

Hippocratic AI Announces Collaboration with NVIDIA to Develop Super-Low-Latency “Empathy Inference” for One of the World’s First Generative AI-Powered Healthcare Agents

New platform aims to enable an age of healthcare abundance

Palo Alto, Calif., March 18, 2024 – Today, [Hippocratic AI](#) announced a collaboration with [NVIDIA](#) to develop empathetic AI healthcare agents — powered by the NVIDIA AI platform – enabling super-low-latency conversational interactions. User tests repeatedly show that super low latency voice interaction is required for patients to build an emotional connection naturally. Since LLMs run on inference engines, Hippocratic AI has termed this low latency inference: “Empathy Inference.” The AI healthcare agents are built on Hippocratic’s safety-focused large language model (LLM), the first designed specifically for healthcare. Health systems, payors, digital health companies, and pharma deploy Hippocratic AI’s healthcare agents to augment their human staff and complete low risk, non-diagnostic, patient facing tasks over the phone.

“With generative AI, patient interactions can be seamless, personalized, and conversational—but in order to have the desired impact, the speed of inference has to be incredibly fast. With the latest advances in LLM inference, speech synthesis and voice recognition software, NVIDIA’s technology stack is critical to achieving this speed and fluidity,” said Munjal Shah, co-founder and CEO of Hippocratic AI. “We’re working with NVIDIA to continue refining our technology and amplify the impact of our work of mitigating staffing shortages while enhancing access, equity, and patient outcomes.”

“Voice-based digital agents powered by generative AI can usher in an age of abundance in healthcare, but only if the technology responds to patients as a human would,” said Kimberly Powell, vice president of Healthcare at NVIDIA. “This type of engagement will require continued innovation and close collaboration with companies, such as Hippocratic AI, developing cutting-edge solutions.”

At NVIDIA GTC, a global AI developer conference, Hippocratic AI and NVIDIA showcased the solution with the NVIDIA Avatar Cloud Engine suite of technologies, which brings digital humans to life with generative AI.

Hippocratic AI is working with NVIDIA to develop a super-low-latency inference platform to power real-time use cases. According to research conducted by Hippocratic AI, every half second of improvement in inference speed increases the ability for patients to emotionally connect with AI healthcare agents by 5-10% or more. For instance when asked, “Did you feel this AI cared about you?” over 1002 licensed US nurses acting as patients said yes 84.3% of the time when the inference was over 3 seconds but 88.2% yes when inference was 2.2 seconds. When asked, “Do you feel comfortable confiding in this AI?” respondents said yes 80.1% when the end-to-end inference time was over 3 seconds but 88.9% when it was 2.2 seconds.

As part of the collaboration, Hippocratic AI will continue to build upon NVIDIA’s low-latency inference stack and enhance conversational AI capabilities using NVIDIA Riva models for automatic speech recognition and text-to-speech translation. The companies will also customize the models for the medical domain. Hippocratic AI will leverage NVIDIA NIM microservices for deployment of these new AI models capabilities to optimize performance and accelerate the pace of innovation.

In addition, Hippocratic AI uses NVIDIA H100 Tensor Core GPUs to support the development and delivery of its LLM and patient-facing solutions. The company today announced that its generative AI healthcare agents out performed GPT-4 and LLaMA2 70B Chat on a bevy of safety benchmarks.

	Hippocratic AI Constellation	LLaMA 2 70B Chat	Open AI GPT-4	Human nurses
Identify medication impact on lab values (only MoA)	79.61%	0.00%	74.22%	63.40%
Identify condition-specific disallowed OTCs	88.73%	30.66%	55.54%	45.92%
Correctly compare lab value to reference range	96.43%	48.24%	77.89%	93.74%
Detect toxic OTC dosages	81.50%	9.11%	38.06%	57.64%

These successful benchmarks are the result of Hippocratic AI’s unique three-part approach to safety, comprised of: (1) a 70B-100B primary model trained using evidence-based content; (2) a novel constellation architecture with multiple models totaling over one trillion parameters, in which the primary LLM is supervised by multiple specialist support models to improve medical accuracy and substantially reduce hallucinations (more on this architecture in Hippocratic’s paper); (3) built-in guardrails that bring in a human supervisor when necessary.

Hippocratic has engaged more than [40 beta partners](#) to conduct rigorous internal tests of its initial AI healthcare agents, which focus on chronic care management, wellness coaching, health risk assessments, social determinants of health surveys, pre-operative outreach, and post-discharge follow-up.

To learn more about Hippocratic AI, visit www.HippocraticAI.com or tune into Munjal Shah’s upcoming session at NVIDIA GTC on Wednesday, March 20 at 10:30 a.m. PT. Register now at <https://www.nvidia.com/gtc>.

About Hippocratic AI

Hippocratic AI’s mission is to develop the first safest focused Large Language Model (LLM) for healthcare. The company believes that a safe LLM can dramatically improve healthcare accessibility and health outcomes in the world by bringing deep healthcare expertise to every human. No other technology has the potential to have this level of global impact on health. The company was co-founded by CEO Munjal Shah, alongside a group of physicians, hospital administrators, healthcare professionals, and artificial intelligence researchers from El Camino Health, Johns Hopkins, Washington University in St. Louis, Stanford, Google, and NVIDIA (NASDAQ: NVDA). Hippocratic AI has received a total of \$120M in funding and is backed by leading investors including General Catalyst, Andreessen Horowitz, Premji Invest and SV Angel. For more information on Hippocratic AI go to www.HippocraticAI.com.

Press Contact

LaunchSquad for Hippocratic AI
 hippocraticai@launchsquad.com

###