News and Perspectives

# Beyond Kimi K3: Spillover Effects of Frontier AI Competition on Health Care

Tejas S Athni, JMIR Correspondent

## Abstract

The July release of Kimi K3, an open-weight frontier AI model, is the latest milestone in the heated competition between AI developers in China and the United States. In this *News and Perspectives* article, JMIR Correspondent Tejas S Athni reports on how health care systems in the United States may benefit from the race to develop frontier AI.

**Key Takeaways:**

- Competition among frontier AI developers can generate downstream benefits for health care through spillover effects on capability, cost, and customizability.
- Health care organizations can capture these benefits without developing frontier foundation models themselves.
- As global frontier AI competition accelerates, health care leaders should anticipate how these spillovers, among others, may expand the range and scale of clinical AI applications.

*Tejas S Athni, MS, is an MD-PhD Candidate at Harvard Medical School. In this op-ed, he reflects on the potential health care implications of the intense global competition to develop leading-edge frontier AI models.*

The race to build the world's most capable generative AI models is accelerating and has become one of the defining competitions of this decade. Successive releases of new frontier models from both American and Chinese AI companies promise stronger reasoning, greater efficiency, and lower costs. The past two years have seen a dizzying cadence of blockbuster frontier model releases from American firms including OpenAI, Anthropic, and Google DeepMind, alongside Chinese firms such as DeepSeek, Moonshot AI, Alibaba, and Tencent. Each has pushed forward the state of the art. The consequences of this rapid pace of innovation extend far beyond Silicon Valley and Shenzhen. Health care has emerged as one of the largest downstream beneficiaries, deploying AI at greater than twice the rate of the broader economy.

Health care organizations rarely develop frontier foundation models themselves. Instead, they commercially procure finished AI products, much like they do with medical devices and pharmaceuticals, or co-develop them with technology and health-tech firms. In this setting, advances achieved by frontier AI can rapidly diffuse into health care. Providers, hospitals, and patients all benefit from the improving capabilities of AI models. The recent release of Kimi K3 in July 2026 by the Chinese startup Moonshot AI, rivaling the capabilities of flagship models from Anthropic and OpenAI, provides a timely reminder that the accelerating wave of advances in frontier AI generates spillover effects for health care.

## Kimi K3 and the Accelerating Wave of Innovation

Founded in 2023 by Yang Zhilin, Beijing-based Moonshot AI has quickly ascended among the ranks of China's foremost frontier AI companies as one of the country's 6 "AI Tigers." Following the Kimi K3 release, Moonshot AI surpassed its initial funding goals and has reportedly raised approximately US $3.5 billion in funds, with backing from several groups including Alibaba, Tencent, and the state-backed National Artificial Intelligence Industry Investment Fund. The new Kimi K3 release is a 2.8 trillion parameter, open-source, mixture-of-experts model designed to maximize performance. The model approaches the performance of the latest US flagship models—Anthropic's Claude Fable 5 and OpenAI's GPT 5.6 Sol—while outperforming several previous-generation frontier models on benchmarks such as coding, reasoning, and agentic tasks. Kimi K3 joins the ranks of Chinese AI models, including DeepSeek-R1, Qwen3, and GLM-4.5, that are rapidly narrowing the performance gap with US peer models, despite constraints on access to cutting-edge NVIDIA chips resulting from US export restrictions.

Whether Kimi K3 ultimately ranks above GPT or Claude is irrelevant, as benchmark leadership changes quickly. The more important trend is that competition between global AI companies, namely American and Chinese firms, is accelerating the pace of innovation. Health care organizations could rapidly benefit as successive generations of frontier AI are incorporated into commercial health care products, use cases, and clinical workflows. These benefits are likely to diffuse into health care via what I would refer to as capability, cost, and customizability spillover effects.

## Capability Spillover

Advances in general-purpose frontier foundation models could be inherited by downstream health care applications built atop them. For example, ambient documentation platforms such as Microsoft Dragon Copilot, Suki, Abridge, and Nabla continuously incorporate improvements in underlying foundation models (not necessarily Kimi K3) to produce faster, more faithful clinical notes. Similarly, provider-facing clinical AI assistants such as OpenEvidence and Doximity GPT can leverage stronger reasoning and larger context windows. As a result, these tools more effectively synthesize scientific literature and retrieve evidence on demand. Improvements in multimodal reasoning capabilities may also enhance AI-enabled imaging platforms such as PathAI and Paige. As competition drives advances, including successors to OpenAI's GPT 5.6, Anthropic's Claude Fable 5, Google's Gemini 3, and DeepSeek-R1, downstream health care applications inherit these capabilities.

A theoretical example to illustrate the mechanism: suppose that continued competition between US and Chinese AI firms produces a model capable of reliably reasoning across tens of millions of tokens of context—an advancement beyond the approximately 1-million-token context windows offered by leading frontier models today. A clinical AI application could inherit this capability, enabling reliable synthesis of years of longitudinal notes, laboratory results, imaging, medications, and other clinical data—even when these records span tens of millions of tokens—into a coherent, interactive representation of the patient's history. Though the health care application did not independently develop the million-token context windows, the broader foundation model could translate into entirely new capabilities for clinical care.

## Cost Spillover

Competition among frontier AI developers is driving down the cost of deploying state-of-the-art models that undergird clinical applications. Recent Chinese frontier models, including Kimi K3 and DeepSeek-R1, have emphasized near-frontier performance with substantially lower inference costs. This has placed competitive pressure on American developers to improve both performance and affordability in parallel. As a result, this cycle produces increasing capability per unit dollar. This means that clinical AI can more readily be deployed at scale rather than restricted to specific use cases or in certain environments. For instance, across the hospital ecosystem, AI is increasingly used both in the back end and at point of care, whether to draft answers to patient messages on the patient portal, assist in compiling hospital discharge summaries, complete prior authorization requests, or facilitate billing/coding workflows.

Take, for example, a clinical AI task that costs \$1 per patient encounter using a frontier model such as Kimi K3. If competition produces a model with comparable performance at \$0.10 per encounter, a health system could move from deploying this system selectively toward deploying it across thousands of routine patient interactions. Frontier AI competition can expand the scope of health care applications that are economically viable for the system.

## Customizability Spillover

Health care organizations could adapt and configure frontier AI models to their own clinical workflows and institutional priorities. Competition between proprietary and open-weight models is expanding these options. Open-weight models such as Kimi K3 can be deployed outside the developer's original cloud context, giving health care organizations greater flexibility over where models run. This creates an important alternative to relying exclusively on closed models from companies like OpenAI, Anthropic, and Google. For instance, a health care organization could deploy an open-weight model within its own secure computing environment, connect it to internal databases, and tailor its behavior to local contextual information without compromising sensitive patient data or the Health Insurance Portability and Accountability Act.

Importantly, this kind of spillover can provide health care with more choices about where and how these models are deployed. Take, for example, a small academic medical center seeking to deploy an AI assistant for clinical oncology evidence retrieval. This medical center could deploy an open-weight frontier model within its own secure computing environment and connect it to in-house tumor board records, institutional treatment protocols, and clinical trial databases. This flexibility may be particularly valuable for smaller institutions that lack the resources to develop or co-develop their own proprietary models, but still require varying degrees of control over patient data.

> *"Health care systems stand to capture gains in capability, cost, and customizability, even without developing the underlying foundation models themselves.*

## Benefiting From the AI Competition

The significance of Kimi K3 and the AI competition extends beyond frontier AI model rankings. Health care systems stand to capture gains in capability, cost, and customizability, even without developing the underlying foundation models themselves. Recognizing these trends may help health care

leaders anticipate how the continued future of frontier AI competition will reshape the ecosystem of clinical AI.

**Keywords:** artificial intelligence; AI; generative AI; frontier AI; clinical AI; health care; foundation models; AI competition; digital health; Kimi K3; Moonshot AI

*Please cite as:*
*Athni TS*
*Beyond Kimi K3: Spillover Effects of Frontier AI Competition on Health Care*
*J Med Internet Res 2026;28:e109041*
*URL: https://www.jmir.org/2026/1/e109041*
*doi: 10.2196/109041*