Leading the Consumer AI Revolution with Galaxy AI

Ryan Shrout - President, Signal65

Sebastian Peak - Contributing Analyst, Signal65

IN PARTNERSHIP WITH

SAMSUNG

Contents

Introduction

Samsung has been at the forefront of the consumer AI revolution.

In an era where technology is racing to embed AI into every facet of our lives, smartphones have been harnessing AI to revolutionize user experiences for nearly two decades. Today, with billions of smartphones in use—most running Android and all powered by AI—it's clear: the modern smartphone, as we know it, owes its existence to artificial intelligence.

Imagine your smartphone without its sleek touch and voice interactions—AI is the magic behind these essential features. Take the virtual keyboard, for example. It wasn't until AI stepped in to predict keystrokes that typing on a touchscreen became seamless. Both Android and iOS have since relied on such AI-driven innovations, from voice-to-text conversions to smart digital assistants.

Smartphone cameras have also transformed thanks to AI, with features like portrait mode and object removal enhancing our photo-taking experience and object removal fundamentally shifting how consumers expect photography to function and perform. Samsung was one of the original innovators in this area and as you'll see in our experiential testing, continues to push the boundaries of what's possible. Innovations like Circle to Search with Google and Live Translate are pushing boundaries even further. Samsung's Live Translate, for example, uses AI to translate conversations in real time, bridging language barriers with unprecedented ease.

At the forefront of this AI revolution, Samsung leads with groundbreaking technologies centered on communication, creativity, productivity, and health. Their investments in proprietary AI models and compute engines, alongside Google's critical collaboration, drive the relentless advancement of smartphones. This dynamic interplay of competition and cooperation creates the compelling, innovative features that define AI-powered smartphones.

Client AI is About to Change Everything

As pervasive as AI already seems to be, we will undoubtedly continue to see more AI integration in client operating systems, and smartphones in particular. The global smartphone market has reached roughly 70% penetration, and has shown continued growth. The default computing device for most people is in their pocket, everywhere they go.

Nothing is more important than time, and the impact of a powerful, reliable assistant cannot be understated. We have seen continual improvements and advancements to the functionality of digital assistants.

Beyond even improvements to the technologies that we know and understand today, how companies like Samsung integrate AI into our smartphones, PCs,

and wearables will change our interactions with them at a fundamental level. Just as the mouse replaced the keyboard and touchscreens moved us beyond physical buttons, the next revolution will be even more profound. Our devices are evolving into sophisticated sensors—equipped with microphones, cameras, gyroscopes, and more, sensing and interacting with the world in ways we can barely imagine.

Nearly all leaders in the AI industry today agree that the future will be a hybrid of cloud-based and on-device AI. This means, depending on the specific needs for maximum speed or high performance compute, the actual processing and application of AI models will happen in either (or both) the cloud or using the chip on the device itself. This gives the most

flexibility to the software developers.

The advantage of on-device AI comes down to three key things: latency, security, and convenience.

User interfaces need to be fast and responsive. Think about the frustration of a lagging or stuttering touchscreen and how it negatively affects your experience with the device. For AI to be part of our everyday usage patterns, and part of our workflows, it needs to spend less time waiting on a network and cloud computing resources to get the answer it needs.

It also needs access to your personal data to offer real value, but this raises privacy concerns. Local processing means your data stays on your device, protecting it

from potential breaches in the cloud.

The convenience factor for local AI processing talks to the ability to run AI models and features that don't require a networking connection at all. Developers can focus on creating great experiences without needing to integrate cloud-based AI services if they don't need to.

The future of AI across consumer products will be a combination of local compute and cloud services that can expand the scope of any AI compute to larger and more complex data sets. But when it comes to that local AI compute, not all mobile SoCs are created equal.

Executive Summary:

The Device AI Platform for the Future

- The future of the AI smartphone is here now and platforms with investment in hardware for AI will continue to bring benefits to consumers

- Samsung Galaxy S24 Ultra has the strongest performance for local AI processing, making it the ideal platform for AI innovation

- Real world applications of AI are already in use today highlighted by capabilities of image object removal, dynamic photo processing, live translations, and text summarization

- Samsung's strong smartphone family, and expanding ecosystem of AI-enabled mobile technology, creates the best overall solution for consumer AI integration across any ecosystem

While it might seem like the AI revolution in the smartphone and device markets is just beginning, the truth is that companies like Samsung have been integrating AI and machine learning tools into their platforms for years, improving user experiences and creating innovative new features along the way. Both in partnership with powerful vendors like Google, but also in bespoke solutions that are unique to the Galaxy AI set of features, devices like the Samsung Galaxy S24 Ultra (or even the new Galaxy Z Fold6) will continue to evolve and improve with new software updates and applications.

In our testing for this report, the Samsung Galaxy S24 Ultra showed the strongest combination of both performance and user experience. In critical offline testing using benchmarks like MLPerf Inference: Mobile, where we can see the power of both the hardware selected by Samsung for its platform and the work done by both Samsung and Qualcomm on software enablement, the Galaxy was nearly 8x faster

in some tests, and averaged more than 4.5x faster across all the sub-tests. The "AI Benchmark" tool that also runs inference locally across a broad range of models showed a 4.4x performance delta between the Galaxy S24 Ultra and the Pixel 8 Pro at the device level, and 3.8x at the SoC level (minimizing impact of memory speed and capacity).

Advanced benchmarks like MLPerf are a critical piece of the puzzle when comparing what platforms are capable of for AI today and into the future, but analyzing the real-world impacts of AI is important too. In our testing we compared the image object removal feature for both devices and found that not only was the experience of using the feature on the Galaxy S24 Ultra much simpler and more intuitive than on the Pixel 8 Pro, but the results were notably better from an image quality perspective. We added in some analysis of "time saved" using this feature compared to desktop-based applications and found again that not

only was the result on the Galaxy S24 Ultra better, but up to 7x faster.

And while that measurable difference is impressive, the truth is that other key features on these devices like dynamic photo processing, live translations, and text summarization are already differentiating what an "AI smartphone" can and should do compared to legacy phones.

AI might be a pivotal part of key features in our smartphones today, but we expect these advancements to impact essentially everything we do, sooner rather than later. The Samsung Galaxy S24 Ultra, in our testing, proves itself to be the best and most capable platform for personal AI computing today and enables users for the future. If we add in the benefit of Samsung's broad reach in the ecosystem with a host of other AI-enabled accessories and hardware, it is clearer than ever how the vision of Samsung's Galaxy AI will play out.

AI Smartphone Battle

The Galaxy S24 Ultra represents the current flagship smartphone from Samsung, with specs to match. It features a large 6.8-inch Dynamic AMOLED 2X display with 1440x3120 resolution (good for 505 ppi), and is built using the Qualcomm(r) Snapdragon(r) 8 Gen 3 chip. This processor has an 8-core CPU configuration, 8-core Adreno 750 graphics, and Qualcomm's Hexagon AI Engine. The phone offers 12GB of memory, fast UFS 4.0 storage up to 1TB, and a quad camera array featuring a 200 MP main sensor. The list goes on, but this is the fastest Android phone on the planet right now.
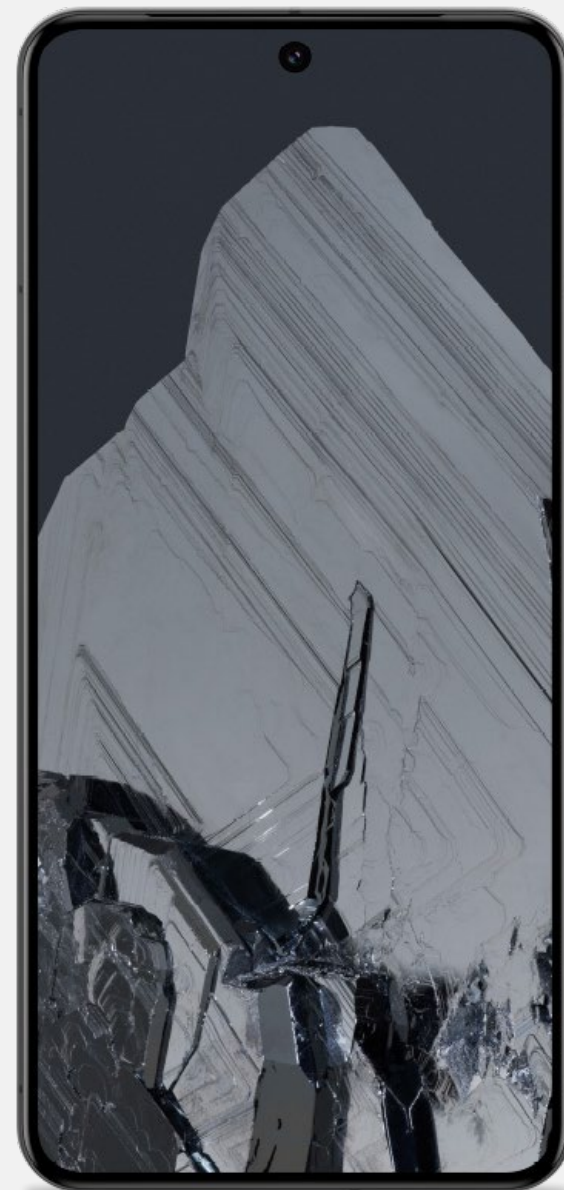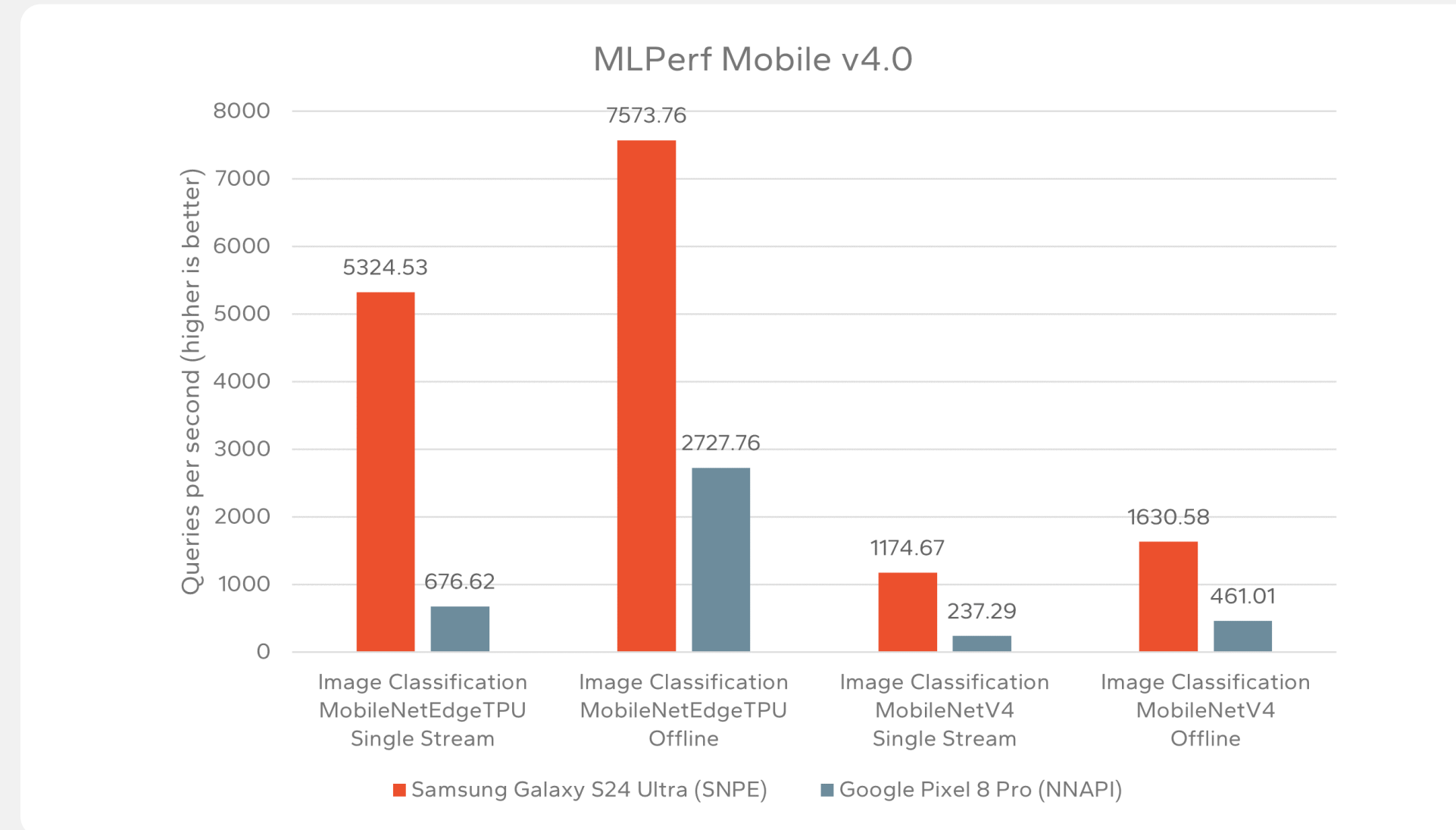
Google's Pixel 8 Pro is the latest flagship phone in the Pixel series, formerly known as Nexus, which offers a "pure" Android experience but lags behind a bit on the hardware side of things. It offers a 6.7-inch LTPO OLED display with 1344 x 2992 resolution (489 ppi), and is built with Google's custom Tensor G3 platform. The Tensor G3 offers a 9-core CPU configuration with 7-core Immortalis-G715 graphics, and a custom Google TPU (Tensor Processing Unit), the IP used for AI acceleration, 12GB of memory, up to 1TB of UFS 3.1 storage, and a triple camera array featuring a 50 MP main sensor.

Even though we have seen newer releases from Samsung and other device OEMs, selecting the Galaxy S24 Ultra and the Pixel 8 Pro was a purposeful decision to allow for the best analysis. Because both have been in the market for several months, any updates or software improvements are likely to have been addressed and corrected, creating an optimal experience comparison.

The first step in determining how these flagship Android smartphones handle modern and future AI workloads is to look at performance across a range of modern, emerging benchmark suites that look at AI/ML (machine learning) capabilities.

MLPerf Inference: Mobile Benchmark

MLPerf has been THE industry standard for AI performance testing in the data center, and it continues to expand its reach into client devices including smartphones and laptops. The MLPerf Inference Mobile Benchmark includes a set of image classification, segmentation, language, object detection and super resolution workloads using relevant AI models. It is a fairly intensive test for these devices but is also the most representative of the expected "real world" performance for AI computing since the MLCommons organization is a consortium of a broad collection of silicon and device companies across the industry. More emphasis and weight is placed at the feet of the MLPerf results than any other AI testing today, and for good reason.

MLPerf Mobile v4.0

- Samsung Galaxy S24 Ultra (SNPE)
- Google Pixel 8 Pro (NNAPI)

Note that the Samsung Galaxy S24 Ultra uses an optimized SNPE (Snapdragon Neural Processing Engine) software stack for these tests, while the Pixel 8 Pro uses Google NNAPI (Android's Neural Network API). Both are the preferred methods from each silicon vendor for best performance.

Using the latest version of the MLPerf Inference: Mobile benchmark, which measures both single-stream and offline AI performance in a variety of tasks, we see a massive advantage in compute power with the Galaxy S24 Ultra. The Single Stream results compare when the workload sends the next query after the current one is complete, measuring 90th percentile latency while the offline scores show performance where all queries are sent at the start of the test, measuring maximum throughput.
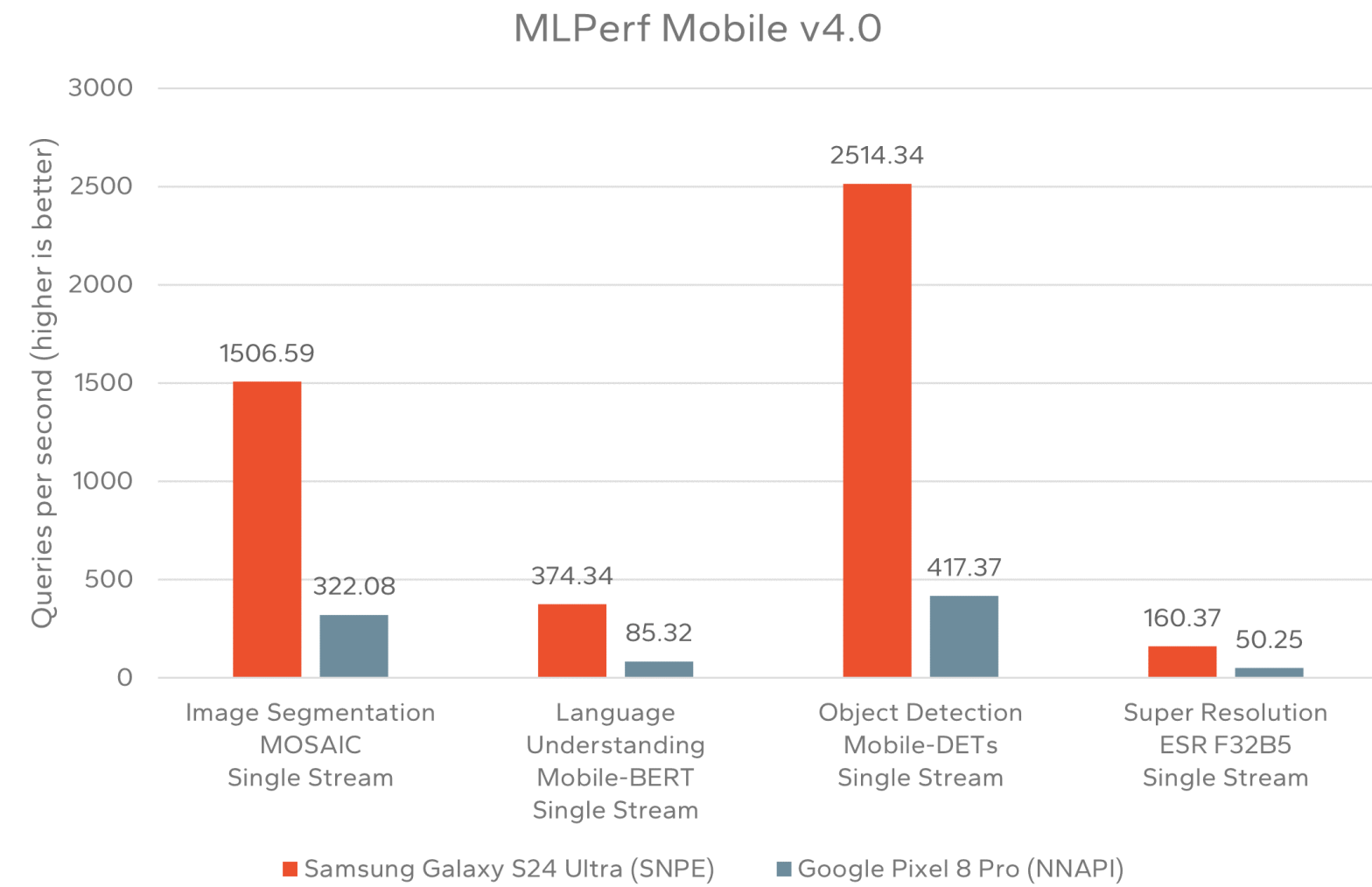
Looking at the first image classification result on the chart, the Samsung Galaxy S24 Ultra had an advantage of more than 4,500 queries per second (QPS) single-stream (nearly 8x faster), and an advantage of more than 4,800 QPS offline (nearly 3x faster) compared to the Pixel 8 Pro.

The Galaxy S24 Ultra offers up to 7.8x more AI inference performance than the competition

MLPerf Inference: Mobile Benchmark

MLPerf Mobile v4.0

Queries per second (higher is better)

Image Segmentation MOSAIC Single Stream: 1506.59 / 322.08
Language Understanding Mobile-BERT Single Stream: 374.34 / 85.32
Object Detection Mobile-DETs Single Stream: 2514.34 / 417.37
Super Resolution ESR F32B5 Single Stream: 160.37 / 50.25

■ Samsung Galaxy S24 Ultra (SNPE) ■ Google Pixel 8 Pro (NNAPI)

Changing the model did not change the outcome, nor did switching tasks. All of the MLPerf results speak for themselves, and it isn't close; the Galaxy S24 Ultra is a clear winner in terms of on-device AI performance.

One area to highlight with the MLPerf results here is that the silicon and phone vendors have the ability and are encouraged to optimize the software stack in the benchmark for their hardware. This can sound contrary to other benchmark groups where influence from the hardware vendors is specifically discouraged. But since so much of the current AI ecosystem is dependent on nascent software APIs and SDKs, vendor-specific work is almost always needed for real-world applications as well, so encouraging that work on MLPerf lets us see the expected, peak performance of each platform under test.

AI Benchmark

The aptly named "AI Benchmark" is a set of tests built by a small team in Zurich. Though maybe lesser known than MLPerf or other tests, this benchmark has legacy back to 2018 where a research paper featuring representation from Qualcomm, MediaTek, Arm, and Huawei were significant contributors.

### AI-Benchmark Score

AI-Benchmark Precision Testing
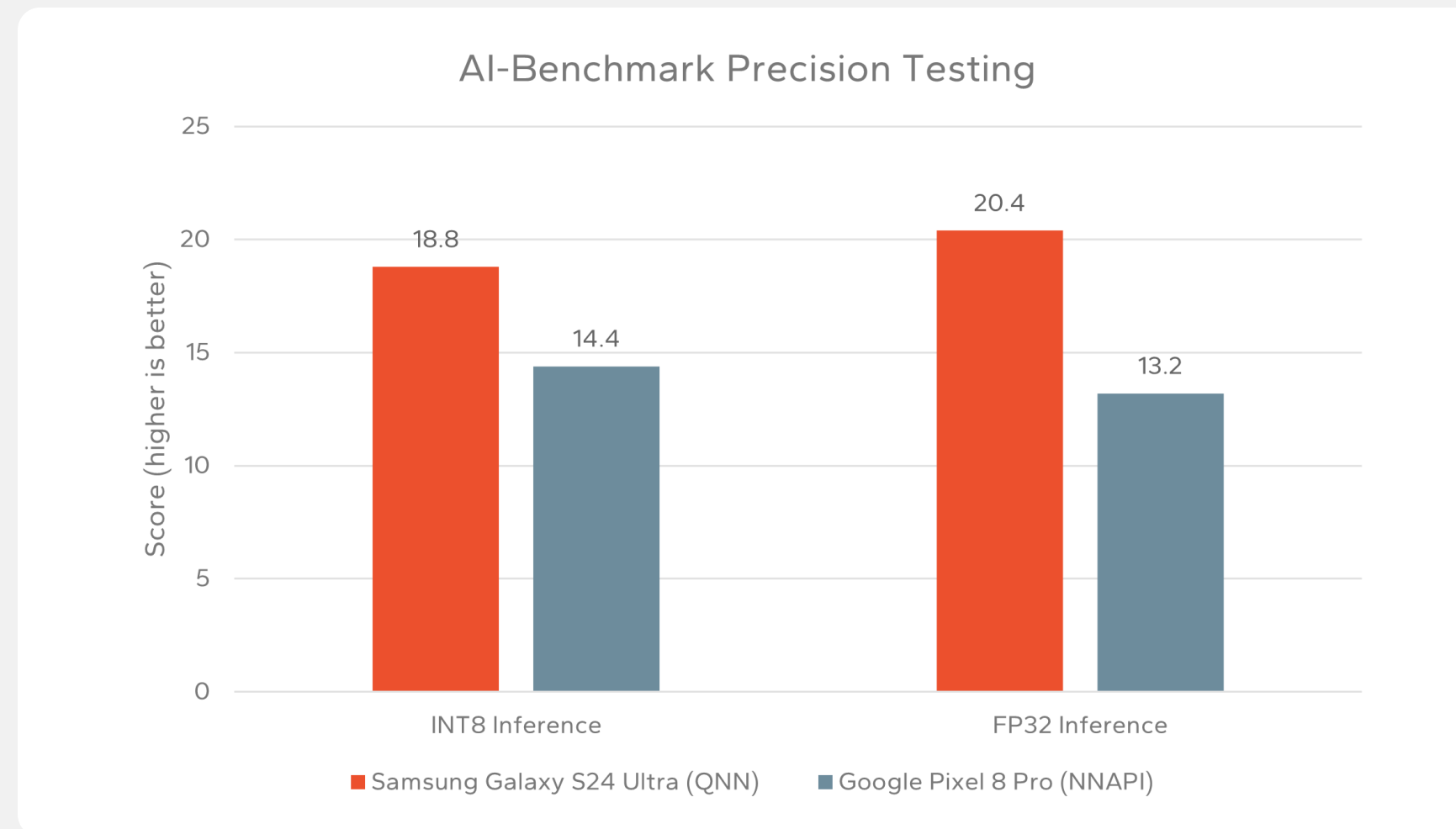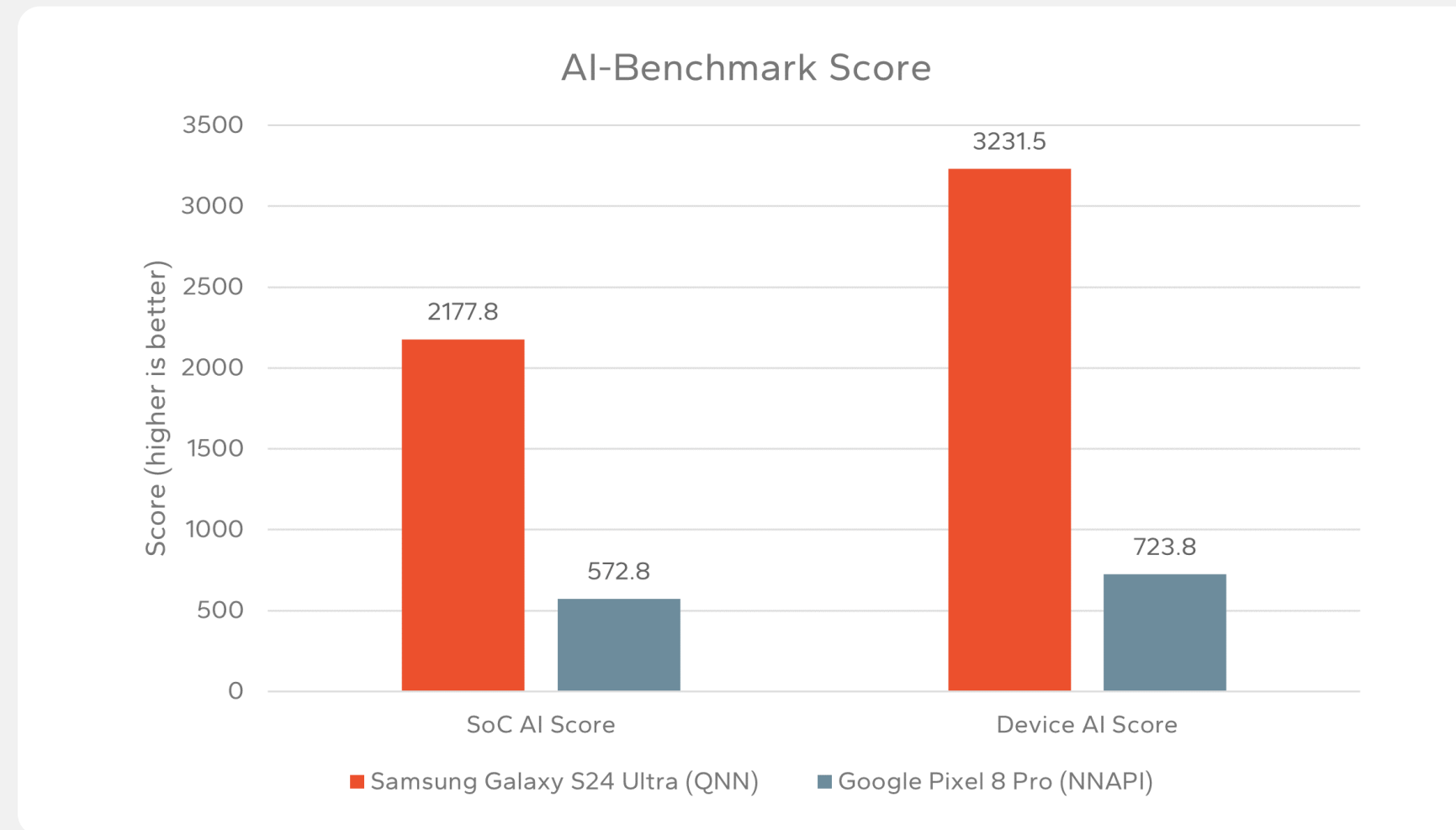
Even as one of the first mobile inference benchmark suites, it is still being updated with new tests and to support new hardware. It provides an overall score like many mainstream Android benchmarks, and this test simply measures latency for each of the workloads in its 26-part test. Looking at the overall benchmark scores we see a lopsided victory from the Galaxy S24 Ultra, with more than 3.5x the performance of the Pixel 8 Pro from the SoC AI Score, and more than 4.4x the performance in the Device AI Score. Note that the SoC score does not take into account the results of recent memory tests, thereby removing the impact from memory capacity on the device. AI Score combines the SoC score with the memory test result.

And while the Pixel 8 Pro has a more respectable showing in the individual INT8 and FP32 precision inference tests, the Galaxy S24 Ultra still offers 35% higher overall scores in this test - and that is with the Pixel 8 Pro accelerated using Google's preferred NNAPI on the backend.

This benchmark is another that understands that in order to properly represent the performance of each platform, individual changes must be made to the runtime for them. The Galaxy S24 Ultra uses QNN, the Qualcomm AI Engine SDK to best optimize for the CPU, GPU, and NPU on the Snapdragon® processor. The Google Pixel 8 Pro uses NNAPI (Android's Neural Network API) that Google still promotes but most other vendors have deprecated.
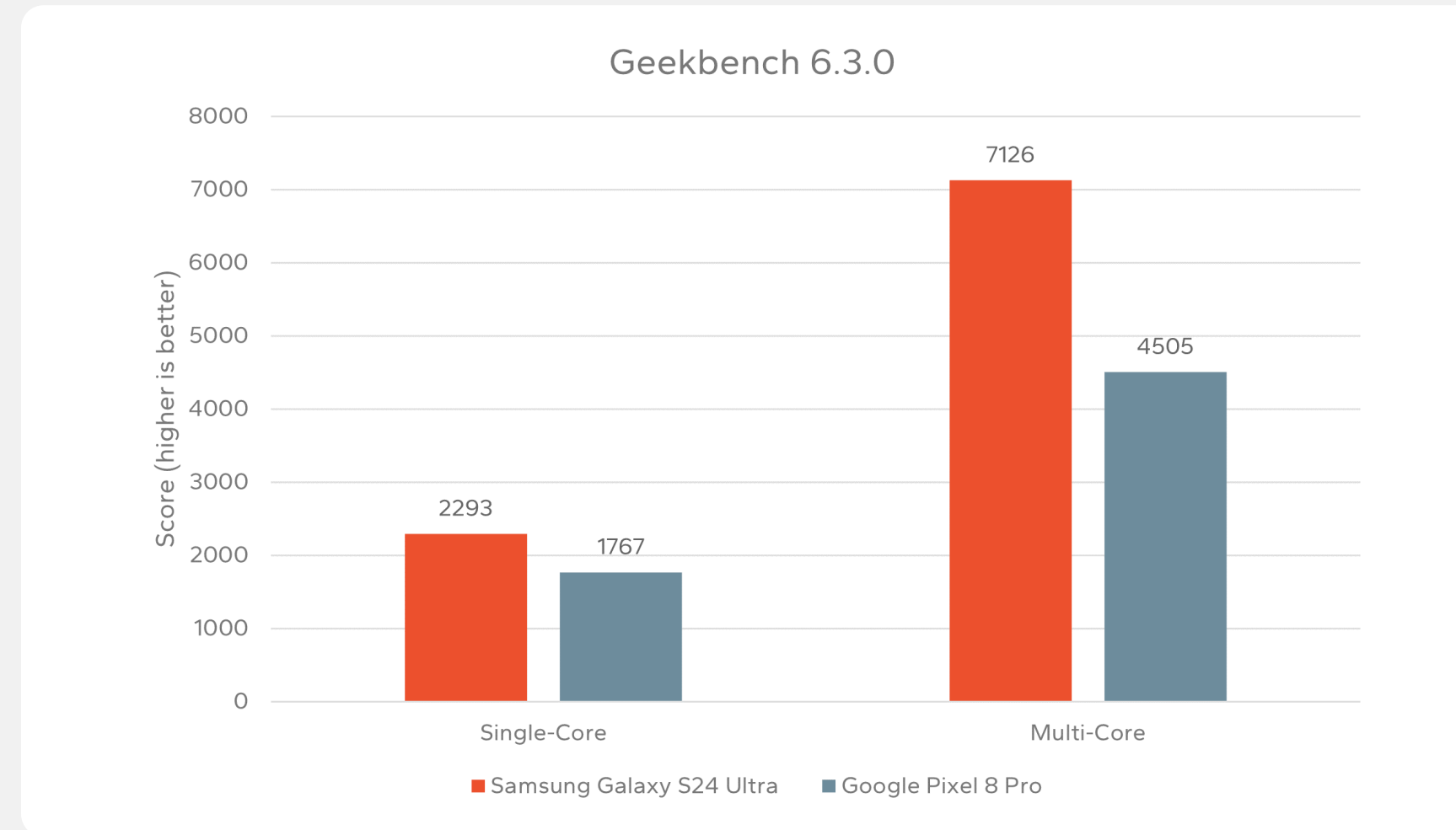
Geekbench ML

The Geekbench benchmark suite has become a staple of client computing testing and evaluation, and the Geekbench ML tool is a cross-platform AI benchmark that uses real-world machine learning tasks to evaluate AI workload performance.
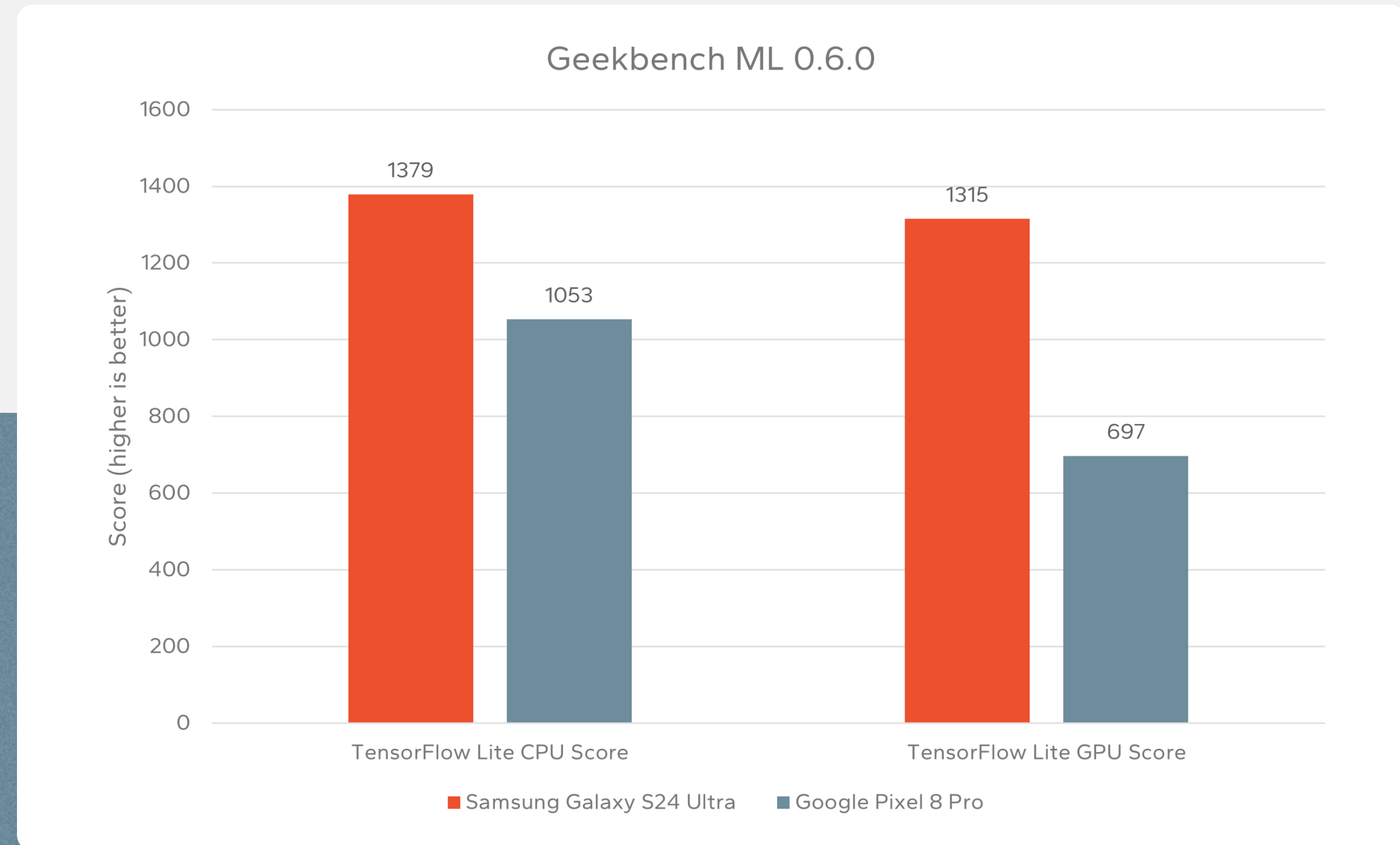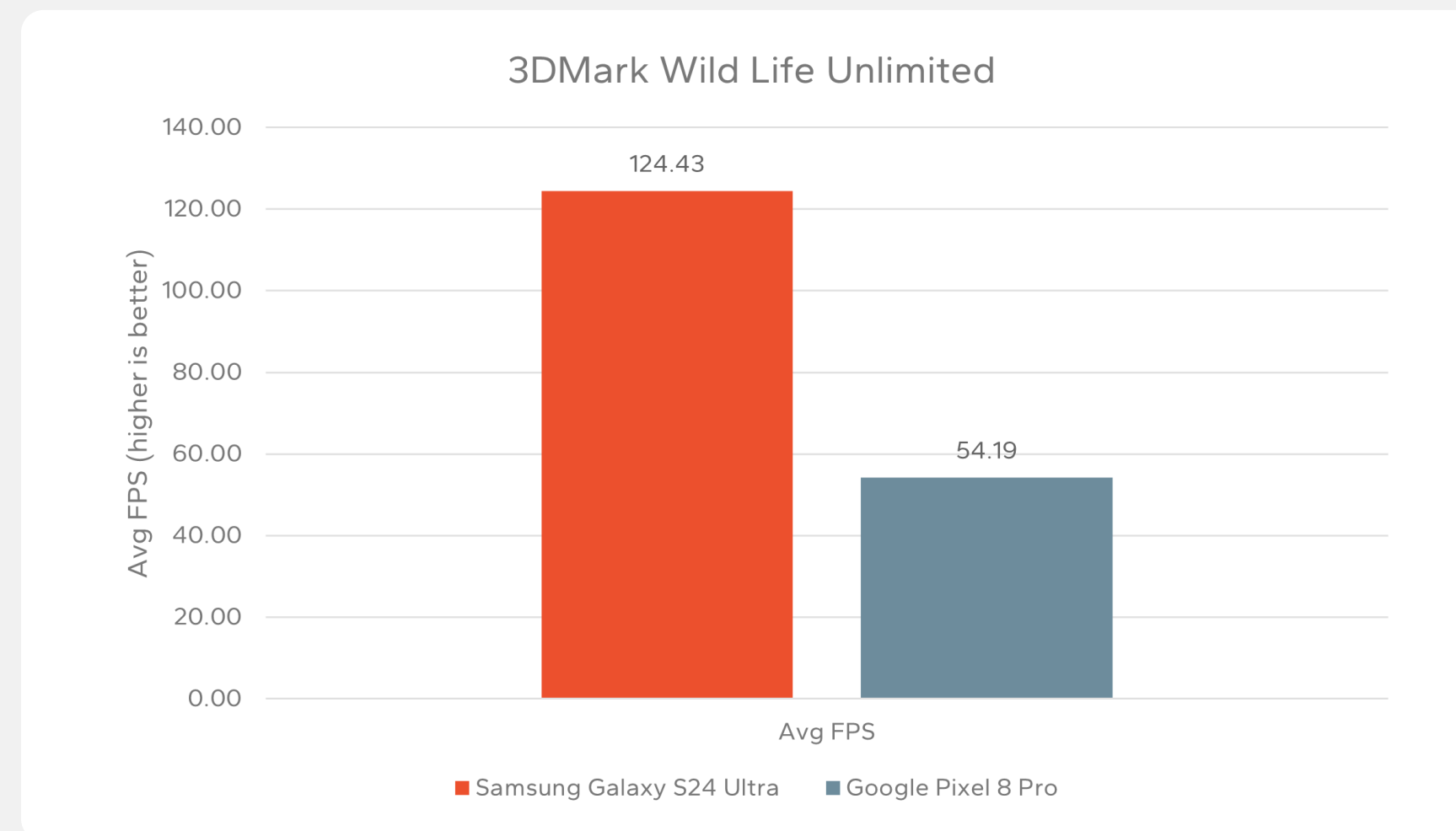
Geekbench ML 0.6.0

We tested the latest version of Geekbench ML, 0.6.0. It uses an internal version of TensorFlow Lite, and can be used to measure both CPU and GPU performance but lacks the ability to target the NPUs on these devices today. Here we see an overall performance advantage from the Samsung Galaxy S24 Ultra of about 31% from the CPU score, and 89% on the GPU score over the Pixel 8 Pro.

General Performance: Geekbench and 3DMark

Because the performance differences in our AI performance analysis were so heavily in favor of the Galaxy S24 Ultra, both on the NPU focused testing and the CPU/GPU testing, we wanted another, more standard set of tests to validate that the gaps seemed inline with expectations. For this we used the latest version of Geekbench and 3DMark Wild Life Extreme to spot check how the Galaxy S24 Ultra and the Pixel 8 Pro compared.

Geekbench 6.3.0

Score (higher is better)

	Single-Core	Multi-Core
Samsung Galaxy S24 Ultra	2293	7126
Google Pixel 8 Pro	1767	4505

■ Samsung Galaxy S24 Ultra   ■ Google Pixel 8 Pro

The single threaded CPU performance for the Snapdragon® 8 Gen 3 in the Galaxy S24 Ultra is nearly 30% faster than the Pixel 8 Pro, and despite having one less core (though the configuration of mobile CPU cores today is rather complex!), is nearly 60% faster in the multi-threading test.

3DMark Wild Life Unlimited

Avg FPS (higher is better)

	Avg FPS
Samsung Galaxy S24 Ultra	124.43
Google Pixel 8 Pro	54.19

■ Samsung Galaxy S24 Ultra   ■ Google Pixel 8 Pro

The Adreno graphics from the Snapdragon® processor within the Galaxy S24 Ultra made easy work of the 3DMark Wild Life benchmark, offering almost 2.3x the average frame rate compared to the Pixel 8 Pro.

Performance Benchmark Summary

The Samsung Galaxy S24 Ultra offers a clear and easily noticeable advantage over the Google Pixel 8 Pro, with a decided victory from the Galaxy S24 Ultra's Qualcomm Snapdragon® 8 Gen 3 platform over the Google Tensor G3 in the Pixel 8 Pro. This shows up in both our general performance testing but also most strongly in our peak AI performance measurements.

In some of the MLPerf Inference: Mobile Benchmark results we are seeing nearly an eightfold increase in AI compute on the Galaxy S24 Ultra than the Pixel 8 Pro! Yes, those results are due to both the hardware differences AND software optimization, but getting the right SDKs, APIs, and runtime environments working for application developers is the most critical part of enabling this AI revolution we are starting on consumer hardware. Samsung (and Qualcomm) should be applauded for this work and effort.

The Galaxy S24 Ultra offers a clear and noticeable performance advantage versus the competition

Real World AI Usage Testing

Performance testing and benchmarks are a critical part of the analysis of any kind of computing platform, client or data center. But at the end of the day, those results aren't what consumers will experience on their devices. For that, we wanted to spend some time going through a powerful and compelling real-world usage example of AI on both of these devices, comparing how fast and how effective they were to determine how these two smartphones compared.

# Fast, Efficient Object Removal in Photos

Both Google and Samsung offer their own object removal function when editing photos, but Samsung's approach rivals desktop photo editors - and does so without requiring the user to have any skill with photo editing. This is well implemented on the Galaxy S24 Ultra and our team kept going back to test the feature with photos taken in different environments, and the accuracy of the object selection was uncanny.
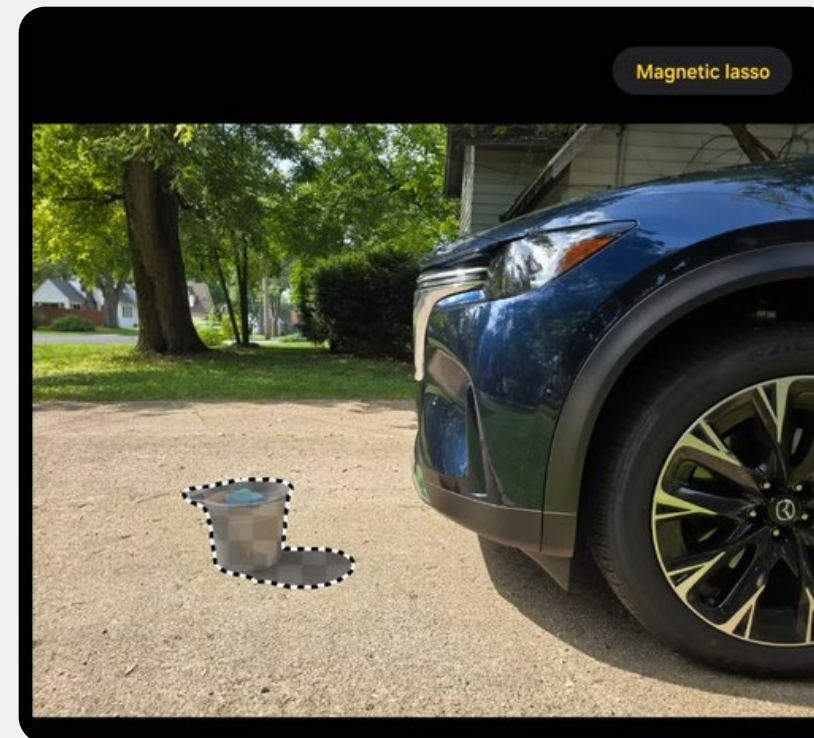
Samsung Galaxy S24 Ultra
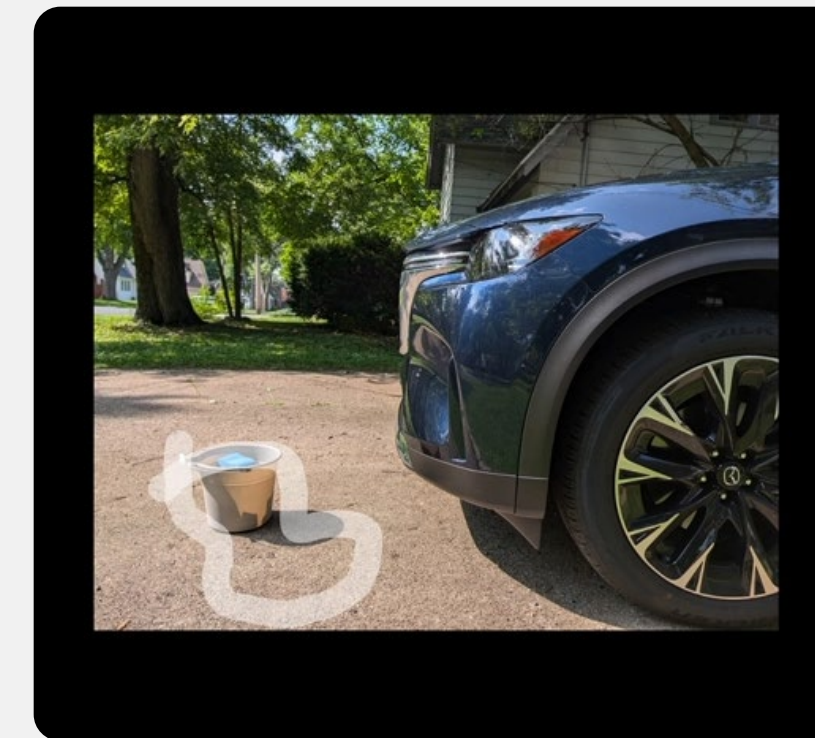
Google Pixel 8 Pro
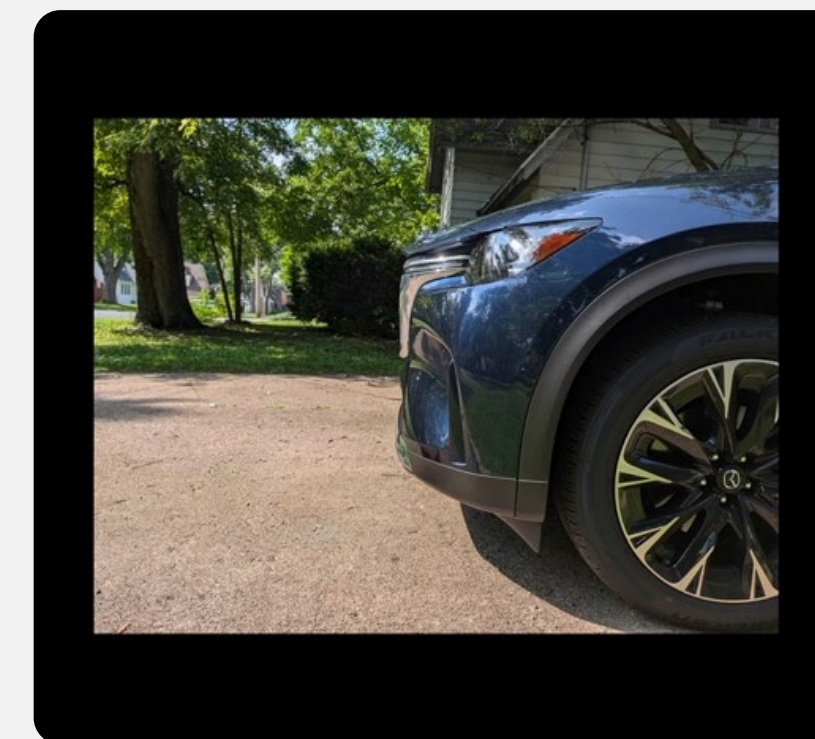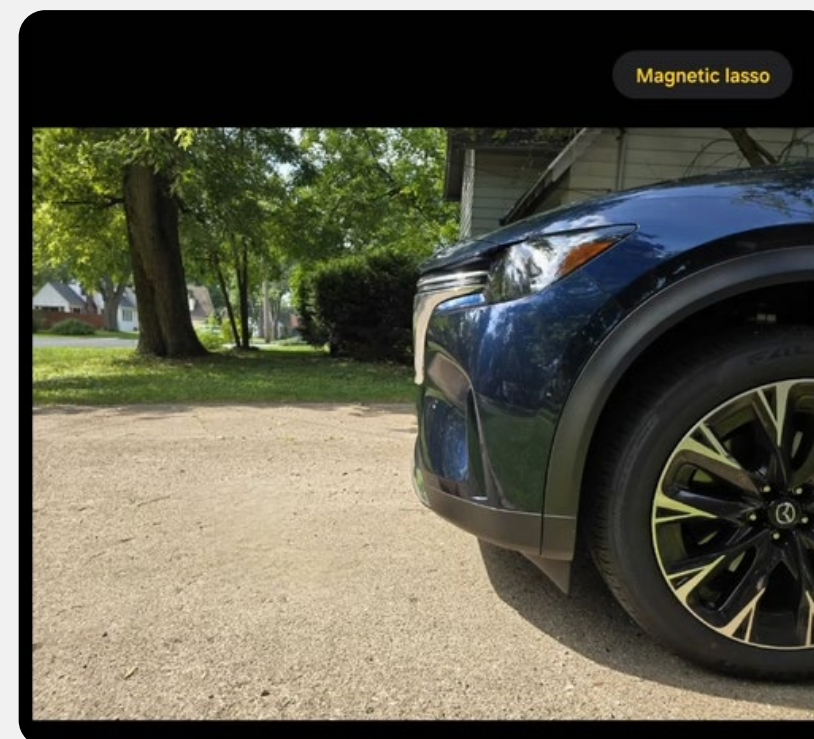
Rather than simply working on an area encompassing the item to be removed, Samsung's AI-powered selection and masking of circled objects places a perfect boundary around the object itself, and in practice the objects are cleanly removed with little to no visible impact on the surrounding area. In contrast, the Pixel 8 Pro's eraser tool is not able to mask as precisely, and was found to sometimes resort to blending parts of selected objects with the background, or cloning areas of the background, and these techniques often produced an unconvincing result on the Pixel.

Samsung's approach can be directly compared to full-scale desktop photo editing, and the speed with which circled objects are precisely selected cuts down a significant amount of time compared to traditional methods from applications such as Photoshop - and that assumes that the user is experienced at photo editing in the first place. Another advantage of Samsung's precision selection/masking was perfect removal of objects near corners and straight edges, while the Pixel's removal could introduce wavy lines as images lost their original edges either due to imperfect masking or blending during the full process. The result and experience provided by the Galaxy S24 Ultra is excellent.

Measuring Time Savings of Samsung Object Removal

Our team wanted to test what the real-world impact of having this kind of powerful, effective photo editing feature on the Galaxy S24 Ultra would be for a consumer. When an AI feature like this feels so intuitive it's easy to gloss over the actual time savings and quality improvements it brings to the table.
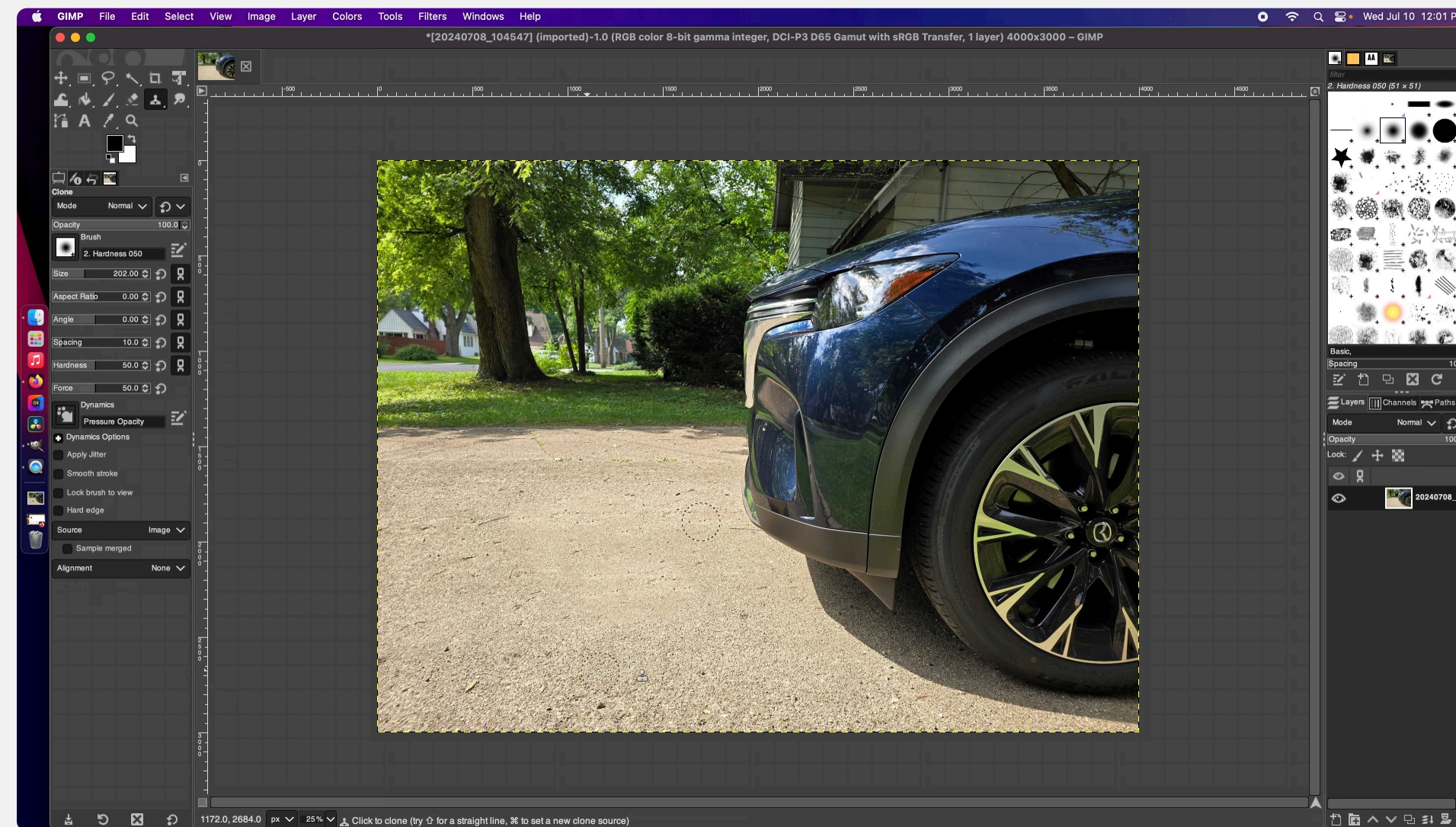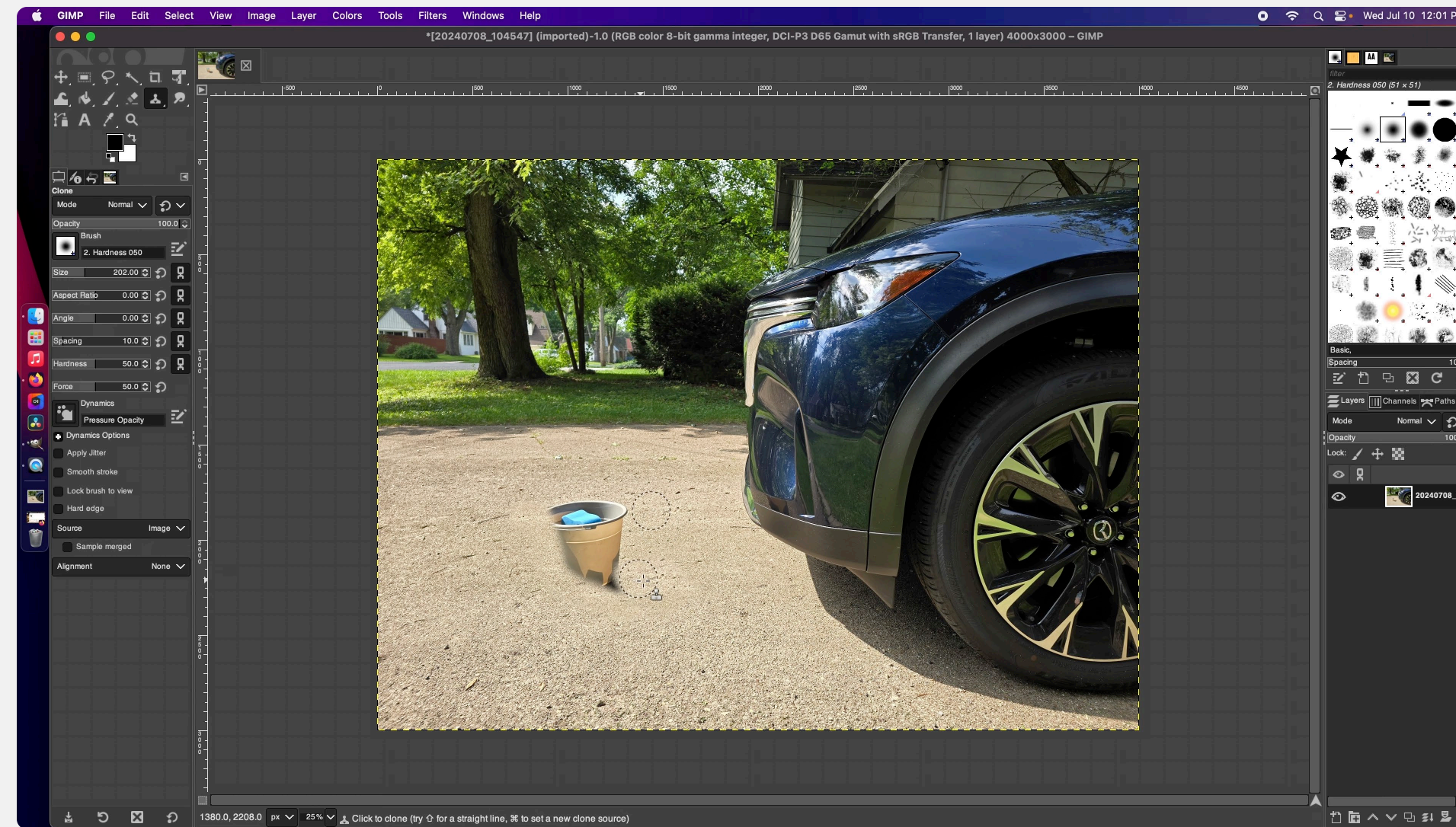

Using a Mac Studio M2 Max desktop we imported the photo used in the phone object eraser test and tried to remove the same object as quickly as possible. For this testing we recorded the process, and was able to precisely timestamp the results. Not counting the time spent emailing the photo from the phone (not to mention finding the message and downloading the attachment), the breakdown was as follows:

GNU Image Manipulation Program (GIMP) 2.10.38:

6.5 seconds to import, between 59.5 and 66.5 seconds total depending on method used (photo editing experience required)

- Method one - add 60 seconds
  - 38 seconds for manual lasso tool selection
  - 22 seconds to backfill using layer duplication/background adjustment
- Method two - add 53 seconds from base image to removal using clone stamp tool

Measuring Time Savings of Samsung Object Removal
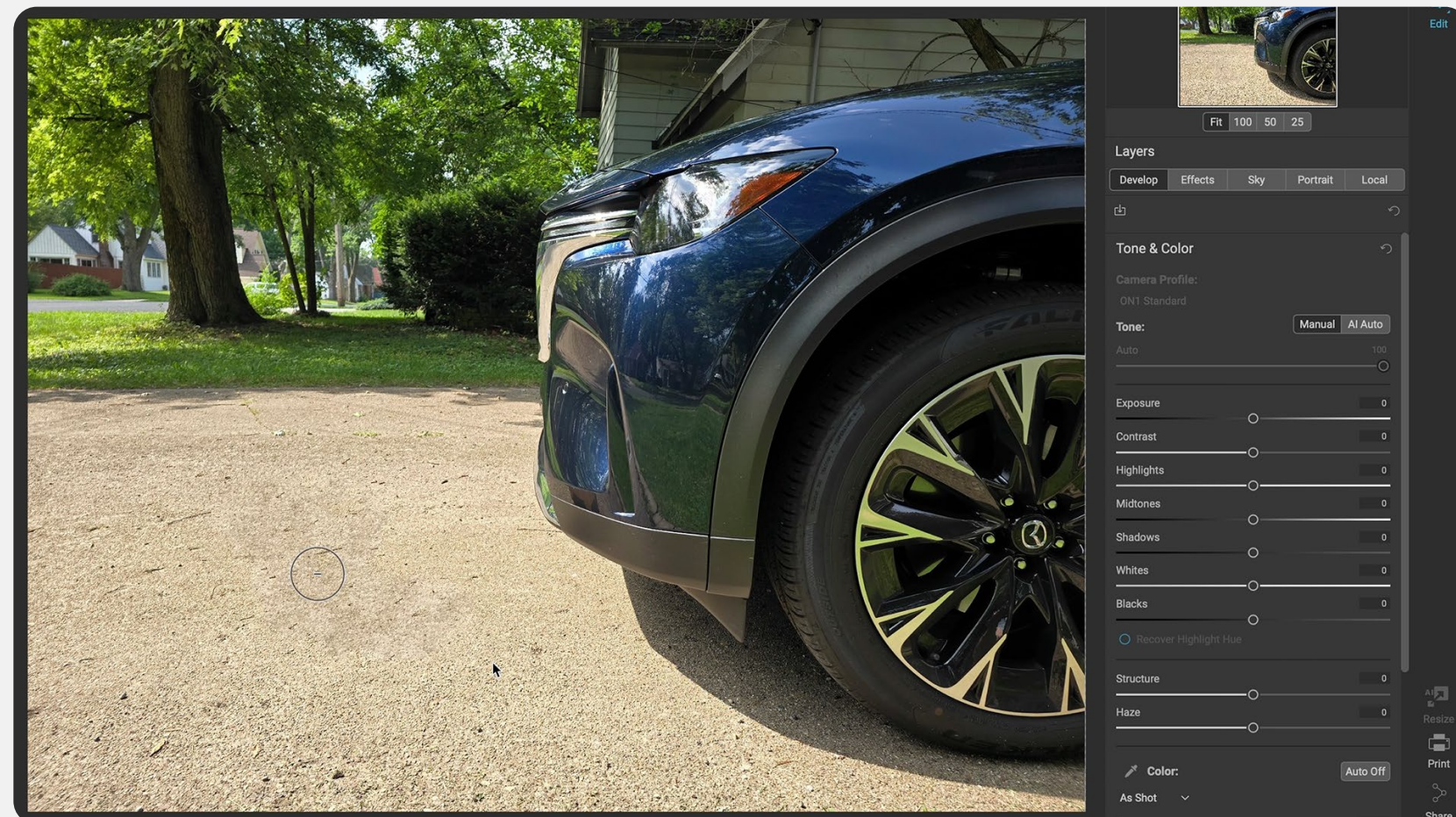
ON1 Photo Raw 2022:

29 seconds from import to completion using "Perfect Eraser" tool

ON1 Photo Raw produced the fastest result, but the quality of the result was rather poor and this is not free software. Better results were possible using GIMP, but only if the user is experienced. The approximately one minute timeframe was only possible since the analyst knew exactly which tools to use, and have years of experience with the application.

Using the Galaxy S24 Ultra, in comparison, we were able to tap into the object eraser function and be ready to select in 4.5 seconds, 2.5 seconds to trace around the object and have the AI software precisely select it, and 1.5 seconds from tapping erase until the final result (which looked perfect). That's a total of 8.5 seconds, or a 3.4x time savings compared to ON1 and 7x time savings compared to the quickest result using GiMP. The Galaxy S24 Ultra also produced a visibly better result than the desktop editing attempts, and of course required no additional hardware or any digital photo editing experience.

Save 3-7x the time needed to edit photos, while generating better quality

Timed Results (Overall)

	GIMP (Lasso Tool)	GIMP (Clone Stamp)	ON1 Photo (Perfect Eraser)	Galaxy S24 Ultra (Object Eraser)
Time (seconds)	66.5	59.5	29	8.5
Galaxy S24 Ultra Time Advantage	7.8x	7.0x	3.4x	--

Important Information About this Report

Contact Information
Signal65 | signal65.com | info@signal65.com

Contributors

Sebastian Peak
Client Performance - Signal65
Ryan Shrout
President & GM - Signal65
Ken Addison
Client Performance Director - Signal65

Inquiries
Contact us if you would like to discuss this report and Signal65 will respond promptly.

Citations
This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

Licensing
This document, including any supporting materials, is owned by Signal65. This publication may not be reproduced, distributed, or shared in any form without the prior written permission of Signal65.

Disclosures
Signal65 provides research, analysis, advising, and lab services to many high-tech companies, including those mentioned in this paper. Research of this document was commissioned by Samsung.

In Partnership with:

SAMSUNG

About Signal65
Signal65 exists to be a source of data in a world where technology markets and product landscapes create complex and distorted views of product truth. We strive to provide honest and comprehensive feedback and analysis for our clients in order for them to better understand their own competitive positioning and create optimal opportunities to market and message their devices and services.

signal65