日本語マルチモーダル生成Al サービス ~Geminiを例に~

国立情報学研究所 コンテンツ科学研究系助教 栗田 修平



栗田修平 (Shuhei Kurita)

国立情報学研究所 助教

研究テーマ: テキストと実世界・物理世界を繋げる

Career

2019年 3月 博士(情報学) 京都大学 黒橋·河原研究室 自然言語処理

2019年 4月 - 2023年3月 理化学研究所 特別研究員

2020年 1月 - NYU訪問研究員

2020年 12月 -JST さきがけ

2023年 4月 - 理化学研究所 研究員

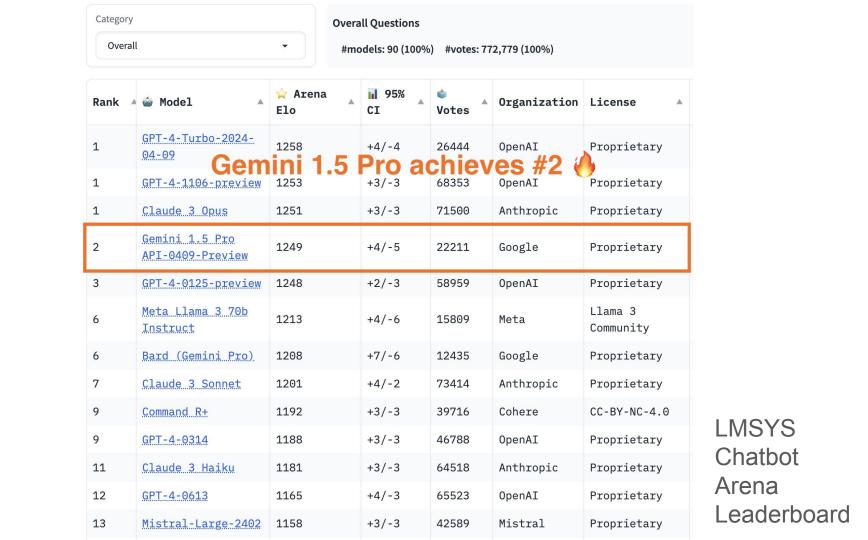
2024年 4月 - 現職

Selected Publications

- 1. RefEgo: Referring Expression Comprehension Dataset from First-Person Perception of Ego4D, *Shuhei Kurita, Naoki Katsura, Eri Onami, (ICCV2023).*
- 2. ScanQA: 3D Question Answering, Daichi Azuma(*), Taiki Miyanishi(*), Shuhei Kurita(*) and Motoaki Kawanabe. (CVPR2022). (*): eq. cont.
- Generative Language-Grounded Policy in Vision-and-Language Navigation with Bayes' Rule, Shuhei Kurita and Kyunghyun Cho, (ICLR2021).
- 4. Reconstructing neuronal circuitry from parallel spike trains, Ryota Kobayashi, Shuhei Kurita, ..., Shinomoto Shigeru, Nature Communications (2019).
- Neural Joint Model for Transition-based Chinese Syntactic Analysis, Shuhei Kurita, Daisuke Kawahara and Sadao Kurohashi, (ACL2017).
 Selected as Out-standing Paper in ACL2017.

Welcome to the Gemini era

Chat with Gemini →



目次: Geminiの特徴

- Long contextへの対応
- 数学能力
- グラフの描画
- 図表の読解
- 動画の理解
- Med-Gemini
- 動画生成 (Veo)

Get more done with Gemini: Try 1.5 Pro and more intelligent features

May 14, 2024 5 min read We're bringing Gemini 1.5 Pro to Gemini Advanced subscribers in over 35 languages, along with a 1 million token context window, a new conversational experience and tools that let Gemini take action on your behalf.



Sissie Hsiao

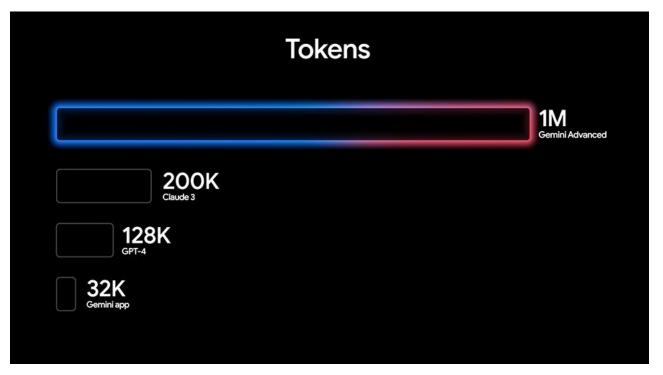
Vice President and General Manager, Gemini experiences and Google Assistant







Geminiの特徴: 非常にLong contextに対応している



言語モデルの入出力トークンサイズ世界最長の1M

Geminiの特徴: 非常にLong contextに対応している

- 実は テクニカルペーパー"Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context" では10Mトークンまで実験している
- Google Al Studio上では、日本語では 1 token = 1.4~1.8文字ほど
 - 夏目漱石『こころ』(青空文庫)全体では 175,126 / 98,654 = 1.8 文字 /トークン(ルビ入り)
 - 夏目漱石『草枕』 (青空文庫) 全体では 98,925 / 65,883 = 1.5 文字 /トークン(ルビ入り)
 - Gemini 1.0 Pro / Gemini 1.5 Flash / Gemini 1.5 Pro で同一トークン数のようである つまりトークナイザは変化なし?
- つまり公開されている1Mトークンで、**日本語で約150万文字対応している**
 - 『こころ』が17.5万文字であることを考えると、厚めの文庫本数冊くらいは入りそう
 - 長編として知られるプルーストの『失われた時を求めて』が原文で 961万=9.6M文字。 人類のたいていの小説は 10Mトークンあれば入ることになる。
- 「内部文書のQA」のレベルではRAG(検索+LMM)もfinetuningさえも 必要なくなる可能性がある。
 - Googleを信頼できる方は、おそらく必要な全ての情報をプロンプトに入力してしまえます
 - (どうしても信頼できない方は Google DriveとGMailの使用も再検討になるかもしれま ...)

Q. そんなに長いcontext、本当に見ているの?

A. 見てはいるらしい。

干し草山の針の実験

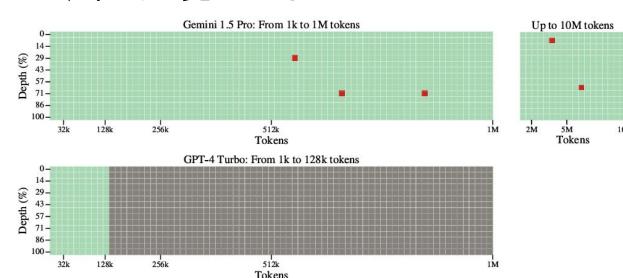
長いテキスト中に埋め込んだ「針」を 探す

https://x.com/aparnadhinak/status/174477129 5940669689

(右図)

GPT-4 TurboもGemini 1.5 Proもトークン限界まで見ている

右のテクニカルレポートでは、マルチモーダル入力(ビデオ・音声)でも同様の結果を報告している。



「針」とは "The special magic {city} numberis: {number}" というような隠し文字列のこと。Depthは「針」の文書中での出現位置を表す。

Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context

Q. そんなに長い context 本当に"読めて"いるの?

A. Geminiに読書感想文を書かせてみました。

手法:

読書感想文の題材として夏目漱石『こころ』**全文**を与えて、作文。 かかった時間はたったの20.8s

比較のため、全文を与えずに事前知識のみと比較。 こちらは8.1s

Prompt 1:

夏目漱石の名作「こころ」の読書感想文を書いてください。日本の高校生になったつもりで、あくまで学校教育で使われるような模範的かつ優等生的な読書感想文を作文してください。あらすじに簡単に触れながら登場人物の感情をまとめる形式が良いでしょう。ただし、小説に対して過度に感情移入する必要はありません。見出しなどを分ける必要はなく散文形式でお願いします。

Prompt 2: (『こころ』全文)

上記の小説を読んで読書感想文を書いてください。日本の高校生になったつもりで、あくまで学校教育で使われるような模範的かつ優等生的な読書感想文を作文してください。あらすじに簡単に触れながら登場人物の感情をまとめる形式が良いでしょう。ただし、小説に対して過度に感情移入する必要はありません。見出しなどを分ける必要はなく散文形式でお願いします。

Q. そんなに長い context 本当に"読めて"いるの?

まず、作品名のみ与えた場合。あらすじは間違ってはいない。

「K」や「先生」という登場人物は出てきたが詳細さにはかける。

(何度か試した中では、「K」と「先生」を取り違えるなど全くだめなケースもあった)

(栗田の主観評価だが) 万人受けする結論の書き方であ

Model 8.1s

夏目漱石の「こころ」を読了し、人間の心の奥底に潜む複雑な感情に触れることができた。物語は、語り手である「私」が、学生時代に偶然出会った「先生」との交流を通して、先生自身の過去、そして人間のエゴイズムや倫理観について深く考えさせられる過程を描いている。先生は過去に、親友であるKを死に追いやったという罪悪感を抱えて生きている。Kは先生と同じ女性を愛しており、その葛藤の中で先生は自己中心的になってしまったことを深く後悔していた。

先生はKへの贖罪として、妻との静かな生活の中で孤独を味わいながらも、過去の罪に向き合おうとしていた。しかし、過去の出来事は先生の心を蝕み続け、最終的に自殺という道を選んでしまう。先生の手記を通して明かされる彼の苦悩は、人間の心の脆さ、そして愛と罪悪感の複雑な絡み合いを浮き彫りにする。

「こころ」は単なる恋愛小説ではなく、人間存在の本質、倫理、そして罪と向き合うことの難しさを問いかける作品であると言える。現代社会においても、他者との関係の中で生じる葛藤や罪悪感は普遍的なテーマであり、先生の苦悩は現代人の心に深く訴えかけるものがある。漱石の巧みな文章と心理描写を通して、人間の心の奥深さを改めて認識させられた。

Model 20.8s 全文入力した場合 夏目漱石の「こころ」は、語り手の「私」が「先生」と出会った鎌倉での避暑から、先生とその妻との複雑な関係、そして先 牛の死に至るまでの過程を、内面的な視点から丁寧に描いた作品です。物語は「私」の視点を通して展開され、先牛の複雑な あらすじがやや詳細に 心情や、周囲の人物との関係が少しずつ明らかになっていきます。 なっている。 「私」は、鎌倉で出会った「先生」という謎めいた人物に強い興味を抱き、次第に親密になっていきます。しかし、先生は 地名・人名が正しく出 「私」に対して常に距離を置き、心の内をなかなか明かしません。先生は、過去の恋愛の失敗から深い傷を負い、人間不信に 現している。 陥っている様子がうかがえます。「私」は先生の心の奥底にある悲しみや苦悩に気づきながらも、その真相をなかなか理解す ることができません。 (栗田の主観評価) 物語が進むにつれて、先生の過去が断片的に明かされていきます。先生の妻である「静」との結婚、そして親友の死という衝 読書感想文としての洞 撃的な出来事、そして叔父からの裏切り…。これらの出来事が、先牛の心を大きく傷つけ、孤独と厭世的な思想へと導いたこと 察は深くはならないか がわかります。 もしれない。「衝撃的な 事実」など、ぼかして 一方、「私」は、先生を通して人間の複雑さや心の奥深さを学び、自身の成長を促されていきます。特に、先生の死後、先生 いる箇所もある。やは が残した手紙を読むことで、先生の心の闇、そして「私」への複雑な感情を知ります。それは、「私」にとって衝撃的な事実 り万人受けする結論 であり、同時に大きな学びとなりました。 の書き方である 「こころ」は、単に恋愛小説や人間関係を描いた作品ではなく、人間の心の複雑さ、そして人生における孤独や苦悩といった 普遍的なテーマを深く掘り下げています。先生という人物を通して、漱石は当時の社会に対する批判や、人間の生き方の問い 871文字で回答してき を投げかけていると言えるでしょう。 た。 この作品を読んだ私は、自分自身の心の奥底に眠る感情や、周囲の人々との複雑な関係について改めて考えるようになりまし た。漱石の言葉を通して、人間は誰しも孤独を抱え、様々な苦しみや葛藤を経験しながら生きていることを実感しました。 「こころ」は、現代社会においても色褪せない普遍的なテーマを扱った作品であり、今後も繰り返し読み返したい名作です。

Q. そんなに長いcontext本当に"読めて"いるの?

- 追加でちょっと細かいことを聞いてみる
- さすがに本文なしではお手上げのようです:

User

叔父からどんな裏切りを受けたのですか?教えてください。

Model 2.4s

申し訳ありませんが、私は「こころ」のあらすじを要約しただけで、登場人物の具体的な行動や詳細な人間関係については言及していません。「叔父」や「裏切り」といった情報は、私の回答に含まれていません。

「こころ」は登場人物の複雑な感情や人間関係を描いた作品です。詳細を知りたい場合は、実際に小説を読んでみることをお勧め します。

Q. そんなに長いcontext本当に"読めて"いるの?

● 追加でちょっと細かいことを聞いてみる (本文+Gemini Pro 1.5)

User

叔父からどんな裏切りを受けたのですか?教えてください。

Model 16.6s **⚠**

先生は両親を早くに亡くし、叔父の世話になっていました。先生は自分の家の財産を叔父に任せきりにして東京で学生生活を 送っていましたが、夏休みで帰省した際に叔父の態度がおかしいことに気づきます。

実は叔父は、先生の家の財産をこっそり自分のものにしていたのです。先生は叔父と話し合いをしようとしますが、叔父はそれをはぐらかし、結婚を急がせるなどして事態をうやむやにしようとしたため、先生は激しく憤りを感じます。

最終的に先生は、親戚の仲裁もあり、減額された財産を受け取り、故郷を去る決意をします。叔父のこの裏切り行為は、先生にとって大きな精神的ショックとなり、後の人間不信へとつながる一因となりました。

● 下線部はHallucinationかと思ったが、おおまかには正解しているようだった...

Q. そんなに長いcontext本当に"読めて"いるの?

- 題材に使用した文章には、節ごとに番号がついています。
- さらに詳しい質問もしてみた結果、根拠の箇所のパラグラフの番号を間違わずに出力できたこともありました。(当然ながら、間違えることや、明らかなHallucinationもありました)
- しかし、長い文章を読みこなす能力はタスクによっては非常に有益ではないかと考えます。

目次: Geminiの特徴

- Long contextへの対応
- 数学能力
- グラフの描画
- 図表の読解
- 動画の理解
- Med-Gemini
- 動画生成 (Veo)

数学能力:積分

以下の問題に答えなさい。回答は日本語で記述しなさい。

$$\int \frac{1}{\cos x} dx =$$

※Google AI Studioではlatex 形式の数式の表示には対応し ていないようです。 このスライドはlatex形式を数式 に変換して表示しています。 また一部を省略しています。 1/cos x の積分

$$\int \frac{1}{\cos x} dx$$
 を計算するには、少し工夫が必要です。直接積分する方法がないため、以下のような手順を踏みます。

1. 分母分子に cos x をかける

まず、積分を以下のように変形します。

$$\int rac{1}{\cos x} dx = \int rac{1}{\cos x} \cdot rac{\cos x}{\cos x} dx = \int rac{\cos x}{\cos^2 x} dx$$

2. cos^2 x を 1 - sin^2 x に置き換える

次に、三角関数の恒等式 \$\cos^2 x = 1 - \sin^2 x\$ を用いて、積分を以下のように書き換えます。

$$\int \frac{\cos x}{\cos^2 x} dx = \int \frac{\cos x}{1 - \sin^2 x} dx$$

3. 置換積分法を用いる

ここで、置換積分法を用います。 $\sin x = u$ 置くと、 $\cos x dx = du$ 」ます。これにより、積分は以下のように書き換えられます。

$$\int \frac{\cos x}{1 - \sin^2 x} dx = \int \frac{1}{1 - u^2} du$$

数学能力:積分

以下の問題に答えなさい。回答は日本語で記 述しなさい。

$$\int \frac{1}{\cos x} dx =$$

- 解答の方針はよい しかし、最後に係数忘れを している
- これは有名な積分であり、 解法をそこそこ覚えている のだろう

- $\frac{1}{1-u^2}$ は部分分数分解によって以下のように表せます。
 - - $\frac{1}{1-u^2} = \frac{1/2}{1+u} + \frac{1/2}{1-u}$

4. 部分分数分解

- 5. 各項を積分
- 部分分数分解された各項を積分すると、以下のようになります。

$$\int rac{1}{1-u^2}du = rac{1}{2}\int rac{1}{1+u}du + rac{1}{2}\int rac{1}{1-u}du \ = rac{1}{2} \mathrm{ln} \left| 1+u
ight| - rac{1}{2} \mathrm{ln} \left| 1-u
ight| + C$$

- 6. u を sin x に戻す

- 最後に、 $u = \sin x$ F戻すと、最終的な答えが得られます。 $\int \frac{1}{\cos x} dx = \frac{1}{2} \ln|1 + \sin x| - \frac{1}{2} \ln|1 - \sin x| + C$

対数関数の性質を用いて、答えを以下のように整理できます。 $\int rac{1}{\cos x} dx = \ln \left| rac{1 + \sin x}{1 - \sin x}
ight| + C$ ここで係数間違い!

グラフの描画

残念ながらGeminiは グラフの描画に対応し ていないようだ User

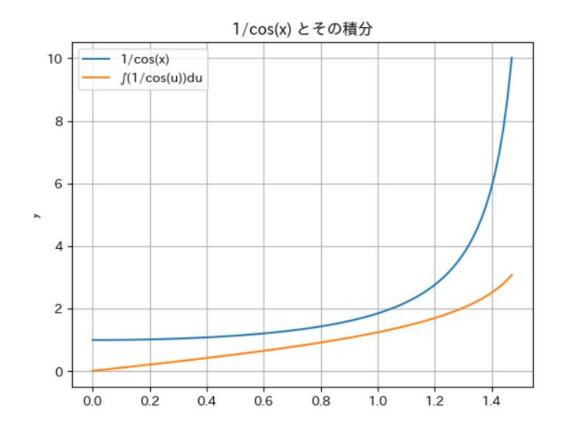
しかし、pythonを利用 してグラフを描くコード を書かせることはできる:

```
Model 9.1s
 import numpy as np
 import matplotlib.pyplot as plt
 # x の値の範囲を定義する
 x = np.linspace(0, np.pi/2 - 0.1, 100) # 0 から pi/2 (90度) まで、特異点を避けるため少し小さい値を使う
 # 1/cos(x) を計算する
 y1 = 1 / np.cos(x)
 # \int 0^x 1/cos(u) du を計算する
 y2 = np.cumsum(1 / np.cos(x)) * (x[1] - x[0]) # 数値積分を近似
 # グラフを描画する
 plt.plot(x, y1, label='1/cos(x)')
 plt.plot(x, y2, label='\( (1/cos(u)) du')
 plt.xlabel('x')
 plt.ylabel('y')
 plt.title('1/cos(x) とその積分')
 plt.legend()
 plt.grid(True)
 plt.show()
```

1/cos(x) と \int O^x 1/cos(u) du のグラフを描くpythonコードを書きなさい。

グラフの描画

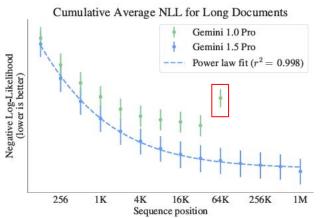
- 残念ながらGeminiは グラフの描画に対応し ていないようだ
- しかし、pythonを利用 してグラフを描くコード を書かせることはできる:

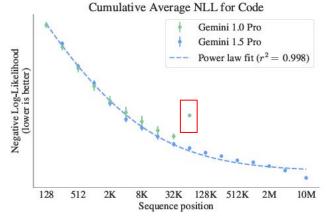


目次: Geminiの特徴

- Long contextへの対応
- 数学能力
- グラフの描画
- 図表の読解
- 動画の理解
- Med-Gemini
- 動画生成 (Veo)

グラフの読み取り





全てが正確ではないがある程度は読めている

Prompt: このグラフから読み取れることを日本語で述べなさい。

<u>Output:</u> このグラフは、Gemini 1.0 Pro と Gemini 1.5 Pro という二つの言語モデルの性能を、長い文章とコードのそれぞれについて、累積平均負の対数尤**阀**LL) を用いて比較しています。

読み取れること

Gemini 1.5 Pro は Gemini 1.0 Pro よりも性能が高い。これは、Gemini 1.5 Pro の NLL が Gemini 1.0 Pro よりも一貫して低いことからわかります NLL は低い方が良い指標なので、低い NLL はより正確な予測ができていることを示唆します。

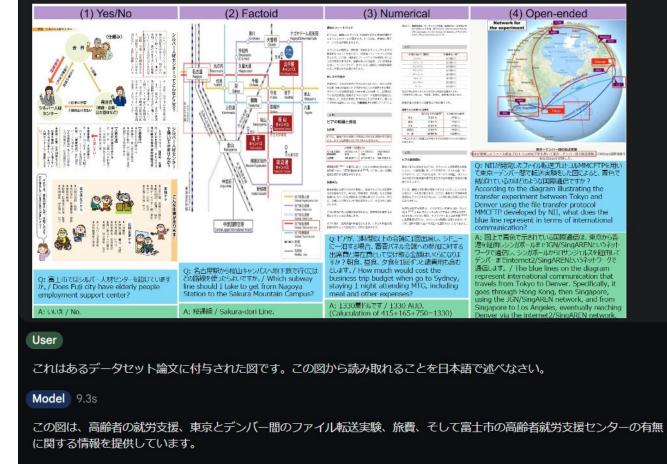
両モデルとも、文章の長さが長くなるにつれてNLLが減少する。これは、モデルが文章のコンテキストをより多く学習し、より正確な予測ができるようになるためです。

両モデルとも、コードの場合よりも長い文章の場合の方がLLが低い。これは、モデルが自然言語処理タスクに比べて、コードの構造を理解するのが得意であることを示唆しています。
NLLの減少はべき乗則に従っている。グラフに示されているように、データ点はべき乗則の曲線によく適合しています。これは、モデルの性能向上が、文章の長さに比例して線形ではなく、べき乗則に従って加速することを示唆しています。

結論として、Gemini 1.5 Pro は長い文章とコードの両方において、Gemini 1.0 Pro よりも優れた性能を示しており、特に長い文章においてその差が顕著です。

図表の読み取り

- JDocQA論文の図 でも試してみまし た。
- やはり、ある程度は 読めていますが、読 み取りに失敗してい る部分もあるようで す



JDocQA: Japanese Document Question Answering Dataset for Generative Language Models, Eri Onami, Shuhei Kurita, Taiki Miyanishi, Taro Watanabe.

目次: Geminiの特徴

- Long contextへの対応
- 数学能力
- グラフの描画
- 図表の読解
- 動画の理解
- Med-Gemini
- 動画生成 (Veo)

動画の理解



https://say-can.github.io/

Prompt: この動画について解説してください。

Output:

この動画は、人間の指示に従って、掃除する様子を映したロボットの動画です。

- 1. ロボットは、まず人間から「テーブルにコーラをこぼしたので、片付けてくれ」と指示を受けます。
- 2. ロボットは、周囲の環境を認識し、コーラの缶を見つけ、 それをゴミ箱に捨てます。
- 3. 次に、ロボットは掃除用のスポンジを見つけ、テーブルに こぼれたコーラを拭き取ります。

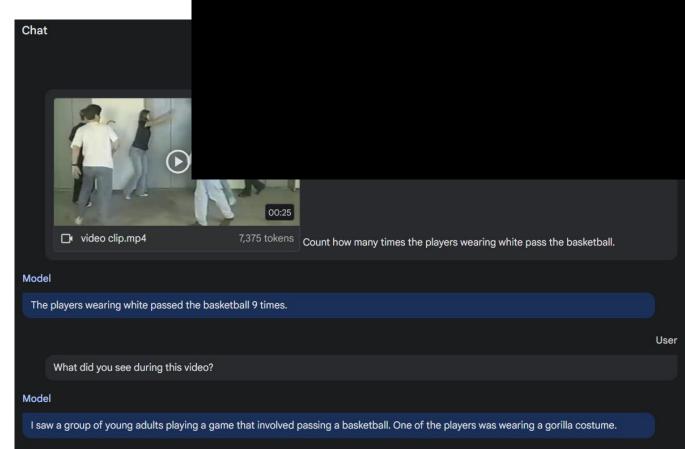
この動画は、ロボットが人間の指示に従って、複雑な作業をこなす能力を示しています。

また、ロボットが周囲の環境を認識し、適切な行動を選択できる能力も示しています。

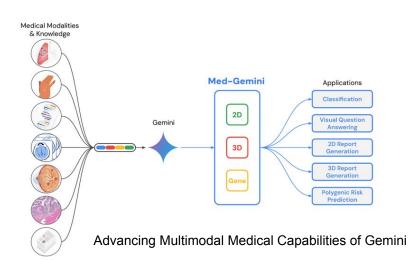
将来的には、このようなロボットが、人間の生活をより便利にするために活用されることが期待されています。

動画の理解

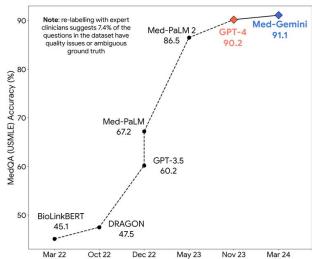
- 動画中の白シャツ の人のパス回数を 数える
- 実際には、パスの 回数はなかなかう まくいかない...
- 動画中にゴリラの 着ぐるみの人が歩 いていることには 気が付いている



Med-Gemini



Med-QA Result



@ Prompt

Given the image, what are the radiology findings?













* Response

FINDINGS: mild cortical atrophy and periventricular white matter changes are noted compatible with microvasculopathy, there is proportionate ventriculomegaly, no bleed, mass effect, midline shift or abnormal intracranial fluid is seen. no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.no abnormal extra-axial fluid collection is identified, no definitive evidence of intracranial hemorrhage or large territorial infarct within the limits of ct imaging. old left basal ganglia infarct no intracranial mass, swelling, hemorrhage, infarction or extra-axial collection is seen. visualized mastoid air cells: well pneumatized, visualized oparanasal sinuses: well pneumatized, atherosclerotic calcification of vasculature at the skull base.

Radiologist Written Report

FINDINGS: the fourth ventricle is midline, the remaining ventricular system is within normal limits, there is mild chronic white matter microangiopathic disease and moderate generalized cortical atrophy, no bleed, mass effect, midline shift or abnormal intracranial fluid is seen, bone windows demonstrate an intact skull, the remaining visualized paranasal sinuses and mastoids are clear, chronic white matter microangiopathic disease, mild generalized cortical atrophy, not significantly changed since 2015, no evidence of acute intracranial hemorrhage, territorial infarct, abnormal fluid or mass effect.

Med-Gemini-3Dのレポートには、放射線科医のレポートでは見逃されていた病理(緑色ハイライト)が正しく含まれているという。

Veo - ビデオ生成モデル



