Deploying LLM on Mobile

Step 1:- Download Termux app APK

Download the relevant version of Termux apk on your phone. I used this version on the Android 8.0 phone.

Download Termux 0.101 Android APK File

Step 2:- Install Termux app APK

- Enable installation from unknown sources: Go to Settings, then Apps & notifications or Security, and toggle on Unknown sources
- Download the APK file: Use a trusted website to download the APK file
- Install the APK: Tap the downloaded APK file

Step 3: Grant Storage Permissions to Termux

1. Open Termux and grant storage permissions with the following command:

```
termux-setup-storage
```

2. Allow Permission: You'll see a prompt requesting permission to access your device's storage. Tap "Allow" to grant access.

3. Verify Access: After granting permission, Termux creates a `storage` directory in your home directory with symbolic links to various folders on your device:
- `storage/downloads` – Access to the Downloads folder.
- `storage/shared` – Access to shared storage, typically the main storage of the device.
- `storage/dcim` – Access to camera images.
- `storage/music`, `storage/pictures`, etc., for corresponding media folders.

With storage permissions in place, you'll be able to read and write files from the device's storage directly within Termux and the Debian environment within `proot-distro`.

Step 4: Update Termux and Install Required Packages

1. Update Termux packages:
- Run the following commands to ensure that Termux is up-to-date:

```
pkg update && pkg upgrade
```

2. Install proot-distro:

- `proot-distro` allows you to manage and run different Linux distributions within Termux using a PRoot environment.

```
pkg install proot-distro
```

Step 5: Install Debian

1. Install Debian:
   - Use the following command to install Debian:

```
proot-distro install debian
```

You can use pd instead of proot-distro to do the same tasks. Like using the command:

```
pd install debian
```

Instead of typing out the whole thing.

2. Launch Debian:

Once installed, you can launch Debian with:

```
proot-distro login debian
```

- This command will bring you into the Debian environment, allowing you to use it as if you're on a regular Debian system.

Step 6: Configure Debian

1. Update and Upgrade Debian Packages:
   - Update the Debian system inside proot-distro to get the latest packages:

```
apt update && apt upgrade
```

2. Install Essential Packages:
   - You may want to install some basic package. In this case, we will be installing 'wget'

```
apt install wget
```

3. To exit (Optional):
- To exit the Debian environment, simply type:

```
exit
```

Step 7: Install ollama

1. Get the install script:

   wget https://ollama.ai/install.sh

2. Run the install script:

   sh ./install.sh

Step 8: Start ollama service

- To start the ollama service in the background, simply type:

   ollama serve &

Step 9: Run a model

Type in the terminal:

   ollama run **<model_name>**

List of models: Tags · llama3.2

Freeing Ram (Optional):

To reduce RAM usage on an Android device via Developer Options, you can adjust a few settings that can help manage memory usage effectively. Here's how:

1. Limit Background Processes:
   - Go to Settings > Developer options.
   - Scroll down to Apps and look for Background process limit.
   - Set the limit to restrict the number of apps that can run in the background. Options include No background processes, 1 process at most, and up to 4 processes.
   - Choosing 1 or 2 processes can free up RAM, but it may slow down task-switching and affect app reload times.

2. Disable or Limit Animations:
   - In Developer options, scroll to Drawing.
   - Adjust Window animation scale, Transition animation scale, and Animator duration scale.
   - Setting these to 0.5x or Off reduces the RAM required for animations, which can be significant in some cases.

3. Don't Keep Activities:
   - There's an option called Don't keep activities in Developer options under Apps.
   - Enabling this option will automatically destroy an app's activity as soon as you leave it, freeing up RAM.

- This can save memory, but it will close apps when you switch between them, potentially leading to slower performance if you need to reopen them often.

4. Force GPU Rendering (Optional):
- Enable Force GPU rendering in Developer options.
- This can offload some of the work to the GPU, which may free up RAM on certain devices, especially in graphic-intensive applications.
- Note that this setting may impact battery life.

5. Limit System Apps (Advanced):
- Some system apps use background memory even when they're not active. Disabling unnecessary system apps (like bloatware) can reduce RAM usage.
- Go to Settings > Apps, select the system app, and Disable it if it's not essential.

These adjustments can help lower RAM utilization, especially on devices with limited memory.