MedLM

A collection of foundation models tuned to follow natural language instructions for tasks in medicine, such as question answering and creating draft summaries.

Overview

MedLM is a family of foundation models fine-tuned for the healthcare industry. Med-PaLM 2 is one of the medically-tuned text-based large language models (LLM) developed by Google Research that powers MedLM. **MedLM-large** is a larger model and has undergone testing by customers in preview. **MedLM-medium** is a newer, smaller, model that offers advantages of a larger context limit and higher throughput. These models have been tuned to follow natural language instructions for tasks in medicine, such as question-answering and creating draft summaries.

Regulatory Information

MedLM Intended Use

MedLM is based on Google Research's medically-tuned large language model, Med-PaLM 2. It is intended to be used for question answering and creating draft summaries from existing documentation - to be reviewed, edited, and approved by the user before use. MedLM is also used for educational purposes for a Healthcare Professional (HCP) to engage in medical questioning and answering to help support the HCP.

⚠ The output of the model(s) is not considered final, and gives only a draft response which the HCP should review. MedLM must not be used for any diagnostic or therapeutic purpose, and is not to be used in direct patient care.

Conditions of Use & Out-of-Scope Applications

- MedLM customers and users must abide by the Generative AI Prohibited Use Policy, Google Cloud Platform Service-Specific Terms, Terms of Service, Acceptable Use Policy, User Guide, and other product documentation.
- As part of Service-Specific Terms, customers may not use MedLM for clinical purposes (for clarity, non-clinical research, scheduling, and other administrative tasks are not

restricted), to provide medical advice, or in any manner that is overseen by or requires clearance or approval from a medical device regulatory agency.
    - Direct patient use is prohibited. The product functions as an assistive tool for a clinician, HCP, or knowledge worker with a high degree of expertise, education, or experience in the healthcare and life sciences industry.
    - Use of MedLM as a Software as a Medical Device is prohibited.
- The intended use for MedLM is to draft documents and responses that would be reviewed by a "human in the loop" before usage.
- We recommend usage of MedLM solely for Medical Q&A and Summarization use cases at this stage:
    - Long form Q&A
    - Multiple choice Q&A
    - Summarizations, such as creation of After Visit Summaries or History & Physical Examination notes
- Examples of medical device uses that are not permitted include (but are not limited to):
    - Analysis of patient records, prescription patterns, geographical data and so forth, to identify patients with possible diagnosis of opioid addiction.
    - Analysis of patient-specific medical information to detect a life threatening condition, such as stroke or sepsis, and generate an alarm or an alert to notify a HCP.
    - Analysis of patient-specific medical information found in the medical records, including the most recent mammography report findings, to provide a list of follow-up actions or treatment options.
    - Analyzing prioritized list of FDA-authorized depression treatment options to an HCP based on an analysis of reported outcomes in a database of clinical studies using medical information (e.g., diagnosis and demographics) from the patient's medical record.

MedLM is currently only available to allow-listed customers in the U.S.

MedLM is not intended to be used as a medical device. Customer use cases must be consistent with the intended use and conditions of use. Q&A should only be used for educational purposes and summarization outputs must always be independently reviewed and verified by the user based on their clinical judgment.

Customer Responsibilities

MedLM has been developed with trained and licensed healthcare practitioner users in mind. Google Cloud customers and end users should understand that LLMs and Generative AI are inherently probabilistic and may not always be accurate. Without adequate consideration or

controls by customers, use of Generative AI models in healthcare may constitute a hazard to patients due to inaccurate content, missing content, misleading content, or biased content.

- Customers should implement appropriate hazard mitigations for all MedLM uses, such as adequate practitioner education, training, assessment of equity, and appropriate technical controls.
- Customers must also perform their own evaluations for performance and safety to ensure prevention of harm for their use cases.

MedLM may produce less accurate results for some groups compared to others depending on the question and how it is posed. Customers should be aware that differing performance of outputs of the model across demographic groups has the potential to exacerbate health inequities and perpetuate harmful biases. Such inaccuracies of outputs are not unique to MedLM and often stem from multiple factors, such as existing social and structural inequities, medical misconceptions, negative stereotypes, and lack of diversity in training data.

- Customers should consider implementing equity-focused evaluations and mitigations. This includes assessing model performance and behavior for intended use cases within various populations (e.g., race/ethnicity, socioeconomic status (SES), geography, gender identity, sexual orientation, age, language preference, caste, etc.); obtaining feedback on performance; engaging interdisciplinary experts and external partners that specialize in defining and addressing social and structural aspects of health; and conducting continuous monitoring efforts to assess and address issues of bias.

Warnings

⚠ Before activating Production use for MedLM, customers must reach out to Google Product Team to discuss usage.

⚠ MedLM has not been designed or developed to be used as a medical device. Any output should be verified by a Healthcare Professional (HCP), and no direct diagnosis should be claimed.

⚠ The generated output may not always be completely reliable. Due to the nature of LLMs and Generative AI, outputs may have incorrect or biased (e.g., stereotypes or other harmful content) clinical information and should be reviewed. All summaries or answers should be considered draft and not final.

⚠ If Vertex AI detects content that violates our policies, including Google Cloud Platform Acceptable Use Policy and Generative AI Prohibited Use Policy, a response is not returned.

⚠ When used by HCPs for Q&A purposes, MedLM is only intended for use as an educational tool for medical training or to reinforce the HCP's prior training.

⚠ LLM output may not follow the exact format laid out in the prompt. The prompt design to extract information for each field should take into account that the format may deviate from the original (e.g., dashes in field names, exact capitalization of letters).

Use cases

- **Question answering**: Provide draft answers to medically-related questions, given as text.
- **Summarization**: Draft a shorter version of a document (e.g. an After Visit Summary or History & Physical Examination note) that incorporates pertinent information from the original text.

Request access

Access to the MedLM models is restricted. To request access, contact your Google Cloud account team.

Provide feedback

Your feedback throughout your experience will help us improve future model versions and ensure that we continue to deliver the best possible experience for our users. Contact medlm-feedback@google.com and copy your Google Cloud account team and Google Cloud Customer Engineer (CE). This email address is not for immediate support. To request immediate support, contact your Google Cloud account team or Google Cloud Customer Engineer (CE).

Email responses will be used as Feedback under the terms of your Agreement for Google Cloud Services and will be collected in accordance with the Google Cloud Privacy Notice. Do not include any personal information (names, email addresses) in this feedback form or other data that is sensitive or confidential. Note that data may be reviewed using both human reviewed and automated processing.

Report abuse

You can report suspected abuse of the MedLM API, any generated output that contains inappropriate material, or inaccurate information in Report suspected abuse on Google Cloud. In the **Google Cloud Platform Service** list, select **Cloud AI**.

Data Overview

MedLM-medium and **MedLM-large** were developed by aligning Google foundation models to the medical domain via domain-specific finetuning. Finetuning was done using datasets of medical question-and-answer pairs and medical summaries. The instruction finetuning was applied using specific finetuning protocols. This finetuning allowed the models to learn the specific vocabulary and relationships that are important in the medical domain.

MedLM models have been trained on large-scale publicly available web information, public datasets and de-identified licenced private datasets with health and medical information – with tasks including question-answering, summarization of medical records and transcripts, and simplification of medical text.

Google Cloud customer data is controlled by the customer and never used to train Google's models.

Detailed information on MedLM datasets, modeling, finetuning and evaluations can be found in Towards Expert-Level Medical Question Answering with Large Language Models and includes:

- MedQA (USMLE)
- MedMCQA
- MMLU Anatomy
- MMLU-College biology
- MMLU-College medicine
- MMLU-Professional medicine
- PubMedQA
- MMLU Clinical knowledge
- MMLU Medical genetics

MedLM was also trained on de-identified licensed datasets that do not contain PII for summarization and document generation from medical notes and transcripts.

MedLM Research Publications

MedLM may be useful for drafting long-form answers to health questions and summaries from existing medical documentation (as judged by panels of physicians and users). MedLM is based on Med-PaLM 2, and the first version of Med-PaLM, published in *Nature* in July 2023, was the first AI system to surpass the pass mark (60%) on U.S. Medical Licensing Examination (USMLE)-style questions. Med-PaLM 2, introduced in March 2023, was the first to reach human expert level on answering USMLE-style questions.

Bias and Equity

Users should be aware that as with every model, bias is a known risk and can be linked to factors such as the size, diversity, and quality of data sets, as well as existing social and structural inequities. The current versions of the MedLM models use datasets as detailed in Data Overview.

Safety

Customers are responsible for ensuring MedLM is only used in accordance with its intended use and the agreed upon terms and conditions (Please also refer to the Customer Responsibilities section above).

Areas of safety risk include but are not limited to:

- Hallucinations
- Inaccurate information
- Bias
- Outputs containing toxic language or content
- Missing information

To help prevent harmful content in the output, content processed through the MedLM API is assessed against a list of safety attributes. If Vertex AI detects content that violates Google's policies, including Google Cloud Platform Acceptable Use Policy and Generative AI Prohibited Use Policy, a response is not returned.

To utilize this technology safely and responsibly, it is also important to consider other safety risks specific to your use case, users, and business context.

To learn more, please reference Google's recommendations for Responsible AI Practices.

Privacy

Google Cloud provides a secure infrastructure that supports storage and processing of PHI in a HIPAA-compliant manner, but the customer is responsible for ensuring that the environment and applications that they build on top of Google Cloud Platform are properly configured and secured according to HIPAA requirements. This is often referred to as the “shared security model” in the cloud. Ultimately customers are responsible for evaluating their own HIPAA compliance.

Google Cloud’s customers retain control over their data. In healthcare settings, access and use of patient data is protected through the implementation of Google Cloud’s reliable infrastructure and secure data storage that support HIPAA compliance, along with each customer’s security, privacy controls, and processes. In addition, MedLM, Vertex AI Search, Healthcare Data Engine, and Healthcare API, along with many other Google Cloud products support HIPAA compliance.

Google Workspace and Google Cloud enable HIPAA compliance for our customers subject to its requirements.

We build tools that a single customer (e.g., a hospital or primary care group) can use with their own patients’ data. The data is siloed, access controlled, and auditable. We do not combine data across partners, and we would not be allowed to under our agreements or the law.

Versions

Resource ID	Release date	Release stage	Description
MedLM-large	2023-12-13	General Availability - Allowlisted	Initial release
MedLM-medium	2023-12-13	General Availability - Allowlisted, Restricted	Initial release