

Pontificia Universidad Católica del Ecuador Sede Esmeraldas
(PUCESE)

**ESCUELA DE HÁBITAT INFRAESTRUCTURA Y
CREATIVIDAD**

CARRERA

INGENIERÍA DE TECNOLOGÍAS DE LA INFORMACIÓN

LÍNEA DE INVESTIGACIÓN

**INGENIERÍA DE SOFTWARE, INNOVACIÓN Y EMPRENDIMIENTO
EN TIC**

TÍTULO DEL ARTÍCULO CIENTÍFICO

**MODELO DE INTELIGENCIA ARTIFICIAL EN
EL BORDE BASADO EN OLLAMA PARA
INFERENCIA LOCAL EN DISPOSITIVOS DE
INTERNET DE LAS COSAS**

TÍTULO PROFESIONAL

INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

AUTOR

González-Castillo Andrew Fernando

ASESOR

Msc. Vargas-Machuca-Del-Salto Adrián Gabriel

ESMERALDAS, 2026

Modelo de Inteligencia Artificial en el Borde Basado en Ollama para Inferencia Local en Dispositivos de Internet de las Cosas

Andrew González-Castillo ^a & Pablo Pico-Valencia ^b & Xavier Quiñonez-Ku ^c

^a Pontificia Universidad Católica del Ecuador, Esmeraldas, Ecuador. afgonzalez@pucese.edu.ec

^b Universidad de Granada, España. pablo.pico@ugr.es

^c Pontificia Universidad Católica del Ecuador, Esmeraldas, Ecuador. xavier.quinonez@pucese.edu.ec

Abstract

El Internet de las Cosas (IoT) ha incrementado la dependencia de la computación en la nube para el procesamiento de datos y la toma de decisiones, introduciendo riesgos de latencia y seguridad, particularmente en dominios sensibles como la vida asistida y los hogares inteligentes. Este trabajo presenta la Arquitectura de Borde Integrada con Ollama (OLLiE), que permite el razonamiento local en dispositivos IoT embebidos mediante modelos de lenguaje de gran tamaño (LLM) cuantizados. Los experimentos realizados en un escenario de hogar inteligente asistido demuestran que los LLM desplegados localmente logran un rendimiento competitivo en razonamiento, preservando al mismo tiempo la privacidad de los datos. El modelo cuantizado phi-3.5-mini (Q2_K) alcanzó un F1-score de 0.45 y un recall de 0.69, superando a varias alternativas basadas en la nube. Aunque la inferencia en el borde presenta una latencia mayor que la ejecución en la nube (76–115 s frente a <5 s), estos retrasos son aceptables para aplicaciones asistivas tolerantes a la latencia. En conjunto, los resultados muestran que la cuantización de modelos equilibra la precisión del razonamiento, las limitaciones de hardware y la preservación de la privacidad, posicionando a OLLiE como una arquitectura viable para entornos IoT de borde seguros y autónomos.

Palabras clave: Internet de las cosas; dispositivo embebido; inteligencia artificial; modelo extenso de lenguaje; computación en el borde.

INFORME DEL DOCENTE-DIRECTOR DEL TRABAJO DE INTEGRACIÓN CURRICULAR

CARRERA INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

Esmeraldas, 10 de abril de 2026

Mgt. Homero Velasteguí

COORDINADOR DE CARRERA INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE ESMERALDAS

De mis consideraciones:

Se envía el informe correspondiente a la tutoría realizada al Trabajo de Titulación que se detalla a continuación:

TITULO DEL TRABAJO DE INTEGRACIÓN CURRICULAR	Modelo de Inteligencia Artificial en el Borde Basado en Ollama para Inferencia Local en Dispositivos de Internet de las Cosas	
DIRECTOR	Nombre	Cédula
	Vargas-Machuca-Del-Salto Adrián Gabriel	1803131190
ESTUDIANTE(S)	Nombre	Cédula
	González-Castillo Andrew Fernando	0804552206

Se informa que el trabajo ha cumplido con todos los parámetros establecidos, mediante el cual el estudiante demuestra el desarrollo de competencias en el campo de conocimiento de su profesión y presenta una propuesta en el área de conocimiento, con un nivel de argumentación coherente.

Dando por concluida esta tutoría de trabajo de titulación, **CERTIFICO**, para los fines pertinentes, que el (los) estudiante(s) está(n) apto(s) para continuar con el proceso de **LECTURA**.

Atentamente,

DIRECTOR/TUTOR DE TRABAJO DE TITULACIÓN

C.I. 1803131190

NOMBRE: Msc. Vargas-Machuca-Del-Salto Adrián Gabriel

FECHA: 10-04-2026

1 Datos del medio científico enviado a revisión por pares o ya publicado

Para artículos en proceso de publicación. Un artículo está en proceso de publicación cuando se ha enviado a la plataforma de la revista científica seleccionada para que el editor inicie su análisis y luego proceda a iniciar el proceso de revisión por pares.

Campo	Detalle
Nombre de la revista científica	DYNA
Enlace (URL) de la revista	https://revistas.unal.edu.co/index.php/dyna
ISSN de la revista	ISSN: 0012-7353 (Printed), eISSN: 2346-2183 (Online)
Medio(s) de indexación	<input type="checkbox"/> Scopus <input type="checkbox"/> Scielo Ecuador <input checked="" type="checkbox"/> Emerging Source Citation Index (ESCI - Web of Science) <input type="checkbox"/> Google Scholar <input type="checkbox"/> Base Search <input type="checkbox"/> Copernicus Index <input type="checkbox"/> EBSCO. Applied Science & Technology Source Ultimate <input type="checkbox"/> Latam + <input type="checkbox"/> Scientific Indexing Services <input type="checkbox"/> Europub <input type="checkbox"/> ScienceGate <input type="checkbox"/> Latindex 2.0
Nombre del editor de la revista	Néstor Ricardo Rojas Reyes (Director-Editor)
Correo electrónico del editor de la revista	dyna@unal.edu.co
Fecha de envío del artículo a la revista	05/02/2026

2 Evidencias de envío a medio científico

1. Documento de aprobación del asesor para realizar el envío del artículo científico (formato similar al usado para las tesis donde se especifica el porcentaje de similitud).

2. Captura de pantalla de la plataforma de la revista en la que se sube el artículo.

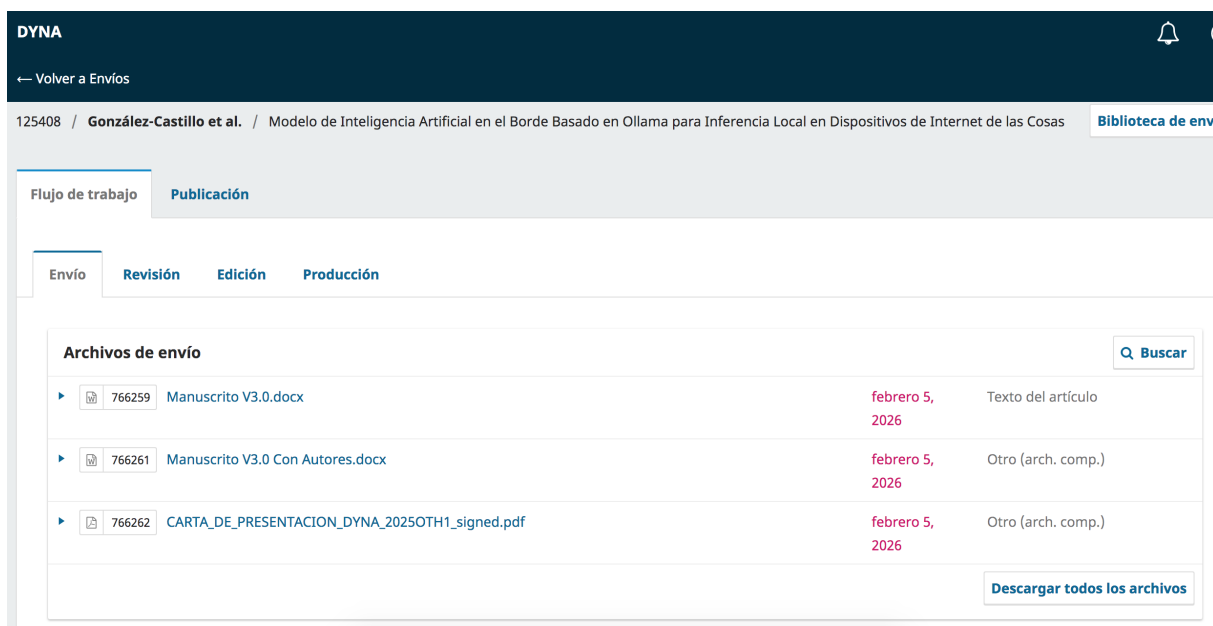


Figure 1: Estado del manuscrito en la plataforma de la revista DYNÁ (Submission ID: 125408)

3. Captura de pantalla del correo recibido por la plataforma o editor de la revista.

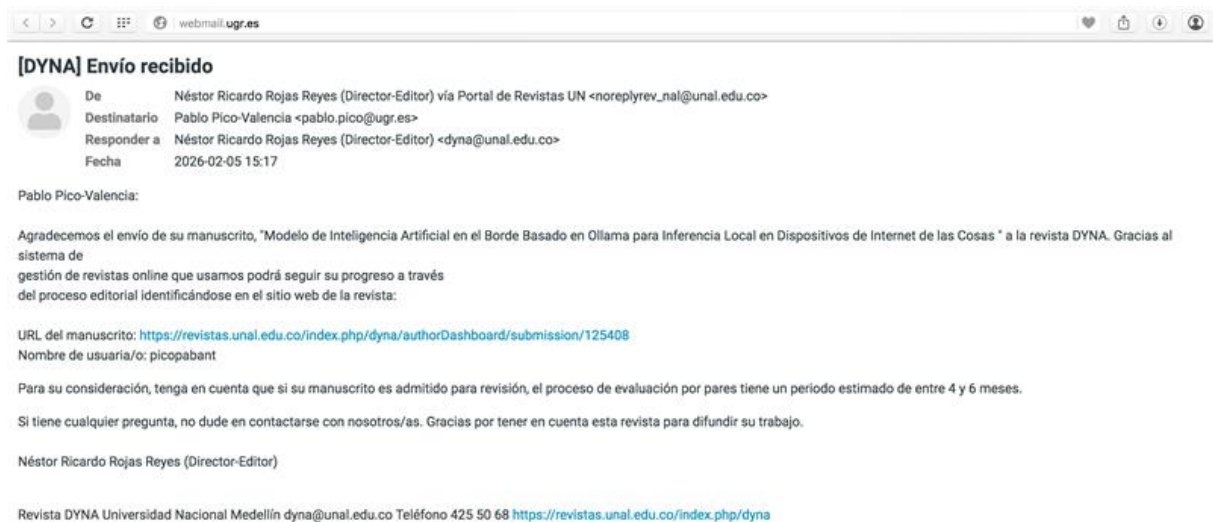


Figure 2: Correo de confirmación de recepción del manuscrito por Néstor Ricardo Rojas Reyes (Director-Editor, DYNÁ)