

2024/5/14 情報科学技術分野における戦略的重要研究開発領域に関する検討会

今後の生成AIに必要な研究開発

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

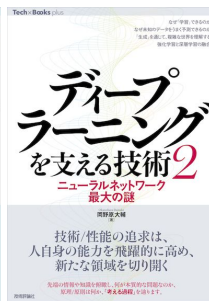
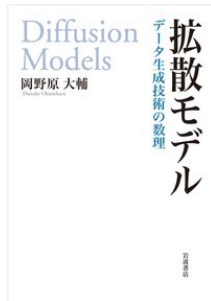
AI事業者ガイドライン 検討会委員



受賞歴

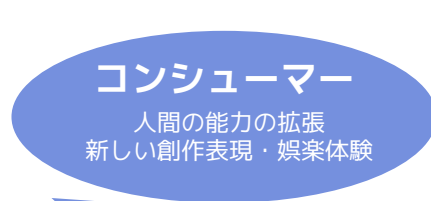
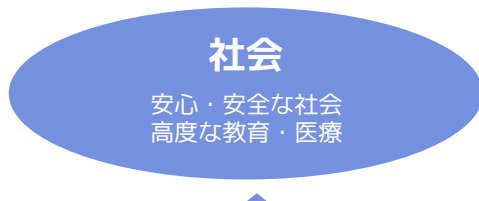
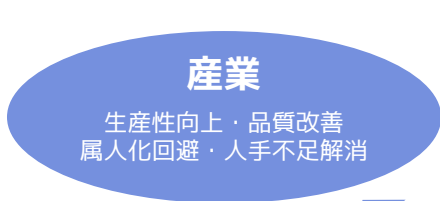
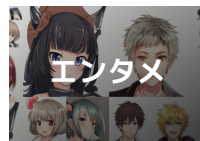
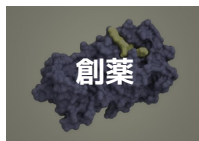
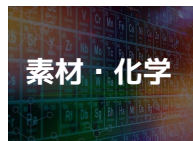
- ・ 令和4年度「現代の名工」データサイエンティスト
 - ・ KDDI Foundation Award 本賞 2022
 - ・ 大川出版賞（拡散モデル）
- など

著書

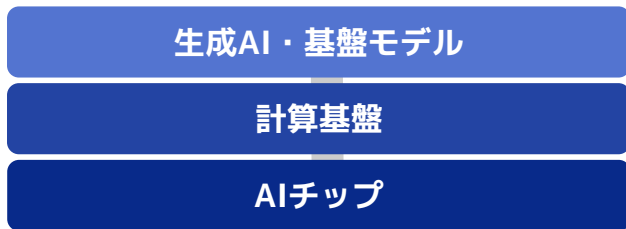


PFNグループは生成AIとそれを支える計算資源を提供する

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

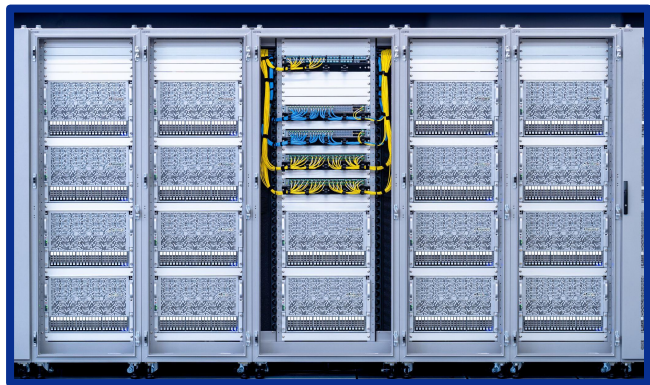


相互作用を引き起こしながら
イノベーションの連鎖へ



多くの仕事や社会生活が
AIによって支えられる

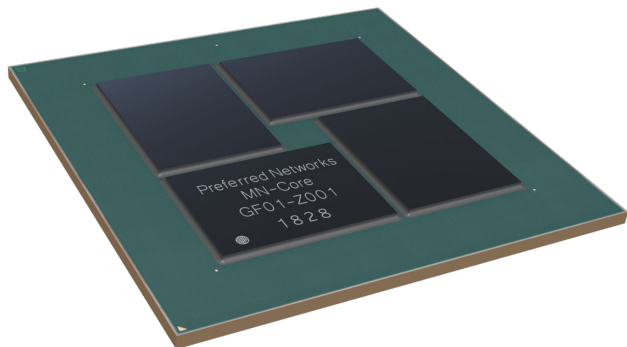
PFNはAI向け半導体、スーパーコンピュータを開発してきた



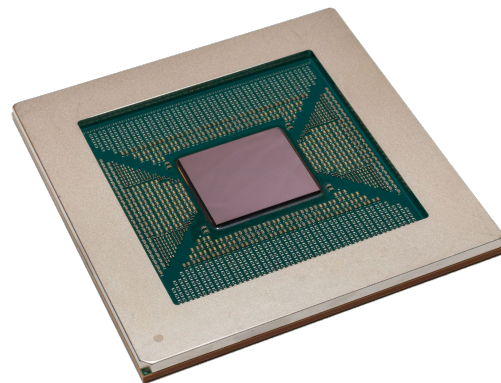
MN-3

省電力性能ランキングGreen500で3度 世界1位
(MN-Core搭載)

MN-Core



MN-Core 2



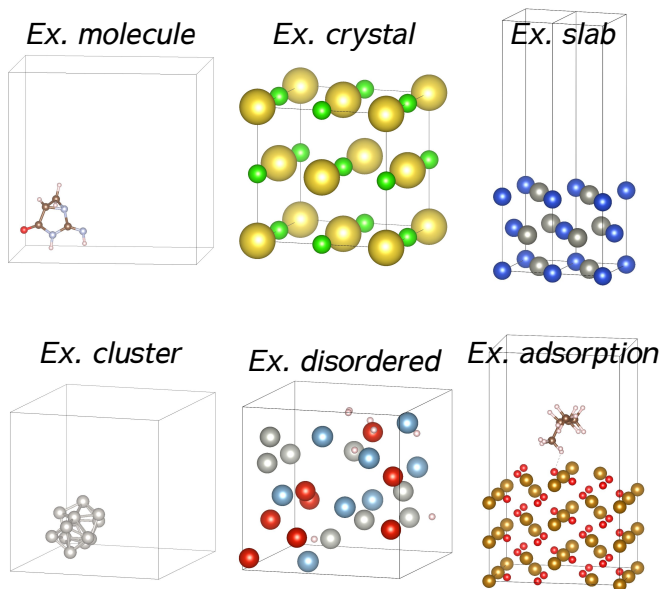
Matlantis = 原子レベルシミュレータ

第一原理計算結果を教師データにして、
独自GNNで学習したNeural Network Potential



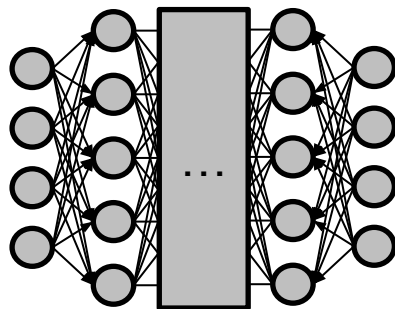
世界最大規模の訓練データを利用

分子・結晶等様々な構造の第一原理計算を実施
PFNの計算設備を活用。2000年分のGPU資源を利用

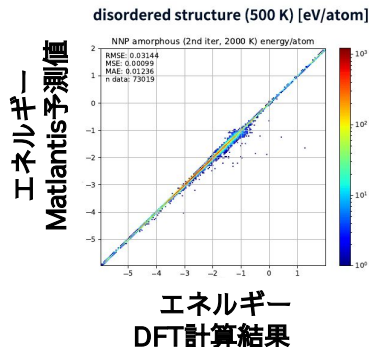


独自NN

量子化学知見を取り込んだ
世界最高レベルのAIモデル



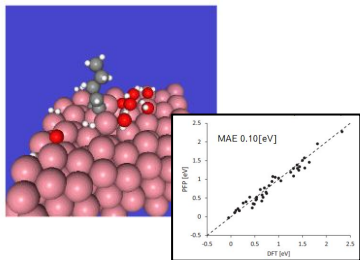
72元素の任意の組み合わせ
でDFT同等精度を
2千万倍高速に実現



PFNとENEOSが2021年よりサービス提供開始。
国内の企業や研究機関で利用で80近くの企業・団体が
利用。1年間で18兆原子がシミュレーションされている。

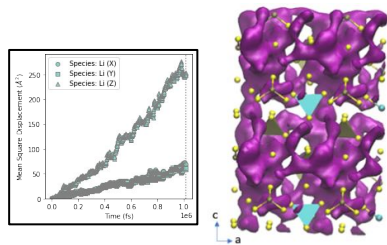
応用範囲

触媒



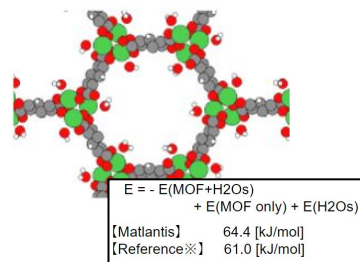
Large scale screening for optimal dopant species

バッテリー



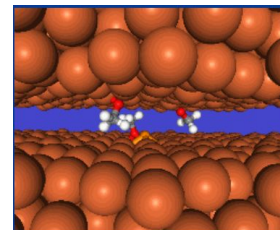
Li diffusion in solid electrolyte

吸着剤



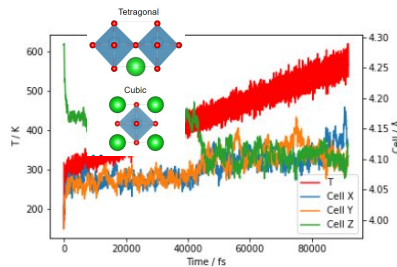
Adsorption energy in MOF

潤滑油



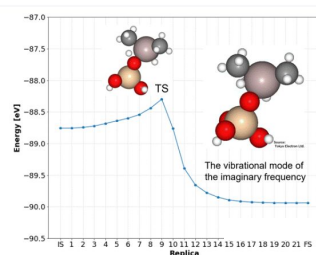
Tribochemical reaction

セラミックス



Phase transition in BaTiO₃

半導体



Reaction analysis of trimethylaluminum

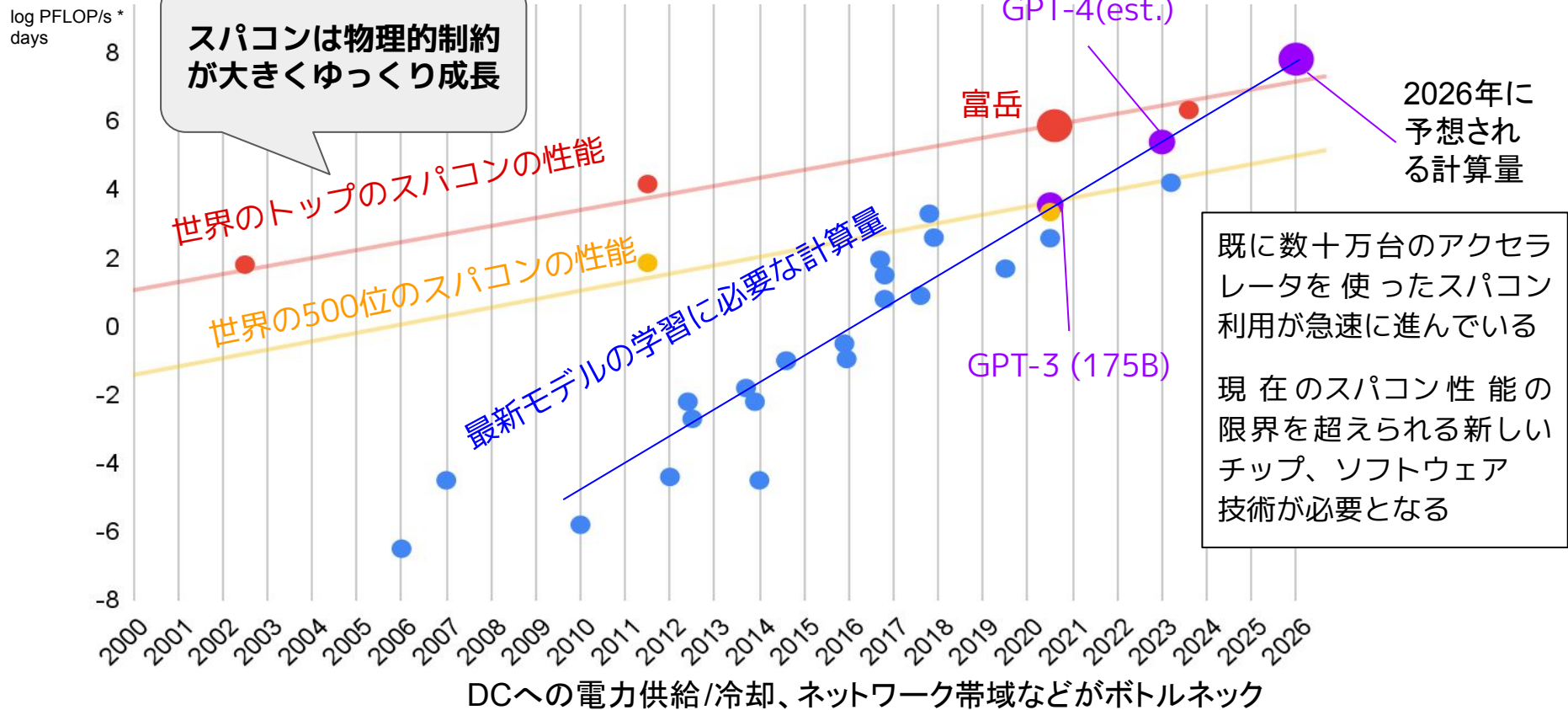
分離膜

金属 & 合金

⋮

- **100B** マルチモーダル基盤モデルを含めた複数モデルを**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において最高性能を目指す
 - アカデミアなどと協力し、安全性を検証した上でモデルを公開予定
- **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**に開始予定
 - 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）
 - エッジ向けに動くような小規模モデル、中規模モデルも同時にリリース予定
- 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

最先端AIの最新モデルの学習に必要な計算資源は増加



生成AIやその利用における研究開発

- 競争が激しいが研究テーマは無数にあり伸びしろは大きい
- 大規模言語モデル
 - 自己回帰に変わる学習目標 / Transformerに変わるアーキテクチャ / 安全性の理論保障 / 記憶の仕組みの改善
- 拡散モデル / フローマッチング / その他
 - 画像、音声、制御などの生成、動画（参考 SORA OpenAI）
- 次世代半導体向けの学習手法、モデルアーキテクチャ
- 生成モデルの理論解析、工学的改善
 - なぜ学習できるのか、できないのか
 - 世界の知識（帰納バイアス、特に対称性）をどのように組み込めるか
- 必ずしも巨大なモデルを作れなくても競争力のある研究は可能
 - パラメータ数あたりの能力も指数的に改善している

AIのためのデータ設計・生成

- 今後は学習以上に学習するためのデータを整備することに力が注がれる
- LLMや生成AIで作られたデータを使い学習することが急速に実用化している
 - 教科書のようなデータを生成したり、既存データをフィルタリング、ラベリングして学習しやすくする (Llama3, Phi3)
 - Wikipediaの数十倍も大きな知識データベースが作られていく
 - 学習データをコントロールでき、問題のある学習データを排除できる
- シミュレーションによるデータ生成や検証が現実的になっている
- データ作成、評価方法をどのように評価するか
 - データの良し悪しを今は人が評価しているが、近い将来はAIが評価するようになる（既にLLMを使った評価が一般的になりつつある）
 - 人間中心を保ったまま、それをAIに教えてスケールできる手法の確立

最先端の研究を支援できるようなAI

- 研究活動の支援
 - AI自身が研究をするより、研究に携わる人・チームを支援し、生産性や競争力を上げられるようにするのが現実的
- 論文、表、グラフを読み、実験結果を理解できるAI
 - 実験・プログラム・データ分析の補助
 - AI研究においても情報が既に溢れているためAIでキュレーションするツールが最新動向のキャッチアップには不可欠になっている
 - 例：AINewsは数百のX(twitter)やdiscord上の議論を要約
- 特に日本では半導体産業が重要である中、その研究開発の競争力を高められるようなAI開発ができるか？

社会にインパクトを与えるようなAIの開発

- 現在のLLMはまだ専門領域に使うためには能力が足りない
 - 推論能力、専門知識、常識、人との適切なインターフェースの欠如
- 3年で専門知識が必要なタスクをこなせるAIを開発する
 - 以下の全ての試験を3年以内に全て通過することは現実的な目標といえる
 - 医師・弁護士・司法書士・弁理士・公認会計士・税理士・社会保険労務士・行政書士・薬剤師
 - 一方、試験通過と実際の業務で使える部分には大きなギャップがあるため、それらを埋める、もしくは人と協調して作業できるようにする部分の研究開発が必要
- 仕事（Job）を置き換えるのではなく、特定の業務（Task）をAIがサポートする
 - 10年以内に上記業務の5割をAIでサポートできるようにするために必要な研究開発
 - 個人で多くの問題解決ができるようにする、専門職の方が今の仕事を半分の労力で終わらせられるようになる

世界レベルの競争力を持つための提言

- 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

補足資料

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

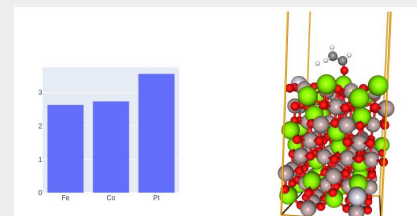
class NGLManager:
    def __init__(
        self, title: str, width: str = "400px", height: str =
    ) -> None:
        self.title = title
        self.ngl_viewer = RawNGLViewer(width=width,
        height=height)
        self.wrap = wrap
        self.recent_log = None
```

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]

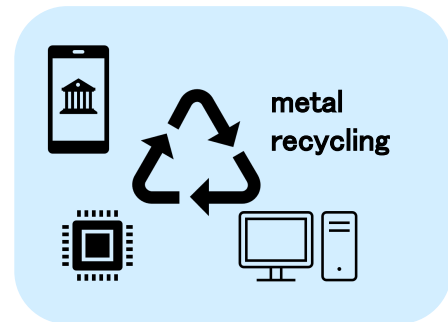
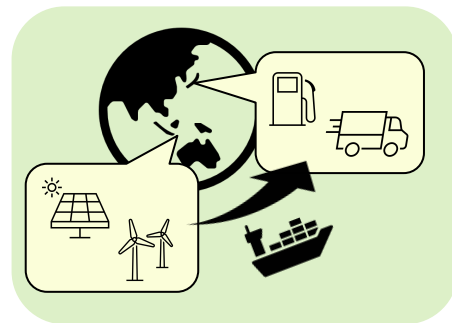
<https://wayve.ai/thinking/scaling-gaia-1/>

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

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なれば、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