OS-COPILOT: TOWARDS GENERALIST COMPUTER AGENTS WITH SELF-IMPROVEMENT

Zhiyong Wu[◇,*], **Chengcheng Han**[♣,*], **Zichen Ding**[♣], **Zhenmin Weng**[♣],
Zhoumianze Liu[◇], **Shunyu Yao**[♡], **Tao Yu**[♠], **Lingpeng Kong**[♠]
[◇]Shanghai AI Laboratory [♣]East China Normal University
[♡]Princeton University [♠]The University of Hong Kong
wuzhiyong@pjlab.org.cn
https://os-copilot.github.io/

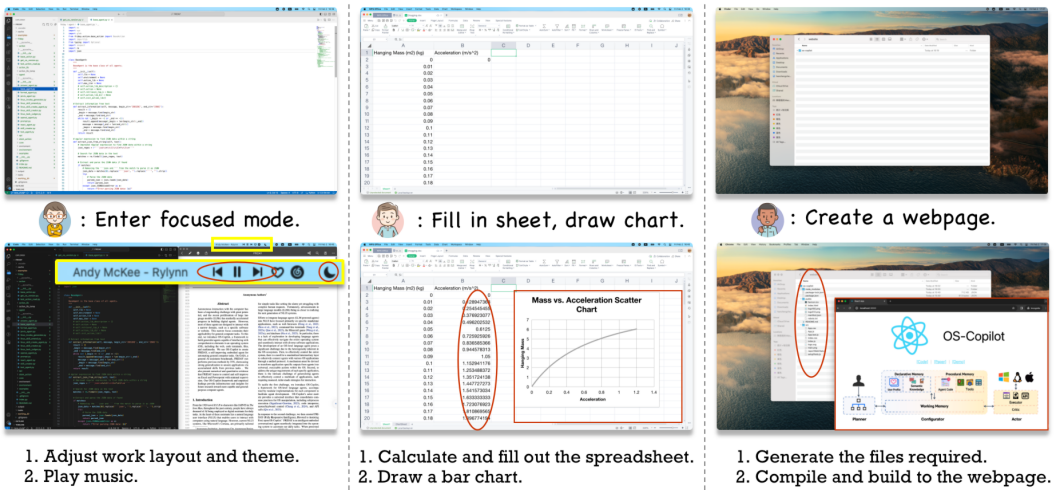

Figure 1: Running examples of FRIDAY when deployed on MacOS and tasked with (1) preparing a focused working environment, (2) Calculating and drawing a chart in Excel, and (3) creating a website for OS-Copilot. The text at the bottom illustrates the subtasks taken by FRIDAY. For each set of examples, the figure at the top represents the initial OS state, while the one at the bottom depicts the final state after execution. Boxes/Ovals highlight the changes made by FRIDAY.

ABSTRACT

Autonomous interaction with the computer has been a longstanding challenge with great potential, and the recent proliferation of large language models (LLMs) has markedly accelerated progress in building digital agents. However, most of these agents are designed to interact with a narrow domain, such as a specific software or website. This narrow focus constrains their applicability for general computer tasks. To this end, we introduce OS-Copilot, a framework to build generalist agents capable of interfacing with comprehensive elements in an operating system (OS), including the web, code terminals, files, multimedia, and various third-party applications. We use OS-Copilot to create FRIDAY, a self-improving embodied agent for automating general computer tasks. On GAIA, a general AI assistants benchmark, FRIDAY outperforms previous methods by 35%, showcasing strong generalization to unseen applications via accumulated skills from previous tasks. We also present numerical and quantitative evidence that FRIDAY learns to control and self-improve on Excel and Powerpoint with minimal supervision. Our OS-Copilot framework and empirical findings provide infrastructure and insights for future research toward more capable and general-purpose computer agents.

[*] Equal Contribution.