

Single-Step

Phone



User instruction:
"The initial April 1 view shows no data. Navigating back reveals the dashboard with zero activity and a timeline graph."

200 Samples

Computer



User instruction:
"The first screenshot shows a cluttered Windows desktop. In the second, the Zoom client window is open with options like Join, Start, and Schedule, indicating the app was launched."

Multi-Step

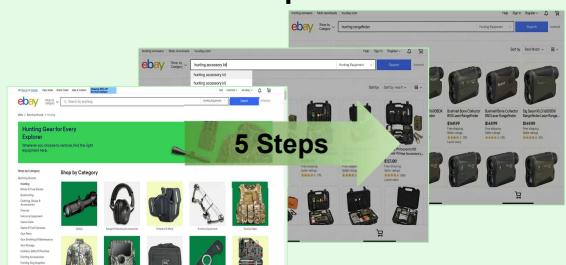
Phone



User instruction:
"What's on the menu at Papa Murphy's?"

200 Samples

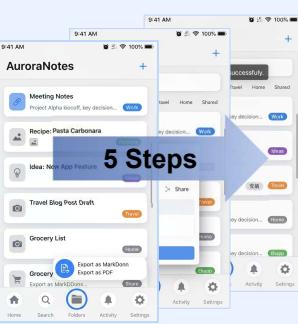
Computer



User instruction:
"How do I create an account?"

Fictional App

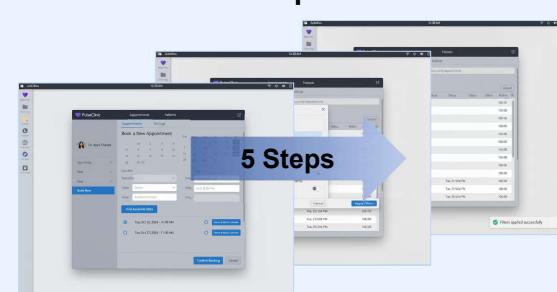
Phone



User instruction:
"Create a recurring bill reminder with notifications."

100 Samples

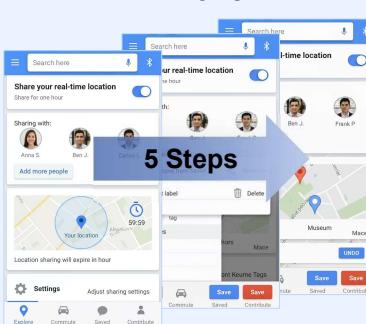
Computer



User instruction:
"Book an appointment and add it to the calendar."

Real App

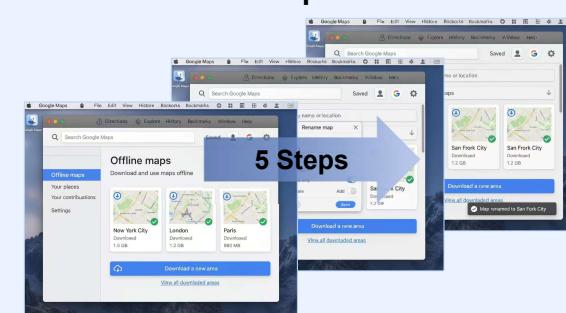
Phone



User instruction:
"Share your location for one hour."

100 Samples

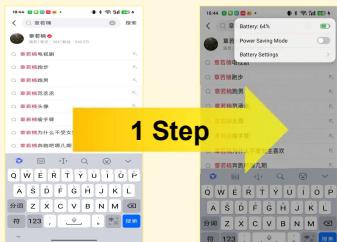
Computer



User instruction:
"Download an area for offline use."

Grounding

Phone



User instruction:
"Given a grounding point [938, 61], the model predicts the subsequent frame transition after a click is performed at that point."

100 Samples

Phone



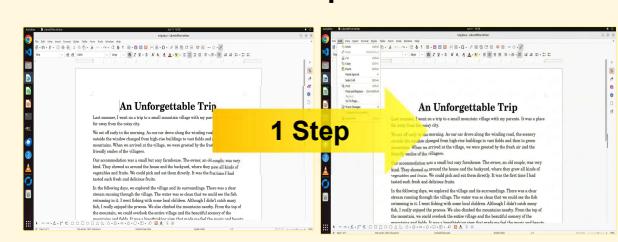
User instruction:
"Given a grounding point [866, 80], the model predicts the subsequent frame transition after a click is performed at that point."

Computer



User instruction:
"Given a grounding point [940, 40], the model predicts the subsequent frame transition after a click is performed at that point."

Computer



User instruction:
"Given a grounding point [119, 78], the model predicts the subsequent frame transition after a click is performed at that point."