

# MedGPT - AI Digital Twin of Doctors

**Category:**

Best Digital Health Solution

**Company Name:**

Future Doctor Inc

**Number of employees:**

51-200

**Turnover and/or Funding:**

2025 expected revenue of US\$20 million

**Product/Solution Name:**

MedGPT - AI Digital Twin of Doctors

**Corporate Name:**

Future Doctor Inc

**Date of Approval:**

2024-01-08

**Indications:**

It is indicated for use in primary care, chronic disease management, and digital health settings, where it provides expert-level diagnostic reasoning, personalized treatment suggestions, and longitudinal patient follow-up

**Therapeutic Areas:**

Leveraging AI-powered digital twin of doctors, we provide full-spectrum AI medical services covering all specialties and the entire care journey across all clinical departments

**General Information File Document upload:**

N/A

## **Background information and need for drug / device:**

### Corporate History

Founded in 2014 and originally built as a professional doctor community, Medlinker Inc. ("Medlinker") was among the pioneers of doctor community and internet hospitals in China and has grown into the country's largest online chronic disease management platform.

As a current subsidiary of Medlinker, Future Doctor Inc., is a world leading AI-driven healthcare services company. In 2023, we launched MedGPT, China's first generative AI model for disease diagnosis and treatment, and it's the first generative AI algorithm in the country approved for such functions.

At the core is our commitment to accurate and expert-level diagnosis, which ensures safe, effective, and cost-optimized treatment downstream.

Our mission is to democratize access to high-quality care, enabling every patient to benefit from expert-grade diagnostic services, thereby reducing common issues such as misdiagnosis, underdiagnosis, and inappropriate medication use, and helping build a safer, more efficient, and equitable healthcare system.

Key milestones include:

- Launched MedGPT, the proprietary medical AI model, in 2023
- Conducted the world's first clinical trial comparing with real doctors in Jun 2023, one year ahead of ChatGPT
- Commercialization started in late 2024

### Condition & Need for Solution

Globally, more than half the world's population lack full coverage of essential health services, with 2 billion pushed into or further into poverty due to out-of-pocket medical expenses.

In many regions, primary care systems are absent, overwhelmed, or fragmented. Even in urban areas, low doctor-to-patient ratios and limited consultation times constrain the quality of medical interactions.

Clinicians face growing administrative burdens; for instance, a survey in Germany revealed that senior physicians spend over 5 hours daily on administrative tasks, leaving only about 2.5 hours for direct patient care.

Diagnostic errors are a significant concern, affecting approximately 12 million adults annually in the U.S., equating to 1 in 20 outpatient visits. These errors can lead to delayed treatments and increased healthcare costs. In China, diagnostic errors are similarly widespread. Published data indicate an average misdiagnosis rate of approximately 30%, with significantly higher rates - often exceeding 40% - observed in complex cancers, and over 50% among patients with rare diseases.

These challenges underscore the urgent need for scalable, AI-driven solutions that enhance diagnostic accuracy, reduce administrative workloads, and expand access to high-quality care.

## Reference

<https://www.worldbank.org/en/news/press-release/2023/09/18/billions-left-behind-on-the-path-to-universal-health-coverage>

<https://www.medscape.com/viewarticle/german-doctors-spend-three-quarters-each-day-paperwork-2024a1000isp>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC4145460/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC5502242/>

## **Background File Document upload:**

N/A

## **History of the development of the solution/product:**

Our proprietary model - MedGPT - was built on 2 billion de-identified doctor-patient interaction records, 8 million structured clinical diagnosis and treatment records, and 41,000+ latest global medical practice guidelines. Over 100 medical experts participated in reinforcement learning from human feedback to calibrate the model's reasoning logic.

MedGPT features a fast-and-flow dual system architecture. The fast system integrates a Mixture of Experts (MoE) for natural language understanding and rapid multi-turn dialogue. The slow system focuses on deep knowledge and clinical reasoning, using expert Chains of Thought, RAG and multi-agent collaboration mechanisms. It verifies and reviews the fast system's outputs to ensure accuracy. Through intent recognition, information routing, and multi-turn validation, this architecture simulates medical expert thinking and supports full-scenario AI services.

MedGPT also models expert clinical experience, diagnostic reasoning, and communication methods to create digital twins of medical experts, enabling expert

knowledge to be accessed across time and geography.

When MedGPT connects patients with doctors, it initiates a data flywheel - new interactions feed the training loop, continuously improving the AI doctor twin models. This enhances performance, generalization, and diagnostic reasoning.

In Jun 2023, we conducted the world's first real world study evaluating the consistency and accuracy of MedGPT compared to human-docotors. The entire event was streamed live. 120 real patients and 10 attending physicians from West China Hospital of Sichuan University (Grade3A hospital) participated in the trial.

The patients were prospectively recruited across 7 clinical departments (cardiology, gastroenterology, endocrinology, nephrology, orthopedics, and urology) for parallel and independent performance assessments. A total of 91 valid clinical reports were reviewed by 7 chief experts from Peking University People's Hospital, China-Japan Friendship Hospital, Fuwai Hospital Chinese Academy of Medical Sciences and Beijing Friendship Hospital (all Grade3A hospitals in China). The average working experience of these doctors was approximately 12 years.

Evaluations covered 7 areas, including medical history taking, diagnostic accuracy, treatment recommendation, ancillary test planning etc. The final experimental results showed that the MedGPT achieved a score of 7.2, while the corresponding real physicians scored 7.5. The consistency in scoring results between the AI doctor and the real physicians reached 96%.

Between May - Jun 2025, we further developed the Clinical Safety-Effectiveness Dual-Track Benchmark (CSEDB), a multidimensional framework built on clinical expert consensus, encompassing 30 criteria covering critical areas like critical illness recognition, guideline adherence, and medication safety, with weighted consequence measures. 32 specialist physicians developed and reviewed 2,069 open-ended Q&A items aligned with these criteria, spanning 26 clinical departments to simulate real-world scenarios. Benchmark testing of MedGPT vs. other 5 LLMs (including OpenAI-o3, Deepseek-R1 and etc.) demonstrated a consistant outperformance of MedGPT in performance\*.

To ensure safety and ethics, we also use a dual-layer approach: \"AI Recommendation + Physician Review.\" AI outputs are always reviewed and approved by licensed physicians. Every recommendation is traceable and version-controlled to meet strict regulatory standards, reinforcing safety, accountability, and compliance. This framework not only reinforces patient safety and physician accountability but also ensures compliance with evolving national and international healthcare regulations.

\*<https://arxiv.org/abs/2507.23486>

## **Development File Document upload:**

N/A

## **Why this drug or device is innovative, the broad implications for future research, and/or how it will improve the human condition:**

### **A NEW CARE DELIVERY MODEL THROUGH PHYSICIAN DIGITAL TWINS**

MedGPT represents a paradigm shift in AI for healthcare - moving beyond simple symptom checkers or decision-support tools, to a comprehensive simulation of cognitive process of thinking and workflows of top tier physicians. It is the first large-scale deployed digital twin that can conduct full clinical interactions, enabling accurate, efficient, and personalized care from the first point of contact.

This digital twin system breaks the long-standing \"impossible triangle\" in healthcare - quality, cost, and efficiency - and redefines the patient and provider experience.

Our product is capable of addressing over 95% of common clinical inquiries, while reducing treatment costs by up to 86%. In the clinical studies comparing with real doctors, we achieved a 96% diagnostic consistency with attending physicians from top tier (Grade3A) hospitals in China. Additionally, it enables a 10x increase in physician productivity, significantly increase the supply of physician resources.

Ultimately, MedGPT enables a proactive, full-lifecycle health management to covering prevention, diagnosis, treatment, and rehabilitation and helps democratize access to top-tier medical expertise for every patient, regardless of geography or provider availability.

What makes it innovative:

- Human Digital Twin Paradigm: More than just an AI tool, it serves as a trained, adaptive medical role capable of interacting with both patients and clinicians.
- Multimodal Input/Output: Future-oriented integration of text, audio, imaging.
- Doctor-AI Copilot Model: Enhances physicians' ability for proactive patient management through automated medical record generation, evidence-based treatment recommendations, and active follow-up.
- Hybrid B2B2C Model: Serves patients either through partnerships with enterprises or by direct deployment on online platforms within hospitals, pharmacies, and other healthcare settings.

Implications for the future:

This model can dramatically reduce the human cost of healthcare delivery, especially in under-resourced settings. It also creates a scalable infrastructure for:

- Drug adherence programs
- AI-supported clinical trials
- Remote monitoring and diseases prevention and control

By enabling the replication and scaling of clinical intelligence, MedGPT could redefine how care is delivered globally.

### **Innovation File Document upload:**

N/A

### **Please provide appropriate references (PubMed, Abstract, Website):**

1. Shirui Wang, Zhihui Tang, et al. A Novel Evaluation Benchmark for Medical LLMs: Illuminating Safety and Effectiveness in Clinical Domains (2025)

Large language models (LLMs) hold promise in clinical decision support but face major challenges in safety evaluation and effectiveness validation. We developed the Clinical Safety-Effectiveness Dual-Track Benchmark (CSEDB), a multidimensional framework built on clinical expert consensus, encompassing 30 criteria covering critical areas like critical illness recognition, guideline adherence, and medication safety, with weighted consequence measures. Thirty-two specialist physicians developed and reviewed 2,069 open-ended Q&A items aligned with these criteria, spanning 26 clinical departments to simulate real-world scenarios. Benchmark testing of six LLMs revealed moderate overall performance (average total score 57.2%, safety 54.7%, effectiveness 62.3%), with a significant 13.3% performance drop in high-risk scenarios ( $p < 0.0001$ ). Domain-specific medical LLMs showed consistent performance advantages over general-purpose models, with relatively higher top scores in safety (0.912) and effectiveness (0.861). The findings of this study not only provide a standardized metric for evaluating the clinical application of medical LLMs, facilitating comparative analyses, risk exposure identification, and improvement directions across different scenarios, but also hold the potential to promote safer and more effective deployment of large language models in healthcare environments.

<https://arxiv.org/abs/2507.23486>

2. Maddox TM, Embí P, Gerhart J, Goldsack J, Parikh RB, Sarich TC. Generative AI in Medicine - Evaluating Progress and Challenges. N Engl J Med. Published online April 10, 2025. doi:10.1056/NEJMs2503956

<https://pubmed.ncbi.nlm.nih.gov/40208922/>

3. Tam, T.Y.C., Sivarajkumar, S., Kapoor, S. et al. A framework for human evaluation of large language models in healthcare derived from literature review. npj Digit. Med. 7,

258 (2024).

<https://pubmed.ncbi.nlm.nih.gov/39333376/>

4. Johri, S., Jeong, J., Tran, B.A. et al. An evaluation framework for clinical use of large language models in patient interaction tasks. *Nat Med* 31, 77-86 (2025).

<https://pubmed.ncbi.nlm.nih.gov/39747685/>

5. Wornow, M. et al. The shaky foundations of large language models and foundation models for electronic health records. *NPJ Digit. Med.*

<https://pubmed.ncbi.nlm.nih.gov/37516790/>

6. [2501.09484] Exploring the Inquiry-Diagnosis Relationship with Advanced Patient Simulators

<https://arxiv.org/abs/2501.09484>

7. [2404.18416] Capabilities of Gemini Models in Medicine

<https://arxiv.org/abs/2404.18416>

8. [2411.14487] Ensuring Safety and Trust: Analyzing the Risks of Large Language Models in Medicine

<https://arxiv.org/abs/2411.14487>

9. Wang R, Luo W, Liu Z, et al. Integration of the Extreme Gradient Boosting model with electronic health records to enable the early diagnosis of multiple sclerosis. *Mult Scler Relat Disord.*

<https://pubmed.ncbi.nlm.nih.gov/33276240/>

10. Jia T, Wu C, Hu X, et al. Physicians' Knowledge, Attitude, and Experience of Pharmacogenomic Testing in China. *J Pers Med.* 2022;12(12)

<https://pmc.ncbi.nlm.nih.gov/articles/PMC9783535/>

11. Singhal, K., Tu, T., Gottweis, J. et al. Toward expert-level medical question answering with large language models. *Nat Med* 31, 943-950 (2025).

<https://pubmed.ncbi.nlm.nih.gov/39779926/>

12. Tu T, Schaekermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence. *Nature.*

<https://arxiv.org/abs/2401.05654>

13. Xi Chen, Huahui Yi, et al. Enhancing diagnostic capability with multi-agents conversational large language models. *npj Digital Medicine* volume 8, Article number: 159 (2025)

<https://www.nature.com/articles/s41746-025-01550-0?fromPaywallRec=false>

## References File Document upload:

N/A