

VerificAgent: Integrating Expert Knowledge and Fact-Checked Memory for Robust Domain-Specific Task Planning

Thong Q. Nguyen*, Shubhang Desai*, Yash Jain, Tanvir Aumi, Vishal Chowdhary
Microsoft



INTRODUCTION

Problem statement

GUI agents struggle with complex domain workflows like Office ($\approx 25\%$ success on OSWorld Office tasks)

Gap

Unchecked memory \rightarrow spurious “learnings” degrade planning.

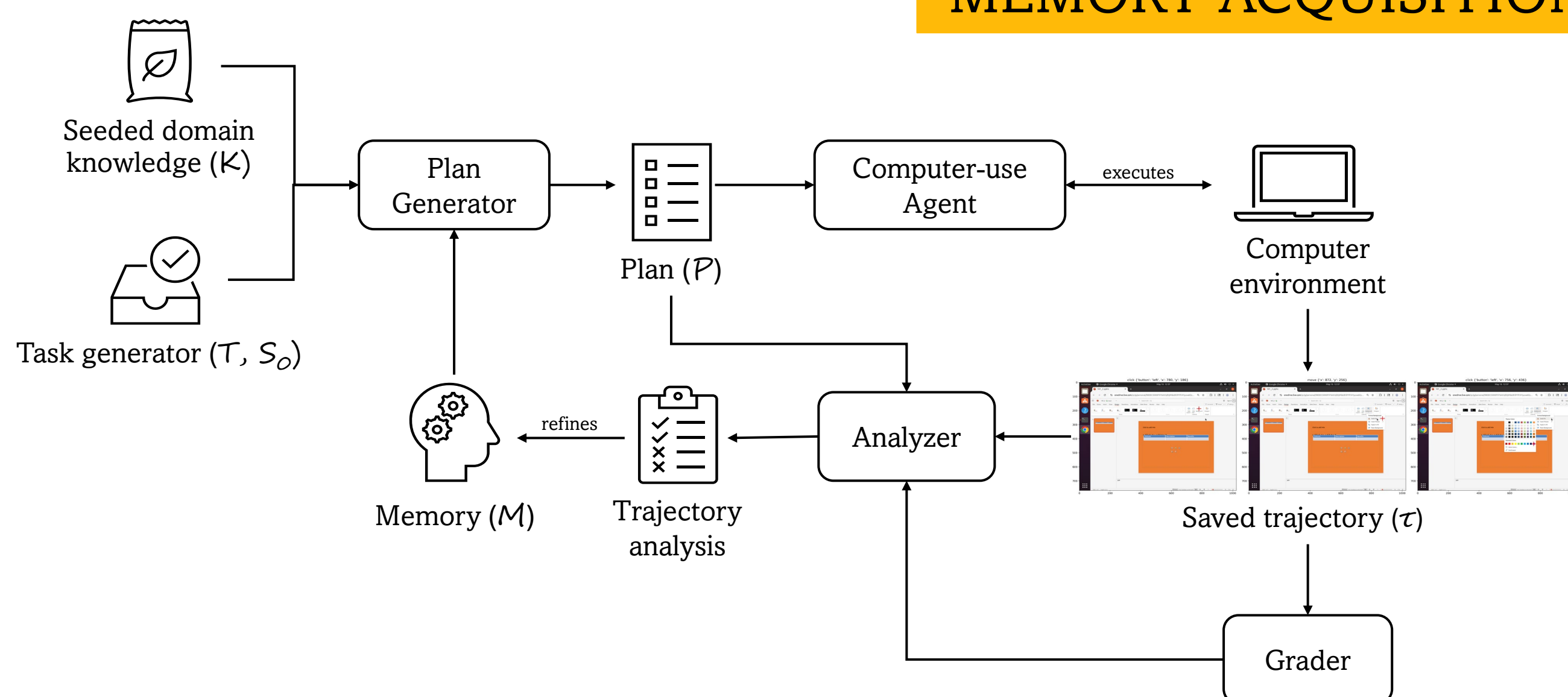
Goal

Design a memory-augmented CUA that remains robust via fact-checked memory.

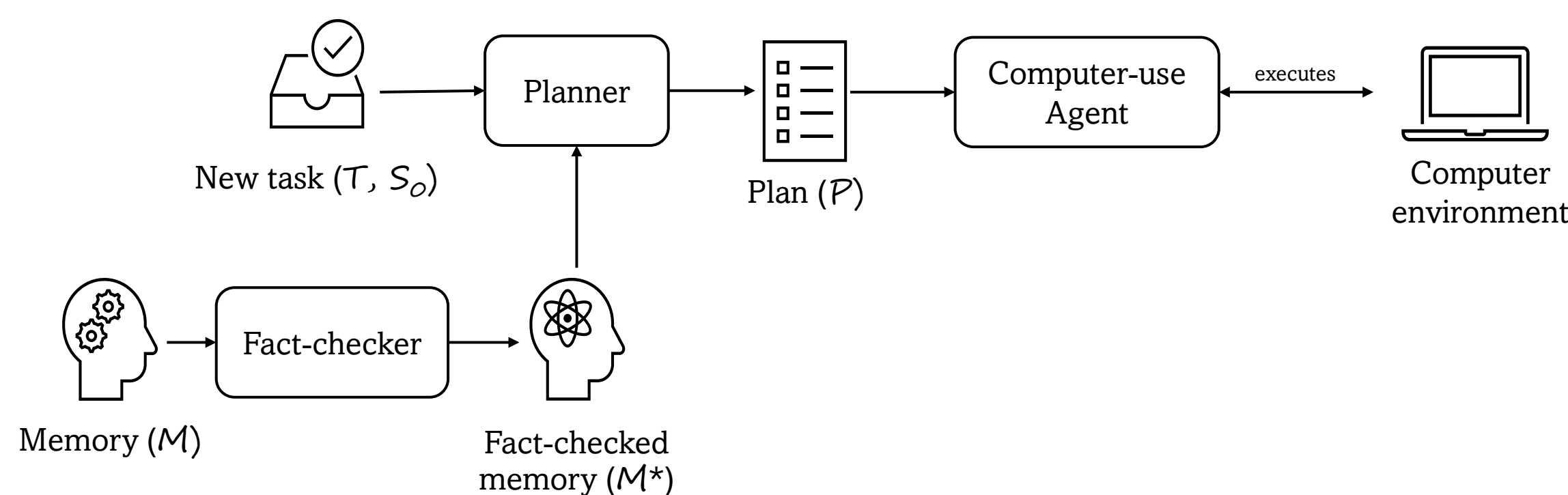
RELATED WORK

- CUA paradigm (Operator, 38% on OSWorld)
- Persistent memory in agents (Agent S, S2)
- Lack of memory sanitization in prior work

MEMORY ACQUISITION

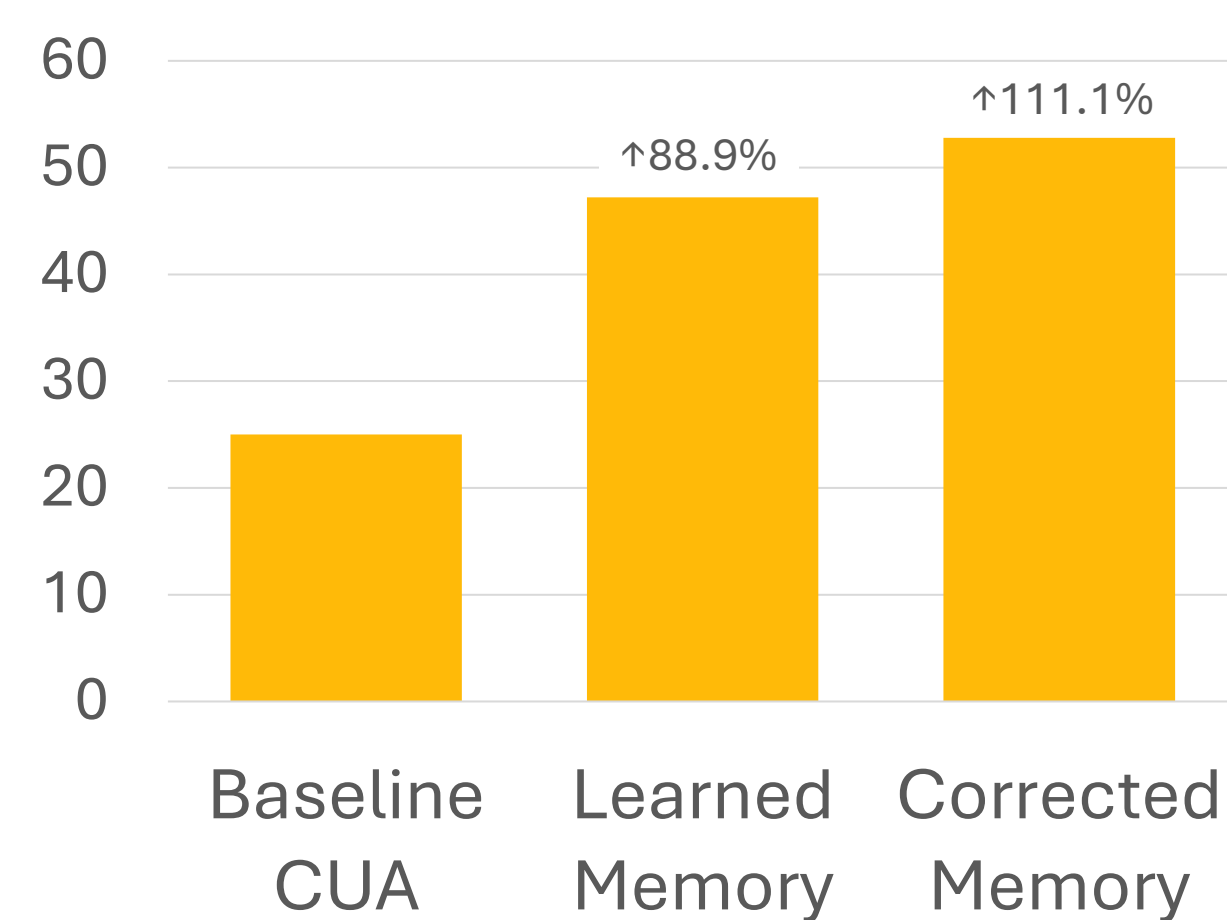


INFERENCE



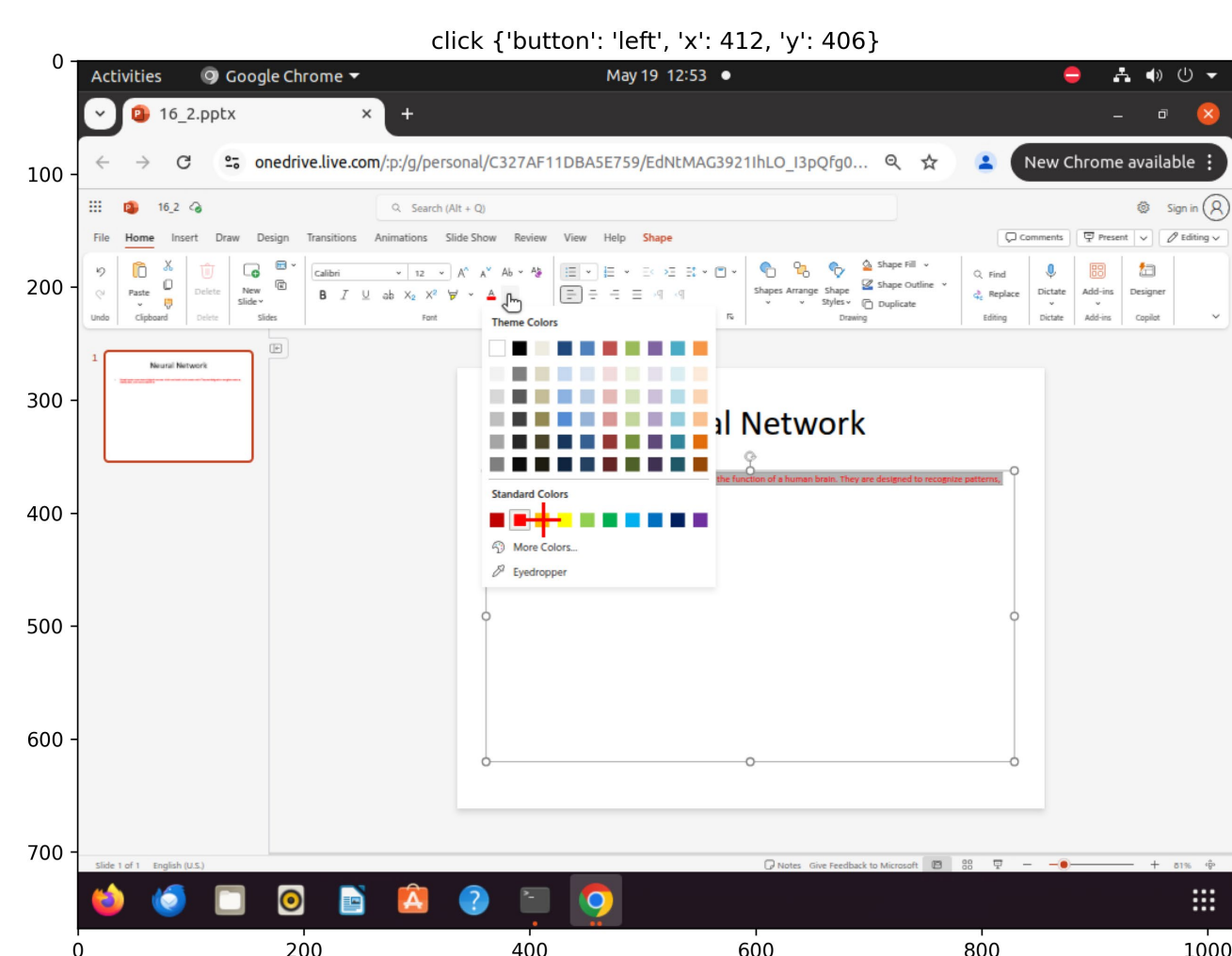
RESULTS

Success Rate Across Memory Settings



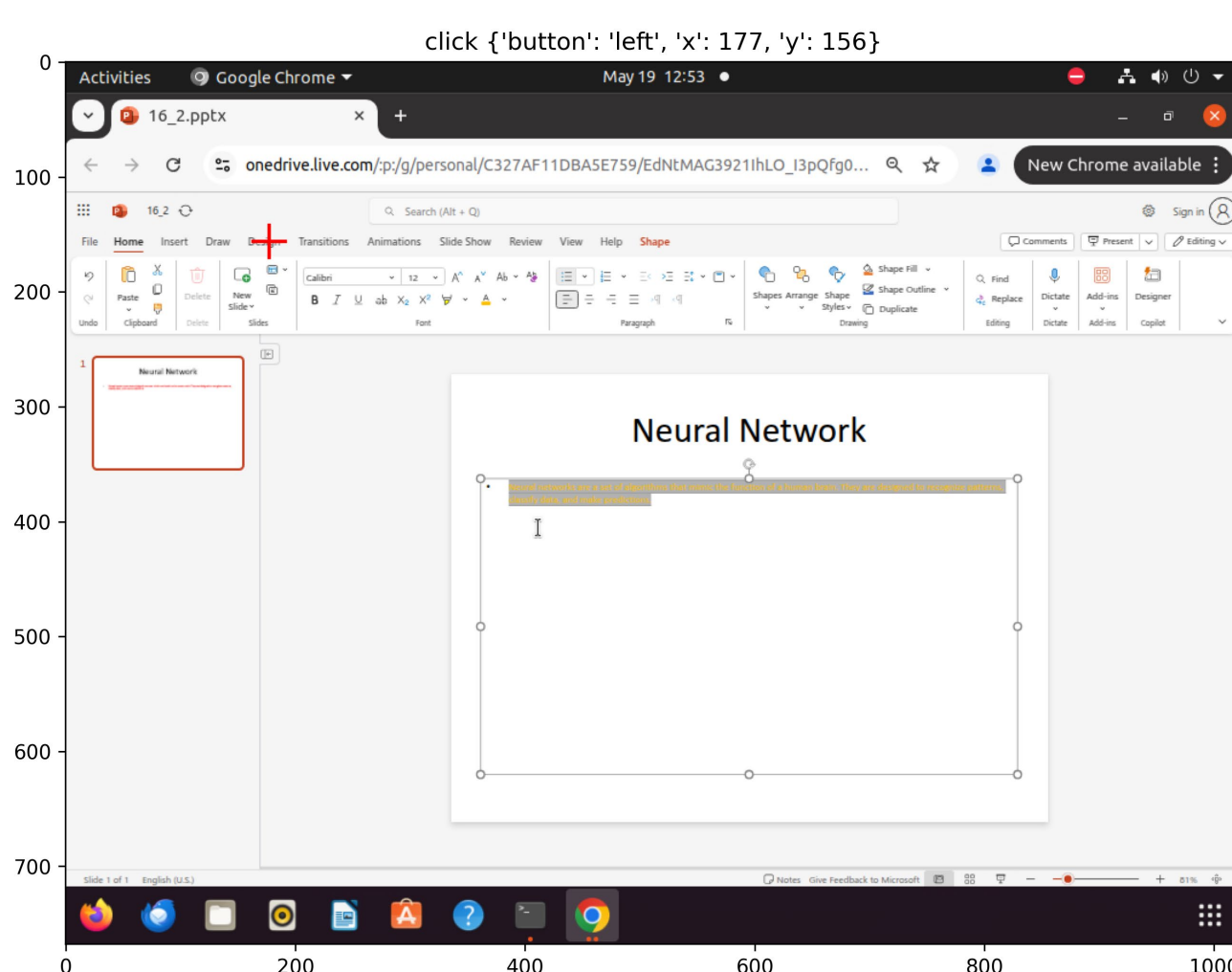
Error type	Learned memory	Corrected memory
Memory	42.11%	—
Planner	15.79%	41.18%
Agent	42.11%	58.82%

QUALITATIVE EXAMPLE



Plan before memory correction

- ✓ Click the text box containing the content.
- ✓ Press Ctrl+A to select all text.
- ✓ Click the “Font Size” dropdown in the Home ribbon, type 12, and press Enter.
- ✗ Click the Font Color button (the “A” icon with a color bar) in the Home ribbon and choose orange.
- ✓ Click the “Design” tab.
- ✓ Click “Format Background”.
- ✓ In the Format Background pane, select “Solid Fill” and choose red.



Plan after memory correction

- ✓ Click the text box containing the content.
- ✓ Press Ctrl+A to select all text.
- ✓ Click the “Font Size” dropdown, type 12, and press Enter.
- ✓ Click the chevron next to the Font Color button in the Home ribbon and select orange.
- ✓ Click the “Design” tab.
- ✓ Click “Format Background”.
- ✓ In the Format Background pane, select “Solid Fill” and choose red.

