OPENCUA: The Open Foundation for Computer-Use Agents

พิมพ์เขียวระบบเปิดสำหรับ AI ผู้ช่วยบนคอมพิวเตอร์แบบ End-to-End

XLANG Lab & Open-Source Contributors

NotebookLM

ปัญหาของระบบปิด (The Black Box Problem)

ระบบปิด (Proprietary Agents)

- ❌ โมเดล CUA ชั้นนำในปัจจุบันล้วนเป็นระบบปิด
- ❌ ไม่มีข้อมูลชุดฝึกสอน (Training Data)
- ❌ สถาปัตยกรรมตัวโมเดลถูกปิดบัง
- ❌ ตรวจสอบความปลอดภัยและข้อจำกัดได้ยาก

ระบบเปิด (The Glass Box Solution)

- ✅ ชุมชนนักพัฒนาต้องการรากฐานระบบเปิดเพื่อวิจัยต่อยอด
- ✅ โปร่งใสทุกขั้นตอนการทำงาน
- ✅ ชุดข้อมูลและโค้ดถูกเปิดเผยทั้งหมด
- ✅ ตรวจสอบ ปรับปรุง และพัฒนาต่อได้แบบ Open-Source

OpenCUA คือเฟรมเวิร์ก Open-Source ตัวแรกที่เปิดเผยสถาปัตยกรรมแบบครบวงจร ตั้งแต่การเก็บข้อมูลไปจนถึงตัวโมเดล

NotebookLM

3 เสาหลักของสถาปัตยกรรม OpenCUA

Pillar 1: Infrastructure
(ซอฟต์แวร์เก็บข้อมูล)

Pillar 1: Infrastructure
(ซอฟต์แวร์เก็บข้อมูล)

AgentNet Tool: เครื่องมือบันทึกพฤติกรรมการใช้งานข้ามระบบปฏิบัติการ (Windows, macOS, Ubuntu) แบบไร้รอยต่อ

Pillar 2: Data
(ชุดข้อมูลคุณภาพสูง)

Pillar 2: Data
(ชุดข้อมูลคุณภาพสูง)

AgentNet Dataset: ข้อมูล Trajectory กว่า 22.6K รายการที่บันทึกการใช้งานจริง ครอบคลุมหลายโดเมน

Pillar 3: The Brain
(กระบวนการคิดและโมเดล)

Pillar 3: The Brain
(กระบวนการคิดและโมเดล)

Reflective Long CoT & Models: กระบวนการสอนให้โมเดลรู้จัก ‘คิดทบทวน’ และ ‘แก้ไขข้อผิดพลาด’ ก่อนลงมือทำ

NotebookLM

โครงสร้างพื้นฐาน: AgentNet Tool ดึงข้อมูลเชิงลึกแบบไร้รอยต่อ

เครื่องมือที่ติดตั้งบนเครื่องผู้ใช้เพื่อบันทึกงานซับซ้อนในเบื้องหลัง โดยไม่รบกวน Workflow ธรรมชาติ

Human Time/Workflow

ScreenVideo:
บันทึกวิดีโอหน้าจอความละเอียดสูง

Keyboard/Mouse Signals:
พิกัดและการคลิก/พิมพ์แบบเรียลไทม์

Accessibility Trees (AXTree):
โครงสร้าง UI เชิงลึก

HTML & Metadata:
โค้ดเบื้องหลังของเว็บหรือแอปพลิเคชัน

NotebookLM

การกลั่นกรองข้อมูล: จากวิดีโอดิบ สู่ State-Action Pairs

Step 1:
Raw Demonstrations
(ข้อมูลดิบ)

วิดีโอต่อเนื่องและสัญญาณเมาส์/
คีย์บอร์ดนับพันรายการ
(หนาแน่นเกินไปสำหรับเทรน AI)

Step 2:
Action Reduction
(บีบอัดการกระทำ)

ยุบรวมการเคลื่อนไหวเมาส์
หรือการพิมพ์ที่ซับซ้อน
ให้เหลือเพียงคำสั่งระดับสูง
(เช่น คลิก, ลาก, พิมพ์ข้อความ)

Step 3:
State-Action Matching
(จับคู่สถานะที่แม่นยำ)

ย้อนหาเฟรมภาพสุดท้ายก่อนที่เมาส์จะขยับ

เพื่อป้องกันไม่ให้ภาพ ‘เฉลย’ ตำแหน่งล่วงหน้า
(No future information leakage)

NotebookLM

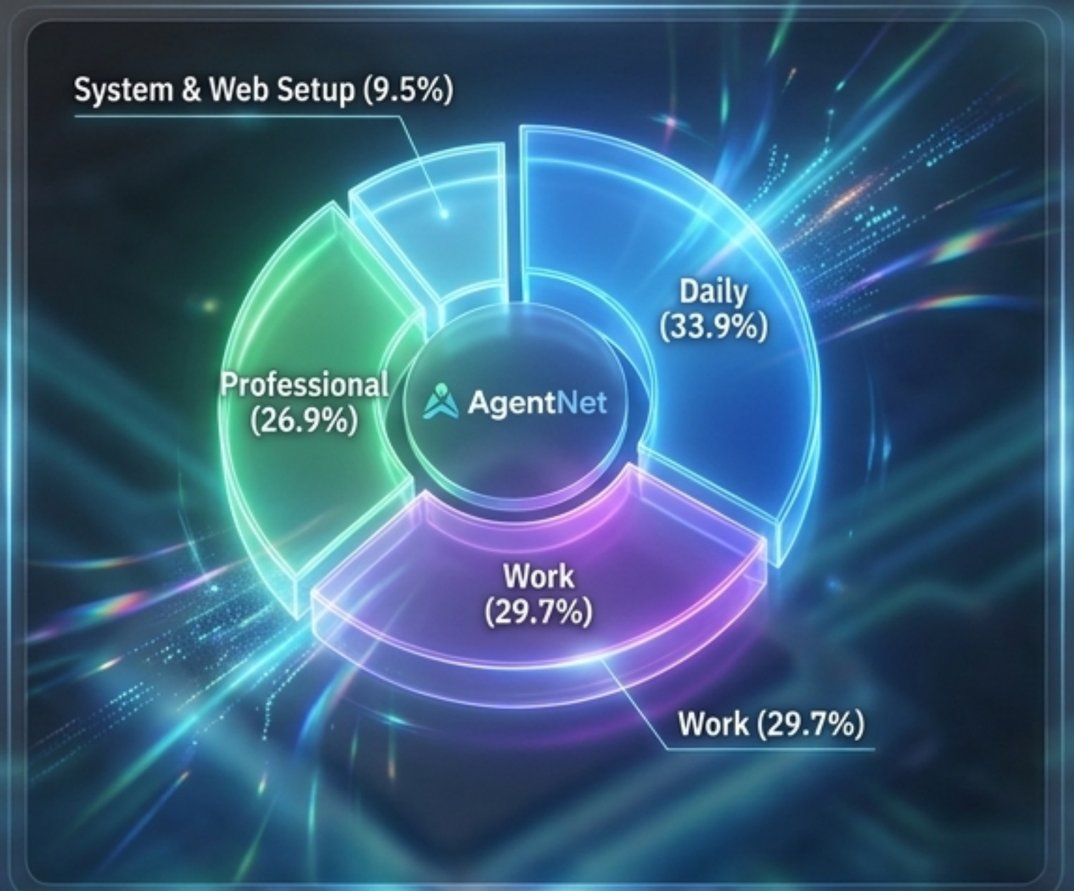
รากฐานข้อมูล: สภาพแวดล้อมที่เหนือกว่า (AgentNet Dataset)

Dataset	Environment	Video	Inner Monologue	Complexity (Avg. Steps)
AndroidControl	Mobile	✓	—	Low (5.5)
Mind2Web	Web	✓	—	Low (7.3)
AgentTrek	Web	✓	Short	Medium (12.1)
AgentNet	Desktop (Win/Mac/Ubuntu)	✓	Long (Reflective)	High (18.6)

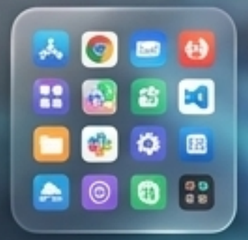
AgentNet คือชุดข้อมูล Desktop Trajectory ระดับสเกลใหญ่ชุดแรกที่สมจริง ซับซ้อน และมี Inner Monologue แบบยาว

NotebookLM

ข้อมูล 226K Trajectories ครอบคลุมทุกการทำงานจริง



ใช้งานบน 3 ระบบปฏิบัติการ: Windows, macOS, Ubuntu (ช่วยเรื่อง Cross-platform Generalization)

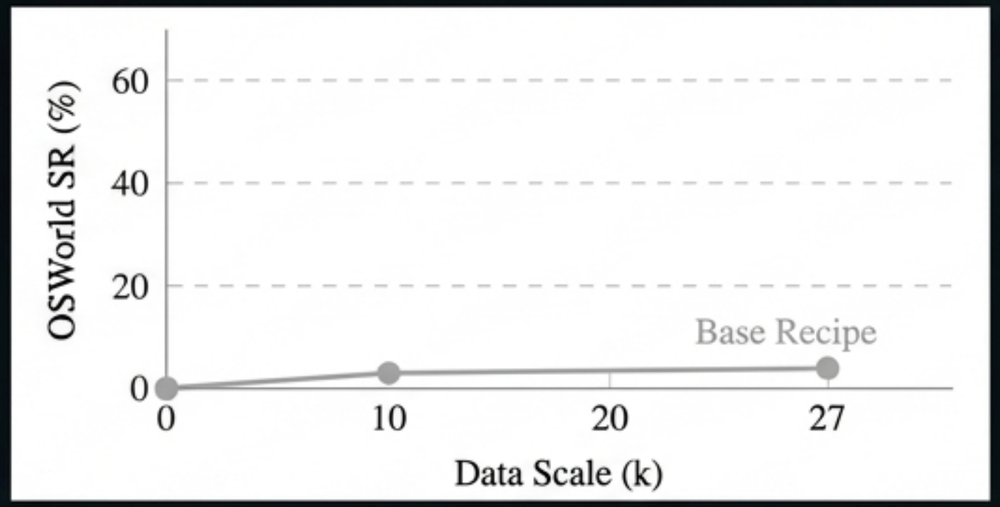


ครอบคลุม **>140 แอปพลิเคชั่น** และ **>200 เว็บไซต์**

Insight: ความหลากหลายนี้บังคับให้โมเดลต้อง 'เข้าใจ' กลไกของ GUI จริงๆ ไม่ใช่แค่การจดจำ (Memorization) เฉพาะแอป

NotebookLM

ปัญหา: ทำไม SFT แบบเดิมถึงล้มเหลว?



The Expectation (ความคาดหวัง)

การป้อนข้อมูล State-Action ล้วนๆ (หน้าจอ + คำสั่งคลิก) น่าจะเพียงพอให้ AI เรียนรู้

The Reality (ความเป็นจริง)

โมเดลได้คะแนน OSWorld เพียง 4.4% (ดูกราฟ Base Recipe)

The Reason (สาเหตุ)

โมเดลขาด 'ความสามารถในการคิดวิเคราะห์' (Inner Monologue) เมื่อเจอปัญหาที่ไม่คาดคิดในสเต็ปที่ 10 หรือ 20 โมเดลจะไม่สามารถกู้สถานการณ์กลับมาได้

NotebookLM

กุญแจสำคัญ: Reflective Long Chain-of-Thought (CoT)

L3: Contextual Observation (กว้างสุด)
สังเกตภาพรวมของหน้าจอ ดึงข้อมูลภาพและข้อความที่สำคัญออกมา (Salient elements)

L2: Reflective Reasoning (แกนกลาง)
วิเคราะห์สิ่งที่เพิ่งเปลี่ยนไป ทบทวนความจำ (Memory) ตรวจสอบข้อผิดพลาด และวางแผนสเต็ปถัดไป (Planning)

L1: Executable Action (จุดบรรจบ)
แปลงความคิดทั้งหมดออกมาเป็นคำสั่งโค้ดสั้นๆ ที่ใช้งานได้จริง (เช่น click(x, y))

การผสมผสาน CoT ทั้ง 3 ระดับเข้าด้วยกัน (Mixture-CoT)
ทำให้โมเดลเข้าใจความเชื่อมโยงตั้งแต่การรับรู้ไปจนถึงการลงมือทำ

NotebookLM

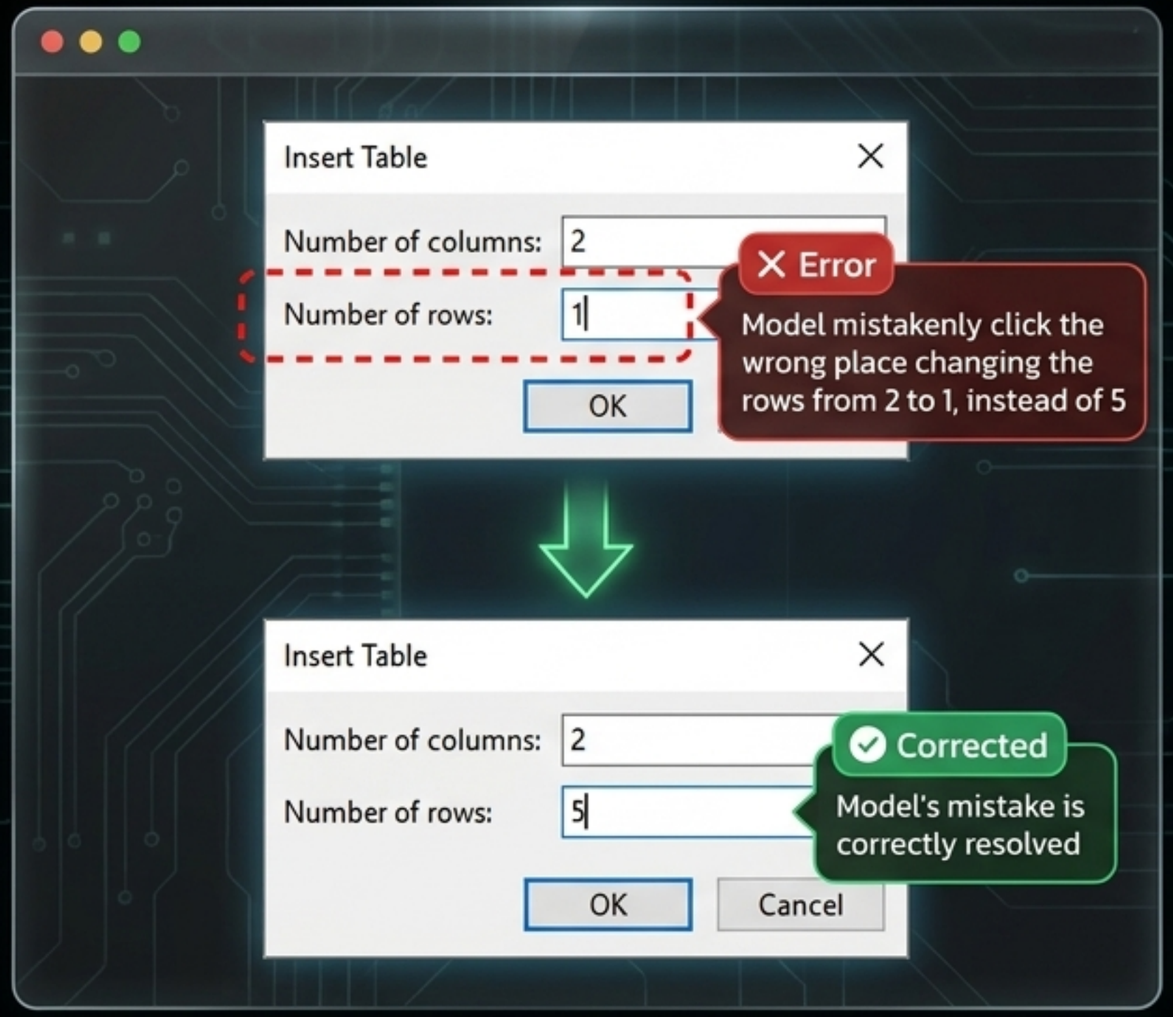
The Reflector: เรียนรู้จากความผิดพลาดของมนุษย์

ข้อมูลจากมนุษย์ไม่ได้สมบูรณ์แบบเสมอไป OpenCUA เปลี่ยน ‘ข้อผิดพลาด’ ให้เป็น ‘บทเรียน’

โมเดลจึงไม่ได้เรียนรู้แค่ ‘วิธีทำงานให้สำเร็จ’ แต่เรียนรู้ ‘วิธีแก้ปัญหาเมื่อทำพลาด’ ซึ่งสำคัญมากในโลกจริง

NotebookLM

เอ็กซเรย์กระบวนการคิด: กลไกการแก้ไขปัญหาของ OpenCUA



Inner Monologue

Amber Reflection (การสะท้อน)

เป้าหมายคือสร้างตาราง 5 แถว แต่ตอนนี้ค่าถูกตั้งเป็น 1

Memory Memory (ความจำ)

ก่อนหน้านี้เรากดเปลี่ยนค่าเป็น 1 ไปแล้ว (เราทำพลาด)

Plan & Reasoning (การวางแผน & การใช้เหตุผล)

เราต้องเลือกค่าปัจจุบันแล้วพิมพ์ '5' ทับลงไป จากนั้นค่อยกด OK

Code Prediction (โค้ด)

```
pyautogui.moveTo(1362, 578); pyautogui.dragTo(...)
```

การคิดอย่างเป็นระบบ ช่วยให้โมเดลนำตัวเองกลับสู่เส้นทางที่ถูกต้องได้ (Self-Correction)

NotebookLM

กลยุทธ์การผสมข้อมูล (Data Mixtures Strategy)

General SFT Data (30-35%)

ข้อมูลทั่วไป (ข้อมูลทั่วไป (คณิตศาสตร์, การเขียนโค้ด, Logic)

Planning & CUA Data (45-50%)

ข้อมูล Trajectory และ CoT จาก AgentNet ให้โมเดลวางแผนเป็น

Grounding Data (20%)

ข้อมูลความเข้าใจพิกัด UI ให้ตำของโมเดลแม่นยำ

Why General Data?

แม้ข้อมูลทั่วไปจะไม่เกี่ยวกับ GUI เลย แต่การผสมข้อมูลชุดนี้ช่วยยกระดับ Generalization และทักษะการให้เหตุผลของ AI ได้อย่างมีนัยสำคัญ

NotebookLM

ผลลัพธ์ SOTA: OSWorld-Verified Leaderboard

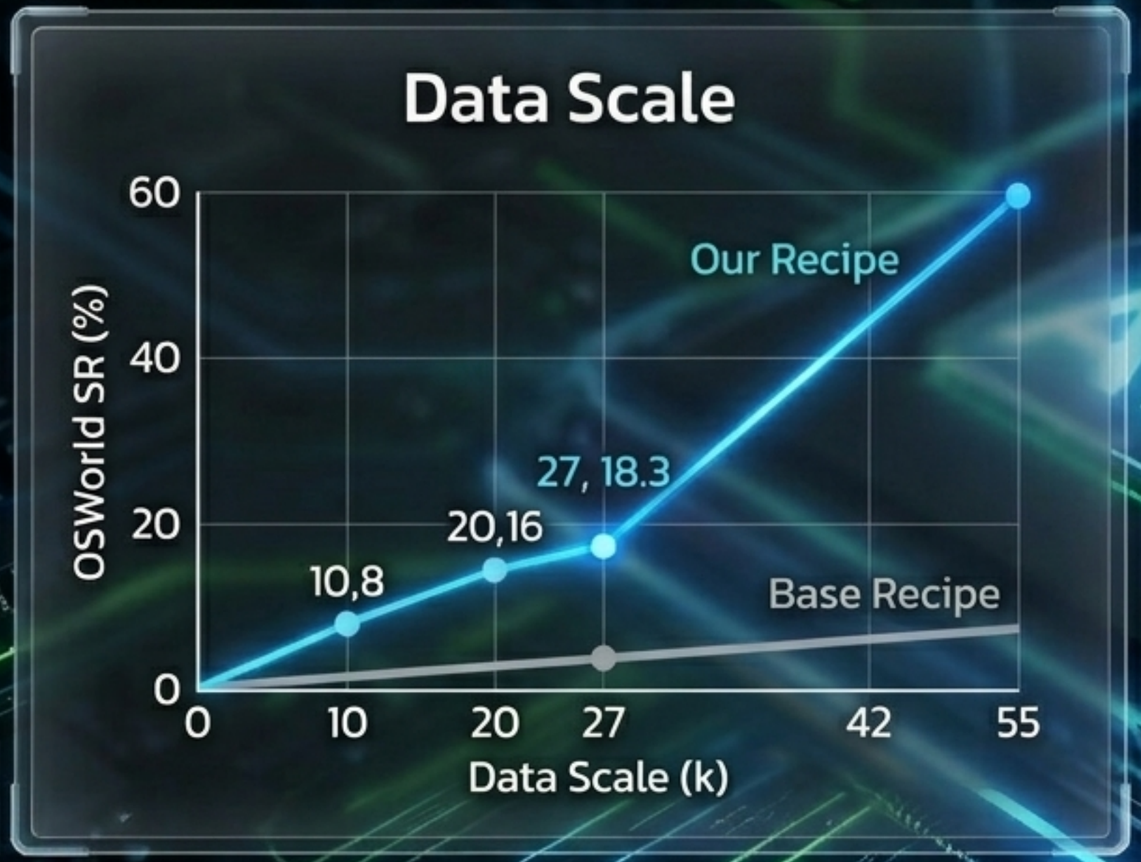
OpenCUA-72B สร้างมาตรฐานใหม่ (State-of-the-art) ในกลุ่ม Open-Source

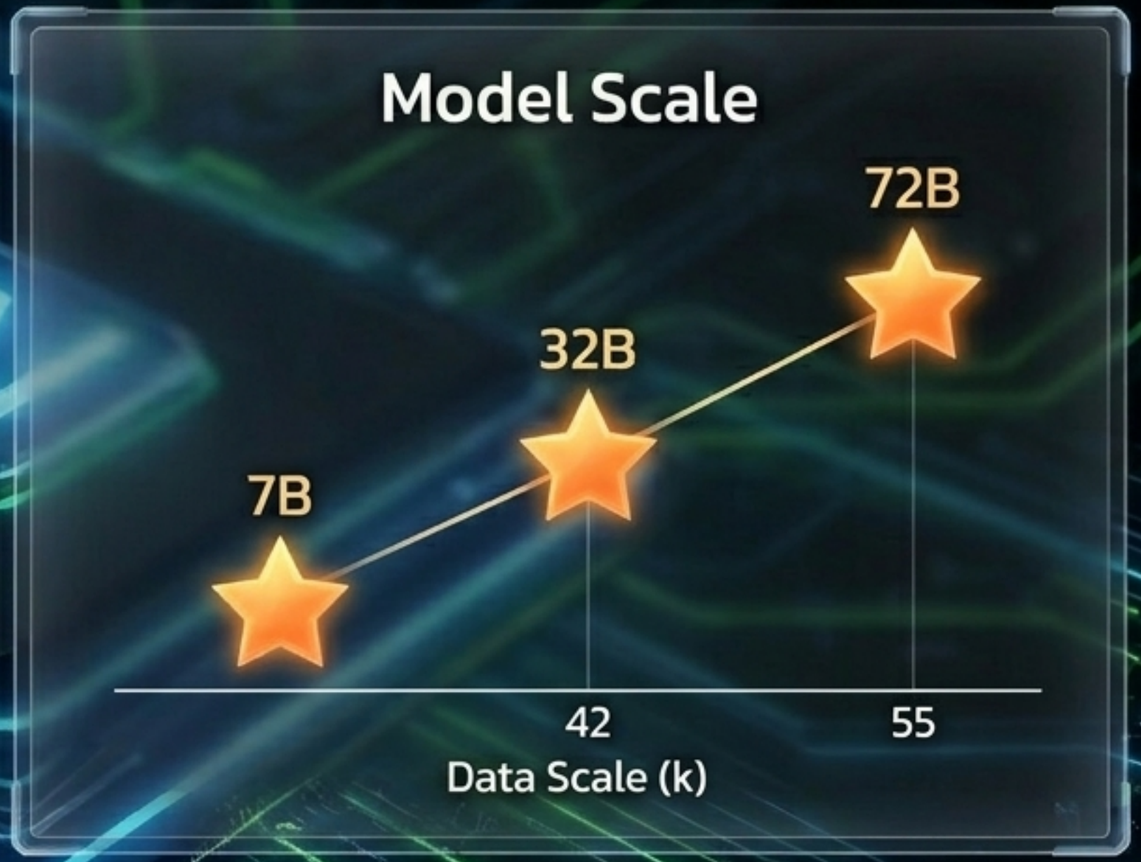
เป็นครั้งแรกที่โมเดลระบบเปิดทำคะแนนได้เหนือกว่า Claude 4 Sonnet และ OpenAI CUA

ปิดช่องว่างระหว่าง Open-Source กับ Proprietary Model อย่างที่ไม่เคยมีมาก่อน

NotebookLM

บทพิสูจน์ศักยภาพ: กฎการขยายสเกล (Scaling Laws)





สูตรการเทรนของ OpenCUA (Reflective CoT)
ปลดล็อกศักยภาพของข้อมูล
ทำให้โมเดลเก่งขึ้นตามจำนวนข้อมูลอย่างชัดเจน

สถาปัตยกรรมนี้รองรับการอัปสเกลโมเดลอย่างสมบูรณ์
ประสิทธิภาพก้าวกระโดดเมื่อเพิ่มขนาดพารามิเตอร์

NotebookLM

ร่วมสร้างอนาคตของ AI Assistant แบบเปิด

OpenCUA ไม่ใช่แค่โมเดล แต่เป็น “พิมพ์เขียว” และรากฐานแบบ Open-Source
เพื่อให้ชุมชนนักพัฒนาทั่วโลกสามารถตรวจสอบ พัฒนา และก้าวข้ามข้อจำกัดของ AI ในปัจจุบัน

The Complete Suite

- ✅ Annotation Tool (ซอฟต์แวร์เก็บข้อมูล)
- ✅ 22.6K Trajectories Dataset
- ✅ Source Code & Training Recipes
- ✅ Model Weights (7B, 32B, 72B)

Project Page: https://opencua.xlang.ai

ปลดล็อกศักยภาพ CUA ของคุณได้แล้ววันนี้ (Unlock the potential of your CUA today)

NotebookLM