

Em Direção a uma Abordagem Baseada em IA Generativa para Automação de Testes End-to-End

Elvis Júnior, Vânia de Oliveira Neves

Instituto de Computação – Universidade Federal Fluminense (UFF)
Niterói – RJ – Brazil

elvisjunior@id.uff.br, vania@ic.uff.br

Abstract. *Software test automation is essential to ensure system quality; however, end-to-end (E2E) testing still faces challenges related to complexity and manual effort. This paper proposes GenIA-E2ETest, a Generative AI-based approach for the automatic generation of executable E2E test scripts from natural language descriptions. The approach was evaluated on two web applications and demonstrated promising results, achieving an average of 77% in element metrics, 85% in execution recall, and minimal manual adjustments (average of 10%). Despite limitations in highly dynamic flows, the results indicate that GenIA-E2ETest is effective in accelerating E2E test automation.*

Resumo. *A automação de testes de software é essencial para garantir a qualidade dos sistemas; contudo, os testes end-to-end (E2E) ainda enfrentam desafios relacionados à complexidade e ao esforço manual. Este artigo propõe a GenIA-E2ETest, uma abordagem baseada em IA Generativa para a geração automática de scripts executáveis de testes E2E a partir de descrições em linguagem natural. A abordagem foi avaliada em duas aplicações web e apresentou resultados promissores, alcançando em média 77% em métricas de elementos, 85% em recall de execução e exigindo ajustes manuais mínimos (média de 10%). Apesar de limitações em fluxos altamente dinâmicos, os resultados indicam que a GenIA-E2ETest é eficaz para acelerar a automação de testes E2E.*

1. Visão Geral da Pesquisa

Garantir a qualidade de sistemas de software é cada vez mais desafiador devido à complexidade das aplicações modernas. O avanço de Modelos de Linguagem de Grande Escala (LLMs) trouxe novas possibilidades para automatizar tarefas a partir de descrições em linguagem natural; apesar disso, sua aplicação em testes *end-to-end* ainda é pouco explorada. Motivado por essa lacuna, este trabalho apresenta a **GenIA-E2ETest**, abordagem baseada em IA generativa que converte descrições textuais de casos de teste em scripts executáveis, reduzindo o esforço técnico e ampliando o acesso à automação.

Embora pesquisas recentes usem técnicas como *web crawling* e aprendizado por reforço para automatizar testes, elas enfrentam desafios de manutenção e escalabilidade, e ferramentas existentes baseadas em IA generativa costumam ser proprietárias e caras, além de focarem em ações isoladas, como testes unitários. A GenIA-E2ETest se destaca por ser gratuita, *open source* e capaz de automatizar cenários completos com múltiplas interações, validações e controle de fluxo.

O trabalho seguiu a metodologia *Design Science Research* (DSR), conforme proposto por Hevner et al. [Hevner et al. 2004], visando integrar modelos de linguagem (LLMs) a frameworks de automação para testes E2E, por meio da criação e avaliação de um protótipo denominado GenIA-E2ETest. O projeto foi estruturado em três etapas: (i) diagnóstico do problema e revisão da base de conhecimento; (ii) desenvolvimento incremental da abordagem e do protótipo; e (iii) avaliação experimental.

Esta primeira etapa do projeto consistiu em mapear os desafios atuais da automação de testes E2E, bem como revisar o estado da arte e o corpo de conhecimento necessário para o desenvolvimento da solução proposta. Além disso, houve a exploração das capacidades dos LLMs como assistentes para geração de código, com foco em técnicas de engenharia de *prompts*.

A segunda fase, com foco no desenvolvimento do protótipo, seguiu um ciclo iterativo de experimentação e melhorias, com a criação de um pipeline capaz de transformar descrições em linguagem natural em scripts automatizados. A abordagem é dividida em três níveis: (1) modularização dos cenários; (2) extração e refinamento dos elementos da interface; (3) geração dos scripts E2E automatizados.

Na última etapa, a avaliação empírica utilizou o modelo *Goal-Question-Metric* (GQM) [Caldiera and Rombach 1994] para medir corretude, completude, esforço de adaptação e robustez. Foram definidos 6 cenários típicos para dois sistemas web distintos. Para garantir determinismo, cada caso de teste foi gerado e executado 3 vezes, com os mesmos *prompts* e configurações, totalizando 36 execuções supervisionadas.

A avaliação empírica da abordagem, realizada em 36 execuções de scripts, mostrou que a ferramenta apresenta alta corretude na identificação de elementos da interface (77% em média) e no *recall* de execução (85%). A completude atingiu 100% de cobertura dos elementos e 104% na cobertura dos passos, devido à decomposição detalhada de ações e verificações adicionais que não comprometem a lógica dos testes. O esforço manual para adaptar os scripts foi baixo, com média de 10% de modificações. Em termos de robustez, a GenIA-E2ETest se mostrou confiável em cenários estruturados, mas sensível à navegação dependente de contexto, elementos dinâmicos, ambiguidades semânticas, interferência de conteúdo externo e localizadores frágeis, indicando, para trabalhos futuros, a necessidade de melhorias para ambientes mais complexos.

Destaca-se que este trabalho faz referência ao estudo mais aprofundado de Junior et al. [Junior et al. 2025], que detalha toda a pesquisa. Além dos avanços técnicos, a pesquisa proporcionou importante experiência acadêmica e profissional ao autor.

Referências

- Caldiera, V. R. B. G. and Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of software engineering*, pages 528–532.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, pages 75–105.
- Junior, E., Valejo, A., Valverde-Rebaza, J. C., and Neves, V. O. (2025). Genia-e2etest: A generative ai-based approach for end-to-end test automation. In *Proceedings of the XXXIX Brazilian Symposium on Software Engineering (SBES)*, Recife-PE, Brazil. To appear.