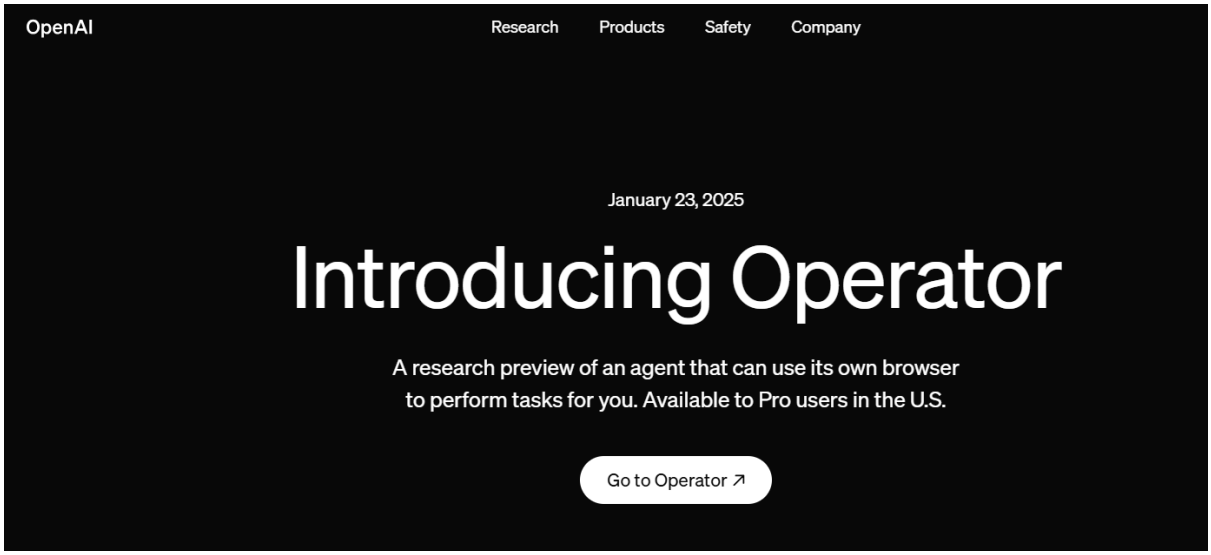
ChatGPT AI Agents – Operator

The Ultimate Guide for FP&A and Finance Teams

By Christian Martinez



OpenAI released Operator – **an AI Agent that can use its own browser to perform tasks for you.**

This is what it means for FP&A and Finance Teams.

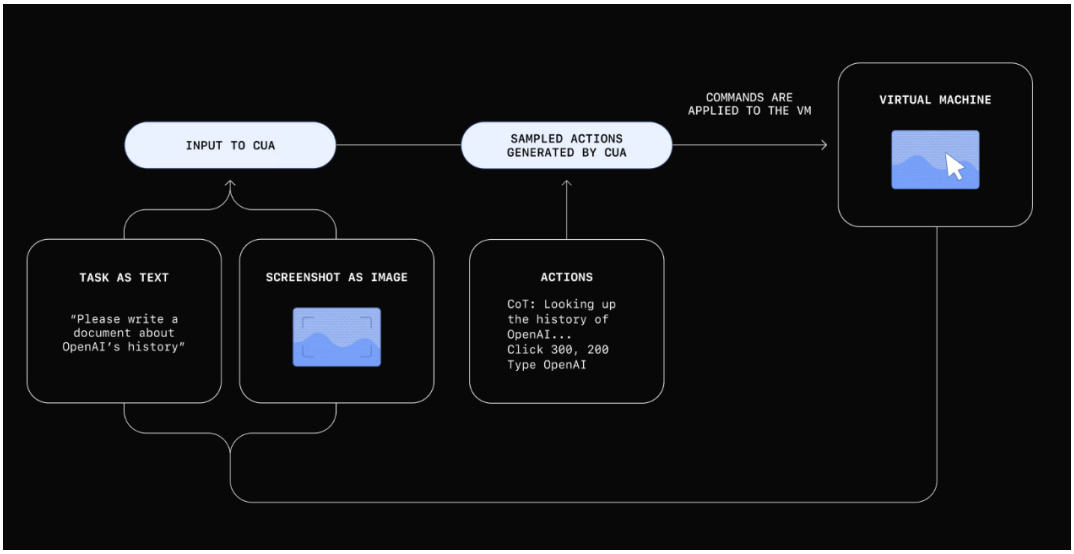
Think of Operator like a highly skilled virtual assistant that works on your computer, navigating websites, filling out forms, and pulling data just like a person would—but faster and more accurately.

To start, Operator will be like having a team member who can handle time-consuming tasks like reconciling financials, updating forecasts, gathering competitive market insights, or creating dashboards, freeing up your team to focus on high-level strategy.

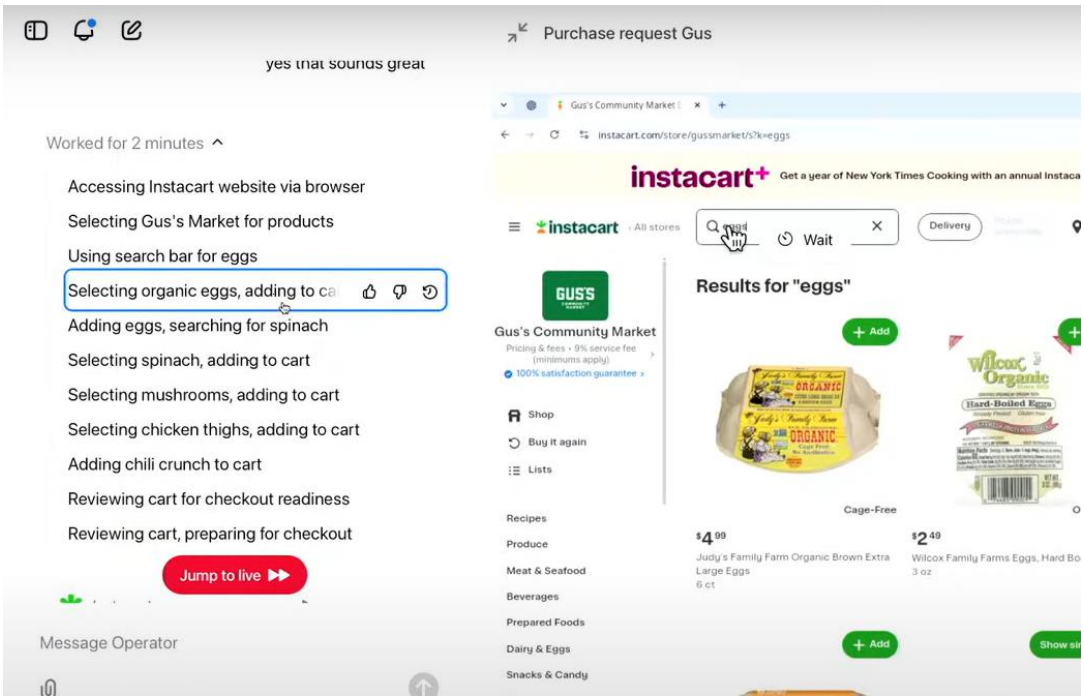
It’s smart, secure, and always works under your guidance, making financial operations more efficient and effective.

How Operator works

Operator is powered by a new model called Computer-Using Agent (CUA). Combining GPT-4o's vision capabilities with advanced reasoning through reinforcement learning, CUA is trained to interact with graphical user interfaces (GUIs)—the buttons, menus, and text fields people see on a screen.



Operator can “see” (through screenshots) and “interact” (using all the actions a mouse and keyboard allow) with a browser, enabling it to take action on the web without requiring custom API integrations.



If it encounters challenges or makes mistakes, Operator can leverage its reasoning capabilities to self-correct. When it gets stuck and needs assistance, it simply hands control back to the user, ensuring a smooth and collaborative experience.

While CUA is still in early stages and has limitations, it sets new state-of-the-art benchmark results in WebArena and WebVoyager, two key browser use benchmarks.

You can read more about evals and the research behind Operator in OpenAI’s research blog post.

Given a user's instruction, CUA operates through an iterative loop that integrates perception, reasoning, and action:

- **Perception**: Screenshots from the computer are added to the model's context, providing a visual snapshot of the computer's current state.
- **Reasoning**: CUA reasons through the next steps using chain-of-thought, taking into consideration current and past screenshots and actions. This inner monologue improves task performance by enabling the model to evaluate its observations, track intermediate steps, and adapt dynamically.
- **Action**: It performs the actions—clicking, scrolling, or typing—until it decides that the task is completed or user input is needed. While it handles most steps automatically, CUA seeks user confirmation for sensitive actions, such as entering login details or responding to CAPTCHA forms.
-

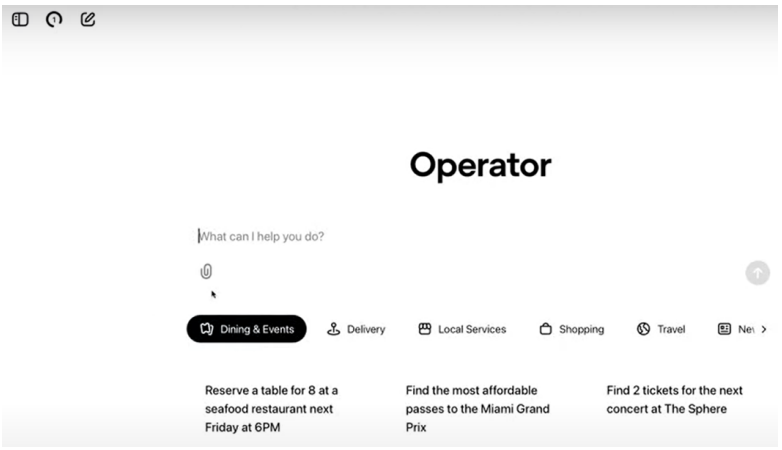
CUA establishes a new state-of-the-art in both computer use and browser use benchmarks by using the same universal interface of screen, mouse, and keyboard.

This is how it performed in AI evaluations

Benchmark type	Benchmark	Computer use (universal interface)		Web browsing agents	Human
		OpenAI CUA	Previous SOTA	Previous SOTA	
Computer use	OSWorld	38.1%	22.0%	-	72.4%
Browser use	WebArena	58.1%	36.2%	57.1%	78.2%
	WebVoyager	87.0%	56.0%	87.0%	-

Where to use Operator?

You can try it here with a ChatGPT Pro license.

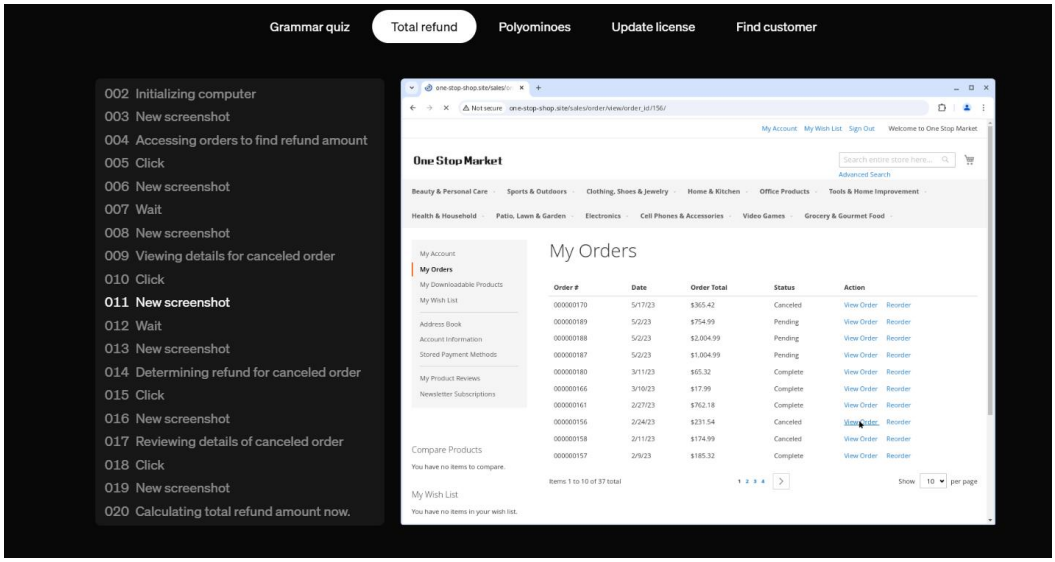


How to use

To get started, simply describe the task you'd like done and Operator can handle the rest.

Users can choose to take over control of the remote browser at any point, and Operator is trained to proactively ask the user to take over for tasks that require login, payment details, or when solving CAPTCHAs.

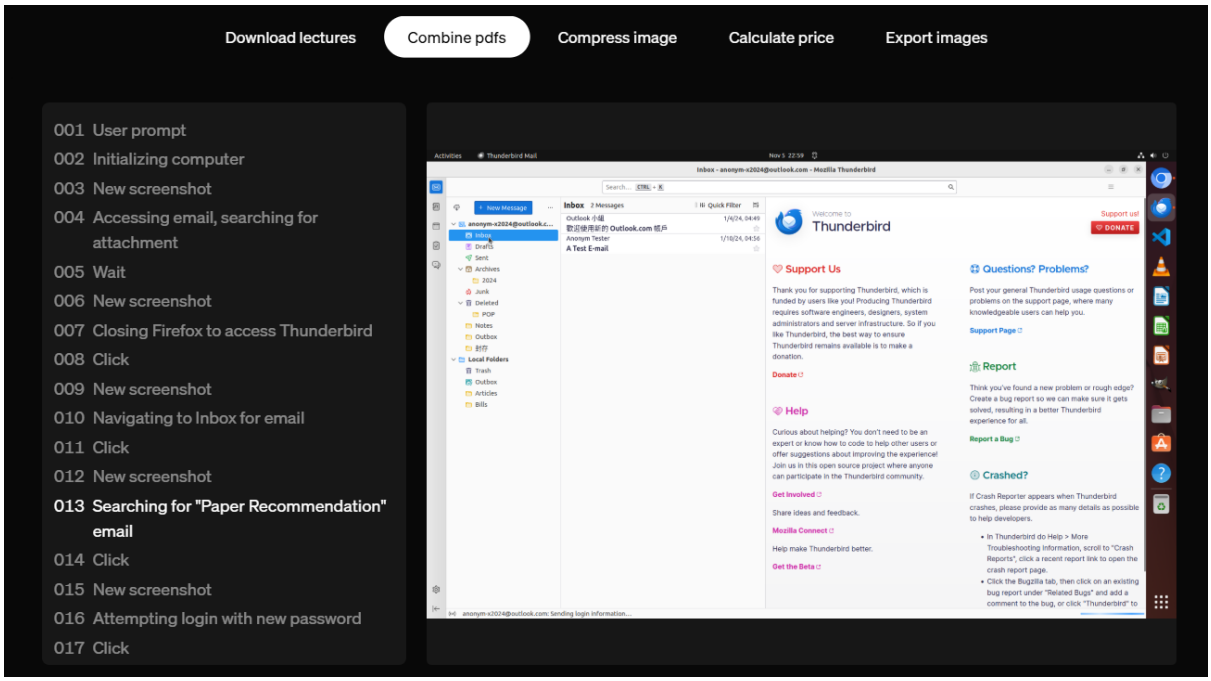
For example, it can handle refunds:



Users can personalize their workflows in Operator by adding custom instructions, either for all sites or for specific ones, such as setting preferences for airlines on Booking.com.

Operator lets users save prompts for quick access on the homepage, ideal for repeated tasks like restocking groceries on Instacart.

Similar to using multiple tabs on a browser, users can have Operator run multiple tasks simultaneously by creating new conversations, like ordering a personalized enamel mug on Etsy while booking a campsite on Hipcamp.



This is a list of Top 10 things to try with Operator for FP&A:

1. **Automate Data Extraction**
   Pull financial reports (e.g., P&L, balance sheets) from accounting systems like NetSuite, SAP, or QuickBooks.
2. **Consolidate Data Across Platforms**
   Gather data from CRM systems (e.g., Salesforce), marketing tools, and other sources into a single financial model or spreadsheet.
3. **Update Rolling Forecasts**
   Automatically input updated metrics (e.g., sales figures, market trends) into forecasting templates.
4. **Expense Tracking and Reporting**
   Retrieve expense data, categorize entries, and generate summaries for variance analysis.
5. **Scenario Analysis**
   Feed macroeconomic or operational variables (e.g., interest rates, commodity prices) into pre-built models to generate alternative forecasts.
6. **Market and Competitive Analysis**
   Scrape competitor earnings reports, pricing data, or market trends for benchmarking purposes.
7. **Generate Leadership Dashboards**
   Collect and visualize KPI data from multiple sources, updating dashboards in PowerPoint, Excel, or Tableau.
8. **Month-End Close Support**
   Automate tasks like reconciliation report generation, discrepancy checks, and journal entry uploads.
9. **Compliance Monitoring**
   Track tax filing deadlines, regulatory updates, or policy changes from government websites.

10. **Vendor and Contract Analysis**
    Compare vendor pricing, retrieve invoices, or extract key contract terms for procurement cost control.

Safety and privacy

OpenAI mentioned that the safe to use is a top priority, with three layers of safeguards to prevent abuse and ensure users are firmly in control.

First, Operator is trained to ensure that the person using it is always in control and asks for input at critical points.

- **Takeover mode:** Operator asks the user to take over when inputting sensitive information into the browser, such as login credentials or payment information. When in takeover mode, Operator does not collect or screenshot information entered by the user.
- **User confirmations**: Before finalizing any significant action, such as submitting an order or sending an email, Operator should ask for approval.
- **Task limitations:** Operator is trained to decline certain sensitive tasks, such as banking transactions or those requiring high-stakes decisions, like making a decision on a job application.
- **Watch mode**: On particularly sensitive sites, such as email or financial services, Operator requires close supervision of its actions, allowing users to directly catch any potential mistakes.

Next, they mentioned they have made it easy to manage data privacy in Operator.

- **Training opt out:** Turning off 'Improve the model for everyone' in ChatGPT settings means data in Operator will also not be used to train our models.
- **Transparent data management:** Users can delete all browsing data and log out of all sites with one click under the Privacy section of Operator settings. Past conversations in Operator can also be deleted with one click.

Lastly, OpenAI mentions they have built defenses against adversarial websites that may try to mislead Operator through hidden prompts, malicious code, or phishing attempts:

- **Cautious navigation**: Operator is designed to detect and ignore prompt injections.
- **Monitoring**: A dedicated "monitor model" watches for suspicious behavior and can pause the task if something seems off.
- **Detection pipeline**: Automated and human review processes continuously identify new threats and quickly update safeguards.
-

Limitations

Operator is currently in an early research preview, and while it's already capable of handling a wide range of tasks, it's still learning, evolving and may make mistakes.

For instance, it currently encounters challenges with complex interfaces like creating slideshows or managing calendars. Early user feedback will play a vital role in enhancing its accuracy, reliability, and safety, helping us make Operator better for everyone.

What's next

CUA in the API: OpenAI plan to expose the model powering Operator, CUA, in the API soon so that developers can use it to build their own computer-using agents.

Enhanced Capabilities: OpenAI will continue to improve Operator's ability to handle longer and more complex workflows.

Wider Access: OpenAI plan to expand Operator to Plus, Team, and Enterprise users and integrate its capabilities directly into ChatGPT in the future once they are confident in its safety and usability at scale, unlocking seamless real-time and asynchronous task execution.