Foundation Models

Enhance features in your by using the on-device model at the core of Apple Intelligence.

Overview

When you want to apply intelligent capabilities to your apps, you can use the same on-device foundation models that power Apple Intelligence to build and improve your features. For example, turn a hard-coded search suggestion list into a generated list of suggestions that is personalized to the moment.

As you begin working with generative models and prompt engineering, it's important to keep design in mind. The HIG provides guidance and best practices to help you create apps that use generative models.

Define the data your app needs for precise output

To integrate generative technologies in your app, look to your app's existing features for ideas. If your app offers people a way to submit restaurant reviews, the model can use custom data types to convert someone's review into a scorecard that lets people visualize how positive the review was. Because you know the type of data your app wants, your custom data types help guide model output to fit your use case. Instead of writing parsing code, this allows you to think about the data your app needs to create a richer app experience.

And once you know the type of data your app needs, focus on writing prompts that produce better results. It takes time and practice to craft a good prompt, so try a variety of requests and test the output the model returns.

Create custom tools for your app

Tool calling allows a model to interact with the code you write to extend the model's capabilities. When you prompt the model with tools, the model can determine whether a tool you provide is available to help complete the request. For example, you could write code in your app that scans the person's calendar events for a dinner reservation, and populates a text to the invitees that includes the name of the restaurant, time of the reservation, and information for nearby parking. When the model encounters a prompt that requests dinner reservation information, it can call the code you write to get up to date information that it uses to complete the request.

Customize further with a custom adapter

For apps that contain tasks that need domain specialization, adapters provide a way to leverage your own training data. Adapters are small modules that you train to enhance — or adapt — the base model's ability to perform a specific task. You write adapters using the Foundation Models Adapter Training Toolkit.