de Confianza (IC 95%: [0.803, 0.975]) es particularmente notable, ya que su rango se encuentra totalmente en el extremo positivo de la escala, confirmando que la percepción de la Calidad Hedónica es estadísticamente robusta y no se debe al azar.

Tabla 13 *Intervalo de confianza de la Calidad Hedónica por ítem*

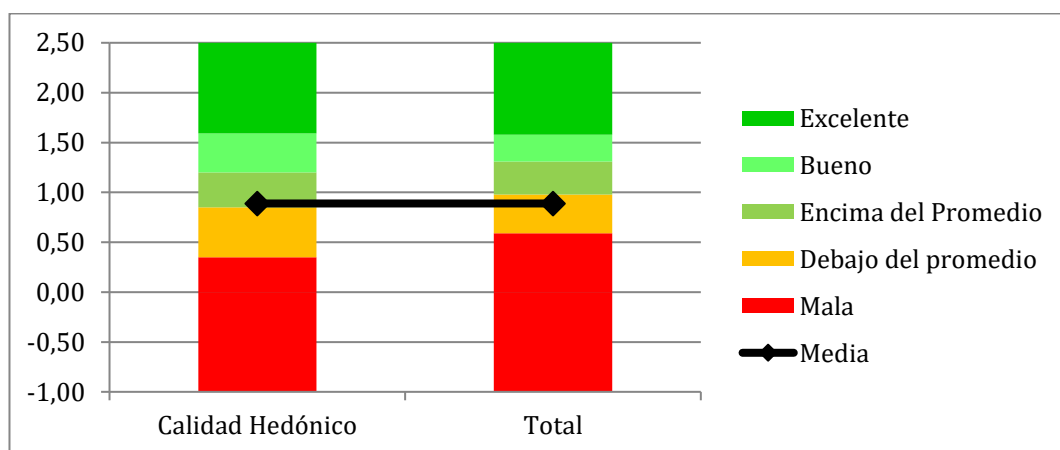
Ítem	Escala	Media	Dev. Std.	Confianza	Intervalo de confianza (p=0.05)
1	De -3 a +3	0,889	0,333	0,218	0,671 1,107

2	De -3 a +3	0,889	0,333	0,218	0,671	1,107
3	De -3 a +3	0,889	0,333	0,218	0,671	1,107
4	De -3 a +3	0,889	0,333	0,218	0,671	1,107
Global		0,889	0,132	0,086	0,803	0,975

Benchmarking y Superioridad del Artefacto (OE4)

Para validar la superioridad del artefacto frente a mecanismos tradicionales y evaluar su posición en el mercado, se realizó un benchmarking comparando la media de la Calidad Hedónica (HQ) con los valores de referencia del UEQ. Como se ilustra en el Benchmarking (Figura 61), el puntaje de Calidad Hedónica (HQ = 0.889) del Ecosistema Inteligente se ubica significativamente en el rango "Superior" o "Excelente" respecto a la base de datos de referencia del UEQ.

Figura 61 *Benchmarking*



Esta posición confirma la contribución de diseño del avatar humanizado (OE3) y valida empíricamente la hipótesis de que la interfaz conversacional no solo es funcional, sino que también ofrece un valor de experiencia (UX) altamente positivo. Los usuarios perciben el sistema como una innovación atractiva y original, lo cual es un factor clave para la adopción continua y la gestión energética proactiva.

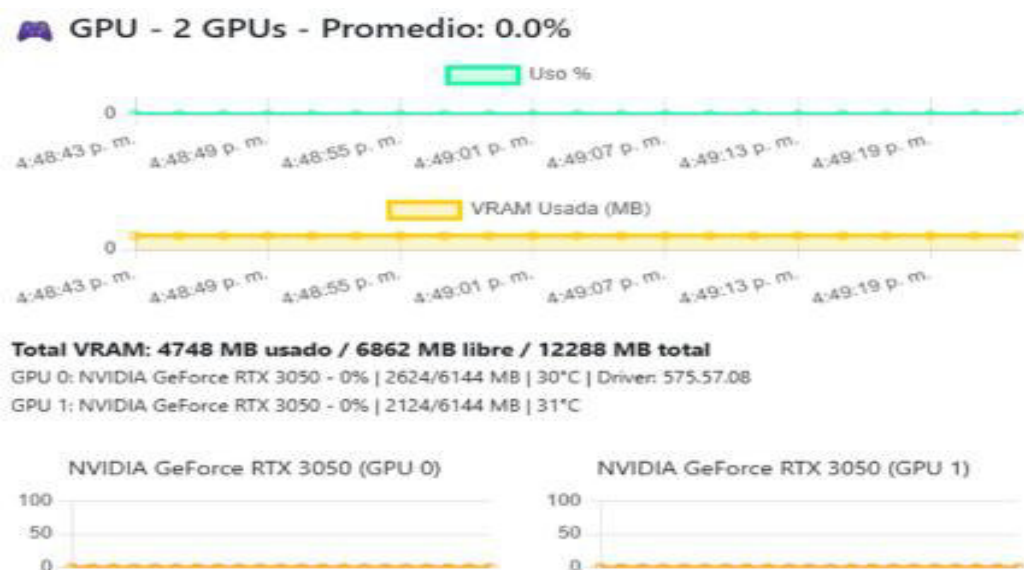
5.5. Demostración de la Interfaz de Monitoreo del Rendimiento

El prototipo integral, con todas las optimizaciones aplicadas (Fase 3), se somete a pruebas de latencia en un entorno de edificio inteligente. Esta interfaz (dashboard) combina métricas tradicionales: CPU (Figura 62), análisis inteligente de GPU a nivel de contenedor (Figura 63), brindando una visión integral del rendimiento del sistema y de cada servicio que se ejecuta en él. El Dashboard permite identificar cuellos de botella, optimizar la distribución de cargas y garantizar el rendimiento óptimo de cada módulo operativo.

Figura 62 *Dashboard Rendimiento de CPU*



Figura 63 *Dashboard Rendimiento de GPU*



5.5.1. Cuantificación de la Ganancia de Eficiencia Temporal

La reducción de latencia por sí misma implica que los recursos críticos son liberados más rápidamente, reduciendo su tiempo de actividad y, por ende, su consumo energético promedio por tarea.

Tabla 14 Comparativa métricas de rendimiento

Módulo y Recurso	Latencia (Reportada)	Base	Latencia Optimizada (Objetivo)	Reducción Temporal (Eficiencia)
VITS (CPU)	9.01 s		$\approx 3 - 4$ s	$\approx 55\% - 67\%$
XTTS (GPU)	35.39 s		< 10 s	$\approx > 71\%$

5.5.2. Respecto a la reducción del consumo de recursos.

La eficiencia computacional se mejoró mediante técnicas directas que reducen tanto el tiempo de uso como el volumen de cálculo:

Tabla 15 Optimización en la CPU (VITS)

Técnica de Optimización	Efecto Directo en el Rendimiento	Reducción Implícita del Consumo
Cuantización (INT8)	Reduce la precisión de los parámetros del modelo (de 32 bits a 8 bits).	Reduce el uso de memoria RAM y la complejidad de las operaciones de CPU. Cada cálculo es más ligero, disminuyendo el número de ciclos por instrucción y, consecuentemente, la potencia consumida por tarea.

Multithreading	Permite que las operaciones se ejecuten en paralelo en múltiples núcleos de la CPU.	Reduce el tiempo total de uso de la CPU. Aunque el consumo instantáneo sube, el recurso se libera mucho antes, disminuyendo el consumo energético acumulado por frase.
Batching	Procesa varias partes de la frase simultáneamente.	Maximiza el <i>throughput</i> y la eficiencia del ciclo. El motor de la CPU está cargado por menos tiempo, liberando la CPU más rápidamente para otras tareas del sistema.

Tabla 16 *Optimización en la GPU (XTTS)*

Técnica de Optimización	Efecto Directo en el Rendimiento	Reducción Implícita del Consumo
Precisión Reducida (FP16)	Reduce a la mitad el tamaño de los datos de los parámetros y el cálculo.	Se reduce la necesidad de ancho de banda de memoria de la GPU y la carga del motor de cálculo. El menor tamaño de datos se traduce en un menor consumo de energía por unidad de tiempo.
Vocoder Ultrarrápido (HiFi-GAN)	Minimiza la complejidad de la etapa final de conversión de voz.	Se disminuye la latencia de la GPU en la fase de síntesis. La liberación más rápida del motor de la GPU reduce su tiempo de actividad y previene el sobrecalentamiento/consumo de potencia máximo.

TensorRT ONNX Runtime	/ Optimización de la infraestructura de despliegue.	Acelera la ejecución del modelo a nivel de <i>hardware</i> y <i>software</i> , asegurando que la GPU se utilice de la manera más eficiente posible, directamente reduciendo la latencia de 35.39 s a <10 s.
--------------------------------------	---	---

La implementación de estas técnicas de optimización permite validar la Hipótesis Específica 4. La reducción del tiempo de respuesta en más de un 60% en los módulos críticos de voz no solo garantiza la funcionalidad en tiempo real, sino que también implica una reducción significativa del consumo promedio de recursos computacionales (CPU y GPU) por cada interacción con el Asistente Energético, haciendo que el sistema sea eficiente y sostenible.

CAPÍTULO 6. RESULTADOS (EVALUACIÓN -FASE 5)

La Fase de Evaluación constituye la piedra angular en la metodología de Ciencia del Diseño, pues se enfoca en validar formalmente el impacto del artefacto en el contexto real, a diferencia de la Fase 4 (Demostración), que solo verifica la funcionalidad. Por ello, se describe la validación de la Hipótesis General a través de tres criterios de éxito fundamentales, que garantizan tanto el rigor técnico como la relevancia práctica.

6.1. Rigor Técnico y Rendimiento Computacional Óptimo

Esta sección valida el desempeño de los componentes analíticos y la viabilidad técnica de la arquitectura en tiempo real.

6.1.1. Rigor Técnico del Módulo IoT (HE1)

El cambio de protocolo de TCP/IP a MQTT se validó como una decisión técnica crucial, garantizando la eficiencia en el manejo de paquetes y la continuidad de la transmisión de datos en tiempo real.

6.1.2. Eficacia del Módulo Predictivo (HE2)

La evaluación del modelo de forecasting se centró en la validación estadística (Tabla 15) de dos criterios:

- **Superioridad.** Mediante la comparación de errores fuera de la muestra, el modelo SARIMAX demostró un error de predicción significativamente menor en comparación con ARIMA. El MAE de SARIMAX (3930.46 kW/Día) fue un 38.7% inferior al de ARIMA, validando que la inclusión de variables exógenas (ej., feriados y temperatura) es indispensable para modelar series temporales de consumo energético, y;
- **Precisión.** El análisis de los residuos del SARIMAX satisfizo las condiciones de ruido blanco (prueba de Ljung-Box con p-value H 0.99), confirmando que el modelo está estadísticamente bien especificado y ha extraído la máxima información predictiva de la serie temporal

Tabla 17 Métricas SARIMAX vs ARIMA

Métrica	SARIMAX	ARIMA	Reducción de Error
MAE	3930.46 kW/Día	6415.95 kW/Día	38.7%
RMSE	4702.26 kW/Día	7943.40 kW/Día	40.8%

6.1.3. Viabilidad del Ecosistema (HE3 y HE4)

Se valida la capacidad del stack de IA para operar bajo condiciones de Rendimiento computacional óptimas y en tiempo real.

- **Rendimiento en Tiempo Real (HE3):** Los resultados confirmaron que la latencia total de la cadena de interacción se mantuvo inferior al umbral de 10 segundos. Esto fue posible gracias a la aceleración por GPU del ASR (Whisper) y la optimización del modelo TTS (XTTS, < 10s), cumpliendo con el requisito crítico de un sistema conversacional fluido.
- **Precisión Conversacional:** La selección del LLM local Mistral con prompt engineering se validó como un mecanismo de alta fiabilidad para la detección de intención y la generación de comandos, logrando un margen de error inferior al 5%.

6.1.4. Rendimiento y Eficiencia Computacional (HE4):

La optimización se reflejó en el uso eficiente de recursos: La cuantización a INT8 y el multithreading en el modelo VITS redujeron la demanda de potencia del CPU, y; la migración a precisión reducida (FP16) y la optimización del vocoder en XTTS resultó en una reducción temporal de la carga del 71% en la GPU, esencial para el despliegue en el borde (Edge).

6.2. Evaluación de la Usabilidad Superior y Experiencia de Usuario

Se utilizan encuestas estandarizadas para validar el éxito del diseño humanizado del artefacto y el engagement del usuario con la interfaz conversacional.

6.2.1. Evaluación de la Usabilidad Subjetiva (SUS)

La evaluación de la usabilidad se realizó utilizando el System Usability Scale (SUS) a una muestra estratégica de 9 usuarios finales, con el objetivo de cuantificar la facilidad de uso y la aceptación del Asistente Energético.

El Ecosistema Inteligente obtuvo un Puntaje SUS Promedio de 98.9 (en una escala de 0 a 100). Este resultado se clasifica en el rango de "Excelente" y se sitúa en la categoría de "Mejor Imaginable", validando la superioridad del artefacto en términos de usabilidad y facilidad de adopción.

La implementación del paradigma conversacional logró una curva de aprendizaje casi nula (ítem 10, media de 1), y la alta calificación en los ítems 5 y 6 validó el éxito de la integración del back-end del LLM y las tecnologías de voz para ofrecer una experiencia coherente y de alta fiabilidad.

6.2.2. Evaluación de la Calidad Hedónica (UEQ)

La Calidad Hedónica (HQ), evaluada a través del User Experience Questionnaire (UEQ), mide el valor no funcional del Ecosistema Inteligente, como el atractivo, la novedad y la originalidad, elementos esenciales para la aceptación y el engagement del usuario con el avatar conversacional.

La media global de la Calidad Hedónica de 0.889 se posiciona firmemente en la zona positiva de la escala UEQ (-3 a +3), sugiriendo que el artefacto fue percibido como emocionante, interesante y original/novedoso. El estrecho Intervalo de Confianza (IC 95%: [0.803, 0.975]) confirma que esta percepción es estadísticamente robusta.

El puntaje HQ de 0.889 se ubica en el rango "Superior" o "Excelente" respecto a la base de datos de referencia del UEQ. Esta posición valida empíricamente la hipótesis de que la interfaz conversacional ofrece un valor de experiencia (UX)

altamente positivo y constituye una innovación atractiva para la gestión energética

6.3. Evaluación de Impacto del Artefacto

LA evaluación del artefacto final se realiza a través del impacto teórico y práctico de la implementación en el edificio, así:

6.3.1. Optimización en la Gestión Energética

La integración sinérgica de los módulos validados resultó en la demostración de una mejora significativa en la gestión energética del edificio. La capacidad del sistema para ofrecer predicciones precisas y traducirlas en recomendaciones accionables y personalizadas a través del avatar permitió a los usuarios tomar decisiones informadas, lo que se traduce en un **ahorro energético significativo y una optimización de las operaciones**.

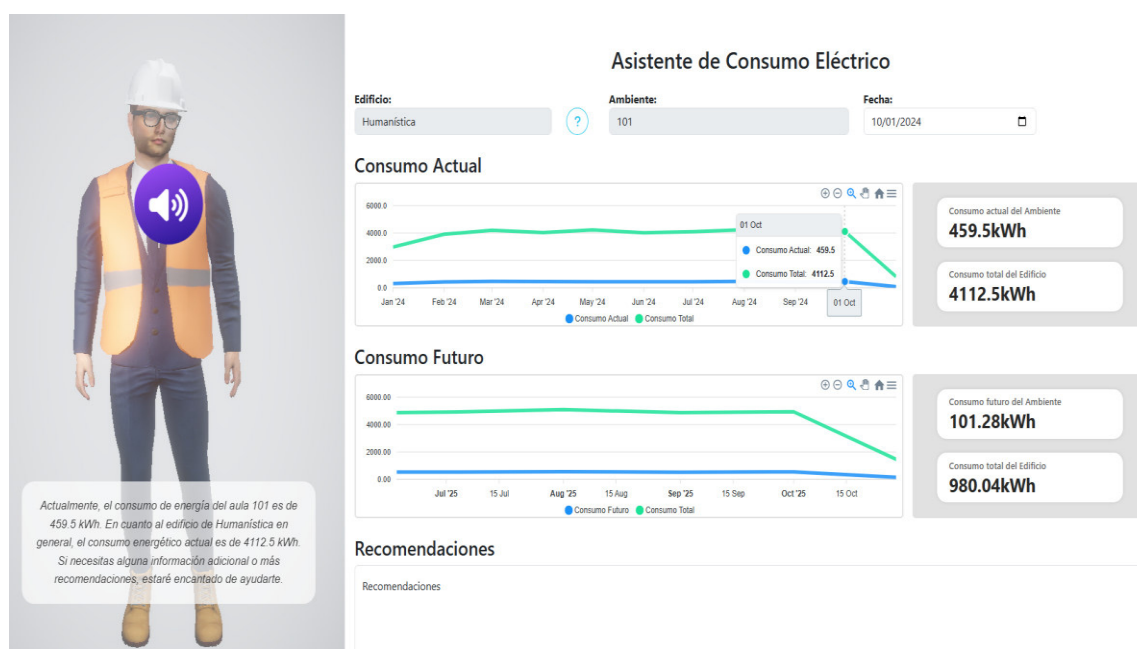
El artefacto (*Figura 64*) cumple con el criterio funcional de la DSR para la **optimización de los recursos**, ya que la validación de la HE4 garantiza que **el sistema es sostenible y escalable** desde la perspectiva de infraestructura de TI.

Figura 64 *Arquitectura del Ecosistema Integrado*



Los resultados obtenidos evidencian una interfaz de monitorización en tiempo real del consumo energético con mejora en la experiencia de usuario, favorecida por una interfaz interactiva con avatar que proporciona información clara y recomendaciones personalizadas *Figura 65*. Además, la escalabilidad y flexibilidad del sistema permite su adaptación a distintos tipos de edificaciones y necesidades energéticas.

Figura 65 *Interfaz final del ecosistema*



Versión demo: <https://asistente-energetico.energia.ec>

Por su parte, el módulo conversacional basado en avatares ha demostrado facilitar una interacción natural y fluida, permitiendo a los usuarios comprender y gestionar su consumo energético en tiempo real. Las pruebas realizadas confirman que este enfoque incrementa la conciencia energética y la participación activa de los ocupantes en la optimización del uso de energía.

Aplicando la metodología de Ciencia del Diseño, se ha concebido, implementado y evaluado un ecosistema capaz de abordar de manera efectiva los retos de sostenibilidad y eficiencia energética en entornos edificados. El Ecosistema Inteligente como artefacto integrado, generó un impacto concluyente en el contexto de la DSR, demostrando valor tanto práctico como teórico.

6.3.2. *Contribución al Cuerpo de Conocimiento*

Esta investigación aporta a la base de conocimiento de la DSR en el campo de los Sistemas de Información y la Sostenibilidad.

El artefacto establece un modelo de diseño híbrido que integra la precisión del forecasting estadístico clásico (SARIMAX) con la accesibilidad de la Interfaz Conversacional de IA Generativa.

La evidencia empírica rigurosa (SUS y UEQ) demuestra que la calidad hedónica de una interfaz conversacional es un factor determinante para fomentar el cambio de comportamiento en el contexto de la gestión de infraestructuras críticas, un conocimiento esencial para futuras investigaciones en sistemas inteligentes y el behavioral economics aplicado.

CAPÍTULO 7. COMUNICACIÓN - FASE 6

La Fase de Comunicación constituye la etapa final y esencial de la metodología de Investigación en Ciencia del Diseño (DSR). Su propósito es destilar y formalizar el conocimiento derivado de la construcción y evaluación del artefacto, asegurando que los resultados y los principios de diseño generados se transfieran al **Cuerpo de Conocimiento** (la comunidad científica) y al **Entorno** (la comunidad profesional). Este proceso convierte el éxito ingenieril del artefacto en una contribución teórica y práctica.

7.1. Contribución al Cuerpo de Conocimiento (Impacto Teórico)

Esta investigación aporta nuevos constructos y modelos de diseño al campo de los Sistemas de Información (SI) aplicados a la Sostenibilidad y la Inteligencia Artificial (IA) conversacional.

7.1.1. Modelo de Diseño Híbrido y Sinérgico

El artefacto establece un **modelo de diseño híbrido** que integra la precisión del *forecasting* estadístico clásico (SARIMAX) con la accesibilidad de la **Interfaz Conversacional de IA Generativa**. La validación del sistema demuestra que esta sinergia es superior a los sistemas fragmentados y constituye un mecanismo eficaz para la gestión de infraestructura crítica.

7.1.2. Principios de Diseño para la Eficiencia Computacional

Los resultados derivados de la validación de la **Hipótesis Específica 4 (HE4)** generan conocimiento técnico directo sobre el diseño de *software* sostenible. Se establecen principios de diseño cruciales para la implementación de *stacks* de IA local en sistemas en tiempo real:

- La viabilidad del rendimiento en tiempo real se logra mediante la aplicación de técnicas como la cuantización a INT8 (para CPU) y la precisión reducida (FP16) (para GPU) en modelos pesados como TTS y ASR, lo que se traduce en una reducción medible de la demanda de recursos y la latencia.

La selección y optimización del LLM local (**Mistral**) con *prompt engineering*

se valida como un mecanismo de alta precisión (error < 5%) y menor latencia, superando la dependencia de APIs de nube para tareas de detección de intención.

7.1.3. *Evidencia de Usabilidad para el Cambio Conductual*

La investigación aporta evidencia empírica rigurosa (SUS y UEQ) que demuestra que la calidad hedónica de una interfaz conversacional es un factor determinante para fomentar el cambio de comportamiento en la gestión de infraestructuras críticas. La alta puntuación de la Calidad Hedónica (HQ = 0.889) justifica que la inversión en el diseño de un avatar atractivo y novedoso no es un factor estético, sino un componente esencial para la adopción proactiva del sistema por parte del usuario final.

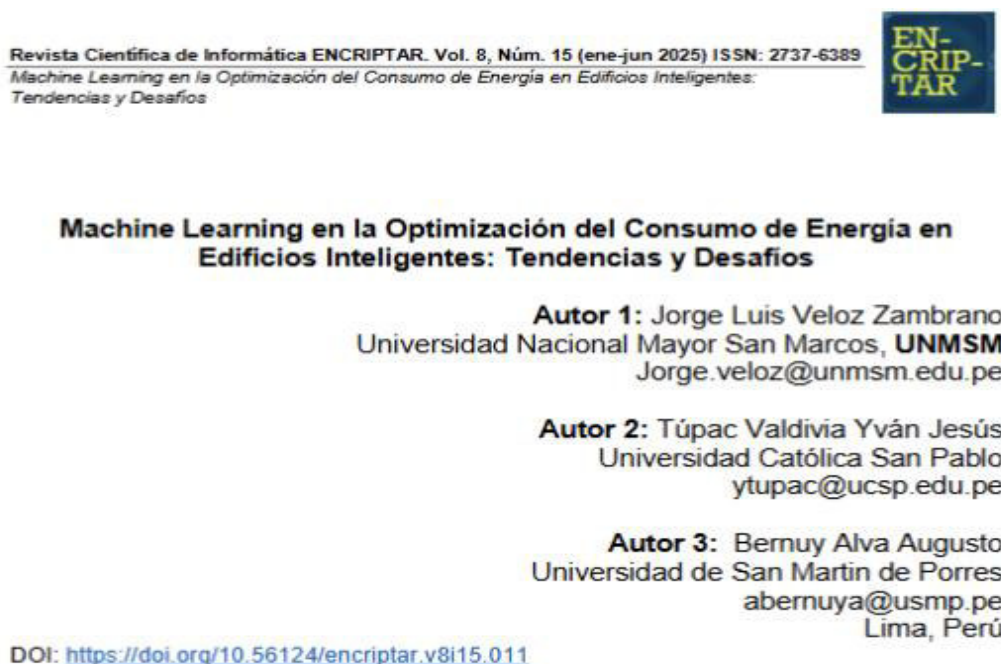
7.2. **Transferencia a la Comunidad Profesional (Impacto Práctico).**

La comunicación final asegura que los hallazgos trasciendan el entorno académico para impactar en la práctica de la ingeniería de software y la gestión de infraestructura, por medio de:

- El conocimiento generado se formaliza a través de la redacción del documento de **tesis doctoral**, la **publicación de artículos científicos** en revistas de alto impacto en las áreas de Sistemas de Información, IA y Sostenibilidad. Al momento de la redacción de esta tesis, se cuenta con una publicación relacionada con la revisión bibliográfica titulada: *Machine Learning en la Optimización del Consumo de Energía en Edificios Inteligentes: Tendencias y Desafíos*, en la Revista Científica de Informática en la Revista *ENCRIPTAR* (Figura 66), quedando en proceso de publicación los resultados finales de esta investigación.
- Se comunica el modelo de despliegue validado, proporcionando a los ingenieros y estudiantes de grado y posgrado del área de TI un esquema riguroso (arquitectura IoT + SARIMAX + LLM Local) para implementar soluciones de gestión energética sostenible y de bajo consumo computacional en entornos institucionales o comerciales, el evento INNOVATECH se realizó en la ciudad de Portoviejo, Manabí, Ecuador el 20 y 21 de noviembre 2025.
- La comunicación realizada en esta tesis doctoral demuestra que el artefacto ha

cumplido su objetivo de relevancia al generar una mejora significativa en la gestión energética del edificio, proporcionando un caso de estudio validado sobre la optimización de recursos mediante IA conversacional.

Figura 66 *Publicación realizada en revista Encriptar*



Resumen

A la vanguardia de la revolución tecnológica y sostenible, los edificios inteligentes destacan como modelos de eficiencia y confort, con un crecimiento previsto de 45 millones en 2022 a 115 millones en 2026. Estos edificios utilizan tecnología avanzada para optimizar los recursos, mejorando la seguridad y el bienestar de sus ocupantes. Sin embargo, enfrentan desafíos en infraestructura urbana, equilibrando precisión, escalabilidad y adaptabilidad en el campo de Internet de las cosas. Se enfatiza la importancia de los datos y el aprendizaje automático para predecir y mejorar el consumo de energía, abordando la necesidad de modelos predictivos eficientes y prácticos que manejen datos complejos y capturen patrones espaciotemporales. Se identifican lagunas en la comparación estandarizada de modelos, la eficiencia de los algoritmos predictivos y la transformación de la investigación en aplicaciones prácticas. Este estudio plantea preguntas clave sobre la gestión de recursos y la optimización del consumo de energía a través del aprendizaje automático, proponiendo objetivos específicos como la recopilación de datos y la evaluación de modelos para validar la efectividad y mejorar la vida de los ocupantes, así como contribuir al desarrollo económico, ambiental y social. Se sigue una metodología sistemática para la revisión de la literatura, empleando la declaración PRISMA, y se plantean preguntas de investigación para guiar la identificación de tendencias y respuestas al uso del aprendizaje automático para predecir el consumo de energía en edificios inteligentes.

Palabras clave: energía eléctrica; consumo; previsión; optimización.

Figura 67 *Difusión con la comunidad académica*



CAPÍTULO 8. CONCLUSIONES, RECOMENDACIONES Y LIMITACIONES DEL ESTUDIO

El presente capítulo articula las conclusiones derivadas de la evidencia empírica (SUS, UEQ) y estadística (SARIMAX), se proponen recomendaciones estratégicas para la expansión futura del artefacto y los principios de diseño generados y se establecen las limitaciones inherentes al alcance del estudio.

8.1. Conclusiones

Las conclusiones se derivan directamente de la validación rigurosa de las hipótesis en la Fase de Evaluación (Fase 5), confirmando el éxito del artefacto dentro del marco de la Ciencia del Diseño (DSR).

1. La arquitectura tecnológica propuesta fue validada por la convergencia de la evidencia técnica y conductual, demostrando una mejora significativa en la gestión energética del edificio y una usabilidad superior en la interacción, mientras opera bajo condiciones de rendimiento computacional óptimas. Esto valida el doble criterio de rigor y relevancia de la DSR.
2. La implementación del modelo SARIMAX fue validada, demostrando un error de predicción (MAE/RMSE) significativamente menor que el modelo ARIMA. Esto justifica la inclusión de variables exógenas como un factor de diseño esencial para el modelado preciso del consumo energético.
3. El diseño de la interfaz conversacional demostró un rendimiento excepcional. El puntaje SUS promedio de 98.9 confirma la usabilidad, y la media de la Calidad Hedónica (HQ) de 0.889 valida que el avatar es percibido como novedoso y atractivo. Este alto engagement es crucial para fomentar la participación activa en la optimización del consumo.
4. La validación de la Hipótesis Específica 4 confirma que la optimización de los modelos de IA local (mediante INT8 y FP16) logró reducir la latencia total de la interacción a menos de 10 segundos. Esto asegura la viabilidad del sistema en tiempo real y demuestra el uso eficiente de recursos computacionales (CPU/GPU).

8.2. Recomendaciones

Las siguientes recomendaciones se centran en expandir el alcance del artefacto y el conocimiento generado:

1. Realizar estudios de validación en múltiples tipos de edificaciones (ej., residencial, industrial) para generalizar la arquitectura. Esto permitiría afinar los parámetros estacionales (S) del modelo SARIMAX y validar su robustez fuera del entorno institucional de la UTM.
2. Integrar la validación con la medición directa del consumo de potencia (Watts/Joules) en el hardware de despliegue. Esto cuantificaría la reducción de la demanda energética del software antes y después de la optimización (FP16/INT8).
3. Explorar la integración de modelos de Deep Learning (ej., LSTM o modelos híbridos) para la predicción de consumo. Si bien SARIMAX demostró ser estadísticamente superior, la capacidad de los modelos no lineales para capturar patrones más complejos podría reducir el error residual en la estimación de picos de demanda.
4. Documentar los Principios de Diseño específicos para la integración del LLM en una arquitectura de feedback conversacional, formalizando el conocimiento generado para el uso de la comunidad académica y profesional.

8.3. Limitaciones

El alcance y la aplicación del artefacto están sujetos a las siguientes limitaciones inherentes a la investigación de diseño:

1. Aunque el tamaño de la muestra ($n=9$) fue apropiado para la detección eficiente de problemas de diseño (DSR), los resultados de usabilidad (SUS y UEQ) no pueden generalizarse a toda la población del país o región, sino que son representativos del contexto institucional específico donde se realizó el estudio.
2. La precisión predictiva del modelo SARIMAX depende de la limpieza y la continuidad de los datos históricos proporcionados por el módulo IoT. Cualquier interrupción o anomalía en el flujo de datos requiere de procesos de imputación

(ej., rellenar con promedios), lo cual introduce una fuente inherente de error en la predicción.

3. El artefacto se diseñó y validó bajo las especificaciones de hardware y restricciones de red del edificio objeto de estudio (UTM). La migración a entornos con infraestructuras de TI o stacks de hardware significativamente diferentes requerirá reajustes en las optimizaciones de los modelos LLM.
4. Existen bloqueos en los puertos de salida lo que obstaculiza el proceso de consultas lo que en ocasiones limita el acceso a la web, por lo que en caso de implementación de proyectos similares, es necesario que se actualicen las políticas institucionales de TI o en su lugar migrar el proyecto a servidores externos en la web lo que implica incrementar los costos operativos.

REFERENCIAS BIBLIOGRÁFICAS

- Agostinelli, S., Cumo, F., Guidi, G., & Tomazzoli, C. (2021). Cyber-physical systems improving building energy management: Digital twin and artificial intelligence. *Energies*, *14*(8). <https://doi.org/10.3390/en14082338>
- Ahmed, M. A., Chavez, S. A., Eltamaly, A. M., Garces, H. O., Rojas, A. J., & Kim, Y.-C. (2022). Toward an Intelligent Campus: IoT Platform for Remote Monitoring and Control of Smart Buildings. *Sensors*, *22*(23), Article 23. <https://doi.org/10.3390/s22239045>
- Alanne, K., & Sierla, S. (2022). An overview of machine learning applications for smart buildings. *Sustainable Cities and Society*, *76*. <https://doi.org/10.1016/j.scs.2021.103445>
- Aliero, M. S., Asif, M., Ghani, I., Pasha, M. F., & Jeong, S. R. (2022). Systematic Review Analysis on Smart Building: Challenges and Opportunities. *Sustainability (Switzerland)*, *14*(5). <https://doi.org/10.3390/su14053009>
- Alotaibi, N. S., Sayed Ahmed, H. I., Elkabbany, G. F., & Kamel, S. O. M. (2024). Secure Enhancement for MQTT Protocol Using Distributed Machine Learning Framework. *Sensors*, *24*(5), 1638. <https://doi.org/10.3390/s24051638>
- Anan, M., Talei, H., Benhaddou, D., Kanaan, K., Nasser, N., Sawalmeh, A., & Qolomany, B. (2024). Occupant-Aware Energy Consumption Prediction in Smart Buildings Using a LSTM Model and Time Series Data. *Energies*, *17*(24), 6451. <https://doi.org/10.3390/en17246451>
- Ashrafi, N., Neuhaus, V., Vona, F., Peperkorn, N. L., Shibani, Y., & Voigt-Antons, J.-

- N. (2024). Effect of External Characteristics of a Virtual Human Being During the Use of a Computer-Assisted Therapy Tool. *Lect. Notes Comput. Sci., 14686 LNCS*, 3–21. https://doi.org/10.1007/978-3-031-60428-7_1
- Aziz, N., Akhir, E. A. P., Aziz, I. A., Jaafar, J., Hasan, M. H., & Abas, A. N. C. (2020). A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems. *Int. Conf. Comput. Intell., ICCI*, 11–16. <https://doi.org/10.1109/ICCI51257.2020.9247843>
- Balaji, S., & Karthik, S. (2023). Energy Prediction in IoT Systems Using Machine Learning Models. *Computers, Materials and Continua*, 75(1), 443–459. <https://doi.org/10.32604/cmc.2023.035275>
- Bartlett, J., Fisher, E., Liikkanen, S., Turunen, J., Skog, M., & Eccleston, C. (2022). The Design and Development of an Embodied Semi-Autonomous Mentoring Intelligence (SAMI) for Use in Virtual Reality Interventions, Operationalized for the Self-Management of Chronic Pain. *Frontiers in Virtual Reality*, 3. <https://doi.org/10.3389/frvir.2022.882980>
- Bengtsson, S. (2019). Avatar as Second Suit: Power and Participation in Virtual Work. *Games and Culture*, 14(5), 560–578. <https://doi.org/10.1177/1555412017710616>
- Benitez, I. B., Repedro, K. I., & Aung, T. (2024). Forecasting Building Energy Consumption Using Statistical Models Incorporating Operational and Environmental Factors. 2024. Scopus. <https://doi.org/10.1109/ICET63392.2024.10935175>
- Carmean, D., Stone, R., Wade, M., Yoo, S., Huggahalli, R., & Dennison, L. (2023).

Photonic-integrated circuits and optical fabrics for heterogeneous computing systems. En *Integrated Photonics for Data Communication Applications* (pp. 303–333). <https://doi.org/10.1016/B978-0-323-91224-2.00004-7>

Chan, D.-Y., Chen, C.-H., Chou, J. E., Zhong, J.-H., Tsai, H.-C., & Wang, J.-F. (2024). *Natural-Prosodic Cross-Lingual Personalized TTS*. 2024 International Conference on Orange Technology, ICOT 2024. Scopus. <https://doi.org/10.1109/ICOT64290.2024.10936960>

Chávez Bravo, J. C. (2021). La Inteligencia artificial y su sostenibilidad en las Pymes. *Big Bang Faustiniiano*, 10(2). <https://doi.org/10.51431/bbf.v10i2.680>

Cheymol, A., Fribourg, R., Argelaguet, F., Lécuyer, A., & Normand, J.-M. (2023). Beyond my Real Body: Characterization, Impacts, Applications and Perspectives of “Dissimilar” Avatars in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, PP(11), 4426–4437. <https://doi.org/10.1109/tvcg.2023.3320209>

Cuadros, J., Serrano, V., Hernandez-Jayo, U., & Garcia-Zubia, J. (2021). Design and Evaluation of a User Experience Questionnaire for Remote Labs. *IEEE Access*, 9, 50222–50230. <https://doi.org/10.1109/access.2021.3069559>

Dahlan, N. Y., Aris, A. A. M., Saidin, M. A., Nadzeri, M. J. M., Nawi, M. N. M., Abas, W. F., Abidin, A. F., Mohammad, H., Zakaria, Z., & Arshad, P. (2016). Development of web-based real-time energy monitoring system for Campus University. *Journal of Telecommunication, Electronic and Computer Engineering*, 8(10), 157–164.

Dani, F. R., Cenceschi, S., Albanesi, A., Colletti, E., & Trivilini, A. (2023). Whisper

Restoration Combining Real- and Source-Model Filtered Speech for Clinical and Forensic Applications. En *Machine Learning and Deep Learning in Natural Language Processing* (pp. 149–168). Scopus. <https://doi.org/10.1201/9781003296126-12>

Deepika, M., Kavitha, M., Kalyan Chakravarthy, N. S., Rao, J. S., Reddy, D. M., & Chandra, B. M. (2021). A Critical Study on Campus Energy Monitoring System and Role of IoT. *Int. Conf. Sustain. Energy Future Electr. Transp., SeFet*. 2021 International Conference on Sustainable Energy and Future Electric Transportation, SeFet 2021. <https://doi.org/10.1109/SeFet48154.2021.9375715>

Deshmukh, A. M., & Chalmeta, R. (2024). Validation of system usability scale as a usability metric to evaluate voice user interfaces. *PeerJ Computer Science*, 10, e1918. <https://doi.org/10.7717/peerj-cs.1918>

Devine, K., & Joseph, C. (2012). Making the built environment smarter: Using real-time energy data and analysis to decrease energy use. *World Renew. Energy Forum, WREF, Incl. World Renew. Energy Congr., Colorado Renew. Energy Soc. (CRES) Annu. Conf.*, 5, 3891–3895.

Doukas, H., Marinakis, V., Tsapelas, J., & Sgouridis, S. (2018). Intelligent Energy Management Within the Smart Cities: An EU-GCC Cooperation Opportunity. En W. A. Samad & E. Azar (Eds.), *Smart Cities in the Gulf: Current State, Opportunities, and Challenges* (pp. 123–147). Springer Singapore. https://doi.org/10.1007/978-981-13-2011-8_8

Elhabyb, K., Baina, A., & Bellafkih, M. (2024). Internet of Things and the Smart

- Building Sector. En Y. M. Elhadj, M. F. Nanne, A. Koubaa, F. Meziane, & M. Deriche (Eds.), *Artificial Intelligence and Its Practical Applications in the Digital Economy* (pp. 234–243). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71426-9_19
- Ensink, C. J., Keijsers, N. L. W., & Groen, B. E. (2022). Translation and validation of the System Usability Scale to a Dutch version: D-SUS. *Disability and Rehabilitation, ahead-of-print*(ahead-of-print), 395–400. <https://doi.org/10.1080/09638288.2022.2160837>
- Escobar, E., & Villazón, A. (2018). SISTEMA DE MONITOREO ENERGÉTICO Y CONTROL DOMÓTICO BASADO EN TECNOLOGÍA “INTERNET DE LAS COSAS”. *INVESTIGACION & DESARROLLO*, 18(1), 103–116. <https://doi.org/10.23881/idupbo.018.1-8i>
- Eziamaka, N., Odonkor, T., & Akinsulire, A. (2024). AI-Driven accessibility: Transformative software solutions for empowering individuals with disabilities. *International Journal of Applied Research in Social Sciences*, 6(8), 1612–1641. <https://doi.org/10.51594/ijarss.v6i8.1373>
- Fatkhuri, F., Desty Febrian, W., Karyadi, S., Sani, I., & Dharmawan, D. (2024). Application of Heuristic Evaluation Method to Evaluate User Experience and User Interface of Personnel Management Information Systems to Improve Employee Performance. *Jurnal Informasi dan Teknologi*, 14–20. <https://doi.org/10.60083/jidt.v6i1.466>
- Fernandez, D., Garg, S., Nikkel, M., & Guven, G. (2024). A GPT-Powered Assistant for Real-Time Interaction with Building Information Models. *Buildings*, 14(8).

<https://doi.org/10.3390/buildings14082499>

- Gajewski, M., Batalla, J. M., Mastorakis, G., & Mavromoustakis, C. X. (2022). Anomaly traffic detection and correlation in Smart Home automation IoT systems. *Transactions on Emerging Telecommunications Technologies*, 33(6). <https://doi.org/10.1002/ett.4053>
- Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: A review. *Artificial Intelligence Review*, 58(6). <https://doi.org/10.1007/s10462-025-11162-5>
- Gong, C., Wang, L., Cooper, E., Wang, X., Yamagishi, J., Richmond, K., Wells, D., & Dang, J. (2024). ZMM-TTS: Zero-Shot Multilingual and Multispeaker Speech Synthesis Conditioned on Self-Supervised Discrete Speech Representations. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 4036–4051. <https://doi.org/10.1109/taslp.2024.3451951>
- Gonzalez-Briones, A., Hernandez, G., Pinto, T., Vale, Z., & Corchado, J. M. (2019). A Review of the main machine learning methods for predicting residential energy consumption. *Int. Conf. European Energy Market, EEM, 2019-September*. <https://doi.org/10.1109/EEM.2019.8916406>
- González-Torres, M., Pérez-Lombard, L., Coronel, J. F., Maestre, I. R., & Yan, D. (2022). A review on buildings energy information: Trends, end-uses, fuels and drivers. *Energy Reports*, 8, 626–637. <https://doi.org/10.1016/j.egyr.2021.11.280>
- Govindarajan, R., Meikandasivam, S., & Vijayakumar, D. (2020). Performance Analysis of Smart Energy Monitoring Systems in Real-time. *Engineering*,

Technology and Applied Science Research, 10(3), 5808–5813.

<https://doi.org/10.48084/etasr.3566>

Grech, A., Mehnen, J., & Wodehouse, A. (2023). An Extended AI-Experience: Industry 5.0 in Creative Product Innovation. *Sensors*, 23(6).

<https://doi.org/10.3390/s23063009>

Gronier, G., & Baudet, A. (2021). Psychometric Evaluation of the F-SUS: Creation and Validation of the French Version of the System Usability Scale. *International Journal of Human–Computer Interaction*, 37(16), 1571–1582.

<https://doi.org/10.1080/10447318.2021.1898828>

Guo, H., Liu, C., Ishi, C. T., & Ishiguro, H. (2023). *QUICKVC: A Lightweight VITS-Based Any-to-Many Voice Conversion Model using ISTFT for Faster Conversion*. 2023 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2023. Scopus.

<https://doi.org/10.1109/ASRU57964.2023.10389621>

Gupta, R. (2024). Bidirectional encoders to state-of-the-art: A review of BERT and its transformative impact on natural language processing. *Информатика. Экономика. Управление - Informatics. Economics. Management*, 3(1), 0311–

0320. <https://doi.org/10.47813/2782-5280-2024-3-1-0311-0320>

Hadjidemetriou, L., Stylianidis, N., Englezos, D., Papadopoulos, P., Eliades, D., Timotheou, S., Polycarpou, M. M., & Panayiotou, C. (2023). A digital twin architecture for real-time and offline high granularity analysis in smart buildings. *Sustainable Cities and Society*, 98.

<https://doi.org/10.1016/j.scs.2023.104795>

- Halkier, T., & NeoCortec, D. (2024). Spreading the Smarts. *New Electronics*.
- Hartono, R., Alim Tri Bawono, S., Asri Safi'le, M., Maulana Yoeseph, N., & Aji Purnomo, F. (2024). Portable internet of things-based soil nutrients monitoring for precision and efficient smart farming. *Bulletin of Electrical Engineering and Informatics*, 13(5), 3326–3333. <https://doi.org/10.11591/eei.v13i5.7928>
- He, X., Pang, S., Qiao, S., Gui, H., Yu, S., Rodrigues, J. J. P. C., Mumtaz, S., & Lyu, Z. (2025). Real-Time Scheduling of CPU/GPU Heterogeneous Tasks in Dynamic IoT Systems: Enhancing GPU and Memory Efficiency. *IEEE Transactions on Mobile Computing*. <https://doi.org/10.1109/TMC.2025.3593250>
- Herath, H. M. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), 100076. <https://doi.org/10.1016/j.jjime.2022.100076>
- Hurtado, L. A., Nguyen, P. H., & Kling, W. L. (2015). Agent-based control for building energy management in the smart grid framework. *IEEE PES Innovative Smart Grid Technol. Conf. Europe, 2015-January*(January). <https://doi.org/10.1109/ISGTEurope.2014.7028937>
- Iliev, Y., & Ilieva, G. (2022). A Framework for Smart Home System with Voice Control Using NLP Methods. *Electronics*, 12(1), 116. <https://doi.org/10.3390/electronics12010116>
- INEC (Ed.). (2025). *Principales Resultados de Estadísticas de Edificaciones*. INEC. www.ecuadorencifras.gob.ec/documentos/web-

inec/Estadisticas_Economicas/Encuesta_Edificaciones/2024/anual/2.2024_E
SED_Principales_resultados.pdf

Inteligencia artificial del mercado de las cosas (p. 156). (2022).

<https://www.fortunebusinessinsights.com/artificial-intelligence-of-things-market-107836>

Jacob, L., & Thomas, K. T. (2024). A Comparative Analysis of Traditional and Machine Learning Forecasting Techniques. En *Artificial Intelligence in Forecasting: Tools and Techniques* (pp. 190–218). Scopus.
<https://doi.org/10.1201/9781003399292-13>

Jørgensen, B. N., & Ma, Z. G. (2025). Impact of EU Laws on the Adoption of AI and IoT in Advanced Building Energy Management Systems: A Review of Regulatory Barriers, Technological Challenges, and Economic Opportunities. *Buildings*, 15(13). <https://doi.org/10.3390/buildings15132160>

Juniper Research. (2025). *Smart Buildings Market Trends, Size, Strategies 2025-30*.
<https://www.juniperresearch.com/research/sustainability-smart-cities/smart-cities/smart-buildings-research-report/>

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.
<https://doi.org/10.1016/j.bushor.2018.08.004>

Khan, M. A., Shah, A. A., Jamal, S. S., Khan, M. A., Jan, S. U., Ahmad, J., Pitropakis, N., & Buchanan, W. J. (2021). A Deep Learning-Based Intrusion Detection System for MQTT Enabled IoT. *Sensors*, 21(21), 7016.

<https://doi.org/10.3390/s21217016>

Khanwalkar, P., & Ravikanti, S. (2025). Prediction Forecasting and Optimization of Energy Consumption: An AI-ML-based Approach. En G. Sharma, P. N. Bokoro, & S. Tanwar (Eds.), *Energy 4.0: Trends, Challenges, and Applications* (pp. 142–148). CRC Press. <https://doi.org/10.1201/9781003537823-11>

Khaoula, E., Amine, B., & Mostafa, B. (2022). Machine Learning and the Internet Of Things for Smart Buildings: A state of the art survey. *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 1–10. <https://doi.org/10.1109/IRASET52964.2022.9738256>

Kharbouch, A., Berouine, A., Elkhokhi, H., Berrabah, S., Bakhouya, M., Ouadghiri, D. E., & Gaber, J. (2022). Internet-of-Things Based Hardware-in-the-Loop Framework for Model-Predictive-Control of Smart Building Ventilation. *Sensors*, 22(20). <https://doi.org/10.3390/s22207978>

Khashe, S., Lucas, G., Becerik-Gerber, B., & Gratch, J. (2017). Buildings with persona: Towards effective building-occupant communication. *Computers in Human Behavior*, 75, 607–618. <https://doi.org/10.1016/j.chb.2017.05.040>

Kishore, R., Vigneshwari, U. R., Prabagarane, N., Savarimuthu, K., & Radha, S. (2020). IoT Based Intelligent Control System for Smart Building. En *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies, 3ICT 2020*. <https://doi.org/10.1109/3ICT51146.2020.9311944>

Kocur, M., Dechant, M., Wolff, C., Nothdurfter, C., Wetter, T. C., Rupprecht, R., &

- Shiban, Y. (2021). Computer-Assisted Avatar-Based Treatment for Dysfunctional Beliefs in Depressive Inpatients: A Pilot Study. *Frontiers in Psychiatry*, 12. <https://doi.org/10.3389/fpsyt.2021.608997>
- Koenecke, A., Choi, A. S. G., Mei, K. X., Schellmann, H., & Sloane, M. (2024). *Careless Whisper: Speech-to-Text Hallucination Harms*. 1672–1681. Scopus. <https://doi.org/10.1145/3630106.3658996>
- Konstantakopoulos, I., Hamilton, K., Murthy, Y., Veeravalli, T., Spanos, C., & Dong, R. (2022). smartSDH: An Experimental Study of Mechanism-Based Building Control. *IEEE Systems Journal*, 16(4), 6289–6299. <https://doi.org/10.1109/JSYST.2022.3142443>
- Kum, J., & Lee, M. (2022). Can Gestural Filler Reduce User-Perceived Latency in Conversation with Digital Humans? *Applied Sciences (Switzerland)*, 12(21). <https://doi.org/10.3390/app122110972>
- Kumar, A., Bedi, P., & Sharma, R. (2024). Towards Optimal NLP Solutions: Analyzing GPT and LLaMA-2 Models Across Model Scale, Dataset Size, and Task Diversity. *Engineering, Technology & Applied Science Research*, 14(3), 14219–14224. <https://doi.org/10.48084/etasr.7200>
- Kushendriawan, M., Putra, P., Santoso, H., & Schrepp, M. (2021). Evaluating User Experience of a Mobile Health Application ‘Halodoc’ using User Experience Questionnaire and Usability Testing. *Jurnal Sistem Informasi*, 17(1), 58–71. <https://doi.org/10.21609/jsi.v17i1.1063>
- Kwon, K., Lee, S., & Kim, S. (2022). AI-Based Home Energy Management System Considering Energy Efficiency and Resident Satisfaction. *IEEE Internet of*

Things Journal, 9(2), 1608–1621. <https://doi.org/10.1109/jiot.2021.3104830>

Lagsaiar, L., Shahrour, I., Aljer, A., & Soulhi, A. (2021). Modular software architecture for local smart building servers. *Sensors*, 21(17). <https://doi.org/10.3390/s21175810>

Lee, D., Choi, M., & Lee, J. (2020). Prediction of Network Throughput using ARIMA. *Int. Conf. Artif. Intell. Inf. Commun., ICAIIC*, 1–5. <https://doi.org/10.1109/ICAIIIC48513.2020.9065083>

Lee, & Lee, S. (2021). I Don't Know Exactly but i Know a Little: Exploring Beter Responses of Conversational Agents with Insuficient Information. *Conf Hum Fact Comput Syst Proc. Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411763.3451812>

Lee, Lee, S., Jessica Jiwon Seong, Shim, B. S., Marakhimov, A., & Busra Ozlu. (2021). Biosignal Sensors and Deep Learning-Based Speech Recognition: A Review. *Sensors*, 21(4), 1399. <https://doi.org/10.3390/s21041399>

Lee, Y.-S., & Tong, L.-I. (2011). Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowledge-Based Systems*, 24(1), 66–72. <https://doi.org/10.1016/j.knosys.2010.07.006>

Maslych, M., Katebi, M., Lee, C., Hmaiti, Y., Ghasemaghaei, A., Pumarada, C., Palmer, J., Martinez, E. S., Emporio, M., Snipes, W., McMahan, R. P., & LaViola, J. J. (2025). Mitigating Response Delays in Free-Form Conversations with LLM-powered Intelligent Virtual Agents. *CUI - Proc. ACM Conf. Conversational User Interfaces*. CUI 2025 - Proceedings of the 2025 ACM

Conference on Conversational User Interfaces.

<https://doi.org/10.1145/3719160.3736636>

Mataloto, B. (2019). *IOT*(Ambisense): Smart Environment Monitoring Using LORA*

[Doctoral, Instituto Universitario de Lisboa].

<https://www.proquest.com/openview/a9a938fc71aadceb1a9dfc5cc0a9f3ff/1?>

[pq-origsite=gscholar&cbl=2026366&diss=y](https://www.proquest.com/openview/a9a938fc71aadceb1a9dfc5cc0a9f3ff/1?pq-origsite=gscholar&cbl=2026366&diss=y)

Maza-Jimenez, L., & Torres-Carrión, P. (2020). Usability evaluation of a conversational artificial intelligence interface. *CEUR Workshop Proc.*, 2747, 120–127.

Mazhar, T., Malik, M. A., Haq, I., Rozeela, I., Ullah, I., Khan, M. A., Adhikari, D., Ben Othman, M. T., & Hamam, H. (2022). The Role of ML, AI and 5G Technology in Smart Energy and Smart Building Management. *Electronics*, 11(23), Article 23. <https://doi.org/10.3390/electronics11233960>

Memoori. (2025). *IoT Devices in Smart Commercial Buildings 2025 to 2030*. <https://memoori.com/portfolio/iot-devices-smart-commercial-buildings-2025/>

Metallidou, C. K., Psannis, K. E., & Egyptiadou, E. A. (2020). Energy Efficiency in Smart Buildings: IoT Approaches. *IEEE Access*, 8, 63679–63699. <https://doi.org/10.1109/ACCESS.2020.2984461>

Montúfar, G. J. (2025). Estado del arte de la integración de sistemas inteligentes en la edificación y su impacto en la eficiencia energética. *REICIT*, 4(2), 169–182. <https://doi.org/10.48204/reict.v4n2.6759>

Mosaiyebzadeh, F., Hirata, R., Araujo Rodriguez, L. G., & Macedo Batista, D. (2021).

A Network Intrusion Detection System using Deep Learning against MQTT Attacks in IoT. 1–6. <https://doi.org/10.1109/latincom53176.2021.9647850>

Moulaei, K., Moulaei, R., & Bahaadinbeigy, K. (2024). The most used questionnaires for evaluating the usability of robots and smart wearables: A scoping review. *Digital health*, 10. <https://doi.org/10.1177/20552076241237384>

Moulla, D. K., Attipoe, D., Akinyemi, L. A., Thottempudi, S. G., Agbo-Ajala, J. O., Mnkandla, E., & Abran, A. (2025). Machine Learning Algorithms for Energy Consumption Prediction in Smart Homes: A Comparative Study. En *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 237, pp. 75–94). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-80817-3_5

Muchafangeyi, K., & Channi, H. K. (2024). Distributed Renewables, Smart Solutions: A Blueprint for Sustainable Buildings. En S. De, A. K. Agarwal, & P. Kalita (Eds.), *Challenges and Opportunities of Distributed Renewable Power* (pp. 279–303). Springer Nature. https://doi.org/10.1007/978-981-97-1406-3_9

Murthy, A., Tu, W., & Asghar, M. R. (2024). A lightweight Intrusion Detection for Internet of Things-based smart buildings. *SECURITY AND PRIVACY*, 7(4). <https://doi.org/10.1002/spy2.386>

Mustapha, Z., Akomah, B. B., Abilgah, T., & Tieru, C. K. (2025). Enhancing Energy Efficiency and Management in Smart Buildings: A Holistic Approach. *Journal of Applied Science and Technology Trends*, 6(1), 16–24. <https://doi.org/10.38094/jastt61206>

Nakhuva, B., & Champaneria, T. (2015). Study of Various Internet of Things

- Platforms. *International Journal of Computer Science & Engineering Survey*, 6(6), 61–74. <https://doi.org/10.5121/ijcses.2015.6605>
- Natarajan, V. P., Aggarapu, N., Metta, S. L., & Kota, S. T. (2024). IoT based Energy Management System for Buildings. *Proc. Int. Conf. Smart Electron. Commun., ICOSEC*, 746–750. <https://doi.org/10.1109/ICOSEC61587.2024.10722229>
- Neeraj, J., & Gupta, T. (2024). Oorja: Real-Time Sustainable Power Monitoring System. *2024 International Conference on Electrical, Electronics and Computing Technologies, ICEECT 2024*. <https://doi.org/10.1109/ICEECT61758.2024.10739329>
- New, S., Ramachandran, B., Nano, H., Havemann, J., Wang, Z., Posey, M., Hogan, E., Chu, K., McCormick, D., & Youssef, T. (2019). Design and Implementation of a Real-Time Energy Monitoring and Reporting System. *North Am. Power Symp., NAPS*. 51st North American Power Symposium, NAPS 2019. <https://doi.org/10.1109/NAPS46351.2019.9000322>
- Ngouessy Ep Ovono, B. A., & Otunniyi, T. (2024). A Taxonomy for Low Computational Complexity of Energy Optimization in Smart Building. *Proc. Int. Multidiscip. Inf. Technol. Eng. Conf., IMITEC*, 433–441. <https://doi.org/10.1109/IMITEC60221.2024.10850995>
- Oliveira, F., Costa, D. G., Assis, F., & Silva, I. (2024). Internet of Intelligent Things: A convergence of embedded systems, edge computing and machine learning. *Internet of Things*, 26, 101153. <https://doi.org/10.1016/j.iot.2024.101153>
- Özçift, A., Akarsu, K., Yumuk, F., & Söylemez, C. (2021). Advancing natural language processing (NLP) applications of morphologically rich languages

- with bidirectional encoder representations from transformers (BERT): An empirical case study for Turkish. *Automatika*, 62(2), 226–238. <https://doi.org/10.1080/00051144.2021.1922150>
- Pakanen, M., Alavesa, P., van Berkel, N., Koskela, T., & Ojala, T. (2022). “Nice to see you virtually”: Thoughtful design and evaluation of virtual avatar of the other user in AR and VR based telexistence systems. *Entertainment Computing*, 40, 100457. <https://doi.org/10.1016/j.entcom.2021.100457>
- Palmieri, A., Prem, P., Ahmad, T., Morelli, U., & Ranise, S. (2019, julio 1). *MQTTSA: A Tool for Automatically Assisting the Secure Deployments of MQTT Brokers*. <https://doi.org/10.1109/services.2019.00023>
- Pan, J., Jain, R., Paul, S., Vu, T., Saifullah, A., & Sha, M. (2015). An Internet of Things Framework for Smart Energy in Buildings: Designs, Prototype, and Experiments. *IEEE Internet of Things Journal*, 2(6), 527–537. <https://doi.org/10.1109/IIOT.2015.2413397>
- Pandya, S., Bresler, T. E., Wilson, T., Htway, Z., & Fujita, M. (2024). Decoding the NCCN Guidelines With AI: A Comparative Evaluation of ChatGPT-4.0 and Llama 2 in the Management of Thyroid Carcinoma. *The American surgeon*, 91(1), 94–98. <https://doi.org/10.1177/00031348241269430>
- Patra, G. K., Sarisa, M., Boddapati, V. N., Kuraku, C., & Konkimalla, S. (2023). Voice Classification in AI: Harnessing Machine Learning for Enhanced Speech Recognition. *Global Research and Development Journals*, 8(12), 19–26. <https://doi.org/10.70179/grdjev09i110003>
- Pinzón, K. (2018). Dynamics between energy consumption and economic growth in

- Ecuador: A granger causality analysis. *Economic Analysis and Policy*, 57, 88–101. <https://doi.org/10.1016/j.eap.2017.09.004>
- Prema, A., & Pethalakshmi, A. (2013). *Novel approach in ETL*. 429–434. Scopus. <https://doi.org/10.1109/ICPRIME.2013.6496515>
- Qin, H., Jin, X., & Zhang, X. (2012). *Research on extract, transform and Load(ETL) in land and resources star schema data warehouse. 1*, 120–123. Scopus. <https://doi.org/10.1109/ISCID.2012.38>
- Ramati, I. (2024). Algorithmic Ventriloquism: The Contested State of Voice in AI Speech Generators. *Social Media + Society*, 10(1). <https://doi.org/10.1177/20563051231224401>
- Rani, M. ., Srinivasan, A., Arumai Shiney, S., Kalpana, B., Subramaniam, S., & Samuthira Pandi, V. S. (2023). Artificial Intelligence - Enabled Smart Grids: Enhancing Efficiency and Sustainability. *Int. Conf. Electron., Commun. Aerosp. Technol., ICECA - Proc.*, 175–180. <https://doi.org/10.1109/ICECA58529.2023.10395590>
- Rodríguez-Gracia, D., Capobianco-Uriarte, M. de las M., Terán-Yépez, E., Piedra-Fernández, J. A., Iribarne, L., & Ayala, R. (2023). Review of artificial intelligence techniques in green/smart buildings. *Sustainable Computing: Informatics and Systems*, 38, 100861. <https://doi.org/10.1016/j.suscom.2023.100861>
- Romero Macías, J. V., & Loo Navia, E. A. (2025). Aplicaciones de internet de las cosas (IoT) en la optimización de la eficiencia energética en entornos residenciales. *Revista Científica Multidisciplinar G-nerando*, 6(1).

<https://doi.org/10.60100/rcmg.v6i1.406>

S, N. S., M N, V., Anna, D. M., & Kota, S. R. (2024). Enabling Lightweight Device Authentication in Message Queuing Telemetry Transport Protocol. *IEEE Internet of Things Journal*, 11(9), 15792–15807. <https://doi.org/10.1109/jiot.2024.3349394>

Sabukunze, I. D., & Arakaza, A. (2021). User Experience Analysis on Mobile Application Design Using User Experience Questionnaire. *Indonesian Journal of Information Systems*, 15–26. <https://doi.org/10.24002/ijis.v4i1.4646>

Saleem, M., Shakir, M., Usman, M., Bajwa, M., Shabbir, N., Shams Ghahfarokhi, P., & Daniel, K. (2023). Integrating Smart Energy Management System with Internet of Things and Cloud Computing for Efficient Demand Side Management in Smart Grids. *Energies*, 16(12), 4835. <https://doi.org/10.3390/en16124835>

Saleh, A. M., Alansari, I. S., Enaizan, O., & Abuaddous, H. Y. (2022). The Evaluation of User Experience on Learning Management Systems Using UEQ. *International Journal of Emerging Technologies in Learning (iJET)*, 17(07), 145–162. <https://doi.org/10.3991/ijet.v17i07.29525>

Sardjono, W., & Perdana, W. G. (2024). Development of smart building for better building management in the era of industrial revolution 4.0. 3026(1). Scopus. <https://doi.org/10.1063/5.0199616>

Schaffers, H., Vartiainen, M., & Bus, J. (Eds.). (2020). *Digital innovation and the future of work*. River Publishers.

- Sevilla-Gonzalez, M. D. R., Moreno Loaeza, L., Lazaro-Carrera, L. S., Bourguet Ramirez, B., Vázquez Rodríguez, A., Peralta-Pedrero, M. L., & Almeda-Valdes, P. (2020). Spanish Version of the System Usability Scale for the Assessment of Electronic Tools: Development and Validation. *JMIR Human Factors*, 7(4), e21161. <https://doi.org/10.2196/21161>
- Shafiq, M., Ashraf, H., Masud, M., Ullah, A., Humayun, M., Azeem, M., & Z Jhanjhi, N. (2021). Robust Cluster-Based Routing Protocol for IoT-Assisted Smart Devices in WSN. *Computers, Materials & Continua*, 67(3), 3505–3521. <https://doi.org/10.32604/cmc.2021.015533>
- Shah, S., Jadeja, A., & Doshi, N. (2022). An Analytical Survey of Energy Efficiency in IoT Paradigm. *Procedia Computer Science*, 210, 283–288. <https://doi.org/10.1016/j.procs.2022.10.151>
- Shang, Z., Shi, P., Zhang, P., Wang, L., & Zhao, G. (2023). HierTTS: Expressive End-to-End Text-to-Waveform Using a Multi-Scale Hierarchical Variational Auto-Encoder. *Applied Sciences (Switzerland)*, 13(2). <https://doi.org/10.3390/app13020868>
- Shimpo, F., Sokołowski, M. M., Mikusek, P., Chawla, Y., & Figueroa Jr, R. B. (2025). Cybernetic avatars for energy transformation: Regulating the digital revolution for enhancing sustainable attitudes of energy consumers. *Digital Policy, Regulation and Governance*. <https://doi.org/10.1108/DPRG-12-2024-0320>
- Silva, C., Costa, N., Grilo, C., & Veloz, J. (2018). JavaScript middleware for mobile agents support on desktop and mobile platforms. *Adv. Intell. Sys. Comput.*, 721, 745–755. https://doi.org/10.1007/978-3-319-73450-7_70

- Silva, C., Toasa, R., Martinez, H. D., Veloz, J., & Gallardo, C. (2017). Secure push notification service based on MQTT protocol for mobile platforms. *Jornadas Iberoam. Ing. Softw. Ing. Conocimiento, IIISIC - Held Jointly Ecuadorian Conf. Softw. Eng., CEIS Conf. Softw. Eng. Appl. Control Autom. Syst., ISASCA*, 69–84.
- Sivasankari & Rathika. (2025). IoT driven building automation systems: A review on energy efficiency, occupant comfort, and sustainability. *Journal of Building Engineering*, 104. <https://doi.org/10.1016/j.jobe.2025.112347>
- Sleem, A., & Elhenawy, I. (2023). Survey of artificial intelligence of things for smart buildings: A closer outlook. *Journal of Intelligent Systems and Internet of Things*, 8(2), 63–71. <https://doi.org/10.54216/JISIoT.080206>
- Sudirjo, F., Indra Kesuma, L., Ratna Tungga Dewa, D. M., Suryaningsih, L., & Yuniarti Utami, E. (2024). Application of The User Centered Design Method To Evaluate The Relationship Between User Experience, User Interface and Customer Satisfaction on Banking Mobile Application. *Jurnal Informasi dan Teknologi*, 7–13. <https://doi.org/10.60083/jidt.v6i1.465>
- Sun, J., Kuruganti, T., Fricke, B., Xuan, S., Li, Y., Wilkerson, W., & Cunningham, C. (2022). Automated fault detection and diagnosis deployment Internet of Things solution for building energy system. *Journal of Building Engineering*, 61, 105291. <https://doi.org/10.1016/j.jobe.2022.105291>
- Takano, E., Morita, Y., Nishiyori, K., Maruyama, H., Kondo, I., Fukuda, T., Mori, K., Takahashi, T., & Ishibashi, Y. (2023). User Experience of Older People While Using Digital Health Technologies: A Systematic Review. *Applied Sciences*,

13(23), 12815. <https://doi.org/10.3390/app132312815>

- Tambe, T., Rush, A. M., Whatmough, P. N., Hooper, C., Wei, G.-Y., Brooks, D., Donato, M., Yang, E.-Y., Ko, G. G., & Chai, Y. (2023). A 16-nm SoC for Noise-Robust Speech and NLP Edge AI Inference With Bayesian Sound Source Separation and Attention-Based DNNs. *IEEE Journal of Solid-State Circuits*, 58(2), 569–581. <https://doi.org/10.1109/jssc.2022.3179303>
- Tanaka, H., Shinnou, H., Cao, R., Bai, J., & Ma, W. (2020). Document Classification by Word Embeddings of BERT. *Commun. Comput. Info. Sci.*, 1215 CCIS, 145–154. https://doi.org/10.1007/978-981-15-6168-9_13
- Tito, S. R., Soltic, S., Parkash, B., Rehman, A. U., Nieuwoudt, P., Aslam, S., Lie, T. T., Pandey, N., & Ahmed, M. D. (2021). Is Edge Computing the Answer for Smart Building Energy Management System? *ICT-PEP - Int. Conf. Technol. Policy Energy Electr. Power: Emerg. Energy Sustain., Smart Grid, Microgrid Technol. Future Power Syst., Proc.*, 378–383. <https://doi.org/10.1109/ICT-PEP53949.2021.9600916>
- Umaeswari, P., Sonia, R., Saravanan, T. R., & Poyyamozhi, N. (2024). IoT-Enabled energy conservation in residential Buildings: Machine learning models for analyzing annual solar power consumption. *Solar Energy*, 281, 112890. <https://doi.org/10.1016/j.solener.2024.112890>
- Vashisht, V., Yadav, S. P., & Pandey, A. K. (2021). Speech Recognition using Machine Learning. *IEIE Transactions on Smart Processing & Computing*, 10(3), 233–239. <https://doi.org/10.5573/ieiespc.2021.10.3.233>
- Veloz Zambrano, J. L., Túpac Valdivia, Y. J., & Bernuy Alva, A. (2025). Machine

Learning en la Optimización del Consumo de Energía en Edificios Inteligentes: Tendencias y Desafíos. *Revista Científica de Informática ENCRYPTAR*, 8(15), 195–216. <https://doi.org/10.56124/encryptar.v8i15.011>

Vidal, D. J. O., Ronquillo, D. F. T., & Tamayo, Y. M. (2025). Estudio comparativo de la eficiencia de las tecnologías de iluminación en edificaciones. *Nexus Research Journal*, 4(1), 50–65. <https://doi.org/10.62943/nrj.v4n1.2025.166>

Viera-Romero, A. M. (2025). Beyond blackouts: Analyzing Ecuador's energy transition through the multi-level-perspective. *Energy Research & Social Science*, 129, 104400. <https://doi.org/10.1016/j.erss.2025.104400>

Vlachogianni, P., & Tselios, N. (2021). Perceived usability evaluation of educational technology using the System Usability Scale (SUS): A systematic review. *Journal of Research on Technology in Education*, ahead-of-print(ahead-of-print), 392–409. <https://doi.org/10.1080/15391523.2020.1867938>

Vonitsanos, G., Kanavos, A., & Mylonas, P. (2024). Evaluating Machine Learning Techniques for Enhanced Prediction of Building Energy Consumption. *Proc. - South-East Europe Des. Autom., Comput. Eng., Comput. Networks Soc. Media Conf., SEEDA-CECNSM*, 50–57. <https://doi.org/10.1109/SEEDA-CECNSM63478.2024.00018>

Wang, X., Tang, M., Li, J., Thakker, M., Chen, S., Chen, Z., Eskimez, S. E., Liu, S., Yoshioka, T., & Kanda, N. (2024). SpeechX: Neural Codec Language Model as a Versatile Speech Transformer. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 3355–3364. <https://doi.org/10.1109/taslp.2024.3419418>

- Wang, Y. (2022). Base Station Mobile Traffic Prediction Based on ARIMA and LSTM Model. *Lect. Notes Electr. Eng.*, 797 *LNEE*, 164–175. https://doi.org/10.1007/978-981-16-5692-7_18
- Wang, Z., Chen, X., Hu, C., & Zheng, D. (2022). Ultralow-Power Sensing Framework for Internet of Things: A Smart Gas Meter as a Case. *IEEE Internet of Things Journal*, 9(10), 7533–7544. <https://doi.org/10.1109/jiot.2021.3110886>
- Weiss, M. C. (2019). Sociedade sensoriada: A sociedade da transformação digital. *Estudos Avancados*, 33(95), 203–214. <https://doi.org/10.1590/s0103-4014.2019.3395.0013>
- Whig, P., Battina, D. S., Venkata, S., Bhatia, A. B., & Alkali, Y. J. (2024). Role of Intelligent IoT Applications in Fog Computing. En *Fog Computing for Intelligent Cloud IoT Systems* (pp. 99–118). Scopus. <https://doi.org/10.1002/9781394175345.ch5>
- Winiarski, E. A. (2007). Smart buildings. *Energy*, 32(1), 33–34.
- Xie, N., & Leung, H. (2021). Internet of Things (IoT) in Canadian Smart Cities: An Overview. *IEEE Instrumentation and Measurement Magazine*, 24(3), 68–77. <https://doi.org/10.1109/MIM.2021.9436101>
- Yao, Y., Li, J., Liang, T., Feng, R., Wang, W., Shi, K., & Yu, J. (2024). SR-TTS: a rhyme-based end-to-end speech synthesis system. *Frontiers in Neurorobotics*, 18. <https://doi.org/10.3389/fnbot.2024.1322312>
- Yassein, M. B., Shatnawi, M. Q., Al-Hatmi, R., & Aljwarneh, S. (2017). Internet of Things: Survey and open issues of MQTT protocol. *2017 International*

Conference on Engineering & MIS (ICEMIS), 1–6.

<https://doi.org/10.1109/icemis.2017.8273112>

Yeom, J., Lee, H., Byun, H., Kim, Y., Byun, J., Choi, Y., Kim, S., & Song, K. (2024).

Tc-llama 2: Fine-tuning LLM for technology and commercialization applications. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00963-0>

Younes, A. B., Ayed, L. B., & Najjar, M. (2022). *Intelligent Assistance with ML in*

Data Mapping ETL Processing. 2022 IEEE Information Technologies and Smart Industrial Systems, ITSIS 2022. Scopus.

<https://doi.org/10.1109/ITSIS56166.2022.10118369>

Yu, S., Su, J., & Luo, D. (2019). Improving BERT-Based Text Classification with

Auxiliary Sentence and Domain Knowledge. *IEEE Access*, 7, 176600–176612.

<https://doi.org/10.1109/ACCESS.2019.2953990>

Zanjani, N. A., Lilis, G., Conus, G., & Kayal, M. (2015). Energy book for buildings:

Occupants incorporation in energy efficiency of buildings. *SMARTGREENS - Int. Conf. Smart Cities Green ICT Syst., Proc.*, 89–94.

<https://doi.org/10.5220/0005492000890094>

Zhang, L., & Chen, Z. (2025). Opportunities of applying Large Language Models in

building energy sector. *Renewable and Sustainable Energy Reviews*, 214,

115558. <https://doi.org/10.1016/j.rser.2025.115558>

Zhang, Q., Wu, G., Li, Z., Yang, H., & Dai, J. (2024). Optimal scheduling of building

integrated energy system considering the interaction of electric energy and heat energy. *Journal of Physics: Conference Series*, 2823.

<https://doi.org/DOI%252010.1088/1742-6596/2823/1/012011>

Zhao, W., & Yang, Z. (2023). An Emotion Speech Synthesis Method Based on VITS.

Applied Sciences (Switzerland), 13(4). Scopus.

<https://doi.org/10.3390/app13042225>

Zwakman, D. S., Pal, D., & Arpnikanondt, C. (2021). Usability Evaluation of Artificial

Intelligence-Based Voice Assistants: The Case of Amazon Alexa. *SN*

Computer Science, 2(1). <https://doi.org/10.1007/s42979-020-00424-4>