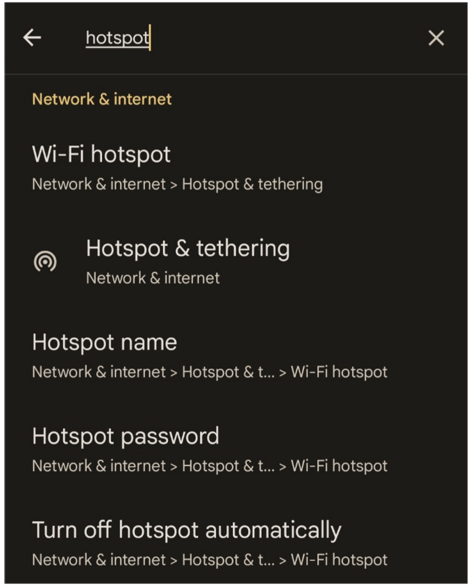
Android settings search with LLM

Research intern: Maozheng Zhao

Background: Android settings search

Search by keywords



hotspot

Network & internet

Wi-Fi hotspot
Network & internet > Hotspot & tethering

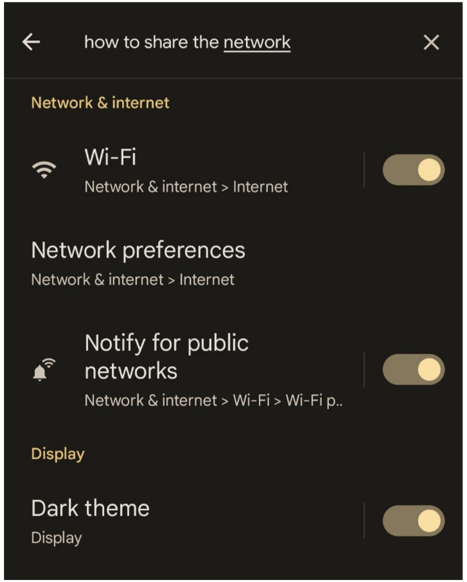
Hotspot & tethering
Network & internet

Hotspot name
Network & internet > Hotspot & t... > Wi-Fi hotspot

Hotspot password
Network & internet > Hotspot & t... > Wi-Fi hotspot

Turn off hotspot automatically
Network & internet > Hotspot & t... > Wi-Fi hotspot

Search by natural language



how to share the network

Network & internet

Wi-Fi
Network & internet > Internet

Network preferences
Network & internet > Internet

Notify for public networks
Network & internet > Wi-Fi > Wi-Fi p..

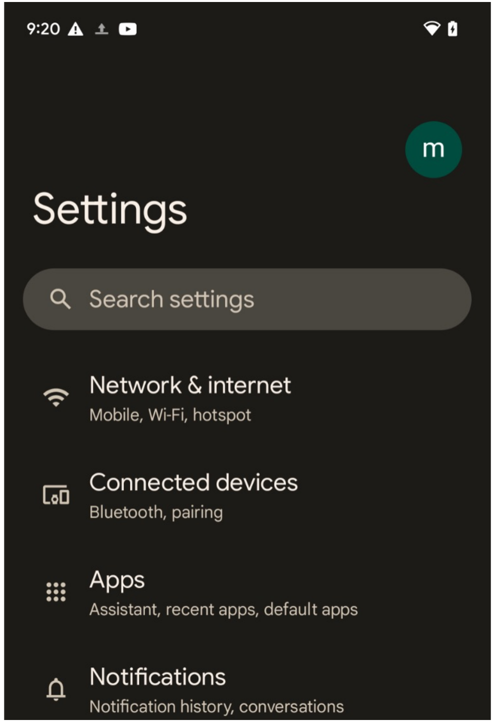
Display

Dark theme
Display

Current search:

1. Only works well when keywords appears
2. Falls short of search by natural language

Settings search with large language model



9:20

m

Settings

Search settings

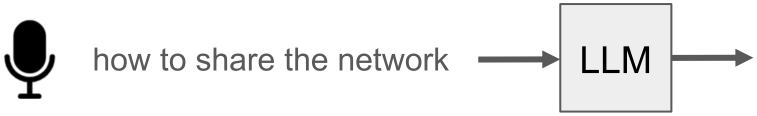
Network & internet
Mobile, Wi-Fi, hotspot

Connected devices
Bluetooth, pairing

Apps
Assistant, recent apps, default apps

Notifications
Notification history, conversations

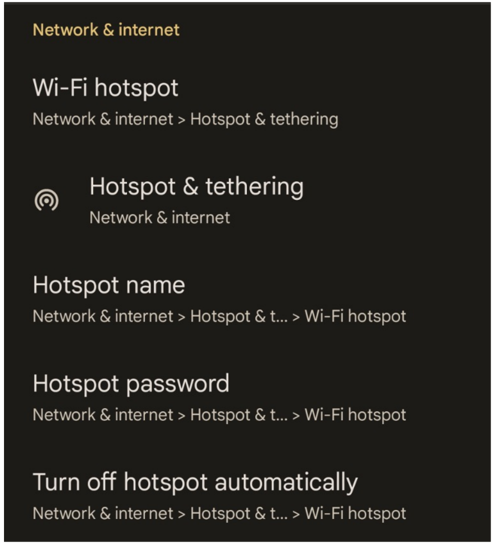
A natural language query.



how to share the network

LLM

Correct setting item.



Network & internet

Wi-Fi hotspot
Network & internet > Hotspot & tethering

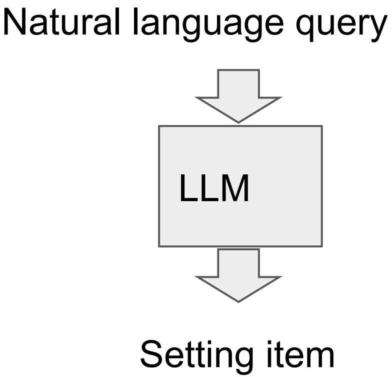
Hotspot & tethering
Network & internet

Hotspot name
Network & internet > Hotspot & t... > Wi-Fi hotspot

Hotspot password
Network & internet > Hotspot & t... > Wi-Fi hotspot

Turn off hotspot automatically
Network & internet > Hotspot & t... > Wi-Fi hotspot

Goal: fine-tune the LLM

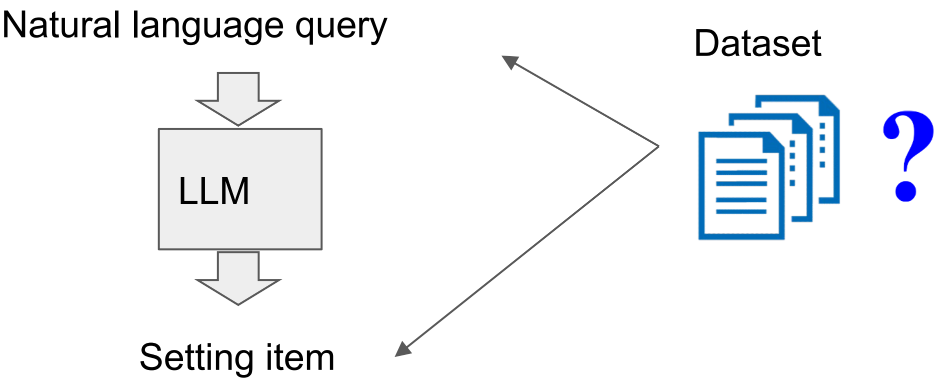


Natural language query

LLM

Setting item

Goal: fine-tune the LLM



Natural language query

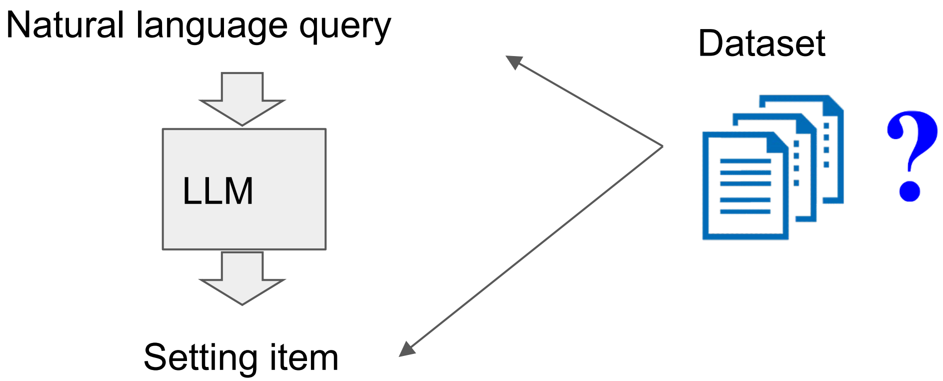
LLM

Setting item

Dataset

?

Goal: fine-tune the LLM



Natural language query

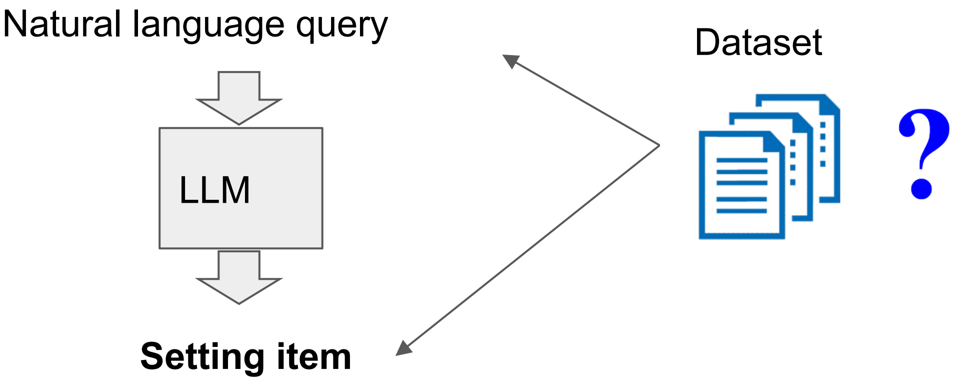
LLM

Setting item

Dataset

?

Goal: fine-tune the LLM



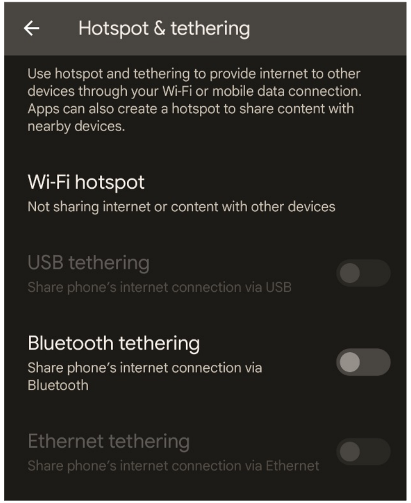
Natural language query

LLM

Setting item

Dataset

?



Hotspot & tethering

Use hotspot and tethering to provide internet to other devices through your Wi-Fi or mobile data connection. Apps can also create a hotspot to share content with nearby devices.

Wi-Fi hotspot
Not sharing internet or content with other devices

USB tethering
Share phone's internet connection via USB

Bluetooth tethering
Share phone's internet connection via Bluetooth

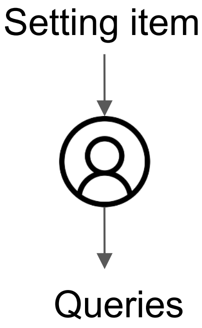
Ethernet tethering
Share phone's internet connection via Ethernet

From settings intelligence team

Item	**Description (may be empty)**	**Key words (may be empty)**	**Category**
bluetooth tethering	share phone's internet connection via bluetooth	usb tether, bluetooth tether, wifi hotspot	Hotspot & tethering
screen lock		slide to unlock, password, pattern, PIN	Security
tap to check phone	to check time, notifications, and other info, tap your screen.	gesture moves	Tap to check phone

1205 items in total

How to get natural language query? ?

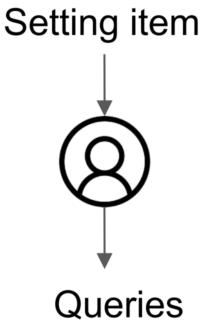


Setting item

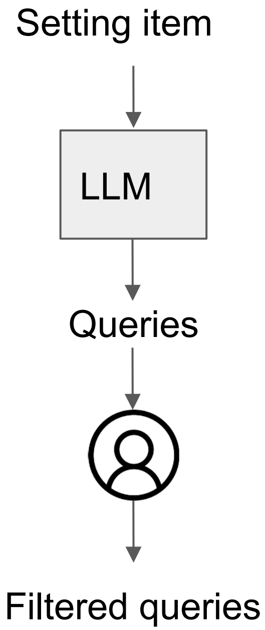
Queries

1. From real users
    a. expensive
    b. time consuming
    c. privacy
    d. Imagined queries not be real queries

How to get natural language query? ?



1. From real users
   a. expensive
   b. time consuming
   c. privacy
   d. Imagined queries not be real queries



2. From language model
   a. Inexpensive
   b. Fast
   c. Need human filtering

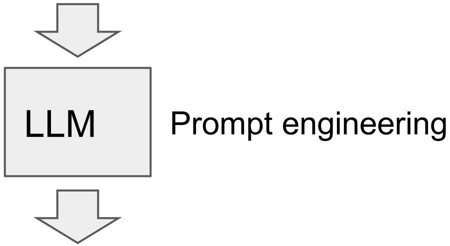
Query generation by LLM

Data form:

Item	Description	Key words	Category
bluetooth tethering	share phone's internet connection via bluetooth	usb tether, bluetooth tether, wifi hotspot	***Hotspot & tethering***

Sentence:

'bluetooth tethering' (usb tether, bluetooth tether, wifi hotspot) means ‘share phone's internet connection via bluetooth’. It's under the title '***Hotspot & tethering***'.



LLM

Prompt engineering

Query:

How to share my network by bluetooth?

Data generation by LLM

Prompt input for the LLM:

Given an Android setting item, please return a possible user's query that can be resolved by this setting item.

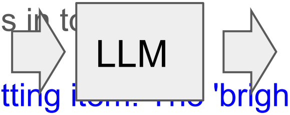
The Android setting item: 'time' It's under the title 'Date & time'.

A possible query: Add New york time.

The Android setting item: The 'hotspot tethering' (usb tether bluetooth tether wifi hotspot) It's under the title 'Network & internet'.

A possible query: how to share my network.

… (17 examples in to



The Android setting item: The 'brightness level' (dim screen, touchscreen, battery bright) It's under the title 'Display'.

A possible query: Make the screen dimmer.

1. For each item, we run the LLM multiple times, it will generate different queries.
2. Run this for all items.
3. Temperature is from 0.5 to 1 with step size 0.1.
4. Generated 30 queries for each of the top 30 items. 900 queries in total.

Data filtering

Items	Commands by LLM	Label by the developer
location	turn on my GPS	1
pair new device	How can I pair my phone with my new bluetooth headset?	1
usb debugging	enable the developer mode on my phone	1
swipe fingerprint	how to unlock phone with fingerprint	0
usb tethering	how do i start the hotspot on my phone?	0

1 : make sense

0: does not make sense

756/900 (84%) were labeled as 1.

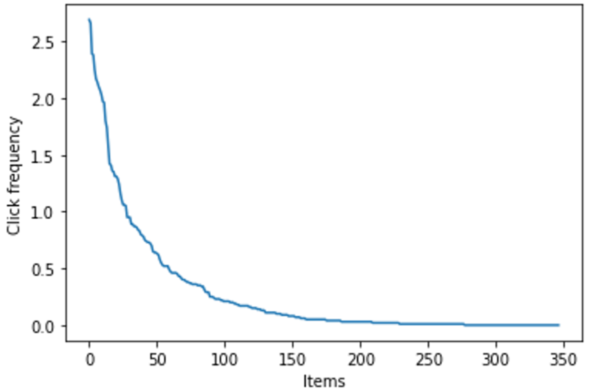
Main reasons for 0:

1. The description of that item is empty. LLM incorrectly guess the function of the item from its name.
2. Not specific enough for that item.

For commands with 1: Diversity of commands can be improved by using higher temperatures.

Data filtering

Top 30 items occupy 51% of the clicks.



Second time of data generation.

1. Added descriptions for 3 (out of 30) items.
2. Generated more data to make sure each item has 30 labeled data.

   In this pass, 96% (348/363) generated examples are labeled as 1.

Dataset

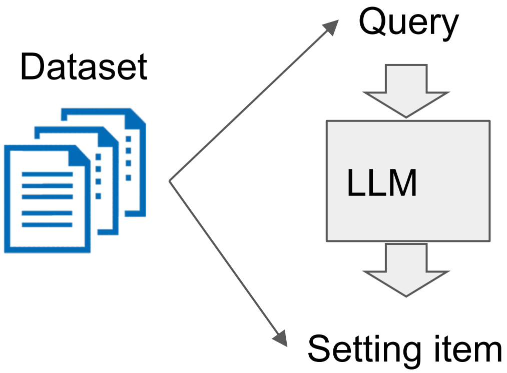
900 labeled examples in total = 30 items * 30 labeled examples for each item

Dataset split:

For each item, 20 for training, 5 for validation, 5 for testing.

The training/validation/testing set has 600/150/150 examples.

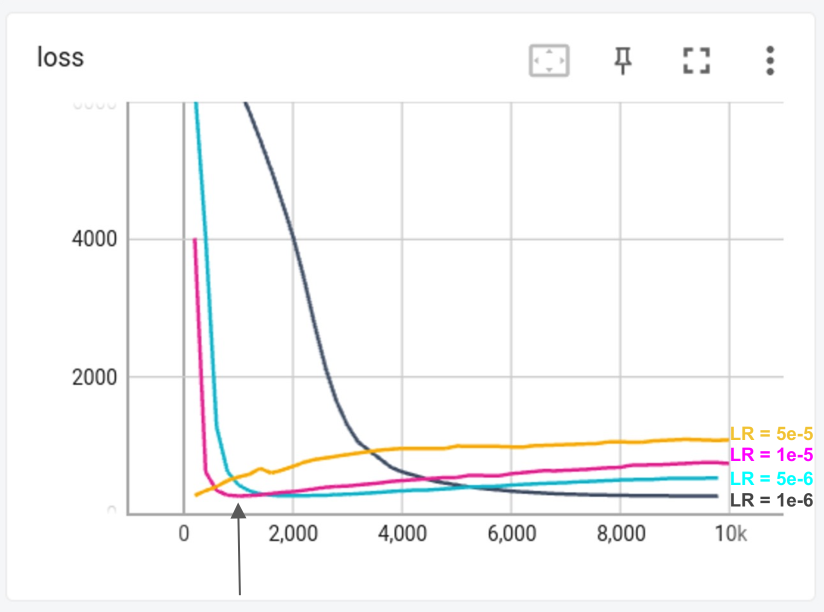
Model fine-tuning



Prepared the dataset by SeqIO.
LaMDA model: 8 Billion and 1 Billion
Learning rate: 1e-6, 5e-6, 1e-5,5e-5

8B LaMDA model fine-tuned with different learning rate

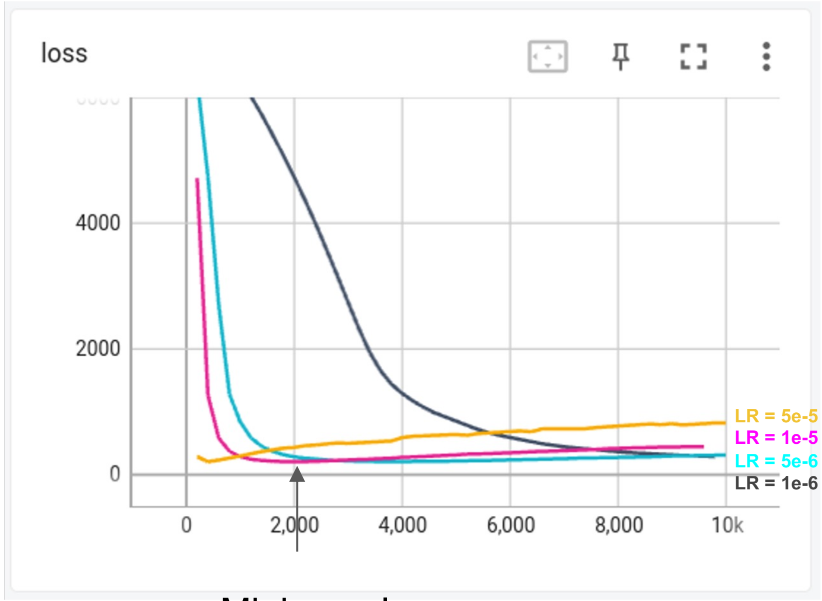
Loss on the test set



Minimum check point:
checkpoint 1000 of learning rate 1e-5

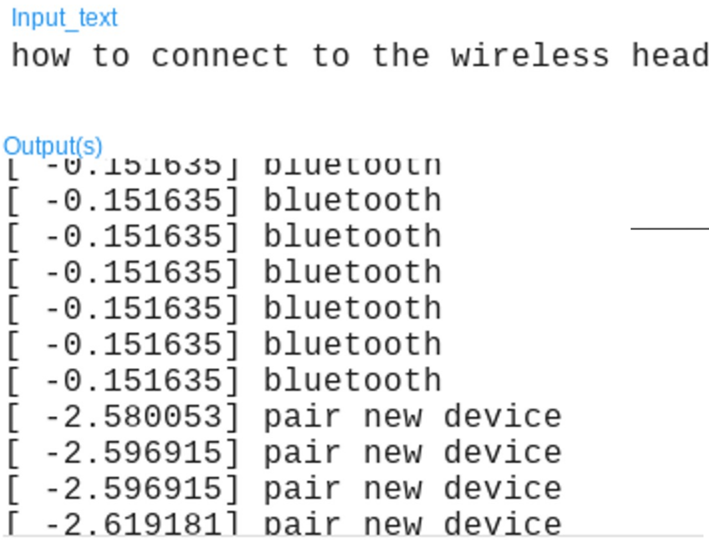
1B LaMDA model fine-tuned with different learning rate

Loss on the test set



Minimum loss:
checkpoint 2000 of learning rate 1e-5

Model evaluation



Input_text

```
how to connect to the wireless headphone
```

Output(s)

```
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -0.151635] bluetooth
[ -2.580053] pair new device
[ -2.596915] pair new device
[ -2.596915] pair new device
[ -2.619181] pair new device
```

Top 5 results:
('bluetooth', 92),
('pair new device', 16),
('wifi', 9),
('wireless headphone', 6),
('connect', 2),

Ground truth: bluetooth

LLM output 128 results. Most of them are repeating. The repeating time is used as the likelihood.

Top k accuracy: If one of the top k result is the ground truth, it’s correct.

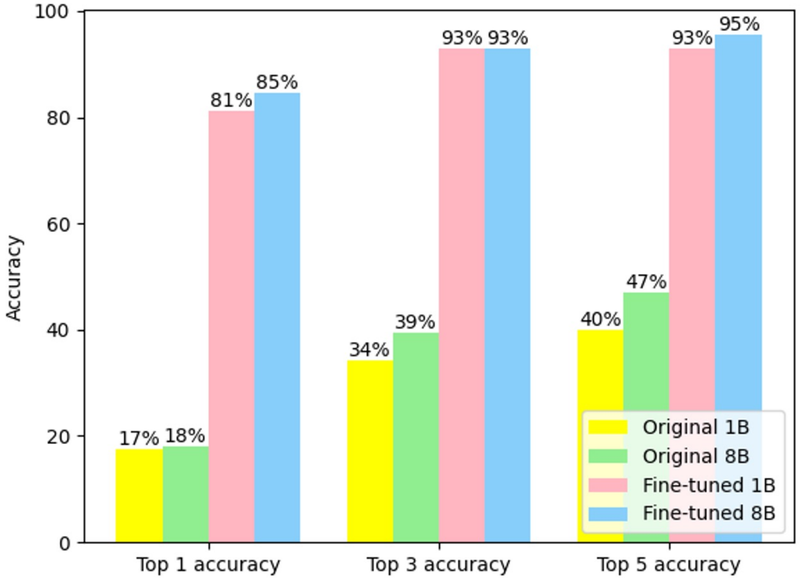
Results on the generated test set

LaMDA models

temperature =1

150 testing queries for 30 items

Used the same few-shot prompting for all models.



The higher the better.
Fined-tuned models outperforms the original models.

Demo of fine-tuned 8B LaMDA for settings search

Web demo page

Examples:

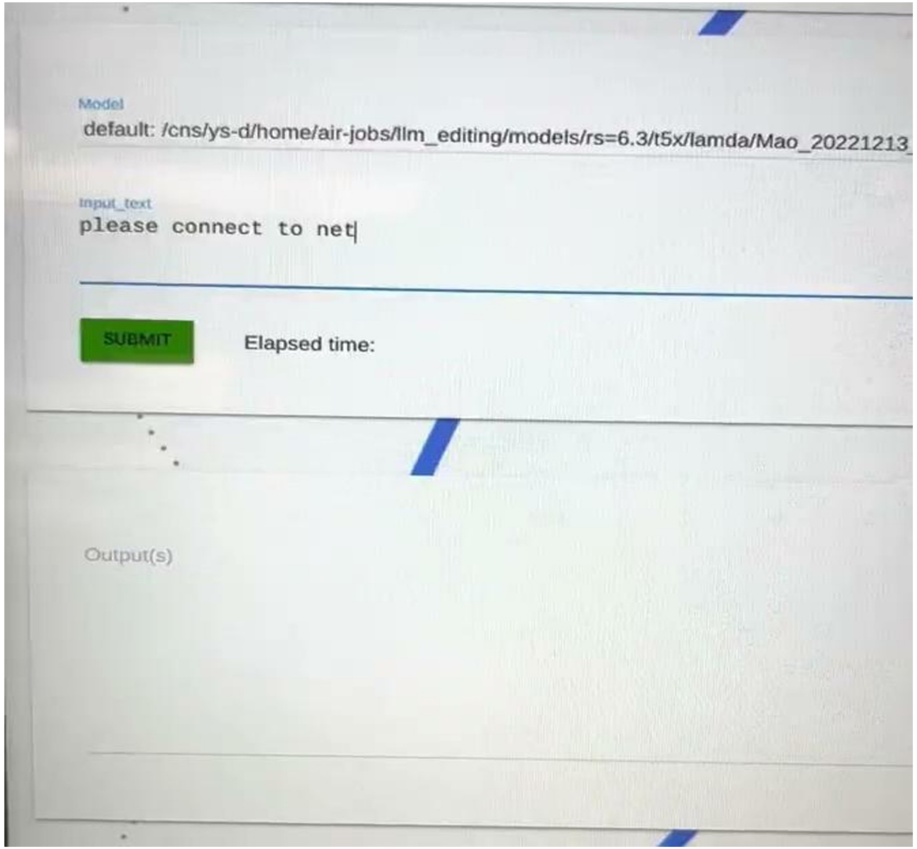
Please connect to network

How to share my networks?

Reset the phone

Add a fingerprint

Change time to eastern time



Model
default: /cns/ys-d/home/air-jobs/llm_editing/models/rs=6.3/t5x/lamda/Mao_20221213_

Input_text
please connect to net

Example results

Success examples:

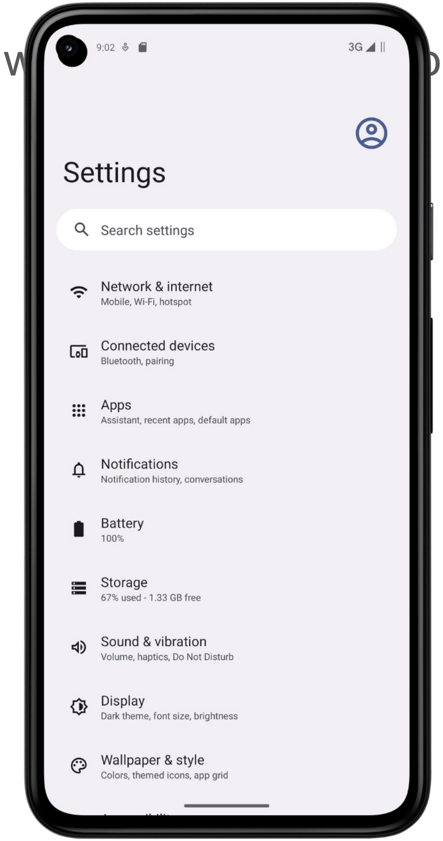
	Ground truth	LLM answer
Please connect to network	wifi	wifi
How to share my networks?	hotspot & tethering	hotspot & tethering
Reset the phone	erase all data (factory reset)	erase all data (factory reset)

Failed examples:

please stop changing the orientation of the phone.	screen lock	use auto-rotate

Dark theme	Display

What w oogle assistant to m ange?



9:02 3G

Settings

Search settings

Network & internet
Mobile, Wi-Fi, hotspot

Connected devices
Bluetooth, pairing

Apps
Assistant, recent apps, default apps

Notifications
Notification history, conversations

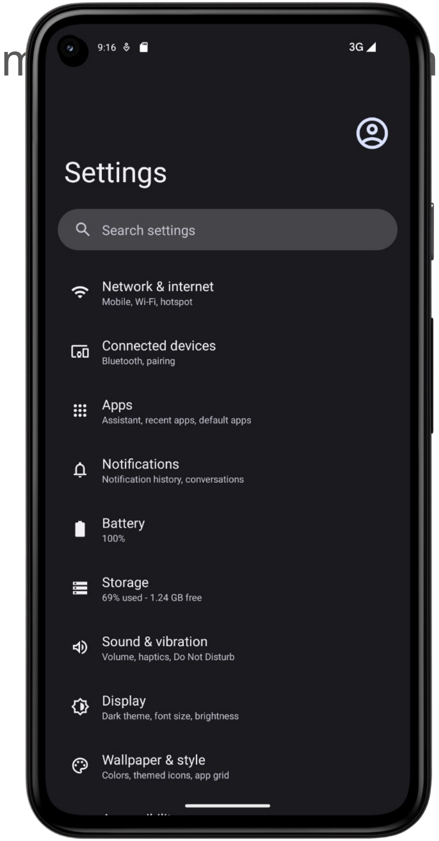
Battery
100%

Storage
67% used - 1.33 GB free

Sound & vibration
Volume, haptics, Do Not Disturb

Display
Dark theme, font size, brightness

Wallpaper & style
Colors, themed icons, app grid



9:16 3G

Settings

Search settings

Network & internet
Mobile, Wi-Fi, hotspot

Connected devices
Bluetooth, pairing

Apps
Assistant, recent apps, default apps

Notifications
Notification history, conversations

Battery
100%

Storage
69% used - 1.24 GB free

Sound & vibration
Volume, haptics, Do Not Disturb

Display
Dark theme, font size, brightness

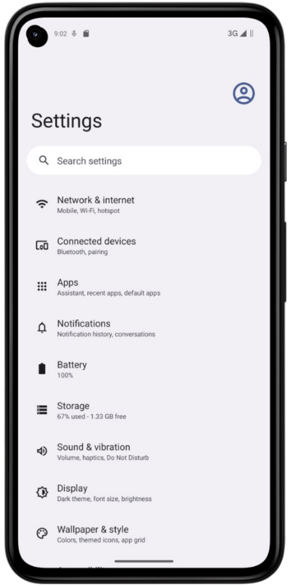
Wallpaper & style
Colors, themed icons, app grid

Elicitation study for data collection

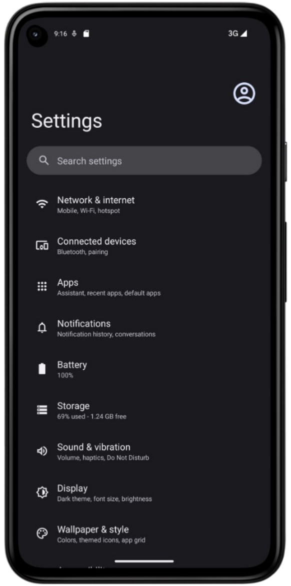
Design rationale:

- Avoiding verbal biases
- Showing pre/post visual effects

Before the setting



After the setting



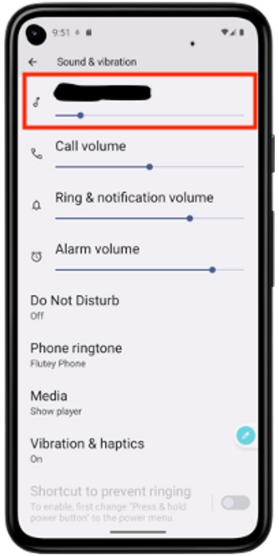
Dark theme <- Display

All images.

Elicitation study for data collection

- Differences are highlighted with red boxes.
- Keywords are covered.

Before the setting



9:51

Sound & vibration

[redacted]

Call volume

Ring & notification volume

Alarm volume

Do Not Disturb
Off

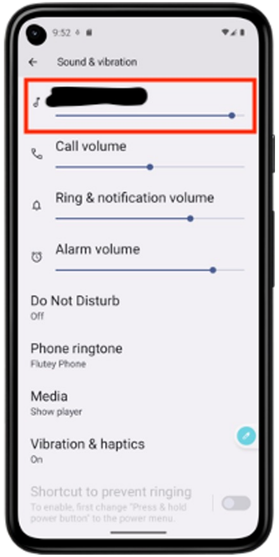
Phone ringtone
Flutey Phone

Media
Show player

Vibration & haptics
On

Shortcut to prevent ringing
To enable, first change "Press & hold power button" to the power menu.

After the setting



9:52

Sound & vibration

[redacted]

Call volume

Ring & notification volume

Alarm volume

Do Not Disturb
Off

Phone ringtone
Flutey Phone

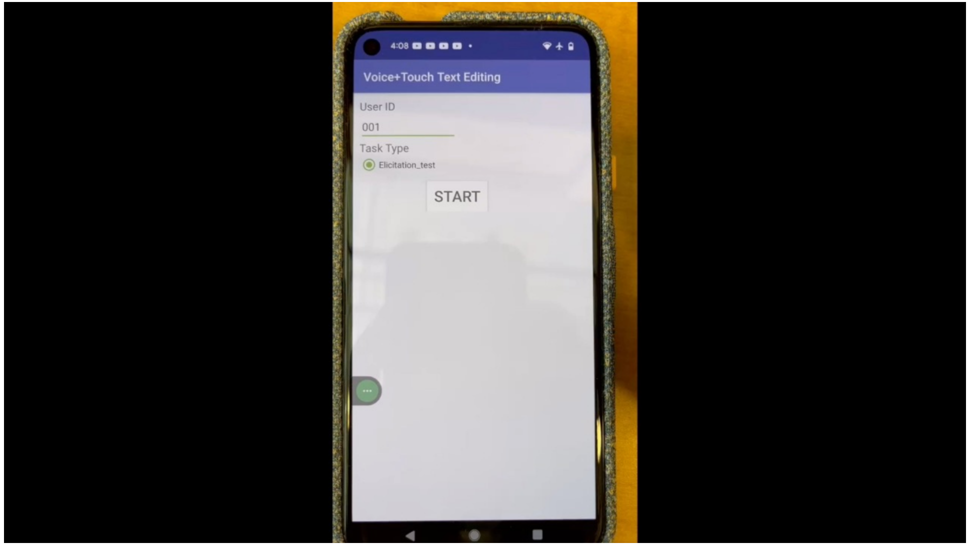
Media
Show player

Vibration & haptics
On

Shortcut to prevent ringing
To enable, first change "Press & hold power button" to the power menu.

Media volume <- Sound & vibration

Demo of the study APP



Voice+Touch Text Editing

User ID

001

Task Type

● Elicitation_test

31 figures in randomized order

5 for warm up.

26 for data collection.

Participants statistics

Number of participants: 20 (10 male, 10 female)

Age: 23.65 +/- 1.136 (average +/- standard deviation)

How often do you use Android phones or tablets?

At least once a day (13), ,At least once a week (3), ,At least once a month (2), Rarely or never (2)

How often do you use smartphones or tablets in general (regardless of whether it’s Android or not)?

At least once a day (20)

Familiarity with Android settings (1-5): 4.35 +/- 0.67 (average +/- standard deviation)

Data statistics

Effective queries: 520 (26*20)

Examples:

Dark theme <- Display:

change the background theme to dark

how to change my phone from light mode to dark mode

Media_volume <- sound & vibration:

increase the phone music volume

increase the sound level

Baseline methods

TF-IDF: (Term Frequency - Inverse Document Frequency) is a handy algorithm that uses the frequency of words to determine how relevant those words are to a given document.

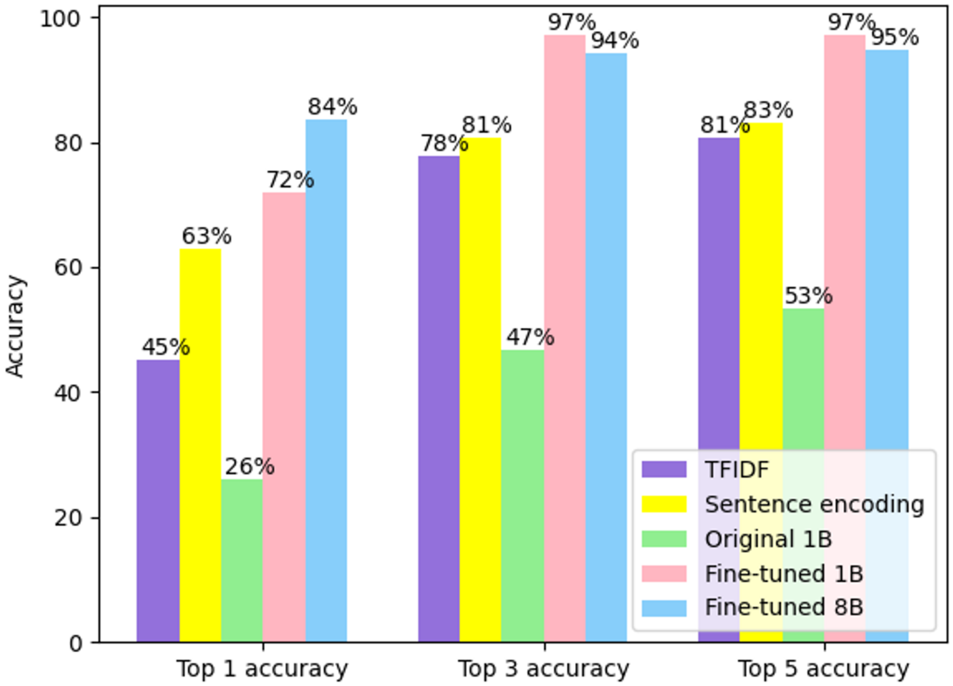
Sentence Encoder encodes text into high dimensional vectors that can be used for text classification, semantic similarity, clustering, and other natural language tasks.

`Sentence_transformers library`

Results on the human dataset

Training set (generated data):
30 items
600 queries

Testing set (human data):
8 items
135 queries



Summary

1. Fine-tuned model outperformed the pretrained model, TFIDF and sentence encoding.
2. Model trained with synthetic dataset still performs well on human data.

Thank you!