浙江大学
ZHEJIANG UNIVERSITY

May 6, 2025

Dear Editor,

We are pleased to submit our manuscript titled L̈LM-Powered GUI Agents in Phone Automation: Surveying Progress and Prospects̈for consideration for publication in *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Motivation and Timeliness: The rapid evolution of Large Language Models (LLMs) has transformed phone automation from rigid script-based approaches to intelligent, adaptive systems. This paradigm shift comes at a critical juncture when commercial applications of GUI agents are rapidly emerging across mobile ecosystems. Despite this technological acceleration, there exists no comprehensive survey that systematically examines LLM-powered GUI agents in phone automation. Our paper addresses this gap by providing a structured framework for understanding both theoretical foundations and practical implementations.

Scope and Contributions: Our survey makes several significant contributions:

1. We provide a comprehensive examination of LLM-powered phone GUI agents, tracing their development from script-based automation to intelligent systems capable of understanding, planning, and executing tasks in dynamic mobile environments.

2. We propose a unified methodological framework that captures various design paradigms (single-agent, multi-agent, plan-then-act), model selection approaches, training strategies, and evaluation protocols.

3. We analyze why and how LLMs enhance phone automation through advanced natural language comprehension, multimodal grounding, and decision-making capabilities that bridge user intent with GUI actions.

4. We review the latest developments in datasets and benchmarks for phone GUI agents, providing a foundation for systematic training and fair performance assessment.

5. We identify key challenges and promising directions for future research, including dataset diversity, on-device efficiency, user-centric adaptation, and security considerations.

Relevance to TPAMI: Our survey aligns with TPAMI's focus on pattern analysis and machine intelligence. The paper examines how multimodal pattern recognition, intelligent decision-making, and adaptive learning are applied to phone GUI agents. It highlights the intersection of computer vision for screen perception and natural language processing for intent understanding, making it particularly suitable for TPAMI's interdisciplinary readership.

Prior Publication: We confirm that this manuscript is not under consideration for publication elsewhere, and all previous work is appropriately cited with clear distinction from our contributions.

We appreciate your consideration and look forward to your response.

Sincerely yours,

Yong Liu Ph.D.
Professor
College of Control Science and Engineering, Zhejiang University.
Hangzhou, 310027, China