

April 24, 2025

Editor-in-Chief  
IEEE Transactions on Pattern Analysis and Machine Intelligence  
IEEE Computer Society

Dear Editor,

We are pleased to submit our manuscript titled “LLM-Powered GUI Agents in Phone Automation: A Survey of Progress and Prospects” for consideration for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).

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

**Scope and Contributions:** Our survey makes several significant contributions to the literature:

1. We provide a comprehensive and systematic examination of LLM-powered phone GUI agents, tracing their development from script-based automation to the current state of 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, including single-agent, multi-agent, and plan-then-act frameworks, as well as model selection, training strategies, and evaluation protocols.
3. We analyze why and how LLMs enhance phone automation through advances in natural language comprehension, multimodal grounding, reasoning, and decision-making capabilities that bridge user intent with GUI actions.
4. We introduce and evaluate 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 novel perspectives for future research, including dataset diversity, on-device efficiency, user-centric adaptation, and security considerations.

**Relevance to TPAMI:** We believe this survey is well-aligned with TPAMI’s focus on pattern analysis and machine intelligence. The paper examines how multimodal pattern recognition, intelligent decision-making, and adaptive learning techniques are applied to the domain of phone GUI agents. It highlights the intersection of computer vision (screen perception), natural language processing (intent understanding), and reinforcement learning (adaptive interaction), 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 of our manuscript and look forward to your response.

Sincerely,

Guangyi Liu, Pengxiang Zhao, Liang  
Liu, Yaxuan Guo, Han Xiao, Weifeng  
Lin, Yuxiang Chai, Yue Han, Shuai  
Ren, Hao Wang, Xiaoyu Liang,  
Wenhao Wang, Tianze Wu, Linghao  
Li, Hao Wang, Guanjing Xiong, Yong  
Liu, Hongsheng Li