



電力曝食いの生成AIと仮想通貨

ナチュラル研究所

工学博士 石川 宏

<http://www.ishikawa-lab.com/>

<mailto:dr.ishikawa@aa.wakwak.com>

2024.8.26

資料はこちらダウンロード可

<https://www.ishikawa-lab.com/pdf/20240826.pdf>



電力需要は再び増加



- 節電・省エネが進み、またコロナの影響で電力需要は減少傾向にあったが、再び増加しはじめたという
- なにか魂胆がある
- データセンターの建設ブーム、また中身が変わってきている



電力暴食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンタは迷惑施設



使って便利 ChatGPT

ChatGPT (Chat Generative Pre-trained Transformer) は、OpenAI社が2022年11月に公開した自然な会話ができる生成AI

- 会話するように答えてくれる
「生成AIとそれ以前のAIの違いはなんですか」
- 文章の要約
「次の文章を200字に要約してください」
- 翻訳
「次の英語を日本語に翻訳して」
- 食事のレシピ
「アスパラガスとヒラメの切り身があります。レシピを教えてください。和風がいいです」
- エクセルで関数を作れる
「エクセルで、A列からJ列までの数字の平均値を出して、小数点2けたで表示する関数を作ってください」
- プログラムの自動生成(プログラミング)

生成AIとそれ以前のAIの違いは？

- 生成AI (Generative AI)
 - 創造性: 生成AIは新しいコンテンツ(テキスト、画像、音声など)を生成する能力
 - 知識の量: 大量のデータを学習している
 - 応用範囲の広さ: テキスト生成、翻訳、画像生成、音楽作成、コード生成など、幅広い応用が可能
 - 自然言語処理の進化: 人間のように自然な言語でコミュニケーションを取る能力
- 以前のAI
 - パターン認識: 機械学習アルゴリズムを使用して特定のタスクを実行するためのもの
 - 制限された応答: 事前にプログラムされた応答やルールベースのシステムが一般的
 - 特化型: 特定のタスクに特化、例えば画像認識、音声認識、推薦システムなど
 - データの利用: 生成AIほど大量のデータを必要とせず、特定のデータセットに基づく
- 技術的な違い
 - アーキテクチャ: 生成AIはトランスフォーマーモデルや自己教師あり学習などの最新技術 以前のAIは主にニューラルネットワークなどのアルゴリズムに依存
 - 計算資源: 生成AIは大規模な計算資源を必要。以前のAIは比較的少ない計算資源で動作

ChatGPTによる

(C) 2024 H.Ishikawa

5

ChatGPTでプログラムを生成

- Pythonのsocketモジュールを使用して1行ごとにデータを受信するプログラムをおしえて

```
import socket

def receive_lines(sock):
    buffer = b'' # 受信データを格納するバッファ
    while True:
        data = sock.recv(1024) # 1024バイトごとにデータを受信
        if not data: # データが空の場合は接続が切断されたと判断してループを終了
            break
        buffer += data # 受信したデータをバッファに追加
        while b'\n' in buffer: # 改行文字が含まれるかチェック
            line, buffer = buffer.split(b'\n', 1) # バッファから1行分のデータを切り出す
            yield line.decode() # 1行分のデータをyieldする(ジェネレータ)
        if buffer: # バッファに残っているデータがある場合はそれを処理する
            yield buffer.decode()

def main():
    # ホストとポート番号
    host = "localhost"
    port = 30005

    try:
        # ソケットを作成し、ポート30005に接続
        with socket.socket(socket.AF_INET, socket.SOCK_STREAM) as s:
            s.connect((host, port))
```

(C) 2024 H.Ishikawa

6

ChatGPTの問題点

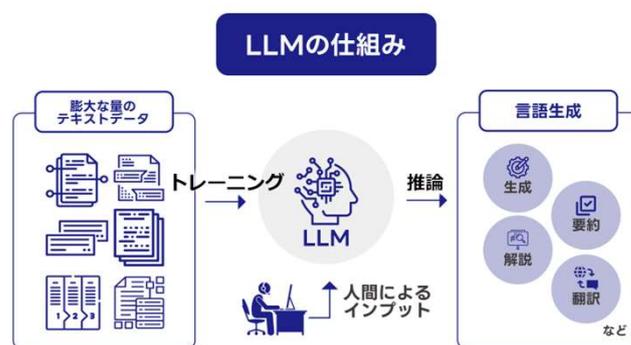
- 知識の最新性: ChatGPTのトレーニングデータは特定の時点までのものであり、最新の情報や最近の出来事についての知識が不足している
- 事実誤認: ChatGPTは、特定の情報に対して確信がなくても回答を生成することがあり、それが事実誤認を引き起こすことがある
- 倫理的問題と偏り: トレーニングデータに基づいているため、データに含まれる偏見や不適切な内容が反映されることがある
- 創造性と独自性の制限: ChatGPTは既存の情報を基にして回答を生成するため、完全に新しいアイデアや独自の視点を提供することは難しい
- 著作権の問題: トレーニング時に著作権で保護されたコンテンツも含まれる可能性がある。生成コンテンツが著作権侵害とみなされる可能性がある
- 電力消費: トレーニングは数週間から数ヶ月かかることがあり、その間に消費される電力は非常に大きい。トレーニング後も、データセンターは冷却や運用のために大量の電力を消費する

ChatGPTによる

(C) 2024 H.Ishikawa

7

トレーニングと推論



LLM:大規模言語モデル

- 生成AIのトレーニングとは、収集した膨大なデータを入力し、そのデータからパターンや関連性をLLMに学習させるプロセスのことを言う。
- 推論とは、トレーニングされたLLMモデルが学習したパターンに基づいて、ユーザーが入力したプロンプトに対し関連性の高いデータを推論・生成するプロセスのこと。

GPTはOpenAIによるLLM

(C) 2024 H.Ishikawa

8

電力曝食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンタは迷惑施設

(C) 2024 H.Ishikawa 9

OpenAIが消費した電力

- MIT Lincoln Laboratoryが2023年11月22日に発表したレポートによると、OpenAI社のGPT-3をトレーニングするのに使用された計算量は、800ペタフロップスの処理能力に相当、約1,300メガワット時の電力が消費された
- これは、1ギガワットの出力を持つ原発が1.3時間稼働した量
- 世界で最も強力な20のスーパーコンピュータを合わせたものに匹敵する。
- さらに、GPT-4のトレーニングデータ量は66倍となり、これを電力換算すると約86ギガワット時となる。これは、米国の平均的な家庭9万5,700世帯1カ月分の電力消費量となる。



アマゾン・ウェブ・サービスのデータセンターは、サスケハナ原子力発電所 BWR、130万kW×2基から直接電力供給を受ける

MIT Sloan Management Review

(C) 2024 H.Ishikawa 10

グラフィックプロセッサが使われる

- ChatGPTのような大規模なAIモデルのトレーニングと実行には、グラフィックプロセッサ(GPU)が多数使われている
- 並列処理能力:
GPUは、数千個のコアを持ち、大量のデータを並列に処理する能力に優れている。これにより、AIモデルのトレーニングや推論に必要な膨大な計算を高速に行うことができる
- 行列計算の効率:
AIモデルのトレーニングは、大量の行列演算を伴う。GPUは、この行列計算を効率的に処理できるように設計されている
- 大規模データの処理:
AIモデルは、膨大なデータベースを使用してトレーニングされる。GPUは、大規模データの処理にも適しており、高速かつ効率的にデータを処理できる
- 消費電力
ChatGPTではNVIDIAの最新GPU「H100」を多数利用する(1万基といわれる)。H100は1つあたり700ワットの電力を消費する。これにデータセンターや冷却用の電力を加え、H100GPU1つあたりの消費電力は1KW
- GPUのとりあい
製造はTSMC。品薄が続く。たくさんGPUを確保した会社が勝者。GPUバブル

ChatGPTによる

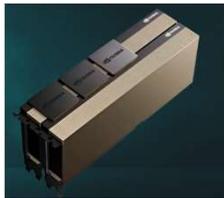
(C) 2024 H.Ishikawa

11

NVIDIAはPCグラフィックステップから



1997年ごろの
グラフィックスアクセラレータ



生成AIに最適化された
NVIDIA H100

- 1993年LSIロジックを退社したジェンソン・フアンがグラフィックステップの会社を設立
- 1997年、SGIに所属していた技術者が続々と参加しグラフィックスアクセラレータの技術的筆頭
- 1999年、グラフィック機能をCPUが内蔵するようになったことで、ローエンド製品から、多大な開発費を要する高性能GPUに
- GPUは高速の浮動小数点演算用の多数のコアを持ち、並列処理
- 2003年ソニーとNVIDIA、PlayStation3向けのGPUを共同開発
- 2008年東工大のスパコンTSUBAMEに多数のNVIDIA GPU
- 2016年頃に起こったAIの波に乗って一気に成長。これは、CPUを遥かにしのぐGPUの並列計算能力が、機械学習やディープラーニングに適していたことが大きな理由
- 2019年、NVIDIAは人工知能、自動運転の分野では圧倒的な地位を占めるようになってきている。
- 2024年6月には、マイクロソフトを抜き時価総額世界首位となった

(C) 2024 H.Ishikawa

12



東工大のTSUBAME

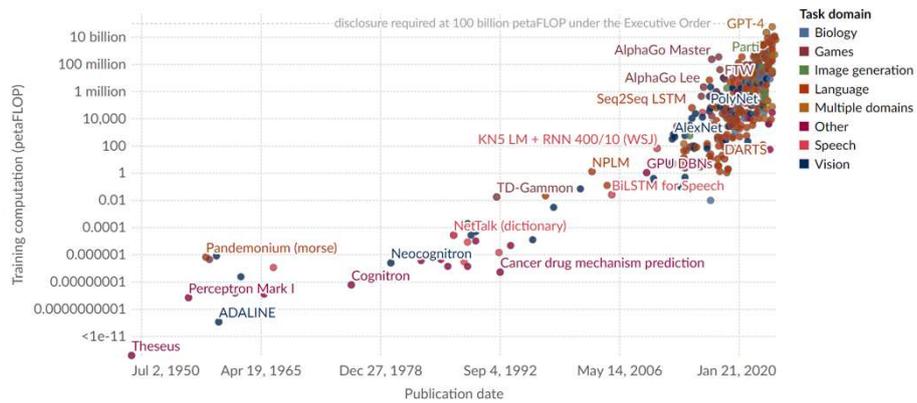


- 2008年 NVIDIAのGPU「Tesla」を170台大
量導入した、世界初のスーパーコンピュータ
- 1台のTesla S1070は、4基のGPUを搭載して
いる
- TSUBAMEの演算能力は約170TFlopsにな
った。その性能はその時点で日本2位
- 松岡 聡教授はTeslaについて、1プロセッサー
で約1TFlopsという優れた性能を持つと評価
- また汎用品のGPUを使う利点について、パソ
コンで一般的に使われているGPUと同じ技術
やソフトウェア開発環境を利用できる
- 汎用のGPUによりスーパーコンピュータが実
用化できることを示したことで、産業界に大き
な影響

2008.1 通信学会東京支部見学会



AIのトレーニングに使用される計算量

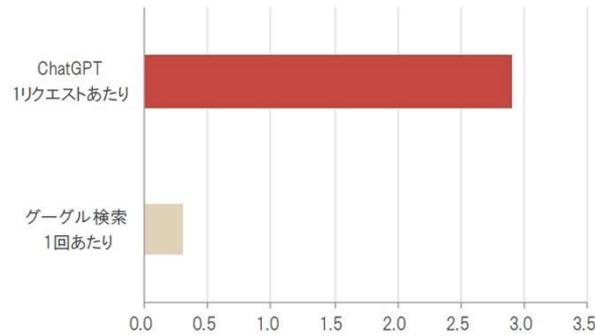


- 1 FLOP は、加算、減算、乗算、除算など、浮動小数点数演算。1 ペタ FLOP は 1 京 FLOP
- 最近の 生成AI システムは、複雑な数学的プロセスとアルゴリズムを伴うため、計算量が多い

2024 Our World in Data

1リクエスト(推論)あたりの消費電力

単位:ワット時(Wh)



出所:国際エネルギー機関(IEA)のデータを基にピクテ・ジャパン作成

- 一般的な Google 検索の平均電力は0.3 Wh、ChatGPT は 2.9 Wh
- 検索ツールに AI が完全に実装された場合、電力需要は 10 倍になる

(C) 2024 H.Ishikawa

15

電力曝食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンタは迷惑施設

(C) 2024 H.Ishikawa

16

暗号資産(仮想通貨)

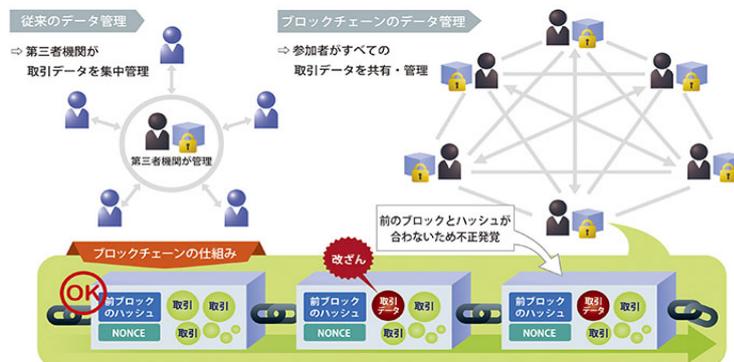
- 「暗号資産(仮想通貨)」とは、インターネット上でやりとりできる財産的価値であり、「資金決済に関する法律」において定義
 - (1) 不特定の者に対して、代金の支払い等に使用でき、かつ、法定通貨(日本円や米国ドル等)と相互に交換できる
 - (2) 電子的に記録され、移転できる
 - (3) 法定通貨または法定通貨建ての資産(プリペイドカード等)ではない
- 代表的な暗号資産には、ビットコインやイーサリアムなどがある
暗号資産は、銀行等の第三者を介することなく、財産的価値をやり取りすることが可能な仕組み
- 「交換所」や「取引所」と呼ばれる事業者(暗号資産交換業者)から入手・換金することができる。暗号資産交換業は、金融庁・財務局の登録を受けた事業者のみが行うことができる
- 暗号資産は、国家やその中央銀行によって発行された、法定通貨ではない。また、裏付け資産を持っていないことなどから、価格が大きく変動する傾向にある点には注意が必要
- また、暗号資産に関する詐欺などの事例も数多く報告されているので、注意が必要

日本銀行ホームページによる

(C) 2024 H.Ishikawa 17

仮想通貨の仕組み

- サトシ・ナカモトの論文(2008年)が原点
- 基本技術はP2Pと暗号
- 仮想通貨は、ブロックチェーンと呼ばれる分散型の公開台帳上で運用され、すべての取引の記録が更新される。



Satoshi Nakamoto "Bitcoin: A Peer-to-Peer Electronic Cash System"
https://www.noc-net.co.jp/blog/2017/07/column_229/

(C) 2024 H.Ishikawa 18

ブロックチェーン

(図表 1)

	A→Bに350異動 (手数料:5)	A→Cに100異動 (手数料:5)	B→Cに50異動 (手数料:5)	C→Aに130異動 (手数料:5)																																																	
預金データのイメージ (データは銀行が管理)	<table border="1"> <tr><td>残高</td><td></td></tr> <tr><td>Aさん口座</td><td>1000</td></tr> <tr><td>Bさん口座</td><td>0</td></tr> <tr><td>Cさん口座</td><td>0</td></tr> </table>	残高		Aさん口座	1000	Bさん口座	0	Cさん口座	0	<table border="1"> <tr><td>残高</td><td>(異動)</td></tr> <tr><td>645</td><td>▲355</td></tr> <tr><td>350</td><td>+350</td></tr> <tr><td>0</td><td></td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	残高	(異動)	645	▲355	350	+350	0		(手数料:5)		<table border="1"> <tr><td>残高</td><td>(異動)</td></tr> <tr><td>540</td><td>▲105</td></tr> <tr><td>350</td><td></td></tr> <tr><td>100</td><td>+100</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	残高	(異動)	540	▲105	350		100	+100	(手数料:5)		<table border="1"> <tr><td>残高</td><td>(異動)</td></tr> <tr><td>540</td><td></td></tr> <tr><td>295</td><td>▲55</td></tr> <tr><td>150</td><td>+50</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	残高	(異動)	540		295	▲55	150	+50	(手数料:5)		<table border="1"> <tr><td>残高</td><td>(異動)</td></tr> <tr><td>670</td><td>130</td></tr> <tr><td>295</td><td></td></tr> <tr><td>15</td><td>▲135</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	残高	(異動)	670	130	295		15	▲135	(手数料:5)	
残高																																																					
Aさん口座	1000																																																				
Bさん口座	0																																																				
Cさん口座	0																																																				
残高	(異動)																																																				
645	▲355																																																				
350	+350																																																				
0																																																					
(手数料:5)																																																					
残高	(異動)																																																				
540	▲105																																																				
350																																																					
100	+100																																																				
(手数料:5)																																																					
残高	(異動)																																																				
540																																																					
295	▲55																																																				
150	+50																																																				
(手数料:5)																																																					
残高	(異動)																																																				
670	130																																																				
295																																																					
15	▲135																																																				
(手数料:5)																																																					
ビットコインデータのイメージ (データは全参加者が共有)	【取引 1】 <table border="1"> <tr><td>原資</td><td>送金先・金額</td></tr> <tr><td>(なし・報酬)</td><td>A 1000(①)</td></tr> </table>	原資	送金先・金額	(なし・報酬)	A 1000(①)	【取引 2】 <table border="1"> <tr><td>原資</td><td>送金先・金額</td></tr> <tr><td>① → B</td><td>350(②)</td></tr> <tr><td>A</td><td>645(③)</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	原資	送金先・金額	① → B	350(②)	A	645(③)	(手数料:5)		【取引 3】 <table border="1"> <tr><td>原資</td><td>送金先・金額</td></tr> <tr><td>③ → C</td><td>100(④)</td></tr> <tr><td>A</td><td>540(⑤)</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	原資	送金先・金額	③ → C	100(④)	A	540(⑤)	(手数料:5)		【取引 4】 <table border="1"> <tr><td>原資</td><td>送金先・金額</td></tr> <tr><td>② → C</td><td>50(⑥)</td></tr> <tr><td>B</td><td>295(⑦)</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	原資	送金先・金額	② → C	50(⑥)	B	295(⑦)	(手数料:5)		【取引 5】 <table border="1"> <tr><td>原資</td><td>送金先・金額</td></tr> <tr><td>③+⑥ → A</td><td>130(⑧)</td></tr> <tr><td>C</td><td>15(⑨)</td></tr> <tr><td colspan="2">(手数料:5)</td></tr> </table>	原資	送金先・金額	③+⑥ → A	130(⑧)	C	15(⑨)	(手数料:5)													
原資	送金先・金額																																																				
(なし・報酬)	A 1000(①)																																																				
原資	送金先・金額																																																				
① → B	350(②)																																																				
A	645(③)																																																				
(手数料:5)																																																					
原資	送金先・金額																																																				
③ → C	100(④)																																																				
A	540(⑤)																																																				
(手数料:5)																																																					
原資	送金先・金額																																																				
② → C	50(⑥)																																																				
B	295(⑦)																																																				
(手数料:5)																																																					
原資	送金先・金額																																																				
③+⑥ → A	130(⑧)																																																				
C	15(⑨)																																																				
(手数料:5)																																																					



(資料) 筆者作成

- ブロックチェーンは改ざん不可能な分散記録システム
- 「取引」データは、コインがどこからどこに行ったのかを記したもので、原資(資金源)と送金先・金額が記載されており、「残高」という視点でデータは蓄積されていない。
- ブロックは10分毎に作られる

<https://www.nli-research.co.jp/report/detail/id=67889?site=nli>

(C) 2024 H.Ishikawa

19

ハッシュ関数

ハッシュ関数はわずかな改ざんも検知



```

改ざん前.txt - メモ帳
-----
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
-----
行 1, 列 1 | 100% | Windows (CRLF) | UTF-8
    
```

SHA-1

0f715bbc07ff11dcbe9ccb8f
05b737cb053b9e59

テキストファイルの1文字を修正しただけで、ハッシュ値は大きく異なる。

```

改ざん後.txt - メモ帳
-----
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKL MNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
ABCDEFGHIJKLMNOPQRSTUVWXYZ
-----
行 1, 列 1 | 100% | Windows (CRLF) | UTF-8
    
```

SHA-1

5fcc52b64b35172fa78fd10
9619c010edd6c4f5a

(注) 下図のテキストファイルでは、英大文字のO(オー)一文字を英小文字にしています。



マイニング処理

一番早く解いた人を見つける

1. 新しい取引は全ノードに送信される
2. 各ノードが新しい取引をブロックに取り入れる
3. 各ノードは10分毎にブロックに対し次の計算を行う
 - ① nonce=0にする
 - ② ブロックデータとnonceを連結してハッシュを計算
 - ③ ハッシュ値の出力256ビットをみて、頭のnビットが0000...となるまでnonceを1ずつ増加し、繰り返す
4. 最も早く見つけたノードはnonce値を全ノードに告知する
5. 各ノードはハッシュを計算し、検算。正しければそれを承認する
6. 各ノードは、承認されたブロックのハッシュを、チェーンの次のブロックに書き込み、つぎのブロック作成にそなえる
7. 一番乗りのノードにはインセンティブが支払われる。nビットの長さは、調整される



仮想通貨のデータセンター

仮想通貨の初期の設計では、取引の検証や新しいコインの生成が、個々のユーザーによる分散処理によって行われることを想定。しかし、その後大型のデータセンターによる集中処理へと移行。なぜか

- 早い者勝ち

参加者は暗号ハッシュを解くために競い合い、成功した参加者は最新のブロックをブロックチェーンに追加することができ、仮想通貨の形で報酬を受る。マイニングという

- 計算難易度の増加

仮想通貨のマイニングには高度な計算能力が必要。個人のPCでは必要な計算量を処理しきれなくなった

- 特殊なハードウェアの登場

効率的なマイニングを行うために、あたらしいGPUなどが開発され、一般のユーザーが手軽に導入することは困難

- 経済規模の拡大

マイニングには初期投資が必要だが、成功すれば報酬が得られるので、大規模な設備を導入し、効率的にマイニングを行うことで、経済的な利益を最大化することが可能。これにより、企業や投資家が大型のマイニングセンターを運営するようになった。



電力曝食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンタは迷惑施設



ビットコインの電力消費

- **全世界での年間消費量**

ビットコインのマイニングに使用される年間の電力消費量は、約100テラワット時 (TWh) 以上に達する。この数値は、アルゼンチンやノルウェーなどの国の消費量に匹敵。世界全体の電力消費量の約0.5%から1%を占めると推定

- **単一のトランザクションあたりの消費量**

ビットコインの1回のトランザクションに必要な電力消費量も非常に大きく、1トランザクションあたりの消費量は数百キロワット時 (kWh) に達することがある。

- **従来のデータセンターとの比較:**

ビットコインのマイニングは、大規模なデータセンターと比較しても、その電力消費量は桁違いに大きい。大規模データセンターが年間に消費する電力量の数倍に達する。

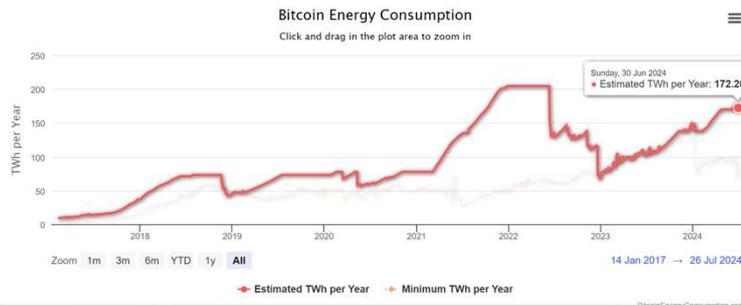
- VISAでの決済取引と比較すると、ビットコインの取引1回に要する環境負荷は、電力消費量では1,542,417回に相当

ビットコインの電力消費

Bitcoin Energy Consumption Index

The Bitcoin Energy Consumption Index provides the latest estimate of the total energy consumption of the Bitcoin network.

NEW RESEARCH: "Bitcoin's growing water footprint" (November 2023); A single Bitcoin transaction could cost as much water as a backyard swimming pool.



- 価値が上がれば上がると、採掘者の報酬は増え、採掘量も増える、そのため消費する電力が増える

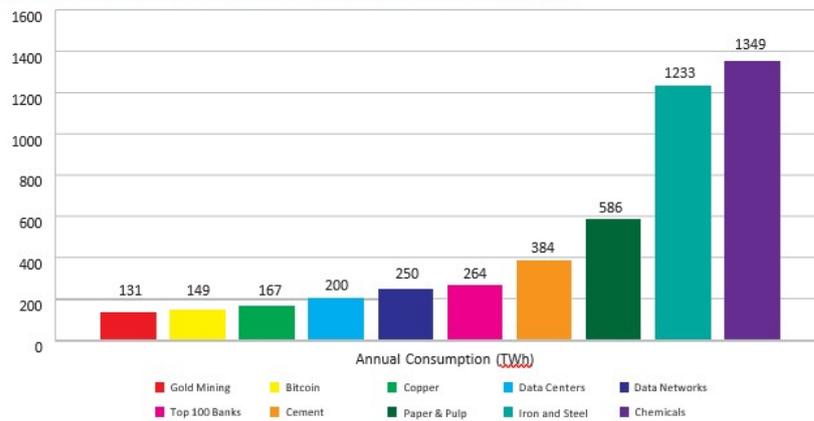
<https://digiconomist.net/bitcoin-energy-consumption>

(C) 2024 H.Ishikawa

25

産業別電力消費

Figure 1: Estimated annual energy consumption for select industries (TWh/yr)



Sources: CCAF Cambridge Bitcoin Energy Consumption Index, Galaxy.com.

Data note: Industry estimates range from 2018 to 2021, depending on the industry. Bitcoin estimate is up to date as of July 10, 2023.

2023 CCAFデータ <https://cei.org/studies/dont-depower-crypto/>

(C) 2024 H.Ishikawa

26



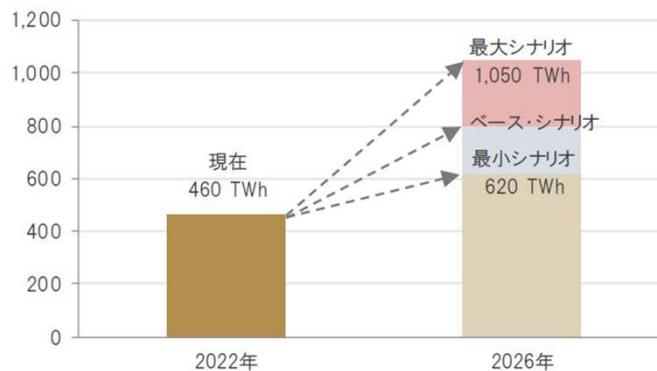
電力曝食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンターは迷惑施設



データセンターの電力需要予測

単位：テラワット時(TWh)、期間：2022年(実績)～2026年(予想)



出所：「IEA Electricity 2024 Analysis and forecast to 2026」

- AI、仮想通貨によりデータセンターの電力消費は急増

日本のデータセンター

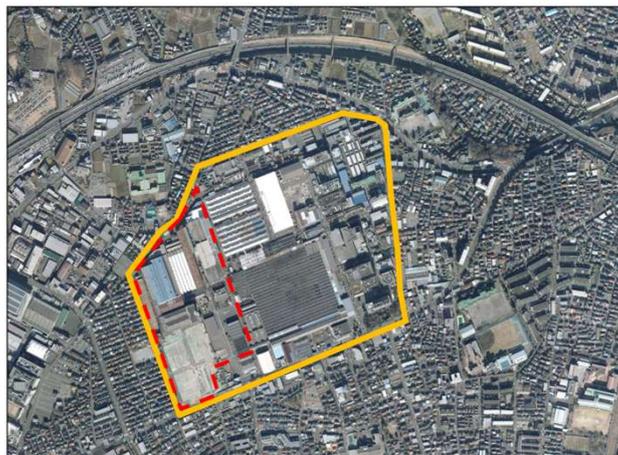
(参考) 2024年以降のデータセンターの新設計画

※コンテナ型データセンターなど、小規模のものや増床案件は除く

デジタルインフラ（DC等）整備に関する有識者会合（第7回）事務局資料



日野自動車工場跡にデータセンター



 日野自動車日野工場全体敷地
 大規模土地取引行為の届出区域

データセンターは迷惑施設

- 生成AIは便利だが、エネルギー大量消費型人工知能
- 仮想通貨はサトシナカモトの非中央集権の理念を離れ、歴史上もっとも無駄な、欲のかたまりのデジタルねずみ講
- データセンターの電力消費量は、さまざまな消費者サービスの中でも最大級で、世界の電力の2%を消費しており、航空業界全体とほぼ同量のCO2を排出している
- データセンターの中身は変化し、その電力消費量は、4年ごとに倍増し続ける
- データセンターは原発並みの排熱、ヒートスポット
- データセンターの排熱を再利用、コジェネ
- 省エネのためのブレイクスルー必須

(C) 2024 H.Ishikawa 31

NTT版生成AI tsuzumi に期待する

学習コスト

GPT-3規模のLLMと比べて、学習コストを約300分の1 (推定値) および約25分の1 (推定値) に低減可能

学習データを300Bトークンとした場合の学習コスト比較 (NTT試算値*)



※試算値
 * 学習データ: 300Bトークン (GPT-3: 175Bトークン、tsuzumi: 7Bトークン)
 * 学習コスト: GPT-3: 4.7億円、tsuzumi: 1900万円 (7B)、160万円 (0.6B)
 * 学習効率: GPT-3: 1/300、tsuzumi: 1/25 (0.6B)
 * 学習時間: GPT-3: 約1000時間、tsuzumi: 約40時間 (7B)、約16時間 (0.6B)
 * 学習電力消費量: GPT-3: 約1000kWh、tsuzumi: 約40kWh (7B)、約16kWh (0.6B)

推論コスト

GPT-3クラスと比べて、推論コストを約70分の1 (推定値) と約20分の1 (推定値) に低減可能

オンプレ環境でのGPU/CPUコスト比較 (NTT試算値*)



※試算値
 * 推論コスト: GPT-3: 1500万円、tsuzumi: 70万円 (7B)、20万円 (0.6B)
 * 推論効率: GPT-3: 1/70、tsuzumi: 1/20 (0.6B)
 * 推論時間: GPT-3: 約1000時間、tsuzumi: 約14.3時間 (7B)、約5時間 (0.6B)
 * 推論電力消費量: GPT-3: 約1000kWh、tsuzumi: 約14.3kWh (7B)、約5kWh (0.6B)

- 非常に軽量なモデル
- 高い日本語処理能力を実現
- 医療機関やコンタクトセンタなどローカル環境での利用が可能

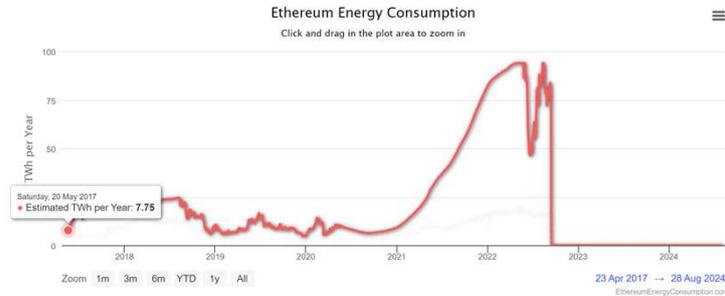
(C) 2024 H.Ishikawa 32

イーサリアムの電力消費

Ethereum Energy Consumption Index

The Ethereum Energy Consumption Index provides the latest estimate of the total energy consumption of the Ethereum network.

NEW RESEARCH: "Cryptocurrencies on the road to sustainability: Ethereum paving the way for Bitcoin" (December 2022); Ethereum has reduced its electrical energy requirement by at least 99.84% by changing its method of production.



Annualized Total Ethereum Footprints

- 2022年12月イーサリアムは方法を変更することで、電力消費を少なくとも99.84%削減
- ブルーフ・オブ・ワーク(PoW)メカニズムをプルーフ・オブ・ステーク(PoS)に変更

<https://digiconomist.net/ethereum-energy-consumption>

(C) 2024 H.Ishikawa

33

電力曝食いの生成AIと仮想通貨

- 生成AIとは
- 生成AIの電力問題
- 仮想通貨
- 仮想通貨の電力問題
- データセンタは迷惑施設

資料はこちらダウンロード可
<https://www.ishikawa-lab.com/pdf/20240826.pdf>

(C) 2024 H.Ishikawa

34