Rancang Bangun Sistem Pembangkit Skenario Pengujian Multi-Persona Berbasis Large Language Model
(LLM)
Tsania Qurrota A'yunin Qulub, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

RANCANG BANGUN SISTEM PEMBANGKIT SKENARIO PENGUJIAN MULTI-PERSONA BERBASIS *LARGE LANGUAGE MODEL* (LLM)

Tsania Qurrota A'yunin Qulub
22/505618/SV/21839

Jaminan kualitas perangkat lunak memerlukan perancangan skenario pengujian yang komprehensif, mencakup skenario positif, negatif, dan *edge case*. Namun, perancangan skenario pengujian secara manual membutuhkan waktu, rentan terhadap inkonsistensi, dan berisiko melewatkan kebutuhan penting. Penelitian ini mengembangkan Test Scenario Generator, yaitu sistem berbasis web yang menghasilkan skenario pengujian otomatis dari dokumen kebutuhan menggunakan Claude API dan empat metodologi pengujian *black-box*, yaitu *Equivalence Partitioning*, *Boundary Value Analysis*, *Decision Table Testing*, dan *State Transition Testing*. Sistem menyediakan mode pengujian *general*, *chatbot*, dan *hybrid*, mendukung keluaran bahasa Indonesia dan Inggris, serta menyediakan persona untuk skenario *chatbot*. Evaluasi dilakukan melalui tiga pengujian, yaitu komparasi Claude Haiku 4.5, Sonnet 4.6, dan Opus 4.8 berdasarkan *requirement coverage* dan *latency* pada lima belas proses pembuatan skenario dengan variasi persona pada dokumen *chatbot*, evaluasi kesesuaian gaya bahasa persona menggunakan *LLM-as-a-Judge* pada sembilan kombinasi persona dan model, serta *User Acceptance Testing* terhadap sepuluh responden. Kedua dokumen kebutuhan berasal dari proyek mitra industri dan tidak dipublikasikan, sehingga replikasi memerlukan dokumen berkonteks serupa dengan konfigurasi yang sama. Hasil evaluasi menunjukkan *requirement coverage* tertinggi diperoleh Sonnet sebesar 88,7% pada dokumen aplikasi *chatbot* layanan keuangan dan Opus sebesar 95,5% pada dokumen aplikasi web *general* berupa asisten audit, sedangkan Opus memiliki rata-rata *latency* terendah sebesar 210,86 detik. Kesesuaian gaya bahasa persona memperoleh skor rata-rata 4,00–4,22 dari skala 5, dengan Opus sebagai model paling unggul. Tingkat penerimaan pengguna mencapai rata-rata 90,36% dan termasuk kategori Sangat Baik, dengan keempat aspek penilaian juga berada pada kategori tersebut. Hasil tersebut menunjukkan bahwa Test Scenario Generator dapat mendukung otomatisasi penyusunan skenario pengujian dari dokumen kebutuhan melalui konfigurasi model, mode pengujian, dan persona yang dapat disesuaikan.

Kata kunci: *black-box testing*, Claude API, dokumen kebutuhan, *Large Language Model,* persona, skenario pengujian, *Test Scenario Generator*

UNIVERSITAS
GADJAH MADA
Rancang Bangun Sistem Pembangkit Skenario Pengujian Multi-Persona Berbasis Large Language Model (LLM)
Tsania Qurrota A'yunin Qulub, Divi Galih Prasetyo Putri, S.Kom., M.Kom., Ph.D
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

AUTOMATED TEST SCENARIO GENERATION PLATFORM MULTI-PERSONA USING LARGE LANGUAGE MODEL (LLM)

Tsania Qurrota A'yunin Qulub
22/505618/SV/21839

Software quality assurance requires comprehensive test scenarios that cover positive, negative, and edge-case scenarios. However, designing test scenarios manually can be time-consuming, inconsistent, and prone to missing important requirements. This research develops Test Scenario Generator, a web-based system that automatically generates test scenarios from requirement documents using Claude API and four black-box testing methodologies, Equivalence Partitioning, Boundary Value Analysis, Decision Table Testing, and State Transition Testing. The system provides general, chatbot, and hybrid testing modes, supports output in English and Indonesian, and uses personas to generate conversations for chatbot scenarios. The system was evalua-ted through three assessments, consisting of a comparison of Claude Haiku 4.5, Sonnet 4.6, and Opus 4.8 based on requirement coverage and latency across fifteen generation runs with persona variation on the chatbot document, an evaluation of persona language style using LLM-as-a-Judge across nine persona-model combinations, and User Acceptance Testing involving ten respondents. Both requirement documents come from a partner company's real projects and are not publicly available, so replication requires the same configuration applied to requirement documents of a similar context. For the documents and configurations evaluated, the highest requirement coverage was achieved by Sonnet at 88.7% on the financial services chatbot document and by Opus at 95.5% on the audit assistant web application document, while Opus had the lowest average latency at 210.86 seconds, so no single model led on both metrics. Persona language style received average scores ranging from 4.00 to 4.22 out of 5, with Opus the most consistent in maintaining persona style. Overall user acceptance reached 90.36% and falls into the Very Good category, with all four assessed aspects also in that category. These results indicate that Test Scenario Generator can support automated test scenario preparation from requirement documents through configurable models, testing modes, and personas.

Keywords: black-box testing, Claude API, Large Language Model, persona, require-ment document, Test Scenario Generator, test scenarios