

---

## WWDC, Apple Intelligence, Apple Aggregates AI (Stratechery Update 6-11-2024)

---

**Ben Thompson** <email@stratechery.com>  
Reply-To: Ben Thompson <email@stratechery.com>  
To: Scott Hannan <scott.hannan@gmail.com>

Tue, Jun 11, 2024 at 3:18 AM

[View in browser](#)



## WWDC, Apple Intelligence, Apple Aggregates AI

Tuesday, June 11, 2024

[Listen to this Update in your podcast player](#)

---

Good morning,

On this morning's episode of [Dithering](#) John Gruber and I gave our initial reactions to yesterday's WWDC keynote.

Tonight Gruber is hosting his annual live episode of The Talk Show in Cupertino; if you have an Apple Vision Pro you can watch it live! Sandwich Vision has released a new Vision Pro app called [Theater](#) and The Talk Show is set to be the first ever stereoscopic livestream on the Vision Pro. For more details check out [this article on 9to5Mac](#) and Gruber's post on [Daring Fireball](#).

On to the Update:

### WWDC

From the [Wall Street Journal](#):

Apple joined the AI arms race, saying Monday it plans to bring a more personalized version of artificial intelligence to its 2.2 billion device users—including striking a deal with ChatGPT-maker OpenAI. The new AI system, which it called “Apple Intelligence,” offered a preview to what many consider to be the holy grail of AI, a voice assistant empowered with enough personal user information to meaningfully help complete an array of tasks. Apple has partnered with OpenAI, and its ChatGPT, for some new AI functions, such as answering more complex queries or composing messages, capabilities that Apple's AI can't handle. The announcement comes after the iPhone maker saw its market value stagnate compared

with rivals that were quicker to incorporate generative artificial intelligence into their core products...

At its Worldwide Developers Conference, which will run throughout the week, Apple said its new software will retrieve information from across apps and scan personal information to help users proofread text, call up photographs of specific family members or gauge traffic patterns ahead of an atypical commute. Users can create images and emojis and even convert rough sketches into polished diagrams. An updated Siri will be able to better understand natural language, process contextual information and take action inside apps. Developers can use new tools to take advantage of the AI systems, a prospect that has excited some investors who have hoped that Apple could leverage its App Store to unveil new products. Most of these new Apple capabilities will be available later this year...

Privacy is at the heart of Apple's new AI capabilities, a feature that could further lock users into its ecosystem. Most processing will be done on a device instead of shipping to the servers in the cloud. But the company said for running larger AI models Apple will keep it private by running its own servers with what it calls Private Cloud Compute. It will only send data relevant to the task to these servers. The data isn't stored or accessible by Apple for further training, the company said.

This was a high stakes presentation, and not just for Apple given the scrutiny around the company's approach to AI; I am the one that took the chance of writing an Article before the event arguing that **Apple Intelligence is Right on Time!** I know I am being solipsistic, but fortunately I also think that I was proven right: I think that Apple crushed this event and completely validated my thesis.

First off, the overall organization of the keynote was mostly perfect: Apple compressed the usual WWDC keynote schedule, which is organized by their various operating systems, into the first hour of the keynote. This was relatively easy to do because there weren't that many new features to announce in any of the operating systems; this was actually an indication of how big the AI section was, because you definitely got the sense that Apple dropped almost everything they were working on over the last year to focus on Apple Intelligence.

That noted, there were bits and pieces of that first hour that touched on not just machine learning generally but generative AI specifically. Safari, for example, will generate summaries of web pages for you; the distinction, as Gruber explained to me on Dithering, is that that capability will be available no matter your device. Everything in the Apple Intelligence section, though, is only available on the iPhone 15 Pro and M-model Macs, which seems to validate my assumption from yesterday that devices need a minimum of 8GB of RAM to run the on-device models at the foundation of Apple Intelligence.

So what is Apple Intelligence, then? To me the explanation flows directly from Strategy 101: Apple Intelligence is the application of generative AI to use cases and content that Apple is uniquely positioned to provide and access. It is designed, to build on yesterday's Article, to maximize the advantages that Apple has in terms of being the operating system provider on your phone; and, on the other hand, what it is not is any sort of general purpose chatbot: that is where OpenAI comes in — and only there.

## Apple Intelligence

As for specifics, Apple's presentation was very organized:

- The capabilities of Apple Intelligence are language, image generation, actions, understanding personal context, and privacy.
- The infrastructure of Apple Intelligence is on-device processing using Apple-designed models (which, **according to Apple**, compare favorably to Microsoft's Phi models, the current gold-standard for small models), and cloud processing via Apple-owned datacenters running servers outfitted in some way with Apple Silicon. **Apple is promising** that the latter is designed in such a way that all requests are guaranteed to be private and disposed of immediately.
- These capabilities and infrastructure are exposed through various experiences, including an overhauled Siri that can take actions on apps (to the extent they support it); writing tools including rewrite, tone changes, and proofreading; summarization of things like emails and notifications; genemoji (i.e. generated emoji in the style of current emoji offerings); a system-level component called Image Playground that developers can incorporate into their apps; and new experiences in Notes and Photos.

The key part here is the “understanding personal context” bit: Apple Intelligence will know more about you than any other AI, because your phone knows more about you than any other device (and knows what you are looking at whenever you invoke Apple Intelligence); this, by extension, explains why the infrastructure and privacy parts are so important.

What this means is that Apple Intelligence is by-and-large focused on specific use cases where that knowledge is useful; that means the problem space that Apple Intelligence is trying to solve is constrained and grounded — both figuratively and **literally** — in areas where it is much less likely that the AI screws up. In other words, Apple is addressing a space that is very useful, that only they can address, and which also happens to be “safe” in terms of reputation risk. Honestly, it almost seems unfair — or, to put it another way, it speaks to what a massive advantage there is for a trusted platform. Apple gets to solve real problems in meaningful ways with low risk, and that's exactly what they are doing.

Contrast this to what OpenAI is trying to accomplish with its GPT models, or Google with Gemini, or Anthropic with Claude: those large language models

are trying to incorporate all of the available *public* knowledge to know everything; it's a dramatically larger and more difficult problem space, which is why they get stuff wrong. There is also a lot of stuff that they don't know because that information is locked away — like all of the information on an iPhone. That's not to say these models aren't useful: they are far more capable and knowledgeable than what Apple is trying to build for anything that does not rely on personal context; they are also all trying to achieve the same things.

### Apple Aggregates AI

I noted above that Apple's organization of the keynote was "mostly perfect"; I think the company made a mistake in sticking the OpenAI integration at the very end, which seems to have given the impression to some that all of Apple Intelligence was driven by OpenAI. For example:



**Elon Musk**    
@elonmusk · [Follow](#)



If Apple integrates OpenAI at the OS level, then Apple devices will be banned at my companies. That is an unacceptable security violation.

8:35 PM · Jun 10, 2024 

 **344K**  **Reply**  **Copy link to post**

[Read 19.3K replies](#)



**Elon Musk**    
@elonmusk · [Follow](#)



It's patently absurd that Apple isn't smart enough to make their own AI, yet is somehow capable of ensuring that OpenAI will protect your security & privacy!

Apple has no clue what's actually going on once they hand your data over to OpenAI. They're selling you down the river.



**Readers added context**



Apple has developed their foundational models which run on-device (locally) and have approximately 3 billion parameters.

For tasks that require more compute, Apple either uses Private Cloud Compute (open to verify for privacy) or OpenAI with an additional confirmation.

[machinelearning.apple.com/research/intro...](https://machinelearning.apple.com/research/intro...)

Context is written by people who use X, and appears when rated helpful by others. [Find out more.](#)

8:50 PM · Jun 10, 2024



**194.3K**



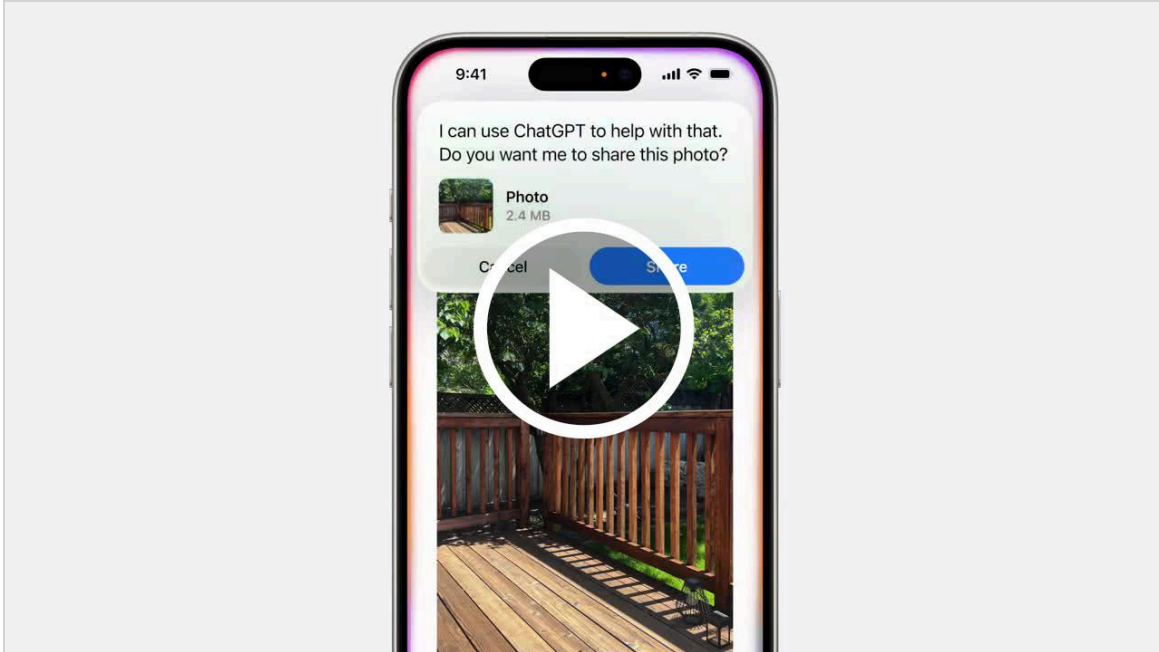
**Reply**



**Copy link to post**

[Read 13.1K replies](#)

Now Musk has **his own grudge against OpenAI**, which, one imagines, is influencing his stridency, but it's worth addressing his concerns head-on. Apple was crystal clear that ChatGPT is in fact plugging into a modular interface that will be available to multiple models; here is how Senior Vice President of Software Engineering Craig Federighi announced the OpenAI "partnership":



Apple Intelligence is available for free with iOS 18, iPadOS 18, and macOS Sequoia, bringing you personal intelligence across the products you use every day.

Still, there are other artificial intelligence tools that can be useful for tasks that draw on broad world knowledge, or offer specialized domain expertise. We want you to be able to use these external models without having to jump between different tools. So we're integrating them right into your experiences, and we're starting out with the best of these: the pioneer and market-leader, ChatGPT from OpenAI, powered by GPT-4o.

First, we built support into Siri, so Siri can tap into ChatGPT's expertise when it might be helpful for you. For example, if you need menu ideas for an elaborate meal to make for friends using some freshly caught fish and ingredients from your garden, you can just ask Siri. Siri determines that ChatGPT might have good ideas for this, asks your permission to share your question, and presents the answer directly.

You can also include photos with your questions; if you want some advice on decorating, you can take a picture and ask what kind of plants would go well on this deck? Siri confirms if it's ok to share your photo with ChatGPT, and brings back relevant suggestions. It's a seamless integration. In addition to Photos, you can also ask questions related to your documents, presentations, or PDFs. We've also integrated ChatGPT into the system-wide writing tools with compose. You can create content with ChatGPT for whatever you're writing about. Suppose you want to create a custom bedtime story for your six year-old who loves butterflies and solving riddles: put in your initial idea and send it to ChatGPT to get something she'll love.

Compose can also help you tap in to ChatGPT's image capabilities to generate images in a wide variety of styles to illustrate your bedtime story. You'll be able to access ChatGPT for free without creating an



account. Your request and information will not be logged. And for ChatGPT subscribers you'll be able to connect your account and access paid features right within our experience. Of course you'll be in control of when ChatGPT is used, and will be asked before any of your information is shared. ChatGPT integration will be coming to iOS 18, iPadOS 18, and macOS Sequoia later this year.

We also intend to add support for other AI models in the future.

First off, Apple Intelligence being free is notable: Apple clearly — and rightly, in my mind — sees Apple Intelligence as a way to differentiate their devices (and potentially sell more expensive devices in the future).

Secondly, Federighi makes the same point I did above: there are lots of things that broad-based models like ChatGPT are good at that Apple isn't even attempting to take on, and why would they? These models cost an astronomical amount of money to train and inference, and whatever differentiation exists is predicated on factors that Apple doesn't have a competitive advantage in. What Apple does have is a massive userbase that the model that wants to win will desire access to, so why not trade that access for the ability to leverage whichever model agrees to Apple's terms?

This gets to the one thing that I think I got wrong in yesterday's Article: my assumption has been that Apple was going to pay for whatever integration it offered, but now I question whether that is the case or not. OpenAI said in [their blog post](#) about the partnership:

The ChatGPT integration, powered by GPT-4o, will come to iOS, iPadOS, and macOS later this year. Users can access it for free without creating an account, and ChatGPT subscribers can connect their accounts and access paid features right from these experiences.

This sounds like a play to acquire users and mindshare, with the potential of upselling those users to a subscription, i.e. the exact same model that OpenAI has on their website and apps. Moreover, if this partnership entails Apple not paying, it also explains why OpenAI is the only option to start: Google, for example, probably wanted to be paid for Gemini, or Anthropic for Claude, and I can imagine (1) Apple holding the line on not paying, particularly if (2) OpenAI is making an aggressive move to build out its consumer business and be a durable brand and winner in the consumer space. In short, my updated current thinking is that both Apple and OpenAI are making the bet that very large language models are becoming increasingly commoditized, which means that Apple doesn't have to pay to get access to one, and OpenAI sees scale and consumer mindshare as the best route to a sustainable business.

To put it another way, and in Stratechery terms, Apple is positioning itself as an AI Aggregator: the company owns users and, by extension, generative AI

demand by virtue of owning its platforms, and it is deepening its moat through Apple Intelligence, which only Apple can do; that demand is then being brought to bear on suppliers who probably have to eat the costs of getting privileged access to Apple's userbase.

In other words, to the extent that Musk hates OpenAI, he should be happy about this partnership: Apple is clearly not sharing private data with OpenAI, and honestly the warnings it throws up every time you access the service are probably going to get pretty annoying pretty quickly; what the company is doing is providing a standardized interface for OpenAI to get access to potential customers for impressive yet commoditized use cases that Apple doesn't need to spend resources on, because OpenAI and any of its would be competitors will be compelled to make the investment and accept Apple's terms in an attempt to find some sort of sustainable advantage.

That's not to say that OpenAI is making a mistake: I've been calling on the company to focus on becoming a consumer brand [for a long time now](#), and acquiring Apple's userbase, even if it costs a lot in inference, is a step in that direction. There should be no question, though, about where the power in the value chain is vested: with Apple.

Indeed, that gets at why I was so impressed by this keynote: Apple, probably more than any other company, deeply understands its position in the value chains in which it operates, and brings that position to bear to get other companies to serve its interests on its terms; we see it with developers, we see it with carriers, we see it with music labels, and now I think we see it with AI. Apple — assuming it delivers on what it showed with Apple Intelligence — is promising to deliver features only it can deliver, and in the process lock-in its ability to compel partners to invest heavily in features it has no interest in developing but wants to make available to Apple's users on Apple's terms.

---

This Update will be available as a podcast later today. To receive it in your podcast player, [visit Stratechery](#).

The Stratechery Update is intended for a single recipient, but occasional forwarding is totally fine! If you would like to order multiple subscriptions for your team with a group discount (minimum 5), please contact me directly.

Thanks for being a subscriber, and have a great day!

Listen to this update and other Stratechery Plus content in your podcast player: [Stratechery](#) | [Sharp Tech](#) | [Dithering](#) | [Sharp China](#) | [Greatest Of All Talk](#)

## Subscription Information



Member: Scott Hannan

Email: scott.hannan@gmail.com

Member since: January 13, 2023

Your subscription renews every year

Renewal date: January 13, 2025

You are receiving this email because you are subscribed to **Stratechery**.

**Click here** to view your account and manage your subscriptions.

**Click here** to unsubscribe.

© 2024 Stratechery LLC, [2093 Philadelphia Pike #9930, Claymont DE 19703](#)