



中国DeepSeekの次世代モデル「R2」、 コスト97%削減で米AI業界に衝撃か

22161412 高橋 和真

R1

今年初め、中国のAI企業DeepSeekが開発したAIモデル「R1」は、OpenAIやGoogleの最先端モデルに匹敵する性能を、格段に低いコストで実現したとされ、AI業界に大きな衝撃をもたらした。これを受けて米国のテクノロジー株は一時急落し、「AI版スプートニクショック」と称賛する声も上がった。



新モデル「DeepSeek R2」

DeepSeek R2で最も注目すべきは、GPT-4 Turboと比較してトークンあたりの推論コストが97.3%も削減される点である。これは、競合モデルを凌駕する圧倒的なコスト効率を実現できることを意味しており、具体的には入力トークン100万あたり0.07ドル、出力トークン100万あたり0.27ドルの価格設定を予定しているという。



Deedy


@deedydas · Follow



Viral rumors of DeepSeek R2 leaked!

- 1.2T param, 78B active, hybrid MoE
- 97.3% cheaper than GPT 4o (\$0.07/M in, \$0.27/M out)
- 5.2PB training data. 89.7% on C-Eval2.0
- Better vision. 92.4% on COCO
- 82% utilization in Huawei Ascend 910B

Big shift away from US supply chain.



R2はR1からなにが進化した？

- パラメータ数の増加：R1の2倍となる1.2兆パラメータを搭載し、活性化パラメータ数は78B。
- 多言語対応の強化：R1は主に英語での推論に優れていたが、R2では英語以外の言語にも流暢に対応できる設計となっている。
- 強化学習の拡大：より大規模なデータセットを用いた強化学習により、論理的かつ人間らしい推論が可能に。
- マルチモーダル機能の追加：テキストだけでなく、画像、音声、さらには動画データの処理にも対応。



コメント

佐藤さん

人間でも学習効率が非常に高い人が、時々います。記憶力が優れていたり、言語能力が高かったり、よい推論方法を身につけていたり。仮に97%も効率よく勉強できるようになるためには、どの能力が最も重要なのでしょうか？

また、学習効率が高いことと賢いことは別です。何を勉強しているかが重要で、よい教育というのはやはり大事です。AIモデルの場合、世の全部を勉強していると主張されるのかも知れませんが、オフセットはあります。DeepSeekが賢く勉強できることと、何を勉強しているかは、分けて評価する必要があります。



感想



- 今なお需要が高まっている生成AIの中で、様々な有料サービスが展開していますが、その中でChatGPTのような性能で、価格が安いものができたなら消費者はR2を購入することになると思います。
- AIによって機能が全く違うものはそれぞれの住み分けができるので発展もしやすいと思いますが、同じような内容のものが重なっていると、価格競争やUIの向上が求められてしまい、成長に繋がらない結果になってしまう可能性が考えられると思います。