卫宁健康
WINNING HEALTH

intel

3X increase

in inference performance compared to 3rd Gen Intel® Xeon® Scalable Processors.[1]

Winning Health Expedites the Implementation of LLMs in Traditional Medical Scenarios

Winning Health Technology Group Co., Ltd. was founded to "enhance science and technology and improve people's health." Building on its leading medical Large Language Model (LLM) WiNGPT, Winning Health collaborated with Intel in graph optimization and weight-only quantization on 5th Gen Intel® Xeon® Scalable processors with Intel® Advanced Matrix Extensions for model inference. As a result, WiNGPT offers healthcare institutions optimized LLM performance, enhanced application experience, and improved cost-effectiveness. Because the solution uses CPUs for inference, healthcare institutions can flexibly allocate computing power between LLM inference and other IT applications as needed, which improves the agility and flexibility of computing power allocation.

"Through our collaboration with Intel, we have found a CPU-based LLM inference solution that not only meets the performance requirements but also offers cost advantages, helping accelerate the deployment of LLMs in hospitals, while providing intelligent knowledge services across various hospital scenarios."

Zhao Daping
Vice President and CTO, Winning Health

Products and Solutions
5th Gen Intel® Xeon® Scalable Processors
Intel® Advanced Matrix Extensions

Industry
Software Development

Organization Size
1,001 – 5,000

Country
China

Learn more
White Paper

1 For more complete information about performance and benchmark results, visit https://www.intel.com/content/www/us/en/customer-spotlight/stories/winning-health-customer-story.html