Pontificia Universidad Católica del Ecuador Sede Esmeraldas (PUCESE)

ESCUELA DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD

CARRERA:
INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN

LÍNEA DE INVESTIGACIÓN:
INGENIERÍA DE SOFTWARE, INNOVACIÓN Y EMPRENDIMIENTO EN TIC

TÍTULO DEL ARTÍCULO CIENTÍFICO:
COMPARATIVE EMPIRICAL EVALUATION OF SOFTWARE TEST AUTOMATION USING ARTIFICIAL INTELLIGENCE AND MANUAL TEST CASE DESIGN

TÍTULO PROFESIONAL:
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

AUTOR:
Jesús José Bone Caicedo

ASESOR:
Msc. José Luis Carvajal Carvajal

ESMERALDAS, 2026

Comparative Empirical Evaluation of Software Test Automation using Artificial Intelligence and Manual Test Case Design

Evaluación Empírica Comparativa de la Automatización de Pruebas de Software mediante Inteligencia Artificial y Diseño Manual de Casos de Prueba

Jesús José Bone Caicedo[1], José Luis Carvajal Carvajal [1], and Victor Xavier Quiñonez Ku[1]

[1]*Systems and Software Department, Pontifical Catholic University of Ecuador Esmeraldas Campus, Esmeraldas Ecuador*
jjbone@pucese.edu.ec, jose.carvajal@pucese.edu.ec, xavier.quinonez@pucese.edu.ec

Abstract

This study presents an empirical comparison between manually designed test cases and those automatically generated by generative artificial intelligence tools—ChatGPT and Diffblue Cover—applied to the Spring PetClinic system, developed in Java and Spring Boot. The fully automated experiment comprised 2,480 runs distributed across 12 test classes (6 human and 6 AI-generated), with 40 iterations per class and a total duration of 6.18 hours. Four main metrics—instruction coverage, branch coverage, mutation score, and execution time—were evaluated using descriptive and inferential analysis (Student's t, Welch, and Mann-Whitney U). At the aggregate level (N = 12), no significant differences were found ($p > 0.05$; $d < 0.30$), while at the complete level (N = 2,480), small effects ($r < 0.15$) were observed in all metrics, indicating marginal differences between the two approaches. The results confirm that generative AI can achieve quantitative performance comparable to that of human testing, albeit with more limited functional reasoning. This work provides empirical evidence on the current capabilities and limitations of AI in automated testing, highlighting its potential to accelerate repetitive tasks and improve productivity without replacing human analytical judgment.

Keywords: automated testing, Diffblue Cover, generative AI, software quality, test case generation

Resumen

Este estudio presenta una comparación empírica entre casos de prueba diseñados manualmente y aquellos generados automáticamente por herramientas de inteligencia artificial generativa —ChatGPT y Diffblue Cover— aplicadas al sistema Spring PetClinic, desarrollado en Java y Spring Boot. El experimento, completamente automatizado, comprendió 2 480 ejecuciones distribuidas en 12 clases de prueba (6 humanas y 6 generadas por IA), con 40 iteraciones por clase y una duración total de 6.18 horas. Se evaluaron cuatro métricas principales —cobertura de instrucciones, cobertura de ramas, mutation score y tiempo de ejecución— mediante análisis descriptivo e inferencial (t de Student, Welch y Mann-Whitney U). En el nivel agregado (N = 12) no se encontraron diferencias significativas ($p > 0.05$; $d < 0.30$), mientras que en el nivel completo (N = 2 480) se observaron efectos pequeños ($r < 0.15$) en todas las métricas, indicando diferencias marginales entre ambos enfoques. Los resultados confirman que la IA generativa puede alcanzar un rendimiento cuantitativo comparable al de las pruebas humanas, aunque con razonamiento funcional más limitado. Este trabajo aporta evidencia empírica sobre las capacidades y restricciones actuales de la IA en el testing automatizado, destacando su potencial para acelerar tareas repetitivas y mejorar la productividad sin reemplazar el juicio analítico humano.

Palabras claves: calidad del software, Diffblue Cover, generación de casos de prueba, inteligencia artificial generativa, pruebas automatizadas

1 Introduction

In the context of digital transformation and the growing incorporation of artificial intelligence in all stages of the software life cycle, test automation has taken center stage in agile development processes due to its potential to increase coverage, reduce time, and improve product reliability. This technical advance not only responds to a need for efficiency in engineering teams, but also to social demand for more reliable, secure, and highly available digital services, whose quality has a direct impact on sensitive sectors such as health, education, and finance. The emergence of generative AI—with large language models (LLMs) and associated tools—has renewed this field by promising faster and assisted test case generation, albeit with mixed and context-dependent results [1, 2, 3, 4]. In ecosystems dominated by Java and Spring Boot, it is pertinent to subject these solutions to rigorous evalua-

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).**

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	COMPARATIVE EMPIRICAL EVALUATION OF SOFTWARE TEST AUTOMATION USING ARTIFICIAL INTELLIGENCE AND MANUAL TEST CASE DESIGN	
DIRECTOR	Nombre	Cédula
	José Luis Carvajal Carvajal	0802476374
ESTUDIANTE(S)	Nombre	Cédula
	Jesús José Bone Caicedo	0850035817

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,

Validar únicamente en FirmaEC.
Firmado electrónicamente por:
JOSE LUIS CARVAJAL CARVAJAL

DIRECTOR/TUTOR DE TRABAJO DE TITULACIÓN
C.I. 0802476374
NOMBRE: Msc. José Luis Carvajal
FECHA: 10-04-2026

2. 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.

Nombre de la revista científica: IEEE Transactions on Software Engineering

Enlace (URL) de la revista: https://www.computer.org/csdl/journal/ts

ISSN de la revista: ISSN: 1939-3520 (Online)
ISSN: 0098-5589 (Print)

Medio(s) de indexación:

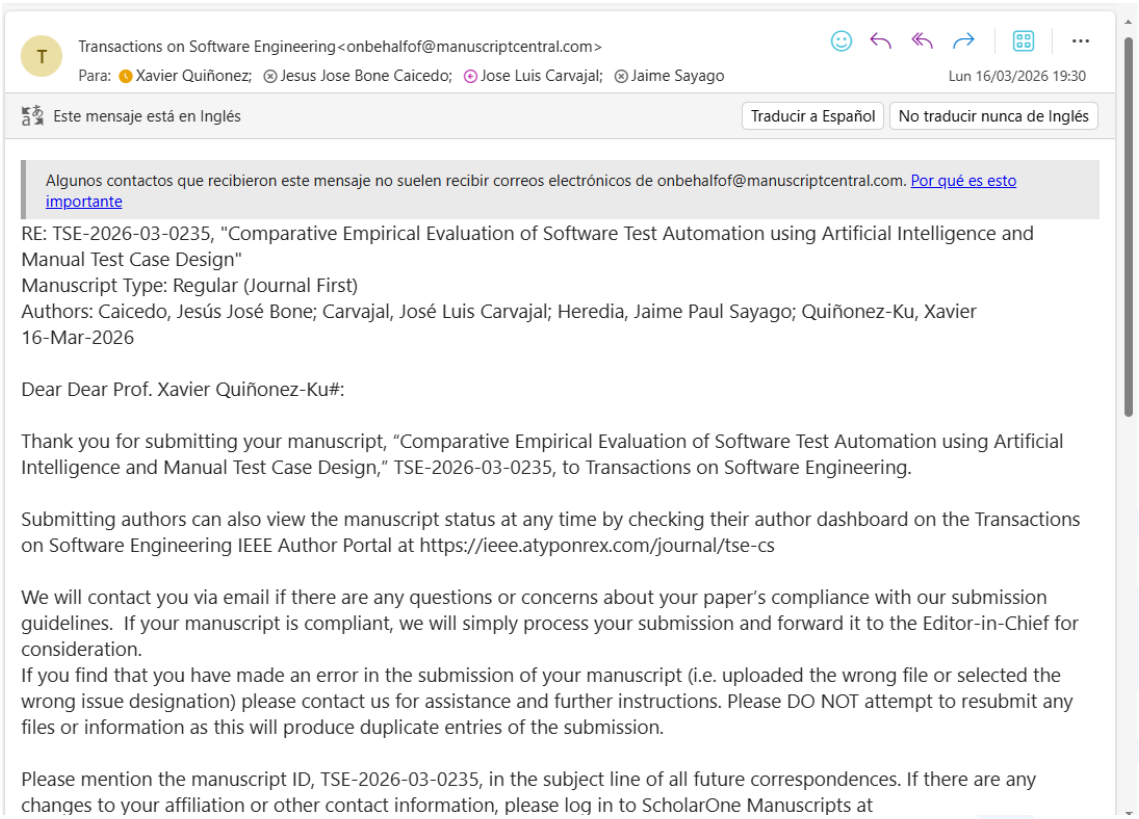
[X] Scopus
[ ] Scielo Ecuador
[X] Emerging Source Citation Index (ESCI - Web of Science)
[X] Google Scholar
[ ] Base Search
[ ] Copernicus Index
[ ] EBSCO. Applied Science & Technology Source Ultimate
[ ] Latam +
[ ] Scientific Indexing Services
[ ] Europub
[ ] ScienceGate
[ ] Latindex 2.0

Nombre del editor de la revista: Dr. Mauro Pezzé

Correo electrónico del editor de la revista: onbehalfof@manuscriptcentral.com

Fecha de envío del artículo a la revista: 17-03-3026

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



Transactions on Software Engineering <onbehalfof@manuscriptcentral.com>
Para: Xavier Quiñonez; Jesus Jose Bone Caicedo; Jose Luis Carvajal; Jaime Sayago
Lun 16/03/2026 19:30

Este mensaje está en Inglés | Traducir a Español | No traducir nunca de Inglés

Algunos contactos que recibieron este mensaje no suelen recibir correos electrónicos de onbehalfof@manuscriptcentral.com. Por qué es esto importante

RE: TSE-2026-03-0235, "Comparative Empirical Evaluation of Software Test Automation using Artificial Intelligence and Manual Test Case Design"
Manuscript Type: Regular (Journal First)
Authors: Caicedo, Jesús José Bone; Carvajal, José Luis Carvajal; Heredia, Jaime Paul Sayago; Quiñonez-Ku, Xavier
16-Mar-2026

Dear Dear Prof. Xavier Quiñonez-Ku#:

Thank you for submitting your manuscript, "Comparative Empirical Evaluation of Software Test Automation using Artificial Intelligence and Manual Test Case Design," TSE-2026-03-0235, to Transactions on Software Engineering.

Submitting authors can also view the manuscript status at any time by checking their author dashboard on the Transactions on Software Engineering IEEE Author Portal at https://ieee.atyponrex.com/journal/tse-cs

We will contact you via email if there are any questions or concerns about your paper's compliance with our submission guidelines. If your manuscript is compliant, we will simply process your submission and forward it to the Editor-in-Chief for consideration.
If you find that you have made an error in the submission of your manuscript (i.e. uploaded the wrong file or selected the wrong issue designation) please contact us for assistance and further instructions. Please DO NOT attempt to resubmit any files or information as this will produce duplicate entries of the submission.

Please mention the manuscript ID, TSE-2026-03-0235, in the subject line of all future correspondences. If there are any changes to your affiliation or other contact information, please log in to ScholarOne Manuscripts at