

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

<https://doi.org/10.35381/i.p.v8i15.5188>

Chatbot con Inteligencia Artificial para la Gestión Proactiva de Permisos, Licencias y Riesgo Operacional

Artificial Intelligence Chatbot for Proactive Management of Permits, Licenses and Operational Risk

Alexander Andres Agurto-Guevara

aagurto4@utmachala.edu.ec

Universidad Técnica de Machala, Machala, El Oro
Ecuador

<https://orcid.org/0009-0004-2823-6653>

Jenner Joan Villalba-Paladines

jvillalba3@utmachala.edu.ec

Universidad Técnica de Machala, Machala, El Oro
Ecuador

<https://orcid.org/0009-0003-9163-8172>

Wilmer Braulio Rivas-Asanza

wrivas@utmachala.edu.ec

Universidad Técnica de Machala, Machala, El Oro
Ecuador

<https://orcid.org/0000-0002-2239-3664>

Edison Luis Lojan-Cueva

elojan@utmachala.edu.ec

Universidad Técnica de Machala, Machala, El Oro
Ecuador

<https://orcid.org/0000-0002-7092-1281>

Recibido: 02 de abril 2026

Revisado: 03 de mayo 2026

Aprobado: 15 de junio 2026

Publicado: 01 de julio 2026

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

RESUMEN

El estudio tuvo como objetivo desarrollar e implementar un chatbot con inteligencia artificial para optimizar la consulta y gestión de permisos del personal del área de talento humano de una institución pública, integrando validación normativa automática y clasificación proactiva de riesgo operativo. Se aplicó la metodología Agile/SCRUM sobre una arquitectura de microservicios. El componente conversacional se construyó con DeepPavlov, complementado por Google Gemini en un rol acotado de enriquecimiento y contingencia; un motor de reglas en Python validó el cumplimiento normativo y un clasificador con scikit-learn estimó el riesgo operativo. La validación se realizó mediante seis casos de prueba funcionales. Los resultados evidenciaron la integración de las cuatro capacidades del chatbot, la clasificación correcta de los tres niveles de riesgo y la continuidad operativa ante fallos de servicios externos. Se concluyó que el enfoque híbrido demostró ser viable para la automatización proactiva de permisos en el sector público, ofreciendo trazabilidad institucional y respuestas adaptativas.

Descriptores: Inteligencia artificial; Procesamiento del lenguaje natural; Aprendizaje automático; Administración pública; Gestión de personal. (Tesauro UNESCO).

ABSTRACT

The study aimed to develop and implement an artificial intelligence chatbot to optimize the consultation and management of staff permits in the human talent area of a public institution, integrating automated regulatory validation and proactive operational risk classification. The Agile/SCRUM methodology was applied on a microservices architecture. The conversational component was built with DeepPavlov, complemented by Google Gemini in a supporting role of enrichment and contingency; a Python-based rules engine validated regulatory compliance and a scikit-learn classifier estimated operational risk. Validation was performed through six functional test cases. The results showed the integration of the chatbot's four functional capabilities, the correct classification of the three risk levels, and operational continuity in the face of external service failures. It was concluded that the hybrid approach proved viable for the proactive automation of permits in the public sector, offering institutional traceability and adaptive responses.

Descriptors: Artificial intelligence; Natural language processing; Machine learning; Public administration; Personnel management. (Tesauro UNESCO).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

INTRODUCCIÓN

El avance de la inteligencia artificial y la ciencia de datos está cambiando la manera en que los gobiernos gestionan los servicios públicos. Cada vez más instituciones incorporan sistemas automatizados y modelos predictivos para ganar eficiencia, reducir errores administrativos y ser más transparentes ante la ciudadanía (CAF, 2021). En Ecuador la adopción avanza, pero de forma heterogénea, condicionada por las brechas de infraestructura y por la cultura organizacional de cada entidad (Velasategui Campoverde, 2023). Convertir datos institucionales en información útil para decidir ya es una capacidad clave para sostener la calidad del servicio en la era digital, y al mismo tiempo es un motor que reconfigura competencias y prácticas dentro del sector público (Celi-Parraga et al., 2021; Gordon Graell, 2023; Padilla-Caballero et al., 2023; Wang et al., 2022).

Dentro de esta línea, los chatbots destacan como una de las aplicaciones con mayor impacto operativo en el sector público: son sistemas conversacionales capaces de atender consultas y ejecutar tareas administrativas en lenguaje natural sin intervención humana constante (Pionce Arteaga et al., 2022). La diferencia entre un chatbot transaccional básico y uno apto para entornos institucionales está en la integración de Procesamiento del Lenguaje Natural (PLN), que permite interpretar consultas complejas, adaptarse a distintos perfiles de usuario y procesar solicitudes con precisión normativa (Gordon Graell, 2023; Wang et al., 2022). En la administración interna de recursos humanos, donde las consultas son frecuentes y están altamente normadas, esta tecnología reduce tiempos de respuesta y libera al personal administrativo de tareas repetitivas, lo que transforma la propia perspectiva del talento humano frente a la automatización (Arias-Chávez et al., 2024; Casazola Cruz et al., 2021; Piña Ferrer, 2024; Salinas Santiago et al., 2024).

En el sector público, la adopción de chatbots enfrenta fricciones específicas, ligadas a las infraestructuras de datos y a las exigencias de transparencia del Estado (Kaun & Männiste, 2025). En el área de talento humano de una institución pública ecuatoriana,

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

esas fricciones aparecen en procesos manuales y descentralizados: las solicitudes de vacaciones y permisos se tramitan por correo electrónico y hojas de cálculo, lo que genera fragmentación de la información, errores en el cálculo de saldos y dificultades para verificar el cumplimiento normativo. A esto se suma la ausencia de un mecanismo centralizado que detecte conflictos entre permisos y eventos críticos (actividades críticas o turnos especiales), lo que incrementa el riesgo operativo y compromete la continuidad de los servicios institucionales. Frente a esta problemática, los frameworks comerciales como Google Dialogflow, IBM Watson Assistant o Amazon Lex quedan cortos en entornos públicos por sus restricciones de personalización del dominio, control de datos sensibles y costos de licenciamiento (Garzón-Quiroz et al., 2025). En contraste, el framework de código abierto DeepPavlov permite construir modelos de PLN ajustados al vocabulario institucional (Burtsev et al., 2018), con más flexibilidad para integrar reglas de negocio y menor dependencia de servicios en la nube que alternativas como Rasa (Espinosa-Luna et al., 2023; Alemán Espinoza et al., 2019). Aun así, una arquitectura basada solo en PLN no alcanza para garantizar normativas que exigen verificar condiciones múltiples (saldos acumulados, anticipación mínima, disponibilidad operativa, ausencia de conflictos con actividades programadas), por lo que el sistema propuesto combina DeepPavlov con un motor de reglas en Python que asegura tanto fluidez conversacional como precisión normativa (Garzón-Quiroz et al., 2025; Pionce Arteaga et al., 2022).

Para superar el carácter reactivo de los sistemas tradicionales de gestión de recursos humanos, el sistema propuesto suma un módulo de clasificación proactiva de riesgo operativo basado en aprendizaje automático (Pedregosa et al., 2011). Este enfoque predictivo resulta valioso en entornos institucionales, donde la coincidencia de permisos con actividades críticas programadas o coberturas mínimas puede comprometer la continuidad del servicio, y ya ha mostrado eficacia en dominios que necesitan anticipar riesgos operativos (Aljohani, 2023). En ese marco, este estudio desarrolla e implementa

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

un chatbot con IA que integra PLN, validación normativa automatizada y clasificación predictiva de riesgo operativo para el área de talento humano de una institución pública ecuatoriana.

MÉTODO

Diseño, contexto y variables

El estudio es de carácter aplicado y tecnológico, con enfoque cuantitativo basado en métricas objetivas de desempeño del sistema. El desarrollo se organizó bajo la metodología SCRUM, elegida por su capacidad de incorporar ajustes continuos ante cambios normativos y de integrar componentes tecnológicos complejos en ciclos cortos de trabajo (Garzón-Quiroz et al., 2025; Gordon Graell, 2023). El objetivo fue diseñar, implementar y evaluar un chatbot con inteligencia artificial para la gestión de vacaciones y permisos del personal del área de talento humano de una institución pública, con verificación normativa automática y anticipación de conflictos con actividades críticas o turnos programados. Por acuerdo de confidencialidad con la institución colaboradora, su denominación específica se omite y se conserva únicamente la descripción funcional del contexto.

La población objetivo comprende al personal operativo y administrativo de esa área que realiza solicitudes de vacaciones y permisos institucionales. Sobre estos sujetos, la variable independiente fue la implementación del chatbot con inteligencia artificial, y la variable dependiente fue la eficiencia operativa en la gestión de permisos, medida por el tiempo de respuesta, la reducción de errores en el cálculo de saldos y la trazabilidad del historial de solicitudes. Se consideraron como variables de control la unidad administrativa, la carga de actividades críticas programadas y los parámetros normativos vigentes. En todos los datos usados para el entrenamiento se aplicaron medidas de protección, sustituyendo la información personal identificable por valores codificados para garantizar la confidencialidad sin afectar su utilidad (Espinosa-Luna et al., 2023).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Instrumentos, datos y procedimiento

El procedimiento se desarrolló de forma secuencial y articuló cuatro instrumentos en el momento en que cada uno resultaba necesario. Primero se preparó un corpus de frases etiquetadas, que recoge ejemplos reales de consultas del personal con la intención que expresan y los datos clave que contienen; este corpus se usó para entrenar el componente de lenguaje natural mediante una partición Train/Test Split de 80% para entrenamiento y 20% para evaluación, lo que permite verificar el desempeño del modelo sobre datos no vistos (Espinosa-Luna et al., 2023). En paralelo se organizó un conjunto de datos históricos de solicitudes, verificando la completitud y consistencia de cada registro, que sirvió para entrenar el clasificador de riesgo bajo el mismo esquema de partición, garantizando un criterio uniforme de evaluación. Junto a ello se implementó el motor de validación normativa con las disposiciones institucionales vigentes y se configuró una bitácora de auditoría que registra de forma automática cada interacción del sistema, dejando constancia de qué solicitudes se procesaron, cuándo y con qué resultado. Finalmente se integraron todos los componentes y se ejecutaron casos de prueba diseñados para verificar el comportamiento del sistema en condiciones normales y ante situaciones límite, aplicando ajustes en cada sprint antes de avanzar.

Para garantizar la replicabilidad del estudio, el código fuente, los scripts de entrenamiento, las configuraciones de DeepPavlov, las colecciones de prueba y los diagramas se conservaron en un repositorio de control de versiones, disponible bajo solicitud al autor de correspondencia; los datos sensibles se reemplazaron por plantillas anonimizadas para preservar la confidencialidad sin afectar la verificación independiente de los resultados.

Arquitectura, flujo operativo e implementación

La solución se construyó sobre una arquitectura de microservicios para garantizar modularidad y trazabilidad (Oyekunle Claudius Oyeniran et al., 2024; Vahdati & Ramsin,

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

2024). Tres servicios independientes se reparten el trabajo. El microservicio de IA/PLN, en Python con FastAPI, ejecuta la comprensión del lenguaje con DeepPavlov sobre BERT (Devlin et al., 2019) y se apoya en Google Gemini para un rol acotado de enriquecimiento y contingencia; sobre esa base aplica la validación normativa mediante un motor de reglas y la clasificación de riesgo con scikit-learn. La capa transaccional GIATH, en Java con Spring Boot, opera como fuente única de verdad sobre la normativa, las solicitudes y los datos del personal. La interfaz Angular captura la entrada por texto o voz y renderiza las respuestas. La comunicación viaja por HTTP REST a través de Traefik y se autentica con una clave compartida (INTERNAL_SECRET_KEY), ya que GIATH no emite JSON Web Tokens. La resiliencia se sostiene con circuit breakers en tres niveles (Falahah et al., 2021; Saleh Sedghpour et al., 2023). Esta integración constituye un enfoque híbrido que combina flexibilidad lingüística, precisión normativa y anticipación predictiva de riesgos operativos (Garzón-Quiroz et al., 2025; Wang et al., 2022).

El procesamiento de cada solicitud atraviesa tres etapas que recorren los componentes Angular, Python y Java. En la primera, el mensaje del usuario llega al microservicio de IA, donde el LanguageToolAdapter clasifica el modo de intención en cuatro categorías (consultivo, operativo, personal o neutro) mediante análisis morfosintáctico, y el clasificador BERT de DeepPavlov asigna la intención con un puntaje de confianza sobre el que se aplican tres umbrales: por encima de 0.65 la intención se acepta, entre 0.40 y 0.65 el modo detectado actúa como criterio de desempate, y por debajo de 0.40 un circuit breaker deriva al motor heurístico local (Figura 1).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

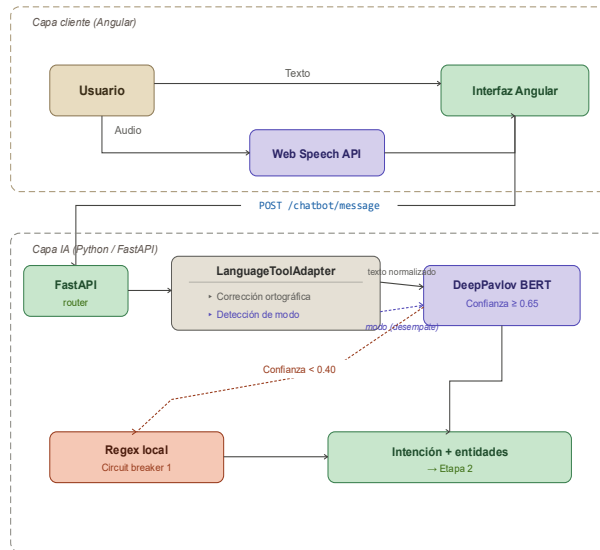


Figura 1. Flujo de intercepción y comprensión.
Elaboración: Los autores.

En la segunda etapa, el orquestador enruta la solicitud según la intención: las de trámite (permiso_horas y licencia) aplican slot filling y recurren al circuit breaker CB3 cuando ese llenado falla, mientras que consulta_normativa apoya su respuesta en Gemini bajo el circuit breaker CB2; en paralelo, la clasificación de riesgo opera como servicio independiente en POST /risk/vacation, y toda la comunicación con GIATH se autentica con INTERNAL_SECRET_KEY y queda asentada en la bitácora de auditoría (Figuras 2 y 3).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

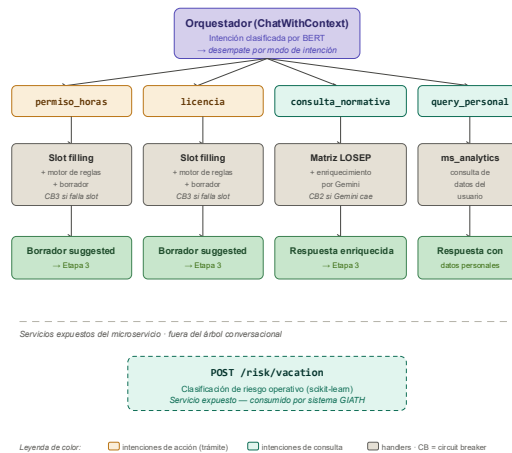


Figura 2. Árbol de enrutamiento por intención.
Elaboración: Los autores.

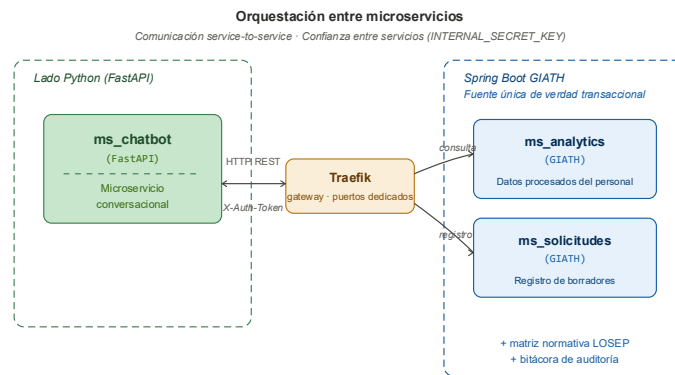


Figura 3. Orquestación entre microservicios.
Elaboración: Los autores.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

En la etapa final, el sistema decide si el tipo de licencia exige documentación de respaldo y, de ser así, emite un recordatorio con la lista de documentos y el plazo de carga; la respuesta se entrega al cliente Angular en formato JSON con un identificador de trazabilidad que permite auditar cada interacción (Figura 4).

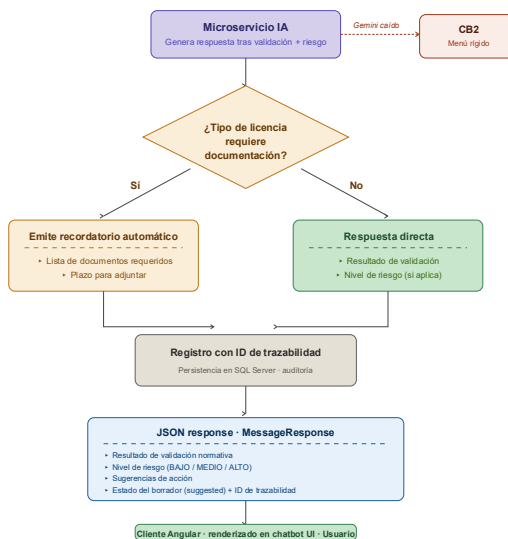


Figura 4. Flujo de construcción y entrega de respuesta.

Elaboración: Los autores.

Internamente, cada mensaje recorre un pipeline de seis fases protegido por los tres circuit breakers: CB1 sustituye al clasificador BERT con expresiones regulares locales, CB3 ofrece un manejador sin conexión cuando falla el slot filling, y CB2 entrega un menú rígido si Gemini deja de responder (Figura 5).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

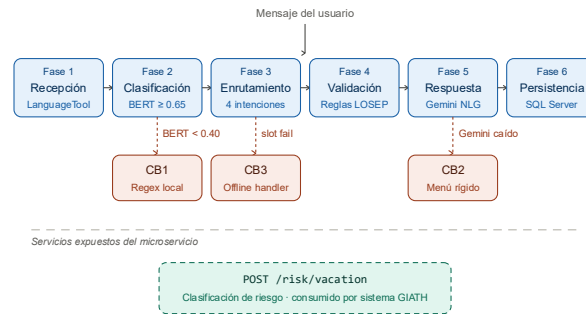


Figura 5. Pipeline interno de procesamiento de mensajes con los tres niveles de circuit breakers.

Elaboración: Los autores.

El microservicio de IA/PLN, desarrollado con FastAPI, expone un conjunto acotado de endpoints REST que reflejan el principio de orquestación interna: la lógica de clasificación de intenciones, validación normativa y enrutamiento por flujo se ejecuta dentro del propio servicio y no como llamadas externas independientes. La Tabla 1 detalla cada endpoint, su método HTTP, el propósito funcional y el componente que lo invoca.

Tabla 1.

Endpoints REST expuestos por el microservicio de IA/PLN.

Método	Ruta	Descripción	Invocado por
POST	/chatbot/message	Punto único de entrada de mensajes del usuario. Ejecuta internamente el pipeline completo: corrección ortográfica, clasificación de intención, validación normativa y generación de respuesta.	Interfaz Angular del chatbot
GET	/chatbot/history	Recupera el historial de conversación asociado a un usuario para preservar el contexto entre interacciones.	Interfaz Angular del chatbot

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

POST	/chatbot/attachments	Gestiona la carga de documentos asociados a un borrador de solicitud cuando el trámite lo requiere.	Interfaz Angular del chatbot
POST	/risk/vacation	Servicio de clasificación de riesgo operativo. Predice el nivel (alto, medio o bajo) y emite una alerta antes de que el trámite continúe.	Formulario de vacaciones del sistema GIATH

Elaboración: Los autores.

La validación normativa no se expone como endpoint independiente, sino que opera como un componente interno (motor de reglas) dentro del flujo orquestado por POST /chatbot/message. Este diseño reduce la superficie de ataque y centraliza la trazabilidad de cada decisión.

El motor normativo aplica de manera automática las reglas que garantizan el cumplimiento institucional: anticipación mínima en la presentación de solicitudes, límites anuales por tipo de permiso, disponibilidad operativa mínima por unidad y ausencia de conflictos con turnos o actividades críticas ya programadas. Estos criterios provienen del marco establecido por la Ley Orgánica del Servicio Público (Asamblea Nacional del Ecuador, 2010) y de las disposiciones internas vigentes en la institución colaboradora. Las excepciones se administran mediante parámetros versionados, lo que permite incorporar cambios reglamentarios sin alterar la trazabilidad histórica.

Como complemento a estas reglas determinísticas, el clasificador de riesgo operativo opera como una capa proactiva especializada en solicitudes de vacaciones. Cuando el usuario inicia el trámite desde el formulario institucional, el chatbot consulta el endpoint POST /risk/vacation, donde el modelo evalúa siete variables (carga del área, historial del funcionario, tipo de cargo, duración de la ausencia y contexto institucional, entre otras) y devuelve un nivel bajo, medio o alto. El resultado no bloquea la solicitud: emite una alerta temprana y, cuando el riesgo es elevado, propone fechas alternativas de menor impacto operativo. La decisión final queda en el solicitante.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Evaluación y métricas

El desarrollo se organizó mediante seis sprints iterativos, cada uno orientado a entregar componentes funcionales y verificables del sistema. La Tabla 2 resume los objetivos y los entregables principales de cada ciclo.

Tabla 2.
Plan de sprints SCRUM.

Sprint	Objetivo principal	Entregable clave
1	Entrenamiento del modelo PLN	Corpus etiquetado y modelo DeepPavlov entrenado
2	Motor de reglas y validaciones normativas	Motor de reglas LOSEP integrado en el chatbot
3	Integración con capas transaccionales	Integración con servicios REST de GIATH y bitácora de auditoría
4	Clasificador de riesgo operativo	Modelo ML validado y endpoint /risk/vacation operativo
5	Interfaz Angular (texto, audio, adjuntos)	Chat UI funcional conectado a los microservicios
6	Integración final, pruebas E2E y optimización	Plataforma lista para despliegue piloto

Elaboración: Los autores.

El plan de pruebas verificó tanto las funcionalidades transversales del chatbot (procesamiento conversacional, entrada por voz, autenticación) como los tres comportamientos críticos del módulo de riesgo: detección de ausentismo elevado, concurrencia de roles críticos y sugerencias prescriptivas para solicitantes de bajo impacto. Cada escenario se construyó sobre configuraciones controladas del backend

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

(mock de personal, asignaciones y solicitudes activas) que reproducen condiciones operativas representativas del sector público. La Tabla 3 documenta los seis casos de prueba con sus escenarios, pasos clave y resultados; los escenarios CF-02, CF-03 y CF-04 evalúan por separado tres dimensiones del riesgo operativo: magnitud del ausentismo, criticidad de los roles ausentes y duración de la solicitud.

Tabla 3.
Casos de prueba funcionales y resultados de ejecución.

ID	Escenario	Pasos clave	Resultado
CF-01	Solicitud válida de permiso por horas	Enviar petición con datos completos, dentro del horario laboral y sin solapamientos	Validación normativa superada; borrador en estado <i>suggested</i> registrado en GIATH con identificador de trazabilidad; respuesta JSON entregada al cliente Angular
CF-02	Riesgo alto por ausentismo elevado (Tormenta Perfecta)	Solicitud de vacaciones de un funcionario crítico (Director) en período con dos ausencias activas que cruzan fechas	Riesgo Alto (80% de probabilidad de rechazo, ausentismo 50%); soft-block activado en la UI con sugerencia automática de fechas alternativas de riesgo bajo
CF-03	Riesgo alto por concurrencia de roles críticos (Choque de Liderazgo)	Solicitud de vacaciones del Director en fechas que coinciden con las del Coordinador, ambos roles críticos	Riesgo Alto (62%); alerta emitida indicando concurrencia de roles críticos en la unidad pese a un ausentismo del 25%
CF-04	Riesgo medio con analítica prescriptiva (Solicitante No Crítico)	Misma configuración que CF-03 pero con solicitante de rol no crítico (Asistente)	Riesgo Medio (40%); recomendación de validación con jefe inmediato y sugerencia automática de fechas alternativas (21–27 oct 2026)
CF-05	Reconocimiento de voz	Enviar audio equivalente al texto de un trámite	Audio transcrito correctamente vía Web Speech API; intención y entidades coincidentes con el

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

			flujo escrito; respuesta JSON consistente
CF-06	Seguridad service-to-service	Petición sin header X-Auth-Token o con clave inválida	Respuesta HTTP 401 Unauthorized en el primer intento sin token; sin exposición de stack trace ni datos del backend

Elaboración: Los autores.

El desempeño del componente de lenguaje natural se evaluó con tres métricas complementarias: la precisión, la recuperación y el valor F1, calculadas por cada tipo de consulta reconocida (Espinosa-Luna et al., 2023). El clasificador de intenciones se evaluó mediante una partición Train/Test Split (80% entrenamiento / 20% prueba) y una matriz de confusión que cruza las predicciones con los valores reales. El clasificador de riesgo operativo, implementado con Random Forest sobre scikit-learn (Pedregosa et al., 2011), siguió el mismo esquema de partición 80/20 para las tres clases de riesgo (alto, medio, bajo); la Figura 6 sintetiza ese flujo de evaluación, desde la preparación de los datos hasta el cálculo de métricas.

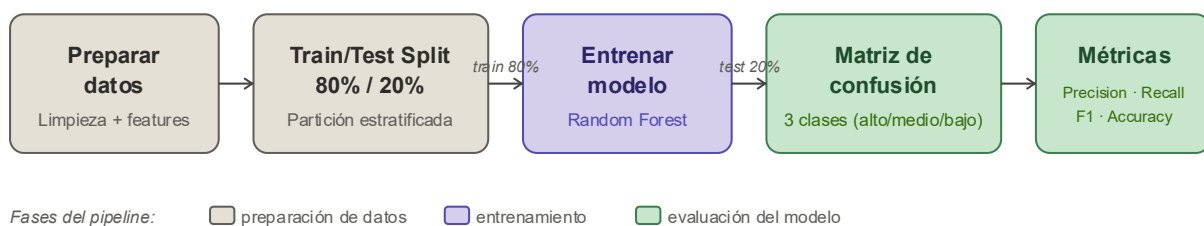


Figura 6. Pipeline de modelos.
Elaboración: Los autores.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

En el plano no funcional, el sistema sostiene latencias p95 de hasta 1.5 s en el componente PLN más reglas y 2.5 s en la orquestación completa, con comunicación cifrada vía HTTPS y autenticación service-to-service mediante INTERNAL_SECRET_KEY. Una capa de caché con Redis acelera las consultas frecuentes a la matriz normativa y reduce los tiempos de respuesta en escenarios de alta concurrencia.

RESULTADOS

La evaluación del sistema GIATH se articula en tres dimensiones críticas: (a) verificación de la arquitectura modular y flujo de gestión de trámites, (b) desempeño del motor de clasificación de riesgos mediante Machine Learning, y (c) precisión del modelo de lenguaje en la interpretación de intenciones y normativa.

Verificación funcional y gestión de trámites

La funcionalidad del sistema se validó ejecutando flujos completos, desde la consulta inicial hasta la confirmación de solicitudes, sobre las dos vertientes de trámite: permisos por horas y licencias multivariantes. La Figura 7 presenta la interfaz de bienvenida con preguntas frecuentes precargadas.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

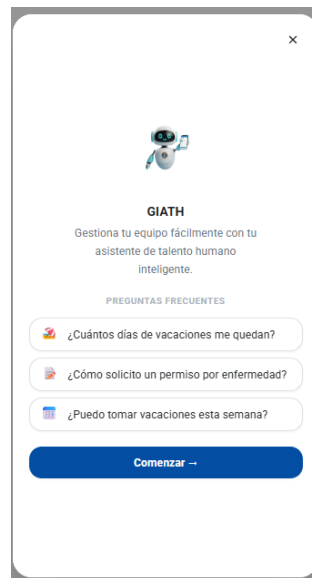


Figura 7. Interfaz de bienvenida del chatbot GIATH con preguntas frecuentes.
Elaboración: Los autores.

La Figura 8 muestra una consulta normativa resuelta: el sistema devuelve los documentos requeridos para una licencia por calamidad doméstica, personalizados con el nombre del funcionario.

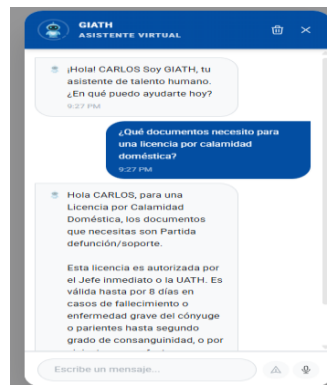


Figura 8. Consulta normativa: documentos requeridos para licencia por calamidad doméstica.
Elaboración: Los autores.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Las Figuras 9 y 10 ilustran el flujo de permiso rápido: la Figura 9 muestra la captura conversacional de los datos con sugerencias de justificación.

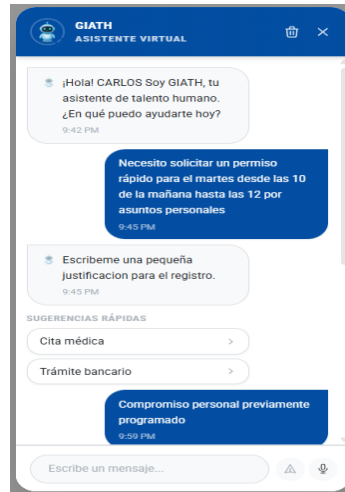


Figura 9. Flujo de permiso rápido: ingreso conversacional de la solicitud con sugerencias de justificación.

Elaboración: Los autores.

La Figura 10 presenta el resumen de confirmación previo al registro, que el usuario puede corregir o aprobar.

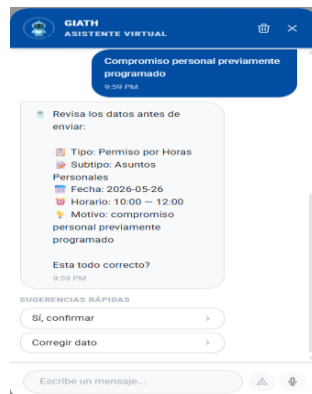


Figura 10. Flujo de permiso rápido: confirmación de datos de la solicitud antes del registro.

Elaboración: Los autores.

Motor de clasificación de riesgo operacional

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

La Figura 11 muestra una alerta de riesgo muy alto (80% de probabilidad de rechazo), donde el sistema sugiere fechas alternativas de menor impacto.

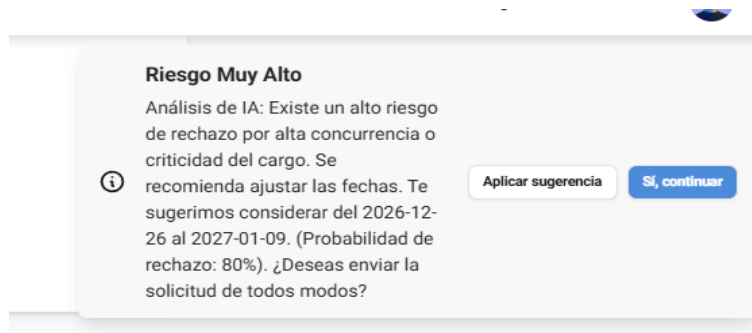


Figura 11. Alerta de riesgo muy alto con sugerencia proactiva de fechas alternativas
Elaboración: Los autores.

La Figura 12 muestra una alerta de riesgo medio (40% de probabilidad de rechazo), donde el sistema recomienda validación con el jefe inmediato sin bloquear la decisión del usuario.

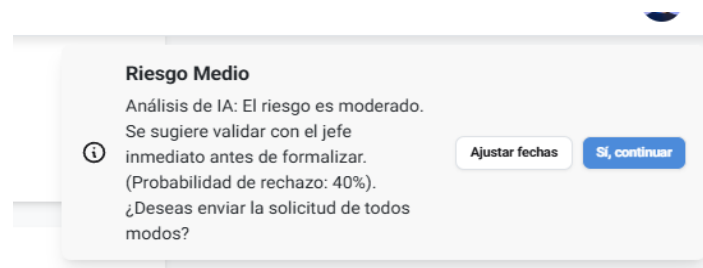


Figura 12. Alerta de riesgo medio con recomendación de validación previa a la formalización.

Elaboración: Los autores.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Evaluación del Desempeño y Precisión del Modelo

El proceso de optimización del chatbot se rigió por un enfoque iterativo que perfeccionó el pipeline NLU híbrido. El sistema combina un clasificador BERT en español (DeepPavlov) con tres capas de heurísticas específicas del dominio institucional, utilizando el modo de intención detectado por LanguageTool como señal de desempate en zonas de confianza ambigua (umbral 0.65). Ante una confianza inferior a 0.40 o un fallo del servicio BERT, un circuit breaker activa automáticamente la detección local de intenciones, garantizando una respuesta útil al usuario en todo momento.

La Tabla 4 resume la evolución del rendimiento del sistema a través de las distintas fases de desarrollo:

Tabla 4.

Evolución comparativa del rendimiento del modelo por fase de desarrollo.

Métrica	Evaluación Inicial	Mejora Anterior	Optimización Final (B1)	Validación (B2)	Stress Test (B3)
Total Test	20	20	20	15	15
Aciertos	15	14	18	15	13
Precisión	75.00%	70.00%	90.00%	100.00%	86.67%

Elaboración: Los autores.

Como se observa en la Tabla 4, el sistema consolidó una alta tasa de aciertos tras el ajuste de pesos semánticos, alcanzando un 90% de precisión en la optimización final (B1) y 100% en la fase de validación (B2), manteniéndose en 86,67% bajo condiciones de prueba críticas (Stress Test B3).

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Para examinar la naturaleza de los errores residuales y la eficacia de la clasificación en condiciones de carga, la Tabla 5 presenta la matriz de confusión obtenida durante el Stress Test B3.

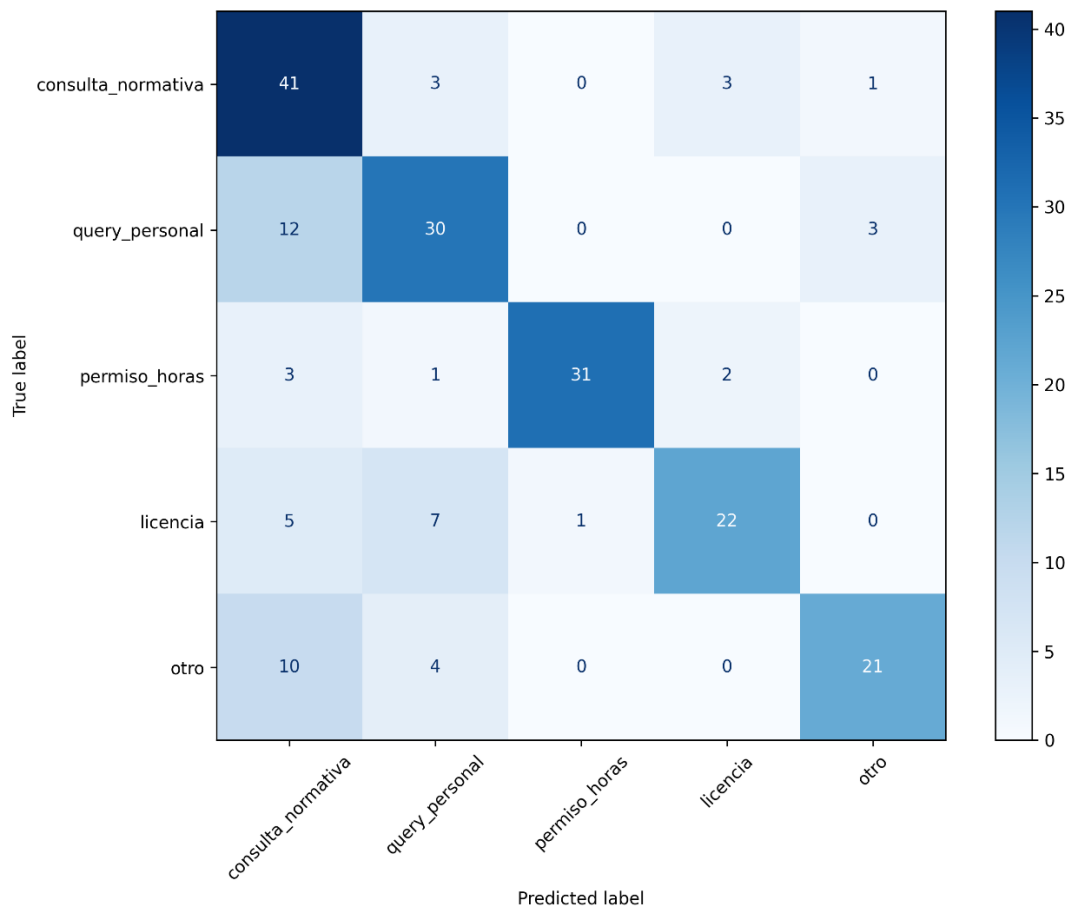


Figura 13. Análisis de desempeño mediante matriz de confusión (N=200)
Elaboración: Los autores.

Tras evaluar 200 casos, los resultados obtenidos son consistentes, lo que demuestra su estabilidad. Su precisión es notable, especialmente en los procesos críticos de concesión de licencias, lo que garantiza la continuidad de la administración.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Se han detectado algunas confusiones entre las consultas normativas y las personales. Esto se debe a la complejidad del lenguaje administrativo. En lugar de considerar esto un error, lo vemos como una oportunidad para mejorar y especializar el sistema en versiones futuras, manteniendo su carácter humano.

DISCUSION

Los resultados confirman la viabilidad del enfoque híbrido en el contexto institucional. Trabajos previos han incorporado chatbots en atención al cliente (Casazola Cruz et al., 2021), educación superior (Espinosa-Luna et al., 2023), mesas de ayuda (Salinas Santiago et al., 2024) y formación educativa (Arias-Chávez et al., 2024; Carbonell-García et al., 2023), pero centrados en flujos abiertos sin marco normativo determinista. El aporte del sistema propuesto es trasladar esa lógica al sector público, un dominio con requerimientos estrictos (saldos acumulados, anticipación mínima, conflictos con actividades críticas) donde no basta entender el lenguaje: hace falta verificar el cumplimiento automáticamente. Este desplazamiento se alinea con la perspectiva del talento humano frente a la automatización (Piña Ferrer, 2024) y con la transformación que la IA introduce en las prácticas del sector público (Padilla-Caballero et al., 2023; Wang et al., 2022).

Los frameworks comerciales como Google Dialogflow, IBM Watson Assistant o Amazon Lex (Garzón-Quiroz et al., 2025) presentan en entornos públicos tres limitaciones críticas: dependencia de la nube de terceros, restricciones de personalización del dominio y costos de licenciamiento difíciles de sostener. La combinación elegida (DeepPavlov para PLN, motor de reglas en Python y clasificador con scikit-learn) las resuelve por tres vías: control institucional del código y los datos, especialización del vocabulario y separación entre fluidez conversacional y precisión normativa (Burtsev et al., 2018). La arquitectura de microservicios con tres niveles de circuit breakers, alineada con patrones de resiliencia

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

recomendados (Norton Stanley & Sebastian, 2024; Vahdati & Ramsin, 2024), evita que la continuidad dependa de servicios externos como Gemini.

El estudio reconoce dos limitaciones: el clasificador de riesgo se entrenó sobre datos mock representativos, lo que da coherencia metodológica al prototipo, pero no sustituye operación real prolongada; el esquema de autenticación con INTERNAL_SECRET_KEY funciona en validación, pero la integración con LDAP queda pendiente para producción. Ninguna invalida el enfoque arquitectónico; sí acotan el alcance de las conclusiones operativas del prototipo.

CONCLUSIONES

El estudio mostró que es posible construir un chatbot útil para la gestión de permisos en el área de talento humano de una institución pública sin recurrir a frameworks comerciales. La combinación de DeepPavlov para la comprensión del lenguaje, un motor de reglas en Python para la validación normativa y un clasificador con scikit-learn para anticipar conflictos resultó en un sistema que entiende las consultas, las valida con criterios institucionales y emite alertas cuando una solicitud puede afectar el funcionamiento del servicio. El módulo predictivo cambió la forma de operar: en lugar de limitarse a tramitar, el chatbot anticipa riesgos, propone fechas alternativas y registra el razonamiento detrás de cada alerta, convirtiéndose en una herramienta de apoyo a la decisión y no solo en un canal de trámites.

La arquitectura mostró además solidez operativa: los tres niveles de circuit breakers sostuvieron la respuesta del chatbot ante fallos de Gemini, ambigüedad del clasificador BERT o errores en el flujo de slot filling, condición indispensable en una institución que opera bajo demanda continua. El uso íntegro de software libre bajo una metodología SCRUM documentada en sprints concretos garantiza la replicabilidad del trabajo en otras instituciones públicas del Ecuador, sin licenciamiento ni dependencia de proveedores.

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Como trabajo futuro queda integrar el directorio activo institucional (LDAP) para reforzar la autenticación en producción y validar el clasificador de riesgo con datos reales de operación prolongada.

FINANCIAMIENTO

No monetario.

AGRADECIMIENTO

A todos los actores sociales involucrados en el desarrollo de la investigación.

REFERENCIAS CONSULTADAS

- Alemán Espinoza, J. W., Ruiz Vargas, M. R., & Aguilar Juarez, E. A. (2019). Retrospectiva al Desarrollo de Chatbots y Procesamiento del Lenguaje Natural. *Realidad y Reflexión*, 190-206. <https://doi.org/10.5377/ryr.v48i0.7095>
- Aljohani, A. (2023). Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation and Agility. *Sustainability*, 15(20), 15088. <https://doi.org/10.3390/su152015088>
- Arias-Chávez, D., Ramos-Quispe, T., & Cangalaya Sevillano, L. M. (2024). Análisis y tendencias en el uso de chatbots y agentes conversacionales en el campo de la educación: Una revisión bibliométrica. *Innovaciones Educativas*, 26(41), 242-260. <https://doi.org/10.22458/ie.v26i41.5135>
- Asamblea Nacional del Ecuador. (2010). *Ley Orgánica del Servicio Público (LOSEP)*. <https://www.gob.ec/regulaciones/ley-organica-servicio-publico>
- Burtsev, M., Seliverstov, A., Airapetyan, R., Arkhipov, M., Baymurzina, D., Bushkov, N., Gureenkova, O., Khakhulin, T., Kuratov, Y., Kuznetsov, D., Litinsky, A., Logacheva, V., Lymar, A., Malykh, V., Petrov, M., Polulyakh, V., Pugachev, L., Sorokin, A., Vikhreva, M., & Zaynutdinov, M. (2018). DeepPavlov: Open-Source Library for Dialogue Systems. *Proceedings of ACL 2018, System Demonstrations*, 122-127. <https://doi.org/10.18653/v1/P18-4021>

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

CAF. (2021). *Experiencia: Datos e Inteligencia Artificial en el sector público*. CAF.
<https://scioteca.caf.com/handle/123456789/1793>

Carbonell-García, C. E., Burgos-Goicochea, S., Calderón-de-los-Ríos, D. O., & Paredes-Fernández, O. W. (2023). La Inteligencia Artificial en el contexto de la formación educativa. *EPISTEME KOINONIA*, 6(12), 152-166.
<https://doi.org/10.35381/e.k.v6i12.2547>

Casazola Cruz, O. D., Alfaro Mariño, G., Burgos Tejada, J., & Ramos More, O. A. (2021). La usabilidad percibida de los chatbots sobre la atención al cliente en las organizaciones: Una revisión de la literatura. *Interfases*, (014), 184-204.
<https://doi.org/10.26439/interfases2021.n014.5401>

Celi-Parraga, R. J., Varela-Tapia, E. A., Acosta-Guzmán, I. L., & Montañño-Pulzara, N. R. (2021). Técnicas de procesamiento de lenguaje natural en la inteligencia artificial conversacional textual. *AlfaPublicaciones*, 3(4.1), 40-52.
<https://doi.org/10.33262/ap.v3i4.1.123>

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>

Espinosa-Luna, B. H., Castillo-Oliva, J., Montañnez-Díaz, B. A., & Mendoza-De-los-Santos, A. (2023). Implementación de un chatbot basado en modelo de lenguaje de inteligencia artificial para responder preguntas frecuentes de estudiantes universitarios. *Revista Científica de Sistemas e Informática*, 3(2), e570.
<https://doi.org/10.51252/rcsi.v3i2.570>

Falahah, Surendro, K., & Sunindyo, W. D. (2021). Circuit Breaker in Microservices: State of the Art and Future Prospects. *IOP Conference Series: Materials Science and Engineering*, 1077(1), 012065. <https://doi.org/10.1088/1757-899X/1077/1/012065>

Garzón-Quiroz, M., Del Campo-Saltos, G., & Loor-Ávila, B. (2025). Análisis sistemático sobre la eficiencia comunicativa entre chatbots basados en reglas y modelos de lenguaje natural. *Universitas*, (42), 167-192. <https://doi.org/10.17163/uni.n42.2025.07>

Gordon Graell, R. D. (2023). Chatbots e inteligencia artificial: Aportes, innovaciones y aplicación en el desarrollo de sistemas de información empresarial. *Visión Antataura*, 7(1), 132-147. <https://doi.org/10.48204/j.vian.v7n1.a3930>

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

- Kaun, A., & Männiste, M. (2025). Public sector chatbots: AI frictions and data infrastructures at the interface of the digital welfare state. *New Media & Society*, 27(4), 1962-1985. <https://doi.org/10.1177/14614448251314394>
- Norton Stanley, S. A., & Sebastian, S. (2024). Elastic circuit de-constructor: A pattern to enhance resiliency in microservices. *International Journal of Computers and Applications*, 46(10), 921-932. <https://doi.org/10.1080/1206212X.2024.2397018>
- Oyekunle Claudius Oyeniran, Adebunmi Okechukwu Adewusi, Adams Gbolahan Adeleke, Lucy Anthony Akwawa, & Chidimma Francisca Azubuko. (2024). Microservices architecture in cloud-native applications: Design patterns and scalability. *Computer Science & IT Research Journal*, 5(9), 2107-2124. <https://doi.org/10.51594/csitrj.v5i9.1554>
- Padilla-Caballero, J. E. A., Naupay-Gusukuma, Á. M., Ruiz-Salazar, J. M., & Poma-Garcia, C. R. (2023). Habilidades investigativas universitarias del futuro: El papel de la inteligencia artificial. *Revista Arbitrada Interdisciplinaria Koinonía*, 8(2), 702-722. <https://doi.org/10.35381/r.k.v8i2.2946>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825-2830. <https://www.jmlr.org/papers/v12/pedregosa11a.html>
- Piña Ferrer, L. S. (2024). La inteligencia artificial desde la perspectiva del talento humano. *Revista Arbitrada Interdisciplinaria Koinonía*, 9(18), 1-3. <https://doi.org/10.35381/r.k.v9i18.4210>
- Pionce Arteaga, M. A., Caicedo Plúa, C. R., Delgado Lucas, H. B., & Murillo Quimiz, L. R. (2022). Chatbots para ventas y atención al cliente. *Journal TechInnovation*, 1(1), 107-116. <https://doi.org/10.47230/Journal.TechInnovation.v1.n1.2022.107-116>
- Saleh Sedghpour, M. R., Garlan, D., Schmerl, B., Klein, C., & Tordsson, J. (2023). Breaking the Vicious Circle: Self-Adaptive Microservice Circuit Breaking and Retry. *2023 IEEE International Conference on Cloud Engineering (IC2E)*, 32-42. <https://doi.org/10.1109/IC2E59103.2023.00012>

Alexander Andres Agurto-Guevara; Jenner Joan Villalba-Paladines; Wilmer Braulio Rivas-Asanza; Edison Luis Lojan-Cueva

Salinas Santiago, J., García Gutiérrez, W. F., Ordoñez Reyes, A. B., & Mendoza De Los Santos, A. C. (2024). Implementación de un chatbot inteligente en la gestión de las mesas de ayuda. *Revista Científica: BIOTECH AND ENGINEERING*, 4(1). <https://doi.org/10.52248/eb.Vol4Iss1.103>

Vahdati, A., & Ramsin, R. (2024). Model-Driven Methodology for Developing Chatbots Based on Microservice Architecture: *Proceedings of the 12th International Conference on Model-Based Software and Systems Engineering*, 247-254. <https://doi.org/10.5220/0012433700003645>

Velastegui Campoverde, E. U. (2023). *Inteligencia Artificial y su potencial adopción en los servicios públicos: Desafíos y oportunidades en Ecuador a partir del período 2021 – 2022*. <https://repositorio.puce.edu.ec/handle/123456789/43320>

Wang, X., Lin, X., & Shao, B. (2022). How does artificial intelligence create business agility? Evidence from chatbots. *International Journal of Information Management*, 66, 102535. <https://doi.org/10.1016/j.ijinfomgt.2022.102535>