
Supplementary Materials for On the Effects of Data Scale on Computer Control Agents

1 The paper and the Appendix of the paper provide details on the dataset (data format, statistics, usage,
2 etc.). For completeness, in the following we include a datasheet based on the format of [1].

3 1 Datasheet

4 1.1 The ANDROIDCONTROL dataset

5 1.1.1 Motivation

- 6 • For what purpose was the dataset created? Was there a specific task in mind? Was there a
7 specific gap that needed to be filled?
 - 8 – The dataset was created to support research in autonomous UI agents that interpret
9 human natural language instructions and perform actions directly on a device UI. Each
10 ANDROIDCONTROL task instance includes both high and low-level human-generated
11 instructions, allowing researchers to explore the level of task complexity an agent can
12 handle. Moreover, ANDROIDCONTROL is the most diverse computer control dataset
13 to date, that facilitates in-depth analysis of the model performance in and out of the
14 domain of the training data.
- 15 • Who created the dataset (e.g., which team, research group) and on behalf of which entity
16 (e.g., company, institution, organization)?
 - 17 – The dataset was created by Wei Li, William Bishop, Alice Li, Chris Rawles, Folawiyo
18 Campbell-Ajala, Divya Tyamagundlu, and Oriana Riva at Google.

19 1.1.2 Composition

- 20 • What do the instances that comprise the dataset represent (e.g., documents, photos, people,
21 countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people
22 and interactions between them; nodes and edges)?
 - 23 – The dataset contains episodes of human demonstrations for mobile device control. It
24 consists of natural language instructions at both the task level and step level, observa-
25 tions (screenshots + UI trees), and actions performed on Android devices. An example
26 instance is shown in Figure 1.
- 27 • How many instances are there in total (of each type, if appropriate)?
 - 28 – The dataset comprises 15,283 episodes spanning 833 apps.
- 29 • Does the dataset contain all possible instances or is it a sample (not necessarily random) of
30 instances from a larger set?
 - 31 – The dataset is a sample drawn from interactions across hundreds of Android apps. It
32 does not cover all possible interactions, but it is representative of many use cases, as
33 the instances were collected from diverse sources and varying conditions to emulate a
34 real system’s experience.

- 35 ● What data does each instance consist of?
 - 36 – Each instance consists of a goal instruction in natural language, a sequence of
 - 37 observation-action pairs describing the execution of the task. Observations consist of
 - 38 full-resolution screenshots along with UI trees extracted from accessibility view hier-
 - 39 archies. Actions are described by: *type* (click, long_press, input_text, scroll,
 - 40 navigate_back, navigate_home, open_app, wait), *target UI element* (only for
 - 41 click, input_text and long_press), and action arguments (text for input_text,
 - 42 app_name for open_app, direction for scroll).
- 43 ● Is there a label or target associated with each instance?
 - 44 – Yes, each episode has a natural language instruction associated.
- 45 ● Is any information missing from individual instances?
 - 46 – No.
- 47 ● Are there recommended data splits (e.g., training, development/validation, testing)?
 - 48 – Yes, there are recommended data splits of training, validation, and testing. The test
 - 49 split contains four sub-splits: 1) *in domain data (IDD)*: randomly pulled episodes from
 - 50 the same distribution as the training data; 2) *unseen-app*: a test split using apps not
 - 51 present in the train split; 3) *unseen-task*: a test split with tasks not present in the train
 - 52 split; and 4) *unseen-category*: a test split with apps from categories not present in the
 - 53 train split. Note that the test splits may contain overlapping episodes.
- 54 ● Are there any errors, sources of noise, or redundancies in the dataset?
 - 55 – We have removed episodes that contained errors and annotators marked as failed. We
 - 56 are not aware of any errors in the released dataset.
- 57 ● Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g.,
- 58 websites, tweets, other datasets)?
 - 59 – It is self-contained.
- 60 ● Does the dataset contain data that might be considered confidential?
 - 61 – No.
- 62 ● Does the dataset contain data that, if viewed directly, might be offensive, insulting, threaten-
- 63 ing, or might otherwise cause anxiety?
 - 64 – No, the dataset does not contain anything offensive, insulting, threatening, or that might
 - 65 cause anxiety.

66 1.1.3 Collection process

- 67 ● How was the data associated with each instance acquired?
 - 68 – The data associated with each instance was acquired by using a physical Android phone
 - 69 (Google Pixel with Android 8.0 or higher) installed with a companion Android app
 - 70 that in turn connects to a web app running on a desktop Chrome browser. Annotators
 - 71 control the phone through the web app, using the WebUSB protocol and Android Debug
 - 72 Bridge (ADB). The web app provides annotators with controls to perform actions on
 - 73 the phone and observe their outcome. An annotator select UI actions to perform on the
 - 74 phone. Before executing each action, the annotator is asked to type in a short natural
 - 75 language description of the action they are about to take.
- 76 ● What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or
- 77 sensor, manual human curation, software program, software API)?
 - 78 – The mechanisms used to collect the data include a hardware apparatus (Android
 - 79 device), manual human curation (annotators collected interaction traces), and a software
 - 80 program (for executing the tasks and capturing data).
- 81 ● Over what timeframe was the data collected?
 - 82 – The data was collected from April 2023 through April 2024.
- 83 ● Does the dataset relate to people?
 - 84 – The dataset does not directly relate to people, however annotators collected the data. No
 - 85 personal information was included in the task specification or demonstration collection.

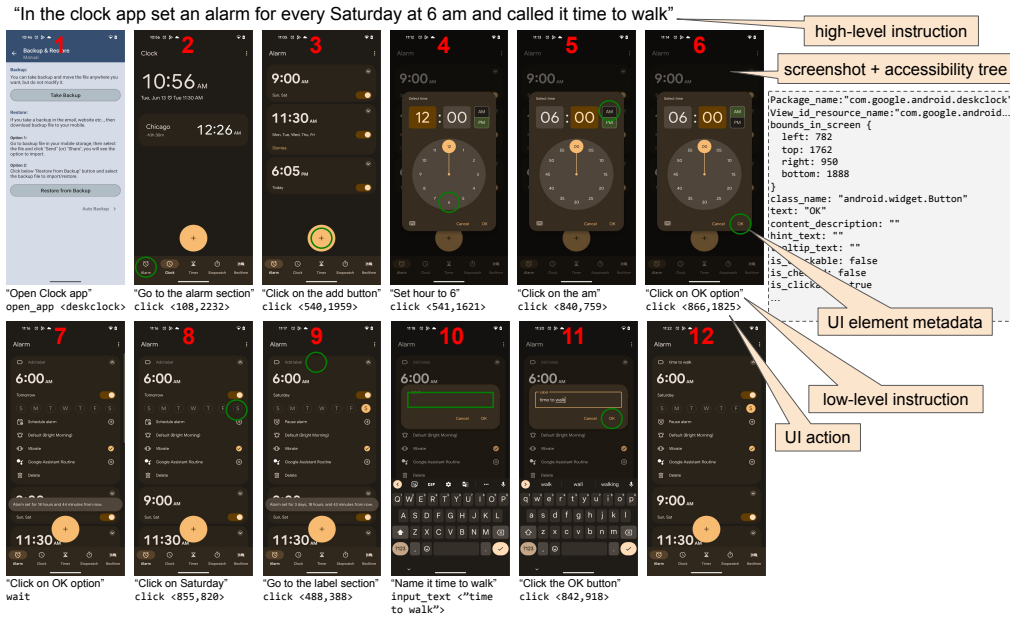


Figure 1: An example task demonstration contained in ANDROIDCONTROL. Green circles/rectangles highlight the action on the screen. Red numbers are added only for illustration purposes.

86 1.1.4 Preprocessing

- 87
- 88 • Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucket-
 - 89 ing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?
 - 90 – Preprocessing of the data was done. Scroll actions were processed to be reduced to
 - 91 four scrolling directions (left, right, up, and down) using information of start position,
 - 92 end position, and speed.
 - 93 • Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to
 - 94 support unanticipated future uses)?
 - 95 – Raw, full-resolution screenshots are provided. We do not provide raw logged events
 - 96 and discarded episodes.
 - 97 • Is the software used to preprocess/clean/label the instances available?
 - 98 – The data preprocessing and cleaning was done using Python. We do not release this
 - 99 part of the code.

100 1.1.5 Use cases

- 101
- 102 • Has the dataset been used for any tasks already?
 - 103 – Yes, we have trained and evaluated several models on this dataset for device automation. These models can serve as baselines for future research.
 - 104 • Is there a repository that links to any or all papers or systems that use the dataset?
 - 105 – Yes, at [https://github.com/google-research/google-research/tree/](https://github.com/google-research/google-research/tree/master/android_control)
 - 106 [master/android_control](https://github.com/google-research/google-research/tree/master/android_control).
 - 107 • What (other) tasks could the dataset be used for?
 - 108 – It is currently used for device automation. Future use cases could include screen
 - 109 understanding, screen generation, question answering, image captioning, and activity
 - 110 recognition.
 - 111 • Is there anything about the composition of the dataset or the way it was collected and
 - 112 preprocessed/cleaned/labeled that might impact future uses?

113 – The set of app categories in ANDROIDCONTROL is still an incomplete representation
114 of all tasks users may ask agents to perform.

115 **1.1.6 Distribution**

- 116 • Will the dataset be distributed to third parties outside of the entity (e.g., company, institution,
117 organization) on behalf of which the dataset was created?
 - 118 – The dataset is free to all under the condition that the dataset is used for non-commercial
119 purposes only.
- 120 • How will the dataset be distributed (e.g., tarball on website, API, GitHub)?
 - 121 – The dataset is available at [https://github.com/google-research/
122 google-research/tree/master/android_control](https://github.com/google-research/google-research/tree/master/android_control).
- 123 • Will the dataset be distributed under a copyright or other intellectual property (IP) license,
124 and/or under applicable terms of use (ToU)?
 - 125 – The dataset is released under the Apache License, Version 2.0.¹
- 126 • Have any third parties imposed IP-based or other restrictions on the data associated with the
127 instances?
 - 128 – N/A.

129 **1.1.7 Maintenance**

- 130 • Who is supporting/hosting/maintaining the dataset?
 - 131 – The dataset is maintained by the authors.
- 132 • How can the owner/curator/manager of the dataset be contacted?
 - 133 – The contact information of the authors can be found at the beginning of the main paper
134 and will be provided in the data repository.

135 **1.1.8 More examples**

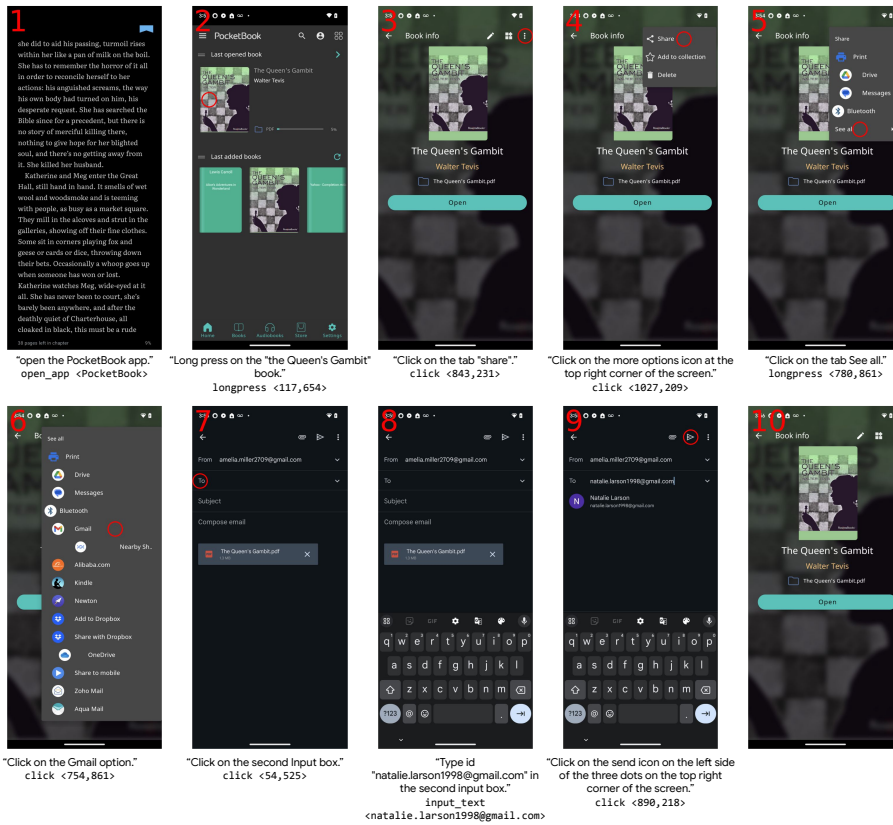
136 Examples of episodes from ANDROIDCONTROL are shown in Figure 2.

137 **References**

- 138 [1] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach,
139 Hal Daumé III au2, and Kate Crawford. Datasheets for datasets, 2021.

¹<http://www.apache.org/licenses/LICENSE-2.0>

Share my favorite Book "the Queen's Gambit" to my Friend Natalie larson over her gmail address -natalie.larson1998@gmail.com from the PocketBook app.



Launch the Amazon app, search for India gate basmati rice, and add 5 Kg of it to your shopping cart.

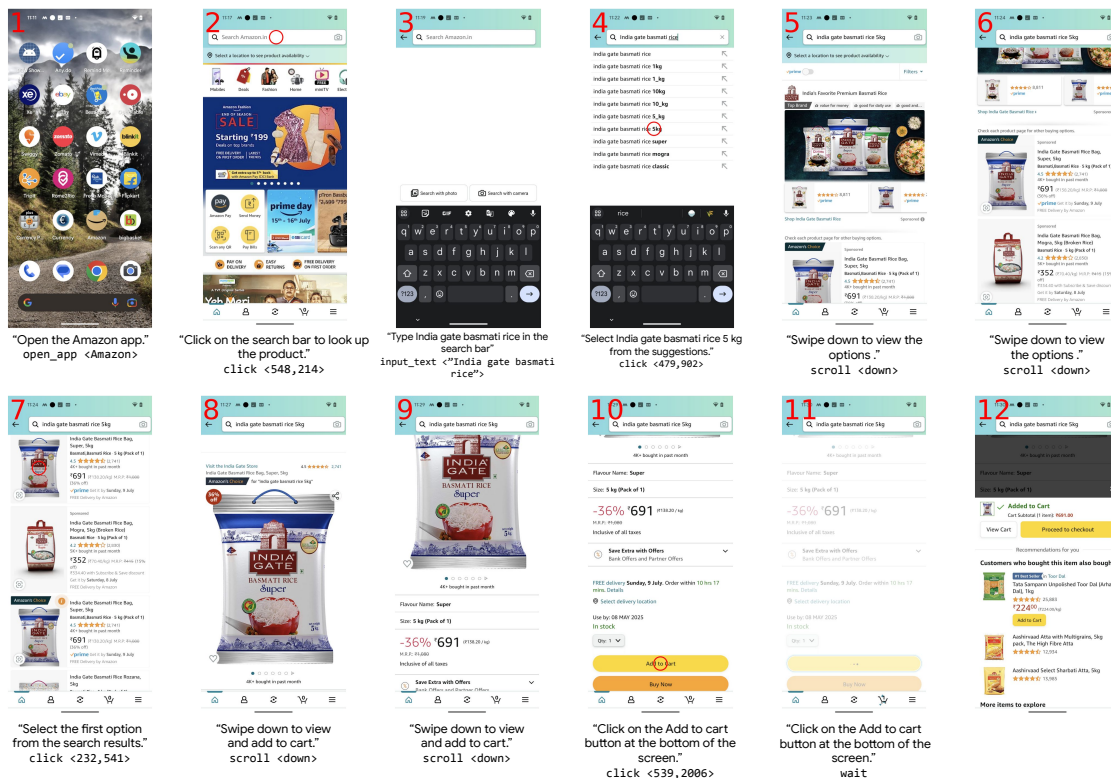


Figure 2: Example episodes contained in ANDROIDCONTROL. Red circles highlight the location of click and long press actions on the screen. Red numbers are added only for illustration purposes.