

**SDS PODCAST
EPISODE 798:
FIVE-MINUTE
FRIDAY:
CLAUDE 3.5 SONNET:
FRONTIER
CAPABILITIES &
SLICK NEW
“ARTIFACTS” UI**



(00:05):

This is Five-Minute Friday on Claude 3.5 Sonnet.

(00:27):

Welcome back to The Super Data Science Podcast. I'm your host, Jon Krohn. Like we often do on Fridays, let's start off with a couple of listener reviews. Our first review today is from Apple Podcasts that is by someone named Harry Mandha. The review is titled "Love it" and Harry goes on to say "This is a great podcast that brings some awesome people, and interestingly they all are very social, patient and want to help others." Harry also included some ideas for the SuperDataScience platform so thanks for that, we did see that.

(00:56):

Our second review comes from Swapnali Patki, who's a Machine Learning Analyst in New York. She says: "Super Data Science Podcast is something I eagerly wait for every week. Listening to SDS episodes during my sketching session is my favorite weekend routine." Cool, first I've heard of someone enjoying sketching while they listen to the podcast, but I'm glad that we're useful for that use case too.

(01:21):

Thanks everyone out there for all the recent ratings and feedback on Apple Podcasts, Spotify and all the other podcasting platforms out there that you use, as well as for likes and comments that you put on our YouTube videos. Apple Podcast reviews are especially helpful to us because they allow you to leave written feedback if you want to and I keep a close eye on those so, I think people who are considering new podcasts to listen to, see those recent positive reviews. And so if you leave one, I think it's helpful for us to grow. Other people learn about the show or just see if this is the kind of show that is something that would interest them.



(01:55):

All right, let's get into the meat of this episode now. And also my apologies for my cold. I waited a few days to record this episode because I thought my cold might get better, but it's only getting worse, and I've run out of time in order to get this edited and published in time, so my apologies for the illness, but let's get right into the meat of this episode, as I said, which is about Anthropic's latest publicly released model, Claude 3.5 Sonnet. This might not seem like a big deal of a release because it's not a "whole number" release like Claude 3 was or Claude 4 eventually will be, or GPT-5 eventually will be. But in fact, it's quite a big deal as this model now appears to actually represent the new state of the art for text-in/text-out generative LLM, outcompeting the other frontier models like OpenAI's GPT-4o and Google's Gemini family.

(02:48):

For a bit of relevant context to tee things up, a quick refresher is that Claude 3 came in three sizes. So there is Haiku which is the smallest, fastest and cheapest in the family. There was the Sonnet size which is the mid-size model that's a solid default for most tasks. And then Opus was the full-size Claude 3 model that was my favorite text-in/text-out model... well, until now.

(03:12):

Anthropic so far has only released Claude 3.5 Sonnet, the mid-size model, So whereas with the Claude 3 released previously, they had those three sizes, Haiku, Sonnet, and Opus. Now, with 3.5, they've only released the mid-sized model, Sonnet, but in my testing as well as on benchmarks, which I'll talk about a little bit later in the podcast, Claude 3.5 Sonnet, the mid-sized model, outperforms the much larger Claude 3 Opus from a capability perspective. This is amazing because Claude 3.5 Sonnet is much smaller than Opus, so not only is it better at complex tasks like code



generation, writing high-quality content, summarizing lengthy documents, and creating insights and visualizations from unstructured data... It's, not only does all those things better than Claude 3 Opus, it's twice as fast because it's probably about half the size.

(04:05):

In terms of quantifying this quality, so we quantified the speed, there was a quantifying quality, and so you're not just trusting my qualitative assessment of it being so great. We've talked on this podcast many times about how benchmarks are not always the most reliable indicator of capabilities because they can be gamed, but alongside my personal qualitative assessment they are potentially helpful. And Claude 3.5 Sonnet's frontier capabilities do indeed set new benchmarks across the most oft-cited benchmark, which is MMLU, that's an assessment of undergrad level knowledge. It also sets new high benchmarks on GPQA, which assesses graduate student level reasoning. That's a particularly challenging test. And in terms of coding proficiency on the HumanEval assessment, 3.5 Sonnet also sets frontier capabilities.

(05:00):

In addition to text in text out, which is primarily how I use generative AI models in UIs. This model is also pretty darn good at machine vision. So this 3.5 Sonnet model is about 10% better than Claude 3 Opus was across a bunch of different vision benchmarks. It performs particularly well and accurately transcribing text out of difficult-to-read photos.

(05:26):

On top of all of the above — the SOTA capabilities, the rapid speed, the low cost and the broad accessibility, because you can get this free. Which is I guess I didn't even mention that yet, but yeah, you can actually get, unlike 3.0 Opus, which was you had to be a subscriber. Which is 20 bucks a



month. I definitely paid that. I think it's 100 % worth it. But this 3.5 Sonnet model is now available for free at claude.ai. So that's cool. I'm also, by the way, I'm not being paid in any way to say any of this stuff. This is genuinely just my plain belief and my plain preferences about particular large language models. And yeah, Claude was already my favorite and now it's even more so the case.

(06:21):

So anyway on top of everything we've talked about already, so SOTA capabilities, rapid speed, low cost and broad accessibility, another super cool item that Anthropic released alongside this Claude 3.5 Sonnet model is something an experimental UI feature within Claude that they've called Artifacts. So when you have Artifacts enabled and you ask Claude to generate content like code, documents or even a functioning website, these artifacts, these outputs appear in a side-by-side panel alongside your text-in/text-out conversation. So this is something new as far as I'm aware of in these generative AI user interfaces. You've got your conversation on the left-hand side, and select your normal kind of conversation that you'd be used to with ChatGPT or Gemini or whatever, that's now on the left and on the right are your big outputs, these artifacts. So it could be code, could be documents, could be a functioning website.

(07:16):

This is a game-changer for me in terms of user experience, because seeing these outputs on the side means that you don't need to scroll up and down through the conversation to get through these various outputs that are just conveniently there in front of you as well alongside the conversation that you're having with Claude. I'm actually recording a video now, so I don't know how well this is going to work out in the audio-only podcast version, but if you want to check out the video on YouTube, you can do that. I think I should be able to narrate this pretty well to make the case, even in a podcast.



(07:44):

So when you want to use this experimental new artifacts capability, there's an experimental feature enabled toggle right next to where you put text in in the Claude user interface, if you turn that to on, then you'll have this artifacts panel on the right-hand side while you type and have a conversation with Claude on the left-hand side.

(08:07):

So to kick things off, I said "Create an 8-bit image of a surfer". And Claude writes back to me to say, "I apologize, I do not have the capability to create, generate, edit, manipulate, or produce images." And it gives a big long explanation and tries to explain how you could style such an image and I just simply wrote back "use code to do it" and in seconds Claude then used SVG code so you can either look at the code which took seconds for it to generate or you can go into the preview mode in the new artifact panel and it renders that SVG code. So in fact, you do end up with an 8-bit style surfer image that is pretty good for a single effort.

(08:55):

Just to test out more capabilities, and I didn't go through lots of different kinds of examples. I just went through a few different ideas off the top of my head, and all of this worked instantly. So the next one that I did was create a simple website design for the Super Data Science podcast. And it in seconds yet again created a whole bunch of HTML and CSS code, which you can view the HTML and CSS code off to the side in the new artifacts panel on the right, or you can toggle to the preview of that HTML and CSS would renders that website for you. And you can scroll through that again in the artifacts panel without losing track of your ongoing conversation with Claude on the left.

(09:42):

As another example, this was the coolest one, I think they came out. I said, “Create a working website of the shell game”. So if you're not familiar with the shell game, Claude tells me that it's also known as the cup and ball game. And so that's where you have three cups on a table, and as you show the person that is guessing, you show them that you put a ball under one of these three cups, and then you shuffle around the three cups on a table. And so I asked, yeah, I gave a very simple command, as I just said, “Create a working website of the shell game.”

(10:20):

And again, in seconds, it generated JavaScript, CSS, an HTML code in order to render the shell game and the very first version of it was overly simplistic I would say but a really good first try, where there are three shells that I can click on as well as a start game button, it says click start game to begin so I click start game and then there's this very simple animation where the shells kind of just bounce around. And you click on a cup and you basically got a one and three chance of kind of randomly finding the ball, you can't actually really play the game. So I just wrote back “Could you please make the animation for shuffling more complex and realistic?”.

(11:03):

And again in seconds it regenerated the JavaScript CSS and HTML code and created a slightly, well a completely more realistic shuffling animation. I don't know, I was about to say slightly there, 'cause it was way more realistic. Now there's tons of shuffling, but the only piece missing was that you couldn't see the ball go in at the beginning of the game. So I said, “Nice. The only thing that would make the game better is allowing the user to see the ball go into one of the cups before shuffling”. And so Claude generated this third version of the game again in seconds with JavaScript, HTML and CSS code, which you can see in the artifacts panel on the right, or you can



click over to preview mode in the artifacts panel on the right, and then you can actually play the game interactively right there in the browser.

(11:50):

So I click start game and you can see a ball go into one of the cups and then it shuffles around. You can see the cup shuffling. I did lose track here, but earlier when I wasn't talking and recording at the same time while playing the game, I was able to track the ball and accurately guess. So, you know, this thing really works.

(12:16):

And then as a final test, just to show some other capabilities, I said “Provide code for a deep learning transformer architecture”. And again, that showed up in the artifacts panel on the right-hand side. So on the left, I get this nice conversation explaining what the code is doing. And this is more than just comments. It's kind of an overview like you would expect a software developer, if a software developer was writing code for you to kind of explain what she was doing as she was going through that work. And so you can see all the code on the right in the artifact panel while going through the kind of conversation on the left explaining what the code is doing, super cool.

(12:57):

And another idea that I had that I would test here live so that people who are viewing this on YouTube could actually see how quickly this renders, is, I noticed that the artifact panel also allows you to track what content you've added into the chat. So I thought I would just try adding an image to test out the machine vision capabilities. And so I put in an image from a website that I control, deeplearningillustrated.com. And on that website, I've got a picture of my dog with a copy of my book. And without me even saying anything, asking anything about the image, Claude just assumes that I'm



going to ask about the image. And it very accurately, again, in seconds, if you're watching the YouTube video, you saw it already happen. And I mean, it took less time to generate this description than it did to upload this small image file. And it says "This image is a small Yorkshire terrier dog sitting next to a book titled Deep Learning Illustrated, A Visual Interactive Guide to Artificial Intelligence by Jon Krohn". And it goes into more detail about my dog and the book and what's in the image and all of it is exactly spot on. Very, very cool. I'm thoroughly impressed.

(14:17):

So for me personally, Claude was already my go-to model for most generative AI tasks. This 3.5 Sonnet release from Anthropic has cemented that position for me even more.

(14:29):

All right, that's it for today's episode. If you enjoyed today's episode or know someone who might, consider sharing this episode with them, leave a review of the show on your favorite podcasting platform, tag me in a LinkedIn or Twitter post with your thoughts, I'll respond to those. And of course, if you're not already subscriber, subscribe. Most importantly, you know, all that other stuff doesn't, it doesn't matter to me, like just keeping on listening. I hope you will continue to and until next time, keep on rocking it out there and I'm looking forward to enjoying another round at the Super Data Science Podcast with you very soon.