Apple Foundation Models is out (in beta)

June 2025

Roger Lam

WWDC25 launched Foundation Models framework

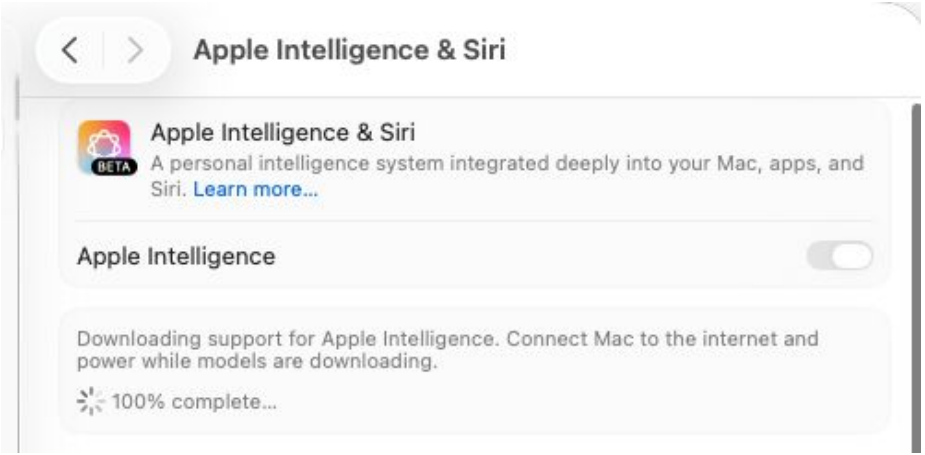
This means

- Private, on-device usage
- Offline usage
- Built into the Operating System

It might seem late but the level of polish and integration is well worth the wait.

Getting started was somewhat annoying

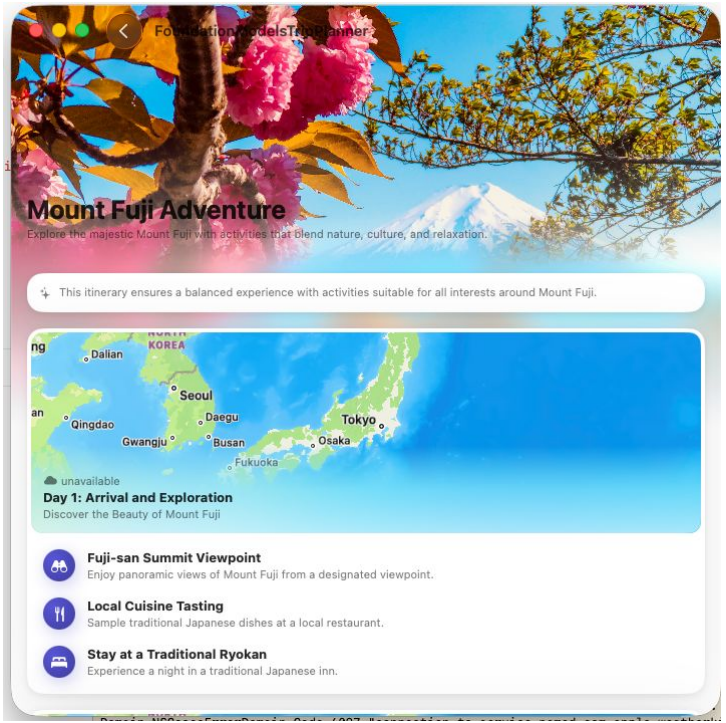
1. Have an M-series Mac
2. Update to OS26
3. Download Xcode-beta
4. Update Apple Intelligence
    - Apple intelligence was down June 18th and 19th



But the interactive demo app is amazing

Hard to show how smooth the generation is.

Imagine the impact of bringing local LLM usage to every app in the App Store



Video: [Code-along: Bring on-device AI to your app using the Foundation Models framework]

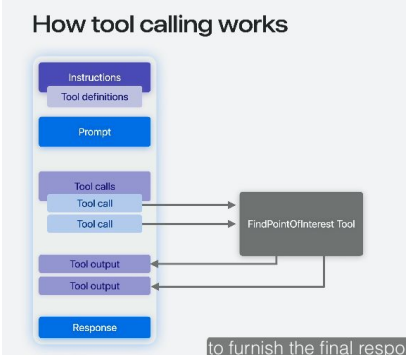
Deep library integration with Swift gives the best API

Swift is already an amazing language - type safe, modern syntax, compiled

Apple added new Swift data structures.

- “Generable” to make **typed LLM responses**, aka structured output.
- “PartiallyGenerated” type makes it easy to stream output.
- Tool calling is first class.
- Sessions for multi-turn interaction

Video: [Meet the Foundation Models framework]




```
// Composing Generable types

@Generable
struct Itinerary {
    var destination: String
    var days: Int
    var budget: Float
    var rating: Double
    var requiresVisa: Bool
    var activities: [String]
    var emergencyContact: Person
    var relatedItineraries: [Itinerary]
}
```

LLM is 3B parameter quantized to 2-bit

3B is device-sized while server models like GPT are 100’s of billions of parameters.

I’m sure it’s trained specifically to be useful in smaller contexts but anything more specific needs “Adaptors”, aka LoRA fine-tuning.

Adaptors each take ~160 MB in the app. Need 100 - 1000 samples for basic tasks, 5000+ to teach complex tasks.

IMO, a good balance of minimal base image and well-scoped fine tuning available

Updates to Apple’s On-Device and Server Foundation Language Models

Getting started with Foundation Models adaptor training

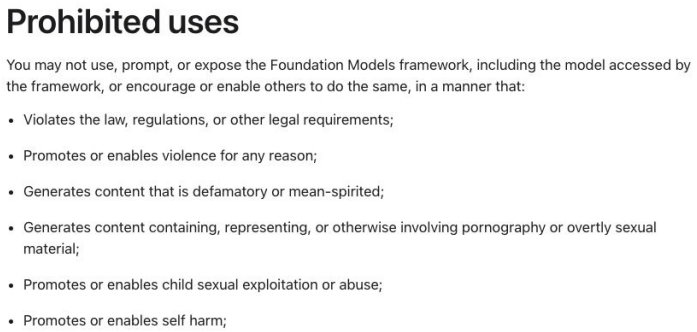
However, you don’t really control the main LLM

There are a few downsides.

Like when I wasn’t able to download the model. You’re relying on Apple and how Apple pushes updates to the model for reliability.

Each foundation model update requires retraining of Adaptors.

There are Acceptable use requirements for the Foundation Models framework.



Prohibited uses

You may not use, prompt, or expose the Foundation Models framework, including the model accessed by the framework, or encourage or enable others to do the same, in a manner that:

- Violates the law, regulations, or other legal requirements;
- Promotes or enables violence for any reason;
- Generates content that is defamatory or mean-spirited;
- Generates content containing, representing, or otherwise involving pornography or overtly sexual material;
- Promotes or enables child sexual exploitation or abuse;
- Promotes or enables self harm;

Follow for more

I’ll be playing around with Foundation Models more so stay tuned!

lamroger.com

https://mathstodon.xyz/@lamroger/

https://www.linkedin.com/in/lam-roger/