

Transcript:

Alex Karargyris: MedPerf is a platform that evaluates medical AI models on diverse real-world medical data.

Jason Lopez: In the public conversation about artificial intelligence, one thing just about everyone agrees on is that healthcare will be a big beneficiary of AI. This is the Tech Barometer podcast. I'm Jason Lopez. In this story, we're going to explore how AI can impact healthcare, how testing in the real world can improve trust and how it can quantify medical AI performance on a platform called MedPerf.

Alex Karargyris: MedPerf can securely distribute the AI models to healthcare organizations and test those models on their local data. No data leaves the premises of the healthcare organization. It stays within the private network and only high-level aggregated results are shared among the participating organizations.

Jason Lopez: Alex Karargyris is co-chair for the medical working group within the nonprofit ML Commons, and he's one of Med Perfs co-founders and developers.

Alex Karargyris: The platform has been designed with patient privacy protection at its very core.

Jason Lopez: During our conversation, Alex circled back to the issue of security and privacy a number of times to emphasize how sensitive the healthcare industry is to the risks of deploying new technologies, especially in the light of HIPAA security rules.

Alex Karargyris: We aim for patient privacy first and foremost, and one more thing because we are developing source code, MedPerf is open source so that everyone can contribute and reuse freely, right?

Jason Lopez: The high-level description of MedPerf you've heard so far was outlined in a 2023 paper Carris authored, entitled, announcing MedPerf Open Benchmarking platform for Medical AI. It lays down what the organization is, what it does, and how to get involved.

Debo Dutta: The reactions after the paper came out were different from different walks of life.

Jason Lopez: Debo Dutta is vice President of Engineering at Nutanix, and he was integral to starting the group ML Commons.

Debo Dutta: When I talked to technologists, they say, wow, this is a cool system. The ML folks were saying, oh, that's a very good application for federated learning. Awesome. Some of my colleagues in computer science, they were quite amazed at the amount of effort it took to get so many institutions together and do clinical trials. I think people were just amazed by the sheer amount of effort and passion that this whole group took to get to this point. Then there are doctors.

Jason Lopez: Debo says, in the area of cancer, some oncologists who've been a bit skeptical about AI say the paper changes things.

Debo Dutta: I have had oncologists come and tell me this, that this is awesome. Oncologists who were skeptical about ai, one of my collaborators on this paper, told me, now that this is done, we can move on to doing even more amazing things that have a bigger clinical impact.

Jason Lopez: The story of the Med Perf platform essentially started in 2020 with the medical AI benchmark project that was done through ML Perf, an organization that benchmarks computing platforms for machine learning workloads, initially focused on brain tumor detection and M R I scans. People from organizations such as Nutanix, Intel, I B M, Google, and the Dana-Farber Cancer Institute wondered if they could replicate the success of it benchmarking in healthcare, they aim to build a platform that could test machine learning models on real-world medical data while fostering trust in AI for healthcare. Over time, the group's membership grew to include people from over 20 organizations across five continents with full-time engineers and volunteers working together to develop the platform and establish best practices within the medical AI community.

Alex Karargyris: We were discussing with other people how the same benchmark philosophy could be applied to medical AI and how to expand the capabilities of the whole community. Some questions we're asking are, how can we validate medical AI efficiently, and mutually, and make sure this privacy is guaranteed? So we set out to test this hypothesis three years ago, and we found out there were more and more people that did the same questions. The group grew from seven people to right now we have people from 20 organizations across five different continents, and we have full-time engineers in the group and also a lot of volunteers contributing to the development of the platform as well as best practices framework around the platform.

Jason Lopez: The Med Perf framework is a system designed to coordinate the evaluation of medical AI models. It consists of two main components, the Med Perf server and the Med Perf client, a lightweight software installed in hospitals and healthcare organizations. The framework allows the initiation of model evaluations on the med perf server, download the model weights, and push them to the client nodes.

Alex Karargyris: And then the clients can execute this particular model against the data by executing the tests.

Jason Lopez: This is the basic workflow that enables medical research using AI at hospitals, universities, and labs around the world. The Med Perf platform makes it as seamless as possible, but arguably the highest priority is security and privacy.

Alex Karargyris: So we did observe during the development and prototyping of MET PERF with our partners that this is the number one. Number two, major concerns, security and privacy in this context with a regulatory framework that has been developed around the world will force even stricter and deeper security considerations. I think that IT teams will have to look for professional solutions that offer tight and holistic security that cover AI as well, because it comes back to resource management. How do you handle heavy workloads?

Jason Lopez: This is one of the challenges. Med Perf is figuring out how to run things in the cloud because many healthcare organizations are not necessarily heavily IT-oriented.

Alex Karargyris: They don't have the infrastructure to run this, but they want to join this AI ecosystem, right? So this is the way to go to be able to execute in a very trusted, secure way on the cloud. All these workloads from training all the way to validation that Met PERF does.

Jason Lopez: Alex says there's a broad diversity of workloads in healthcare, and there's not necessarily a right computing platform, but there is a common thread from an IT perspective,

Alex Karargyris: Better application management on the available resources. These AI systems could be very resource-hungry, GPU to be used accelerators to run these things. I think this is the value that many of the cloud providers can bring because they support this, right,

Jason Lopez: And always a high priority in the health IT conversation, security and privacy

Alex Karargyris: Major major importance like Nutanix, there is secure trust execution as well is important to be able to close the space. You can run these workloads. Depo has supported the group from its inception, and he has been a provider of wisdom, I would say, a strong believer in AI for healthcare, and she has been supported in open efforts like ours. Without email, it would be impossible to do this.

Debo Dutta: I was one of the founding members of the ML Commons organization. It was not called ML Commons before. It was just called ml perf.org. It was just a small organization of mostly AI infrastructure vendors and a bunch of academics who wanted to redefine how machine learning performance would be evaluated because we saw the tide of ML coming very soon, and we said, okay, we need to evaluate performance in a meaningful way, in a repeatable way, and we want to be a nonprofit so that people trust us. And then over time, ML perf evolved into ML Commons. The goal of this organization, it's an amazing organization by the way, is to basically improve the state of the art and accelerate the AI transformation via open artifacts like performance benchmarks, data sets, and best practices.

Alex Karargyris: What we try to achieve with this, with MedPerf, if you're an AI researcher and you run our study, MedPerf can help you collaborate with many institutions and evaluate your model on a much wider, and I would say diverse patient population than you could do possibly right now. Similar to what ML Commons has been doing with its other benchmarks in other fields such as computing power platforms, we are creating benchmarks that are neutral and reproducible and can help improve the effectiveness of medical AI.

Debo Dutta: I'm elated because this has been a long journey to get the leadership on board and then getting hold of all the top-tier oncologists and medical researchers, AI researchers, and companies who do AI to converge on the one cause. That's let us actually build clinically impactful ML platform that respects data privacy and patient privacy. Frankly, to evaluate the next-gen AI models that could change the world of therapeutics, cancer therapeutics in particular. So I'm super excited.

Jason Lopez: Debo Dutta is Vice President of Engineering at Nutanix and a founding member of ML Commons. Alex Karargyris, who led the effort with his collaborators to write the MedPerf paper, is the co-chair of the medical working group in ML Commons. You can find the paper "Federated Benchmarking of Medical Artificial Intelligence with MedPerf" at mlcommons.org. This is the Tech Barometer podcast, produced by the Forecast. I'm Jason Lopez. We'll be doing some deeper dives into the MedPerf platform - a profile of Alex and one of Debo in future podcasts. You'll be able to find those at the forecastbynutanix.com.