Appendix - Building the App

Going into the details of the building the final mobile app was slightly outside the scope of this course. We have build this help guide for you to show the main steps and code samples you need to build the app we saw in the last lesson.

TensorFlow Lite

TensorFlow Lite is a runtime that enables on-device machine learning by helping developers run their models on mobile, embedded, and edge devices. The models produced by TensorFlow Lite work on multiple platform support, covering Android and iOS devices, embedded Linux, and microcontrollers. The toolchain also has diverse language support, which includes Java, Swift, Objective-C, C++, and Python.

TensorFlow Lite guide [+] (https://www.tensorflow.org/lite/guide)

Delegation

In the context of TensorFlow Lite, "delegation" refers to the use of delegates to enable hardware acceleration of machine learning models on mobile and edge devices. Delegates act as a bridge between TensorFlow Lite and on-device accelerators like GPUs and DSPs, optimizing performance an efficiency by leveraging specialized hardware capabilities. This process can significantly improve the speed and power consumption of running machine learning models on such devices. For more details, you can visit the TensorFlow Lite Delegates page.

TensorFlow Lite Delegates page [+] (https://ai.google.dev/edge/lite/performance/delegates)

Qualcomm QNN delegate

Qualcomm QNN delegate allows you to run models on the NPU.

Download Qualcomm QNN Delegate – (Zip 724 MB)
(https://softwarecenter.qualcomm.com/api/download/software/qualcomm_neural_processing_sdk/v2.2

End-to-end examples