

2024/2/2 次世代半導体のアカデミアにおける研究開発等に関する検討会

未来社会での先端半導体の 活用領域

Preferred Networks 代表取締役 最高研究責任者

Preferred Computational Chemistry 代表取締役社長

Preferred Elements 代表取締役社長

岡野原 大輔



自己紹介：岡野原 大輔

Preferred Networks 代表取締役 最高研究責任者 / 共同創業者

Preferred Computational Chemistry 代表取締役社長

Preferred Elements 代表取締役社長

Preferred Robotics 取締役 他

X(twitter): @hillbig

AIに関する次世代リーダーとの車座対話メンバー

AI事業者ガイドライン 検討会委員



著書

受賞歴

- ・ 令和4年度「現代の名工」データサイエンティスト
 - ・ KDDI Foundation Award 本賞 2022
 - ・ 大川出版賞（拡散モデル）
- など



PFNグループは生成AIとそれを支える計算資源を提供する

Preferred Networks (PFN) グループは中長期的に社会基盤を担う大規模基盤モデル、それを支える計算力の提供に貢献していく

半導体・化学



製造・工場



交通システム



教育



医療



エンタメ



ロボット



産業用途

効率・品質改善
イノベーションの加速

社会インフラ

安全・安定・効率的
全ての人に優れた教育と医療を

コンシューマー用途

新しい創作表現を可能に
人手不足の解消

マルチモーダルな大規模基盤モデル

AIを支える計算力

データセンター・
スーパーコンピュータ

AI半導体
電力効率改善・最先端プロセス

相互作用を引き起こしながら
イノベーションの連鎖へ

多くの仕事や社会生活が
AIによって支えられる

社会のインフラとしてのAI

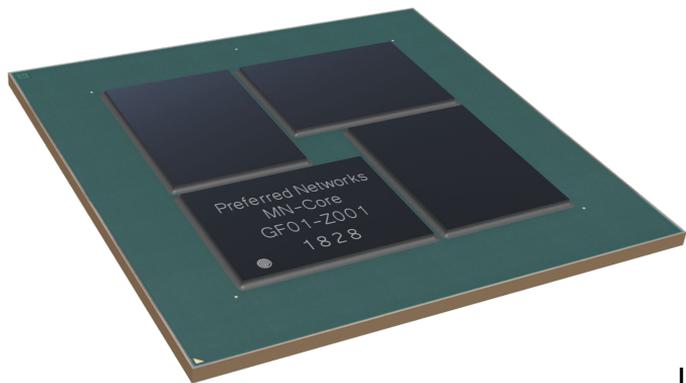
PFNはAI向け半導体、スーパーコンピュータを開発してきた



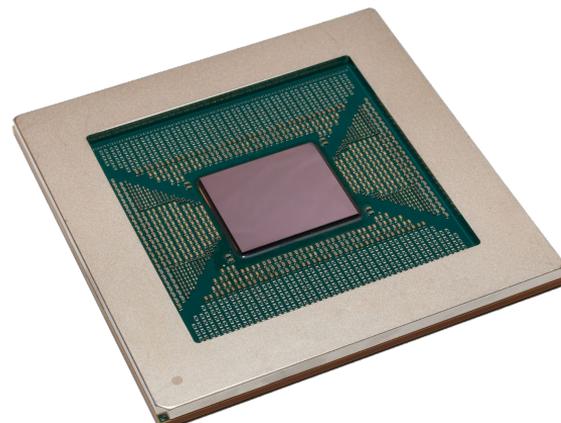
MN-3

省電力性能ランキングGreen500で3度 世界1位

MN-Core



MN-Core 2



最先端AIの最新モデルの学習に必要な計算資源は増加

log PFLOP/s *
days

スパコンは廃熱の問題など物理的制約が大きくゆっくり成長

世界のトップのスパコンの性能

世界の500位のスパコンの性能

最新モデルの学習に必要な計算量

GPT-4(est.)

富岳

GPT-3 (175B)

2026年に
予想される
計算量

2024年で既に数万台のアクセラレーターを集約したスーパーコンピュータの利用が急速に進んでいる

現在のスパコン性能の限界を超えられる新しいチップ、ソフトウェア技術が必要となる

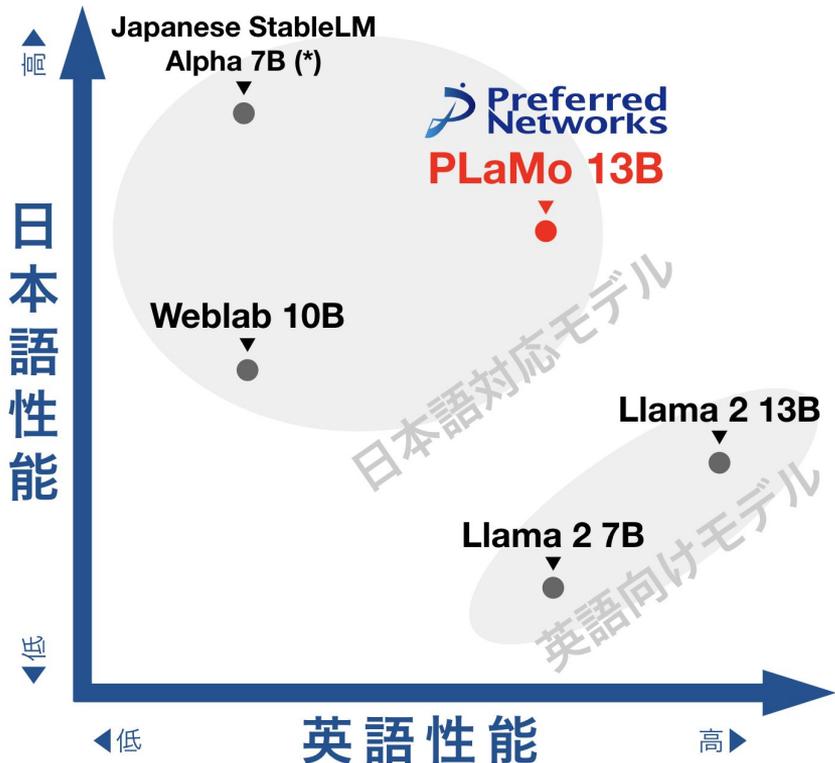
2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026

DCへの電力供給/冷却、ネットワーク帯域などがボトルネック

モデルの学習期間を半年と仮定している

PLaMo-13B PFNのマルチモーダル基盤モデル

PFNは研究/商用利用な130億パラメータの大規模言語モデルを去年9月に公開



- 2023年9月公開当時、日英2言語をあわせた能力で世界トップレベル
- 経産省ABCIのスーパーコンピュータ A100 480GPUで1ヶ月弱利用。日本語、英語の1.4兆トークン（数兆文字）を使って学習
- 学習時の実行効率は41%で世界の他の学習基盤と比べても高い（効率的に学習資源を使える）
- 開発実績を元に基盤モデル開発・提供を行う Preferred Elementsを2023年11月に設立



今後の大規模基盤モデルの開発・リリース計画

- PFN子会社のPreferred Elements（PFE）が開発する**1000億パラメータモデル**のマルチモーダル基盤モデルを**2024年秋頃**に投入予定
 - 言語に加えて**画像・音声**もサポート
 - 日本語性能や特定分野向けでの最高性能を目指す
 - （安全性を検証した上で）事前学習済みモデルや開発ノウハウは公開予定
- PFEは**1兆パラメータモデル**の基盤モデルの検証・学習を**2024年夏頃**から開始予定
 - 世界的にも最大サイズ級（例：GPT-4が1.8兆パラメータと推定）
 - 必要な技術開発・計算力の確保を進めている
- 製造・開発・分析・研究開発など高度なタスクもサポートすることを目指す
 - 例えば半導体においては、素材設計、現象解析、最適化、シミュレーションなどを補助することを目指す

LLMを使った材料探索のデモ

人間の言葉による指示

Fe, Co, Ptについて、反応経路のindexが189(矢印で指した場所)の原子を置換した後に活性化エネルギーを計算してグラフにする

日本語の指示で必要なプログラムを生成

```
from typing import Dict, Optional, List, Literal
from ase import Atoms
import time
import ipywidgets as widgets
import plotly.graph_objects as go
from IPython.core.display import display
from IPython.display import clear_output
from atomsprog.utils.nglviewer import RawNGLViewer

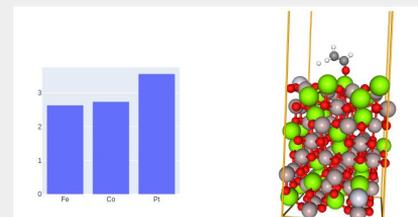
class NGLManager:
    def __init__(
        self, title: str, width: str = "400px", height: str =
    ) -> None:
        self.title = title
        self.ngl_viewer = RawNGLViewer(width=width,
        height=height)
        self.wrap = wrap
        self.recent_log = None
```

計算化学
プログラムの
専門知識を
サポート



シミュレーション
を実行

鉄が有望！



```
[8]: interface = ProgInterface(ProgramGenerator(prompt="assets/prompt_minimum.txt"))
```

ここにテキストを入力してください

コード実行

```
[ ]:
```

PFNのマルチモーダル基盤モデルによる画像認識例 1)

入力
画像



based on provided image,
tell me what is happening in it

画像キャプション

A bowl of food with broccoli and noodles.

what benefits can one get from
the food in the bowl?

VQA

The bowl contains a variety of food items, including noodles, broccoli, and meat. Consuming this meal can provide several health benefits. The noodles are a source of carbohydrates, which are an essential part of a balanced diet. Broccoli is a nutrient-rich vegetable that is high in vitamins, minerals, and fiber, promoting good digestion and overall health. The meat in the dish adds protein, which is crucial for muscle growth and repair. Overall, this meal offers a combination of essential nutrients, vitamins, and minerals that can contribute to a balanced and healthy diet.

物体認識



detect
vegetables

PFNのマルチモーダル基盤モデルによる画像認識例 2)

入力画像



what is going on in the image?

画像キャプション

A man riding a skateboard up the side of a ramp.

what went wrong with the skateboard trick and what could be the aftermath?

VQA

In the image, the skateboarder is attempting to perform a trick, but instead of successfully executing it, he loses his balance and falls off the skateboard. This could lead to various consequences, such as injuries, damage to the skateboard, or even embarrassment. The skateboarder might suffer from bruises, sprains, or even fractures, depending on the severity of the fall. Additionally, the skateboard itself could be damaged, requiring repairs or replacement. The skateboarder might also experience a loss of confidence or self-esteem, as they may have failed to perform the trick properly. This highlights the importance of practicing and mastering skateboarding tricks to avoid such mishaps and ensure safety while enjoying the sport.

物体認識



detect people in the image

生成モデル、実世界への応用



動画データの生成モデルによる「世界モデル」は急速に進展している

(左は車の例)

世界モデルによるシミュレーションによって、実現が加速される

[GAIA-1: 9-billion parameter generative world model for autonomous driving]

<https://wayve.ai/thinking/scaling-gaia-1/>

投入データ量、モデルサイズを大きくしていき、より高度な世界モデルを実現するのは時間の問題

LLMによるロボットへの指示



Preferred Robotics カチャカ（家の中の搬送ロボット）の事例

LLMと組み合わせ、自由に何でも指示をし、それをロボットが処理できるAPIに変換し、実行する

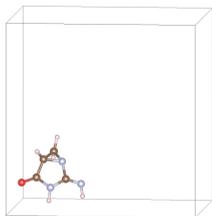
Matlantis = 原子レベルシミュレータ

第一原理計算結果を教師データとし、それを独自GNNで学習モデルを構築したNeural Network Potential (NNP)の一種

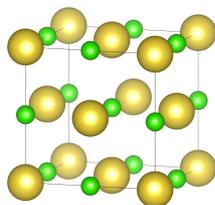
大量の訓練データを利用

分子・結晶等様々な構造の第一原理計算を実施
Preferred Networks社の強力計算設備を活用

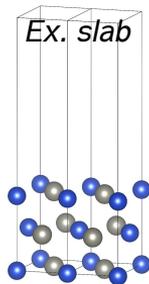
Ex. molecule



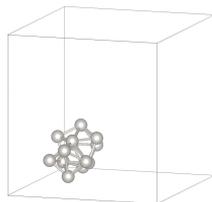
Ex. crystal



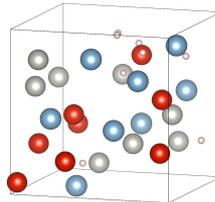
Ex. slab



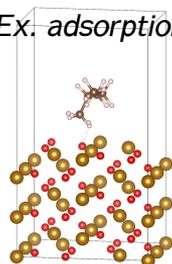
Ex. cluster



Ex. disordered

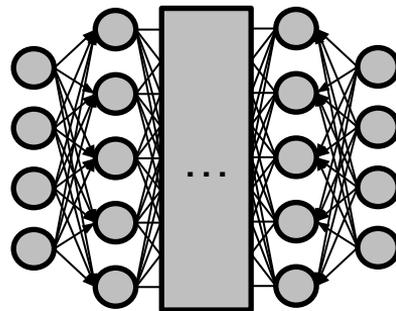


Ex. adsorption



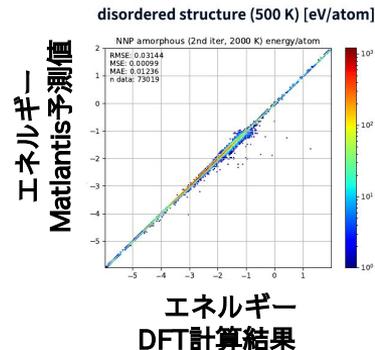
NN

量子化学知見を取り込んだ
独自のAIモデル構築



学習・出力

教師データを再現できるまで
モデルの学習を実施



PFNとENEOSが2021年よりサービス開始
国内外で80近くの企業・団体が利用



MATLANTIS

Matlantis™ by the numbers



Simulation Speed

20M

times faster for 3,000 atoms than conventional DFT method



Max. Input Atoms

20k

atoms for Pt bulk under Professional Plan



Supported Elements

72

elements or 99.9969% of all elements in the earth's crust



GPU Years Spent

1,650

years would be required for 1 GPU to generate the training data with more than 33 million structures



Clients

70+

companies and organizations



Atoms Simulated

2.6T

total atoms simulated in 3 months between January - March 2023

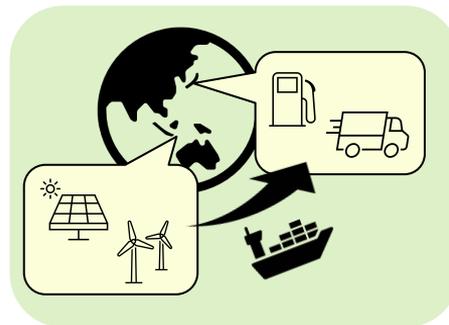
All figures are as of September 2023

The neural network potential powering Matlantis has been developed using National Institute of Advanced Industrial Science and Technology's AI Bridging Cloud Infrastructure (ABCI) and Preferred Networks's in-house supercomputers.

AIによる材料探索

材料探索は持続可能な社会の実現に向けた一つの鍵となっている

- 再生可能エネルギーを水素等で貯蔵する際の効率的な触媒開発
- 既存バッテリーの改良、新規バッテリーの開発
- 環境負荷の大きい材料の置き換え
- 希少な資源（レアアースなど）の置き換え
- リサイクル可能な材料への置き換え
- 製造プロセスの省エネルギー化
- 製造時の温室効果ガスの排出を抑える（CO2キャプチャー）
- 新エネルギー（核融合など）実現するための材料（超電導など）



今後必要な半導体研究

- 電力生産量は大きく変わらない中必要な計算力、データ転送量は指数的に増加
 - 電力あたりの計算量、データ転送量を飛躍的に改善できるような研究開発
 - 本格的にAIが普及した時は全世界で2000TWh/年以上の電力消費増加を予想
- AIを活用した半導体の素材開発・自動設計・シミュレーション技術
 - 既にプログラミングはAI（LLM）によって生産性が劇的に改善されている
 - 設計・課題解決に利用できるようなシミュレーション技術の開発
 - 半導体設計領域においてもAI利用、AIによる専門家のサポートが重要
 - シミュレーションできるようにすれば、自己学習・自動最適化が可能に
- 半導体を活かすためのソフトウェア研究開発
 - ハードウェアの性能を活かせるソフトウェア技術の開発
 - 主要なAIモデル（大規模言語モデルや基盤モデル、サロゲートモデル）を効率的に動かせるような技術（コンパイラ・自動最適化）

国(文科省)がとるべき施策

- 半導体の研究としての魅力を伝える
- 研究者が最先端の計算資源・AIモデルにアクセスできるようにする
 - 各分野での研究競争力を維持するため、これらのアクセスは必須
 - 大規模AIモデルの設計構築・維持は簡単ではない
 - それらの利用を支援できるような体制を作る(ドメインの専門知識がある人、AIモデルを設計・利用できる人との混合チーム)
 - 例えば弊社が様々なモデル構築や維持を支援することは可能
- AIを活用した科学技術研究を推進する
 - AI4Scienceの動きにあわせ、各研究分野での課題をAIを使って解決するための取り組みを進める
- 企業と研究の間を行き来する人を増やしていく
 - 一方通行の場合が多いが、行き来できるようにする。企業でじっくり取り組めない研究課題をアカデミックで行い、逆にアカデミックの知見を企業の活動で活かすようなことを地道に増やしていく