

於 日本化学会 第102春季年会(2022)
イノベーション共創プログラム (CIP)
発表資料改

デジタルトランスフォーメーションに伴う科学技術およびイノベーションの変容
～化学・材料分野への示唆～

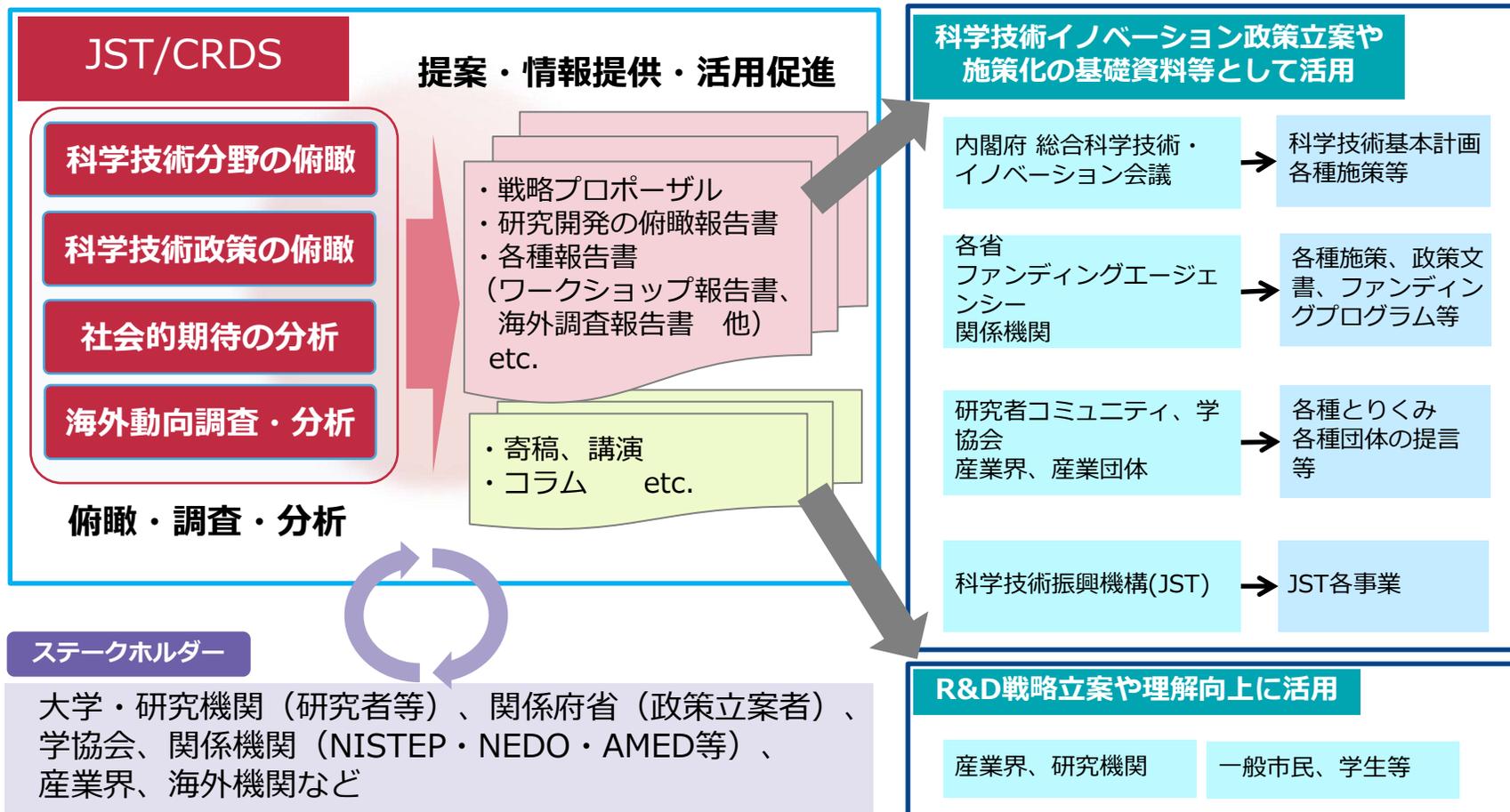
マテリアルズ・インフォマティクスの発展 と今後の展望

2022年4月5日

JST研究開発戦略センター(CRDS)
島津 博基

研究開発戦略センター（CRDS）活動概要

- ①国内外の社会や科学技術イノベーションの動向及びそれらに関する政策動向を把握し、俯瞰し、分析します。
- ②俯瞰報告書や研究開発戦略提言「戦略プロポーザル」をとりまとめ、提言の実現に向けた取組を行います。
- ③ワークショップ等を開催し、関係者の共通認識の醸成を図っています。



CRDSの活動スキーム

研究開発の俯瞰報告書

環境・エネルギー分野
システム・情報科学技術分野
ナノテクノロジー・材料分野
ライフサイエンス・臨床医学分野



俯瞰対象分野の全体像

- ・ 俯瞰の構造と範囲
- ・ 研究開発を取り巻く現状
- ・ 今後の展望と日本の研究開発戦略

研究開発領域（全約120領域）

- ・ 国内外の研究開発動向
- ・ 主要国の国際比較

戦略プロポーザル、調査報告書

CRDS-FY2013-SP-01

2013年

8年ちょっと経過

戦略プロポーザル
データ科学との連携・融合による
新世代物質・材料設計研究の促進
(マテリアルズ・インフォマティクス)
～物質・材料研究を飛躍的に発展させるための新たなパラダイム～

STRATEGIC PROPOSAL
Materials Informatics
Materials Design by Digital Data Driven Method.

CRDS-FY2020-RR-01

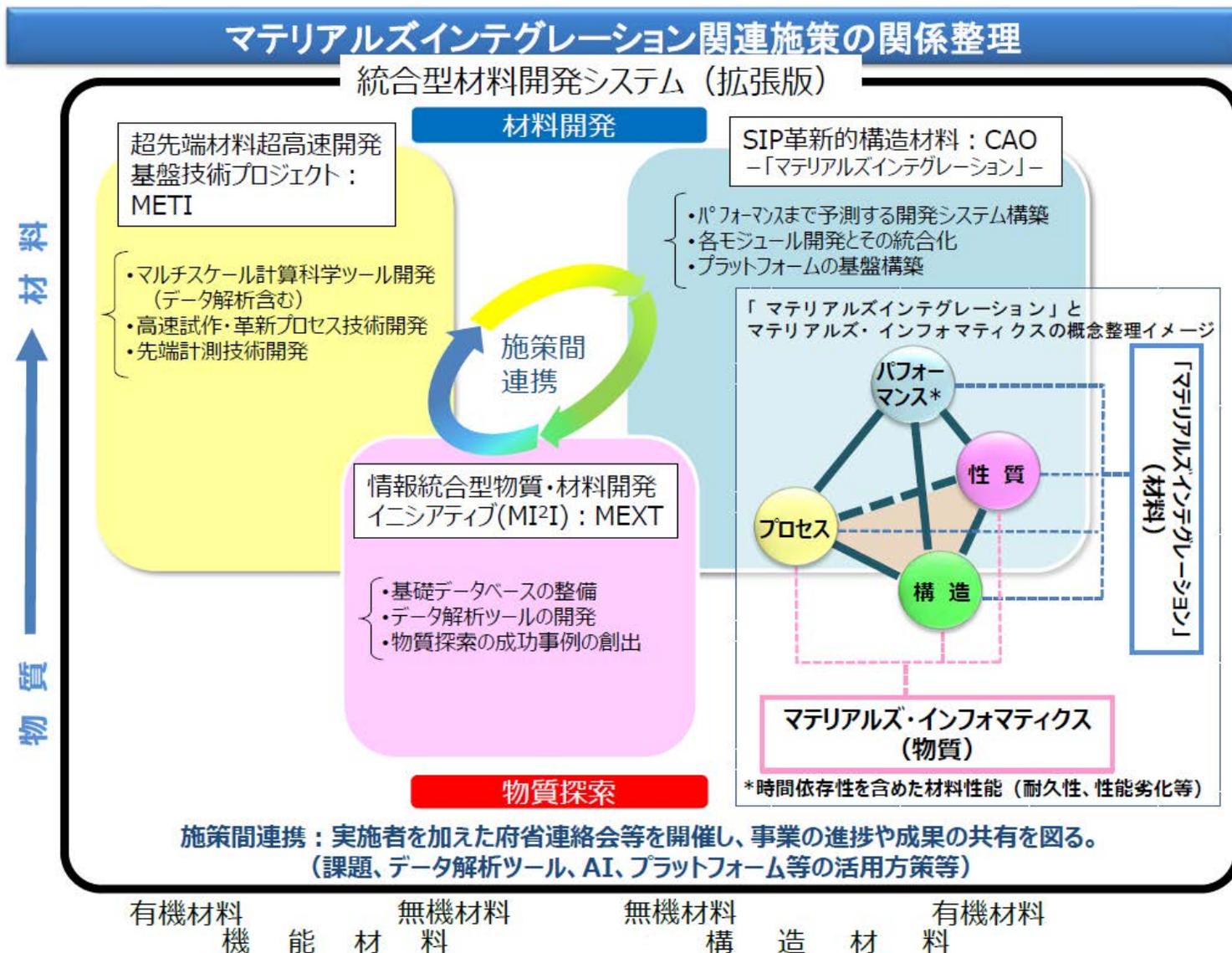
The Beyond Disciplines Collection
デジタルトランスフォーメーションに伴う
科学技術・イノベーションの変容

CRDS-FY2020-RR-03

The Beyond Disciplines Collection
AI×バイオ DX時代のライフサイエンス・
バイオメディカル研究

提言の政策実装（政策立案者や研究者との対話とプロジェクト化）

- JSTさきがけ[マテリアルズインフォ]理論・実験・計算科学とデータ科学が連携・融合した先進的マテリアルズインフォマティクスのための基盤技術の構築（2015～）
- NIMS「情報統合型物質・材料開発イニシアチブ」（2015～）



本日の話題提供の構成

1. AI・DXと科学技術イノベーションの動向

2. MIの発展の歴史

3. MIの研究動向（事例紹介）

① 予測／探索（計算シミュレーションの活用）

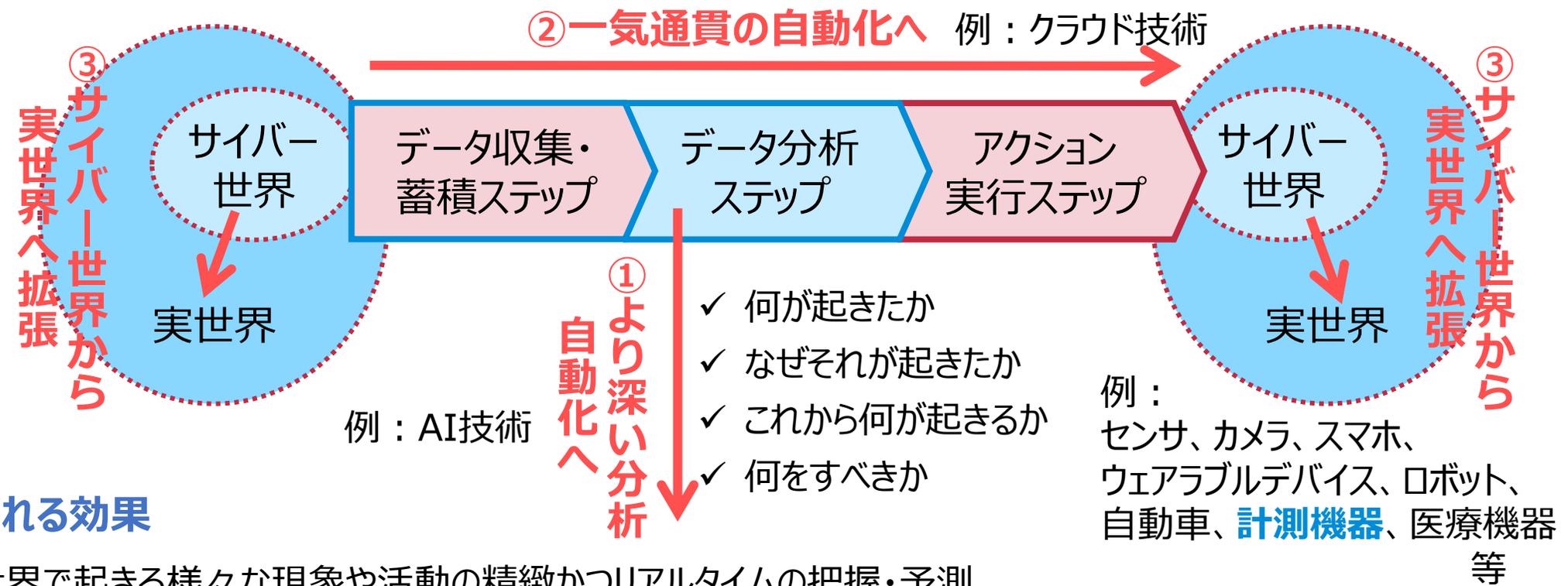
② ディープラーニング（識別モデル、生成モデル、自然言語処理など）の活用

③ 自動化・自律化（実験へのロボット、プログラム制御等の活用）

4. まとめと今後の展望

デジタルトランスフォーメーション（DX） ～データに基づく新たな価値の創出～

- ビッグデータとAI技術の発展・普及によって、データに基づく問題解決プロセスが広く活用されるようになった
- このプロセスは、3方向（AI、クラウド、デバイス等からのRWD）の技術的発展によって進化
- 世界的にあらゆる業界がDXの潮流に直面している。



もたらされる効果

- 実世界で起きる様々な現象や活動の精緻かつリアルタイムの把握・予測
- 膨大な選択肢の網羅的な検証
- 大規模複雑タスクの自動実行等

R&D発見・プロセスの効率化

計測対象：物質・材料・デバイス・
機械、分子・細胞

創薬 バイオ生産
物質・材料 ものづくり

相違点

- ・ 系（開放・閉鎖）
- ・ パラメータの多様性
- ・ ステークホルダーの多様性（データへのアクセス）

顧客起点での新たな付加価値の創出 （ビジネスや社会課題解決）

計測対象：個人・集団、環境、社会

気象・気候 エネルギー
生物多様性 都市
農業
医療・ヘルスケア モビリティ

基盤技術

データ取得技術：
物理・化学的
（量）計測

データ取得技術：
センサ・ロボット
（環境・社会計測）

データ処理技術：
AI・機械学習

共通基盤：
半導体デバイス

データ基盤フレームワーク・
システム

5G通信 セキュリティ

ブロックチェーン

GAFAなどのITベンチャーの躍進（SaaS、Fintechの台頭）

PC

1975 **Microsoft**

1976 **Apple**

1984 Dell（テキサス学生）

1984 MIPS（スタンフォード教員）

1993 **NVIDIA**

インターネット（検索エンジン・EC）

1994 **AMAZON**

1994 Yahoo（スタンフォード学生）

1994 Netscape（スタンフォード教員）

1997 Netflix

1998 **Google**（スタンフォード学生）

1999 Napster（ノースイースタン学生）

ブログ・SNS/スマホ

2003 Android

2003 WordPress（ヒューストン学生）

2003 **Tesla**

2004 **Facebook**（ハーバード学生）

2005 Youtube

2005 Reddit（バージニア学生）

2006 Twitter

2010 Instagram

X-tech/SaaS

2008 Airbnb

2009 Uber

2011 SoFi（スタンフォード学生・fintech）

2012 Lyft

2012 Instacart

2012 Coursera（スタンフォード教員・edtech）

2013 Robinhood

2013 DoorDash（スタンフォード学生）

2013 Databricks（UCバークレー・データ分析基盤）

2010年のDeepMindをはじめAIベンチャーも次々に誕生

- 中国では、アリババ、テンセント、シャオミ、メイトゥアン、バイトダンス等
- これらの企業が世界の企業の時価総額の上位を占める。

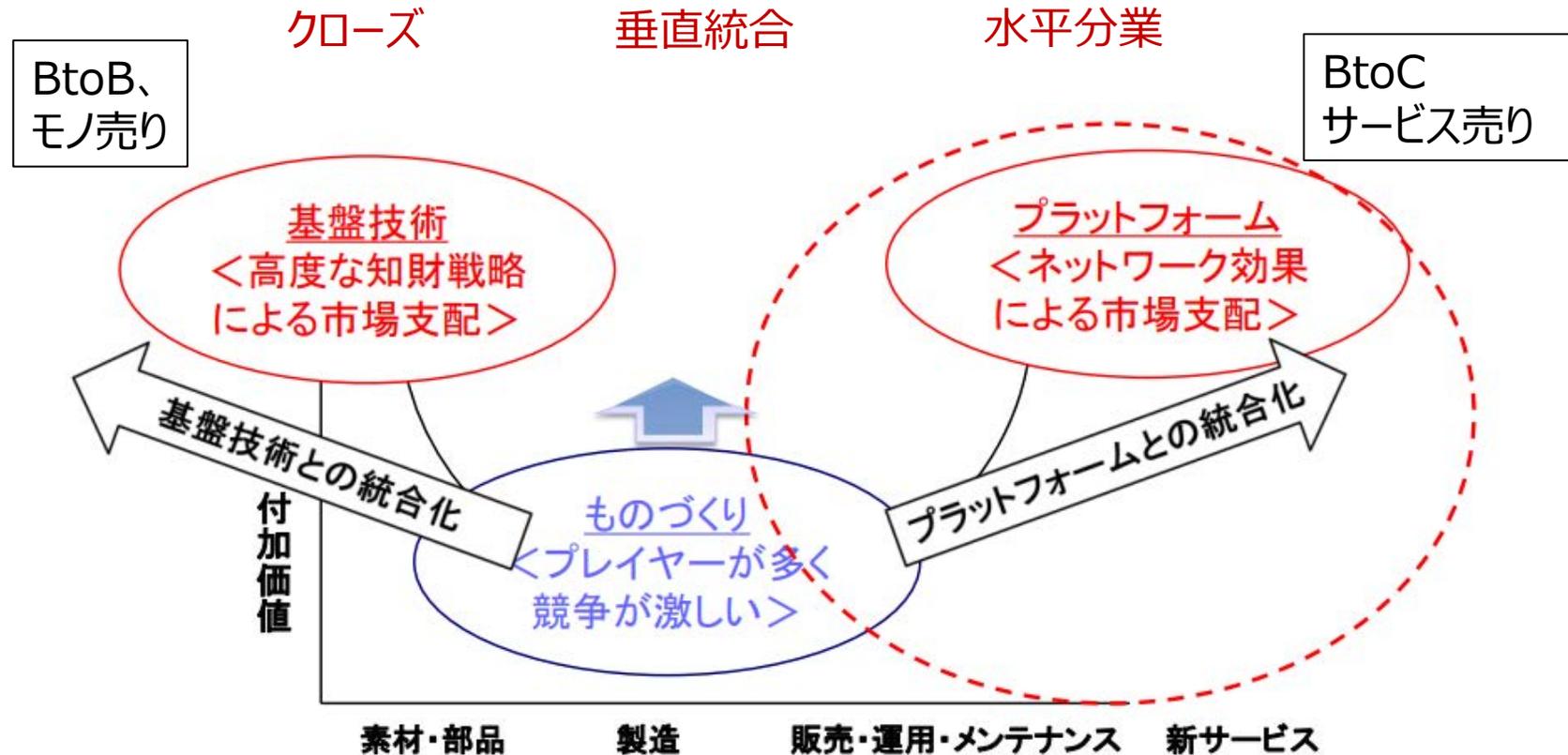
産業構造の変化 ～バリューチェーン上の付加価値の局在（スマイルカーブ）～

- プラットフォームの方が社会に近く産業のパイが圧倒的に大きい。
- 日本が強いのはプロセスと基盤技術

ビジネスモデルの視点が必要

例：素材、半導体、電子部品、医薬
：プロセス

例：SaaS、X-tech
：プラットフォーム

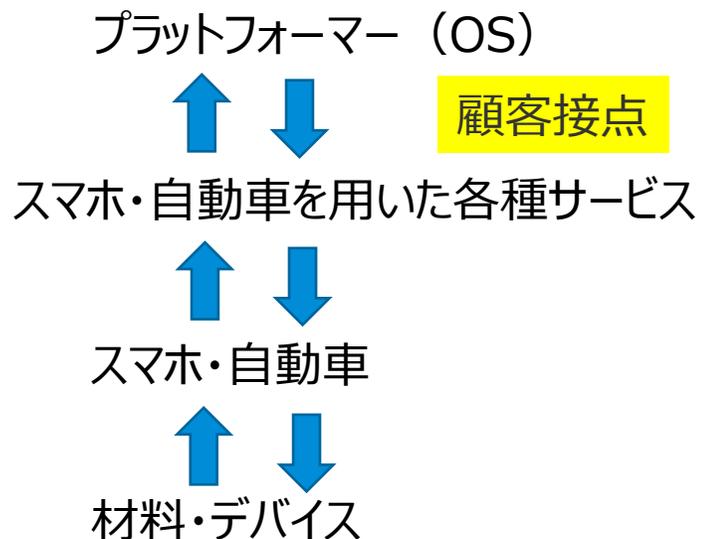


産業構造の変化 ～プラットフォーマーの台頭～

- 2020年4月、GAFAMの時価総額、東証1部2,169社超え560兆円に
- 2020年12月、Elon Muskは、テスラの5540億ドル（約57兆8400億円）という時価総額を利用して、既存自動車メーカーの買収を検討
- 2021年1月、米アップルが電気自動車（EV）市場への参入を目指し複数の自動車メーカーと提携交渉を始めたことが明らかに
- 米国下院司法委員会がGAFAがデジタル市場を独占的に支配し、競争を妨げているとする報告書を公表
- 司法省も独占禁止法に違反しているとして、グーグルを提訴
- EUは2020年12月、テック系企業が欧州でビジネスを行うためのルールを定めた法案、デジタルサービス法とデジタルマーケット法を発表
- 自社のサイトで自社のサービスなどを優先的に扱うことを制限、ユーザーごとに表示されるオンライン広告の基準の開示、差別やテロを促すような不適切な表現を削除することも義務づけ等

垂直統合モデルと水平分業モデル

- 米テスラ・モーターズの台頭や米IT大手のグーグルなどの異業種参入で「自動車業界の競争ルールが大きく変わろうとしている」。
- トヨタは、通信大手3社との協業を推進。NTT（スマートシティ）、ソフトバンク（モビリティサービス）、KDDI（街や住宅、人、クルマがつながる社会）



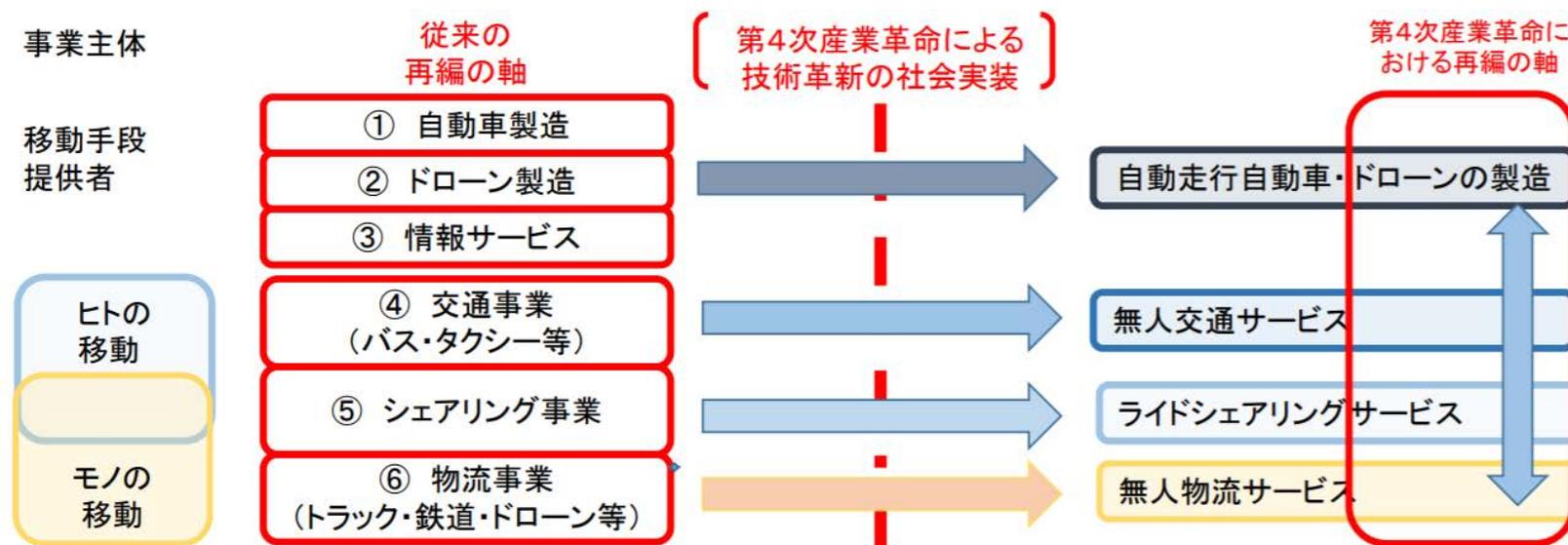
DXがもたらすビジネスモデル変革の事例①

自動化（電動化）によって参入障壁が低下し、従来の業界の枠にとらわれない新しいビジネスが生まれ得る

例：自動車製造→モビリティプラットフォーム

- 豊田社長は、米テスラ・モーターズの台頭や米IT大手のGoogleなどの異業種参入で「自動車業界の競争ルールが大きく変わろうとしている」と指摘。
- 通信大手3社との協業を推進。NTT（スマートシティ）、ソフトバンク（モビリティサービス）、KDDI（街や住宅、人、クルマがつながる社会）

[例] ライドシェアビジネス



出典：経済産業省

DXがもたらすビジネスモデル変革の事例②

個別最適化ではなく全体最適化が可能になることによってシステムが再編され得る

例：ものづくり→サービス→ソリューション

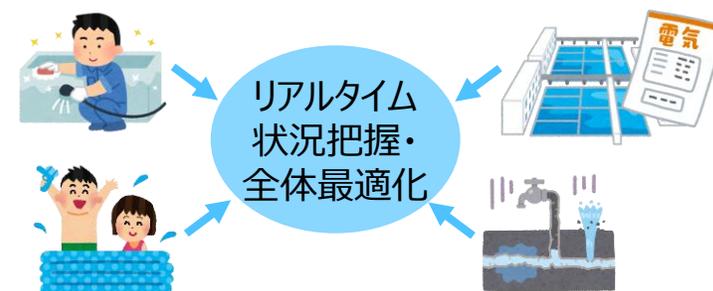
- 建機大手コマツでは、KOMTRAXというシステムを使って、全世界に展開する建機40万台をネットワークで繋いで常時監視、遠隔制御。
- KomConnectによりデータ活用をさらにレベルアップさせて、工事現場にドローンを飛ばして3次元測量データを取得、このデータを読み込んで作業計画やシミュレーションに利用。



- ・センサーから稼働状況を取得。
- ・機械内蔵の端末を通じ、オイルや部品の交換時期を顧客に伝達。
- ・同じ情報をコマツの販売代理店にも同時に発信。

- ・ドローンで実測した3次元データを用いつつ、建機を自動制御し、土木工事の省力化と工期短縮を実現。

[例] 省電力・需要充足・保守費用を同時最適化する社会インフラオペレーション



DXがもたらすプロセス変革の事例

センサなどからの大量のRWD、ロボティクスとAIにより、限りなく現実世界に近い環境を計算機上に構築し運用しているデジタルツイン工場

- BMWには40種類以上のモデルがあり、車両ごとに100種類以上のオプションを選ぶことができるため、車両1台あたり2100通りもの組み合わせが存在
- 工場を設計または再構成する際はバーチャルと現実をリアルタイムでシミュレーションできるオープンプラットフォーム「NVIDIA Omniverse」を活用することで、様々な異なるソフトウェアを使用しながら共同作業を進め、工場の設計や配置を3Dで検討、すべての変更点はリアルタイムで可視化
- BMWの工場（デジタル ツイン）では、製造ラインあたり最大10種類の車種を生産可能にした

5Gプライベート回線による無線ロボットの工場利用

- ボッシュ・レックスロスの工場では、製造装置とロボットアームの間を車輪付きロボットが動き回っている。
- ロボットなどの機器に高速な無線リンクを追加することにより、高精度な連係やキャリブレーション（調整）が可能になり、大きな損害につながる故障や稼働停止の予測に役立つ



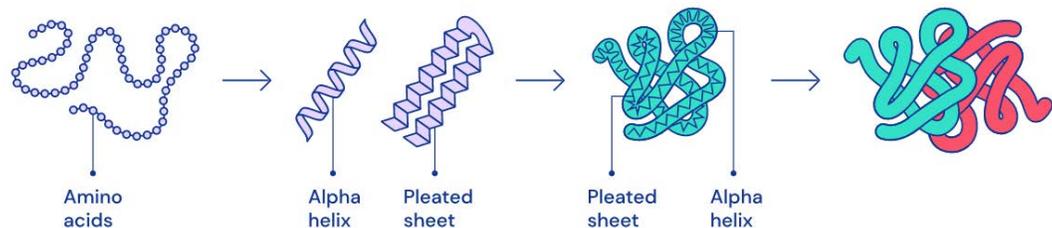
出典：NVIDIA社



出典：CHRISTOPH SCHMIDT/GETTY IMAGES

DX (AI) による科学研究成果の例

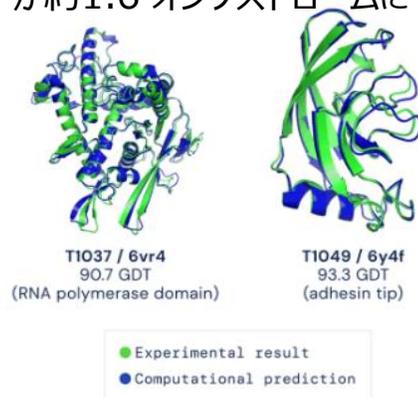
アミノ酸配列からタンパク質の3次元構造（フォールディング）を解析（DeepMind）



事例：AlphaFold

特徴量として、アミノ酸同士の距離、化学結合の間の角度、エネルギー効率の良い配置

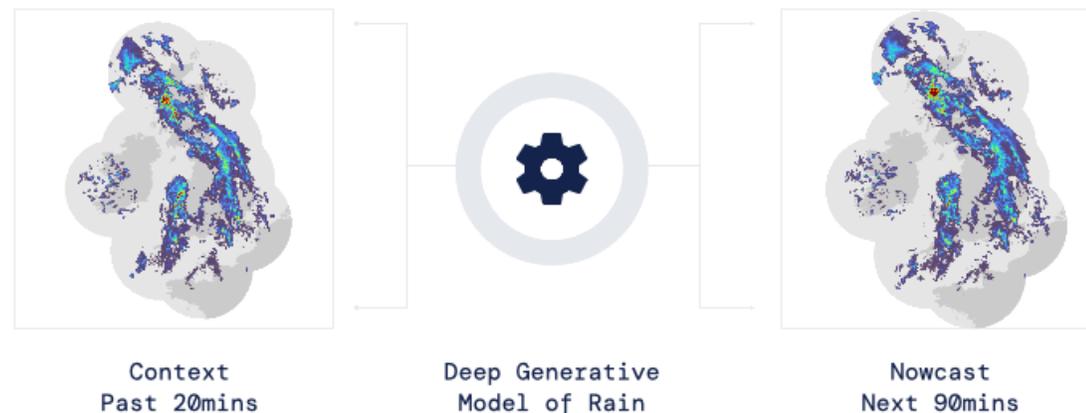
- 2019時点では、予測と実験で6オングストロームの誤差があった（クライオ電顕の解像度は1-2オングストローム）。
- 2020年には予測の平均誤差（RMSD）が約1.6オングストロームに
- 35万個のタンパク質構造をDB化。今後数カ月のうちに、さらに1億個以上のタンパク質の構造を予測し、公開する予定
- Isomorphic Labsを設立：「生体システムを第一原理から理解し、病気治療の新方法を発見する計算プラットフォームを開発する」



Nature volume 596, pages583–589 (2021)

天気予報の革新（DeepMind）

- DeepMindとエクセター大学が協力して「ナウキャスト」と呼ばれるAI主導の短時間天気予報システムを開発
- これまでのシステムでは、6時間から2週間で天気を予測していた。ナウキャストは2時間で天気を予測できる。



過去20分間の観測レーダーは、Deep Generative Model of Rain (DGMR) を使用して、次の90分間の確率的予測を提供するために使用される。

Nature volume 597, pages672–677 (2021)

本日の話題提供の構成

1. AI・DXと科学技術イノベーションの動向

2. MIの発展の歴史

3. MIの研究動向（事例紹介）

① 予測／探索（計算シミュレーションの活用）

② ディープラーニング（識別モデル、生成モデル、自然言語処理など）の活用

③ 自動化・自律化（実験へのロボット、プログラム制御等の活用）

4. まとめと今後の展望

DXと科学技術・イノベーションの変容

R&D発見・プロセスの効率化

計測対象：物質・材料・デバイス・
機械、分子・細胞

創薬

バイオ生産

物質・材料

ものづくり

相違点

- ・ 系（開放・閉鎖）
- ・ パラメータの多様性
- ・ ステークホルダーの多様性（データへのアクセス）

顧客起点での新たな付加価値の創出 （ビジネスや社会課題解決）

計測対象：個人・集団、環境、社会

気象・気候

エネルギー

生物多様性

都市

農業

モビリティ

医療・ヘルスケア

基盤技術

データ取得技術：
物理・化学的
（量）計測

データ処理技術：
AI・機械学習

データ基盤フレームワーク・
システム

データ取得技術：
センサ・ロボット
（環境・社会計測）

共通基盤：
半導体デバイス

5G通信

セキュリティ

ブロックチェーン

コンビナトリアルケミストリー／ケモインフォマティクス（1990年代中頃～）

- コンビナトリアルケミストリー：数十万の化合物を一挙に合成し、一度に評価を行う実験系
- ケモインフォマティクス：特に化学におけるグラフ理論やin silico virtualライブラリを用いることによる化学物質空間のデータ探索や定量的構造活性相関 (QSAR)に関する研究

ハイスループットスクリーニング

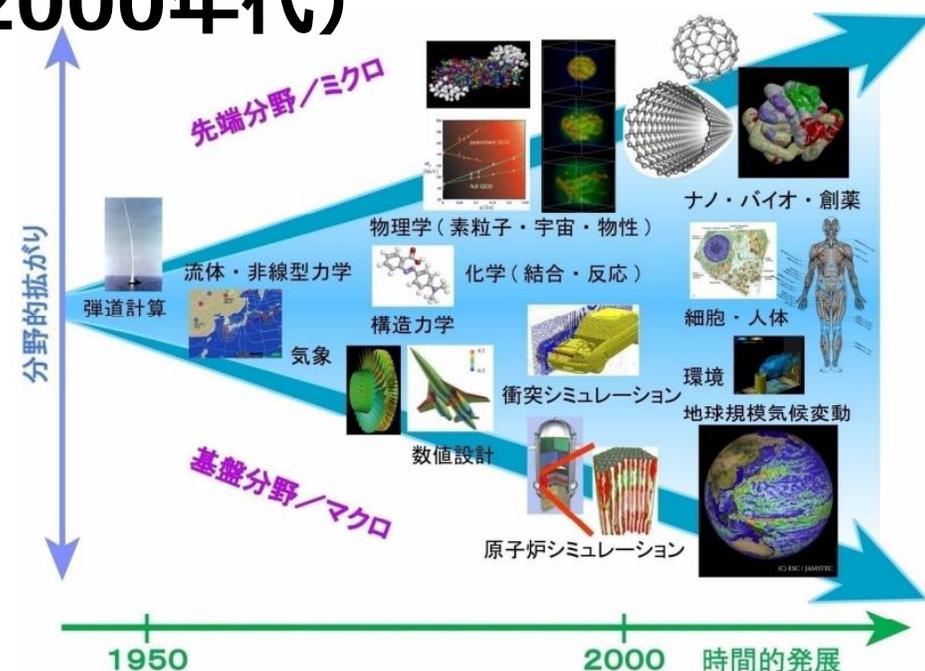
スパコン／DFT計算、計算化学の発展・普及（2000年代）

- 計算機は30年間（1980年～2010年）で1000万倍の高速化
- Natureのチームの調査によると、2014年に最も引用された上位100の論文のうち12がDFTに関するものであった。

『研究者は、新しいデータの急流に対応して組織と慣行を適応させる必要があり、スマートサイエンスをスマート検索で補完する必要がある。』



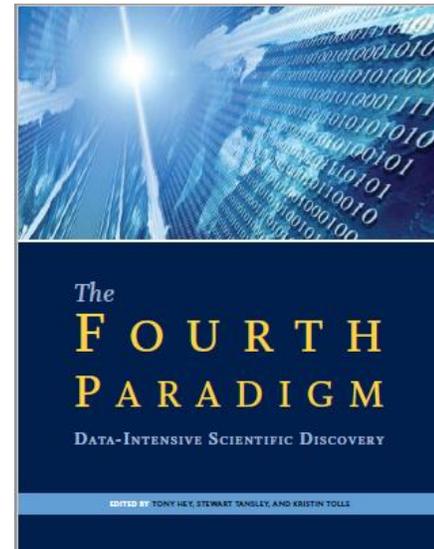
Volume 455 Issue 7209, 4 September 2008



The Fourth Paradigm

Data-Intensive Scientific Discovery (2009年)

- 実験科学、理論科学、計算科学に次ぐ、データ科学（第4の科学）を提唱
- X-Infoの概念を提唱



出典：MICROSOFT RESEARCH

Science Paradigms

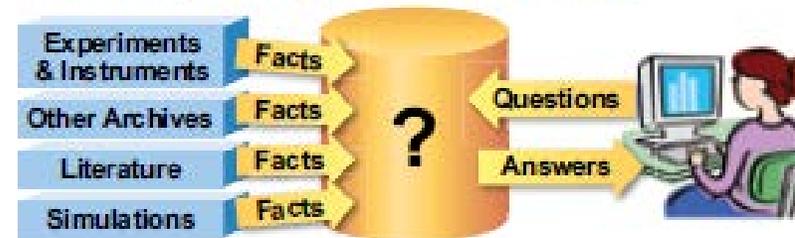
- Thousand years ago: science was **empirical**
describing natural phenomena
- Last few hundred years: **theoretical** branch
using models, generalizations
- Last few decades: a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics

$$\left(\frac{a}{a}\right)^2 = \frac{4\pi G p}{3} K \frac{c^2}{a^2}$$



X-Info

- The evolution of X-Info and Comp-X for each discipline X
- How to codify and represent our knowledge



The Generic Problems

- Data ingest
- Managing a petabyte
- Common schema
- How to organize it
- How to reorganize it
- How to share it with others
- Query and Vis tools
- Building and executing models
- Integrating data and literature
- Documenting experiments
- Curation and long-term preservation

米国 Material Genome Initiative (2011年)

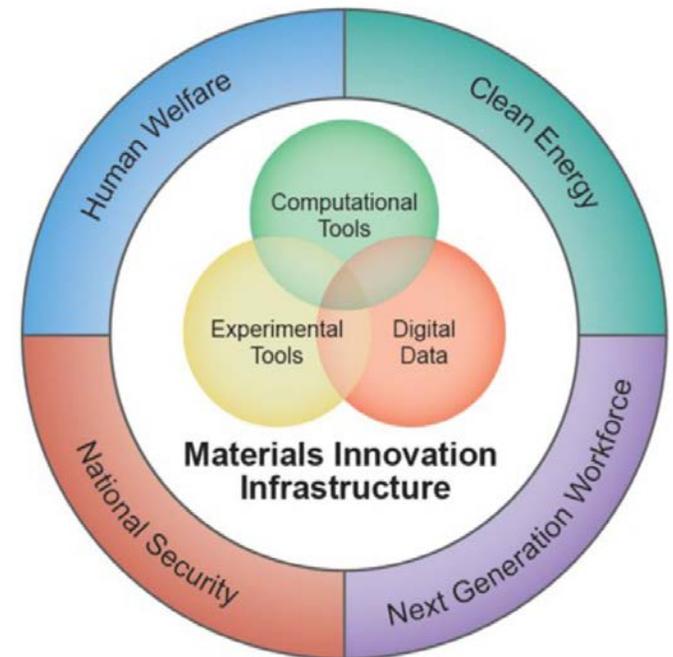
材料開発のスピードとコストを半分に！！

「生物ゲノム」は、4つの塩基の組み合わせ

「材料ゲノム」は、周期表で使用可能な118元素の組み合わせ

1. 物質・材料研究において、計算、データ、実験を連携させた統合アプローチを主流にするための研究者意識の醸成
2. 実験、計算、理論の統合
3. データへの容易なアクセス環境の整備
4. 世界水準の人材育成

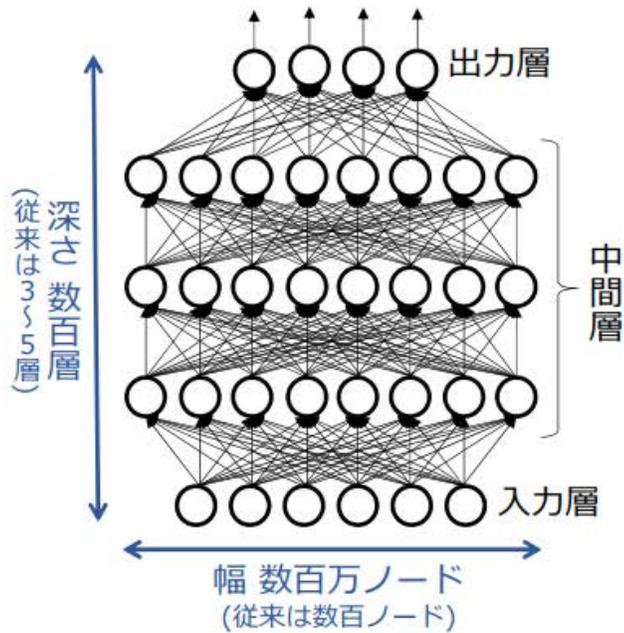
なお、このレポートの中には、AIはおろか、機械学習、データマイニングといった用語は一切出てこない。



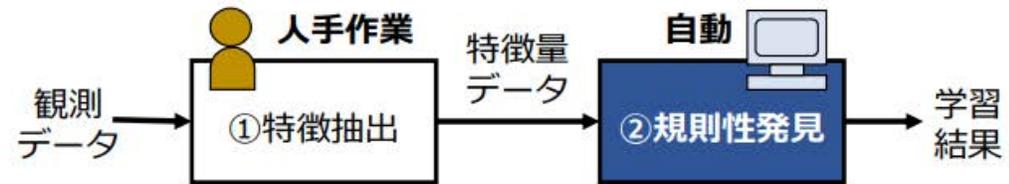
深層学習(Deep Learning)の出現と第3次AIブーム (2012年頃～)

- 機械学習技術の進化×ビッグデータが牽引
- 第2次ブームのときはデータからルール／モデルを人手で記述。第3次ブームは機械学習が作成。

多層ニューラルネットワーク(Deep Neural Network)を用いた機械学習技術
第3次AIブームの中心



(A)従来は①を人手で決めることが多かった



(B)深層学習の場合は①②を通して自動化



出典：福島俊一 (CRDS) 改

AIの技術的進展

- AIによって、画像や音声、文字列などパターンの認識能力が大きく向上した。
- 新しいAI技術が登場して5～6年で応用分野で実装されてきた。

識別モデル

- **2012年**、「畳み込みニューラルネットワーク（CNN）」が画像認識で従来手法と比較して圧倒的精度を発揮（ILSVRC）。
- 2015年、人間による画像認識のエラー率を下回る（ILSVRC）。
- 2016年、17年に網膜や皮膚がんの画像にディープラーニングを用いた論文が出て高被引用に。
- **2018年**、米国FDAは糖尿病網膜症を検出するAIを用いたデバイスを、AIが最終的な診断を下す医療機器として初めて承認。

生成モデル

- **2014年**、画像生成のためのアルゴリズム「敵対的生成ネットワーク（GAN）」発表。
- **2017年**、ディープラーニングを用いた分子生成「VAE(Variational Auto-Encoder)」発表。
- 2017年頃から、世界中でAI（創薬）ベンチャーが続々と大手製薬会社と共同研究を開始。
- 2019年9月、Insilico Medicine社は、通常2～3年を要するリード化合物の特定から前臨床試験までを、AIを用いて、わずか21日で完了、という論文を発表。
- **2020年**1月、大日本住友製薬と英ベンチャーExscientiaはAIを活用して創製した新薬候補化合物のフェーズ1試験を開始と発表。業界平均で探索研究に4年半かかるところ、12カ月未満で完了。

深層強化学習

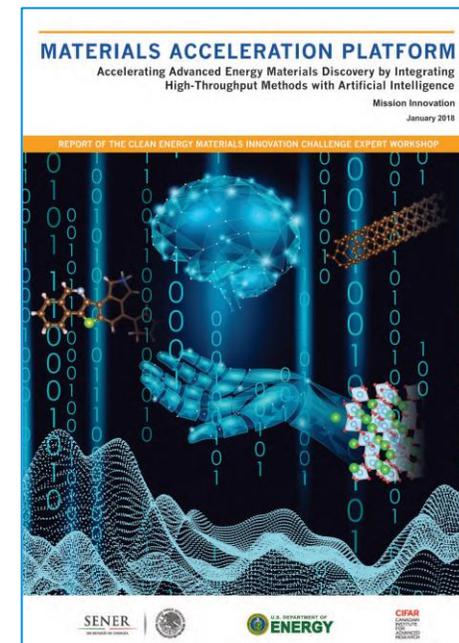
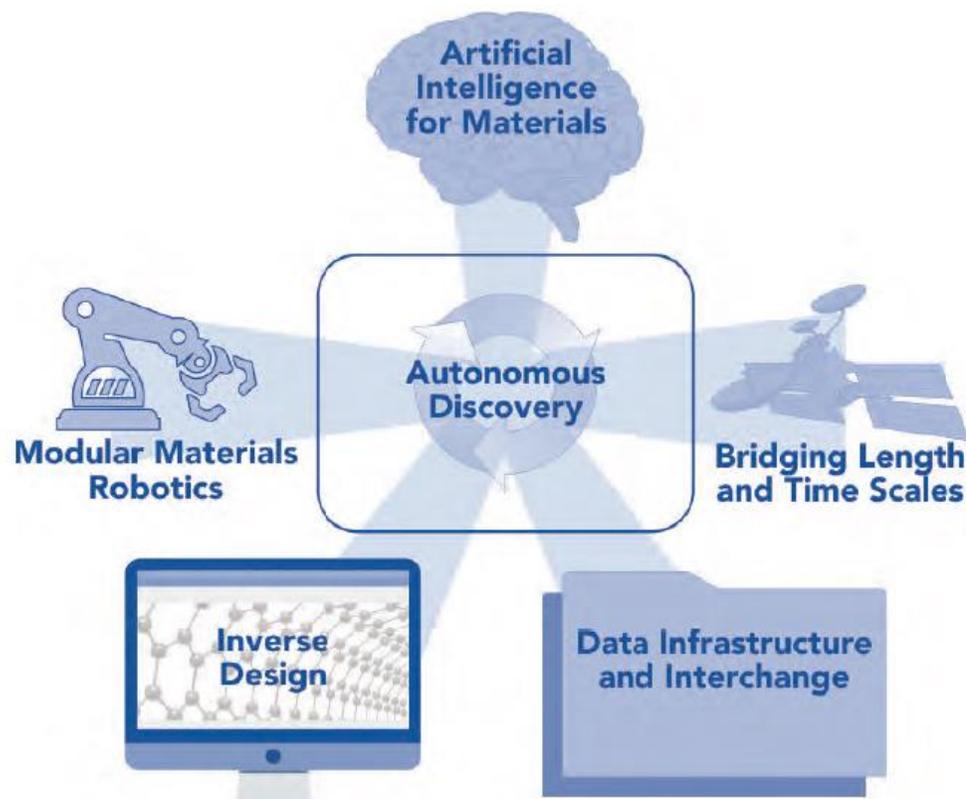
- ディープラーニングを強化学習に応用した学習手法。強化学習とは、試行錯誤を通じて「価値を最大化するような行動」を学習する手法。「現状（状態）の確認」→「アクションと変化」→「評価と報酬」の繰り返しにより、外部からの入力に対応して自律的な制御を行うことを可能に。
- 強化学習法の一つに「モンテカルロ木探索」と呼ばれる「統計的に勝つ確率の高い一手」を計算するアルゴリズムがある。
- **2016年**、Google（Deepmind）が開発した**AlphaGo**がプロ棋士を破る。
- 2017年頃から、化学反応予測や創薬に用いられる。

言語・音声処理モデル

- 電子カルテ、問診などに活用。英Babylon Health（2013～）のスマートフォンによるチャットで、現在の体調や症状を申告すると回答してくれるサービス。Covid-19にも活用された。
- **2018年**、Googleが自然言語処理モデル「BERT」を発表
- 2021年4月、マイクロソフトが音声認識技術を手がけるNuanceを160億ドル（約2兆円）で買収する計画を発表。Nuanceのソフトウェアは医師と患者の会話を聴き、音声での会話を整理された電子的な医療記録へと書き起こす。

Materials Acceleration Platform (2018年)

- 米国DOEがAI等を活用した材料開発の今後の展望について、研究者のワークショップなどで議論した結果を報告
- 「新規物質・材料発見の自律化」に向けて、AI、ロボティクス、逆設計、データインフラ、時空間スケールの統合、が重要な要素となる。
- データインフラの構築には、データ取得の自動化とAIの連携が有効で、そのためには、これまで測定技術・機器セントリックな環境を構築していたものを、サンプルセントリックな研究環境に移行していくことも視野に入れるべき。



Alán Aspuru-Guzik, Harvard University
Kristin Persson, Lawrence Berkeley

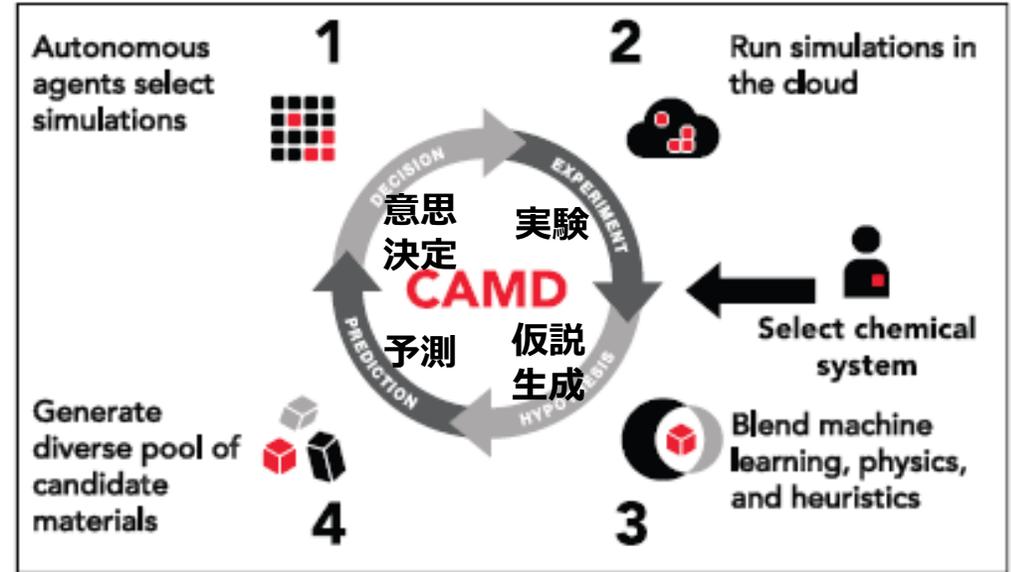
材料研究の自律化に向けて

Toward autonomous materials research (トヨタ・リサーチ・インスティテュート)

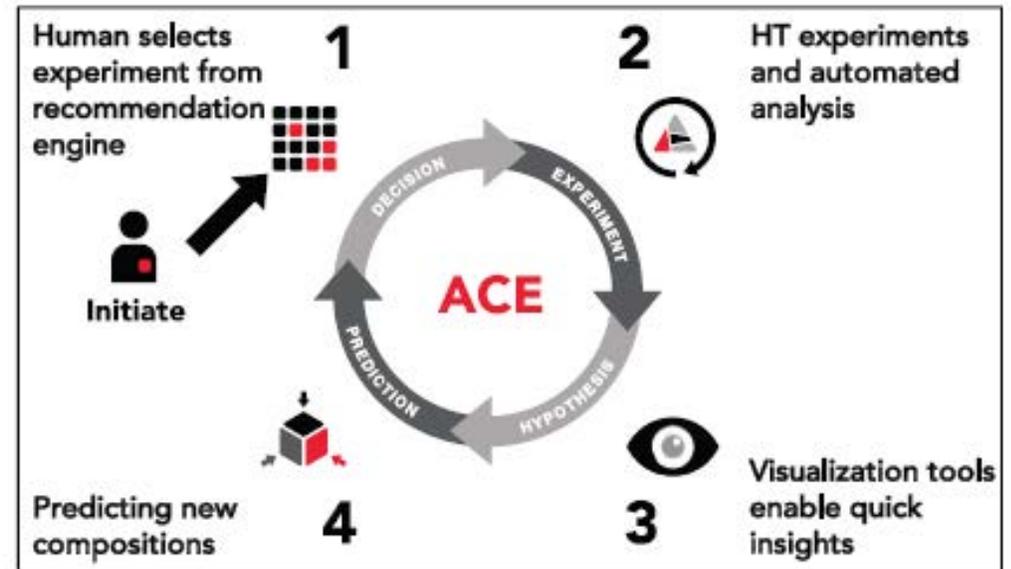
- 最近では、仮説生成、意思決定にAI・機械学習を用いることが可能となり、クローズドループプロセス全体を自動化することが研究のフロンティアとして注目されている。
- 材料科学のための完全自律型研究システムは、もはや遠い未来の話ではない。

Appl. Phys. Rev. 9, 011405 (2022)

①研究者は化学系を選択 例：Mg-V-O

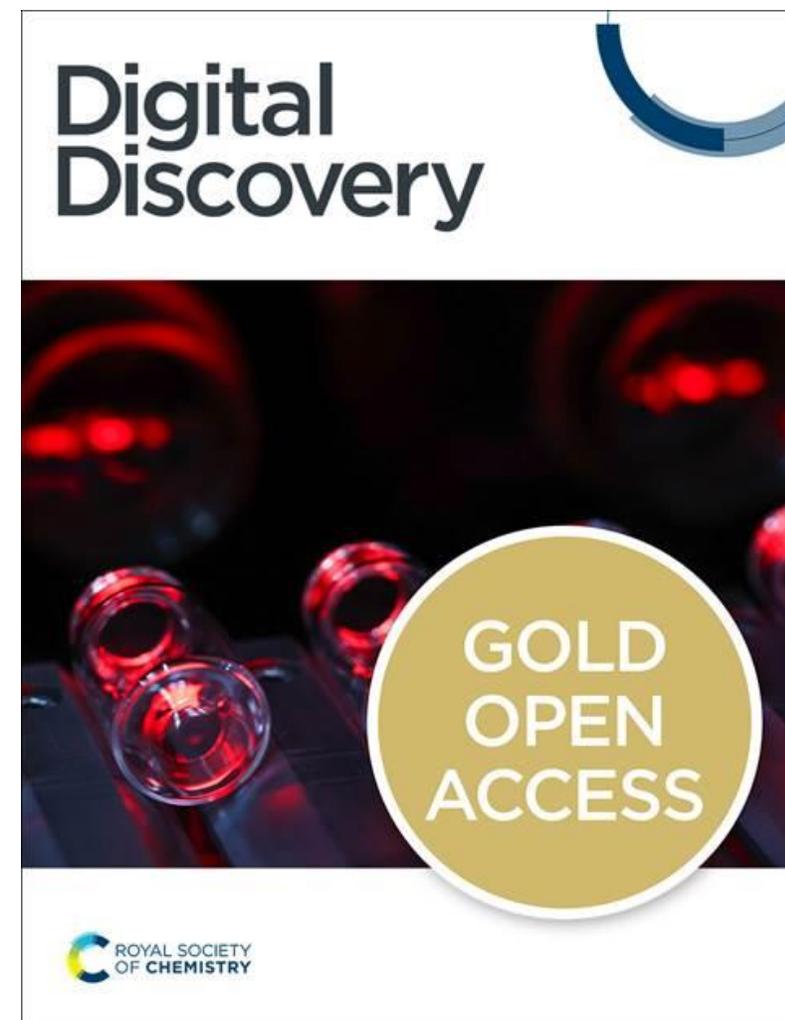


②研究者はどの材料を研究するのか推薦の中から選択



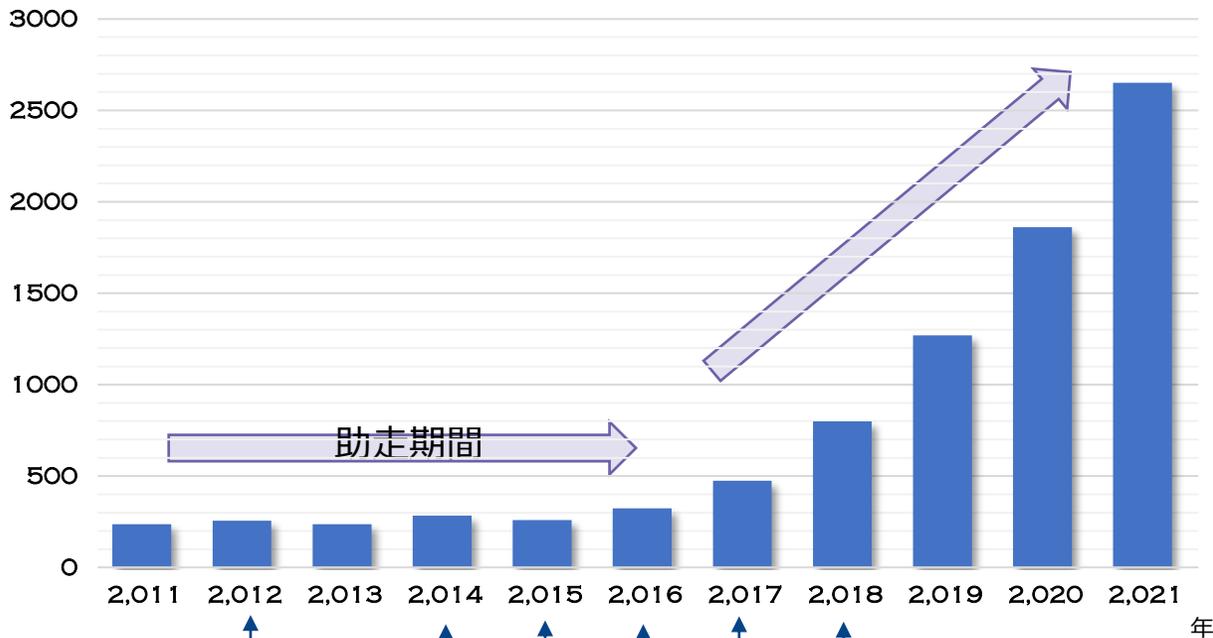
Digital Discoveryの発刊（2022年）

- 化学、材料科学、生化学、生物医学または生物物理学のための人工知能およびデータサイエンスの方法論
- 材料、化学、生物、および生物医学の高度なデータワークフロー
- 化学、生化学、生物医学、および材料科学のための新しい実験的自動化
- 以下のトピックを含む化学および他の科学のインターフェースに関する論文
 - ✓ 指示または加速された進化
 - ✓ DNAエンコードライブラリテクノロジーまたは新しい化学ライブラリテクノロジー
 - ✓ 暗号化学
 - ✓ ブロックチェーン対応の科学



AI・機械学習を用いた論文の動向

化学・材料分野



CNN (登場)

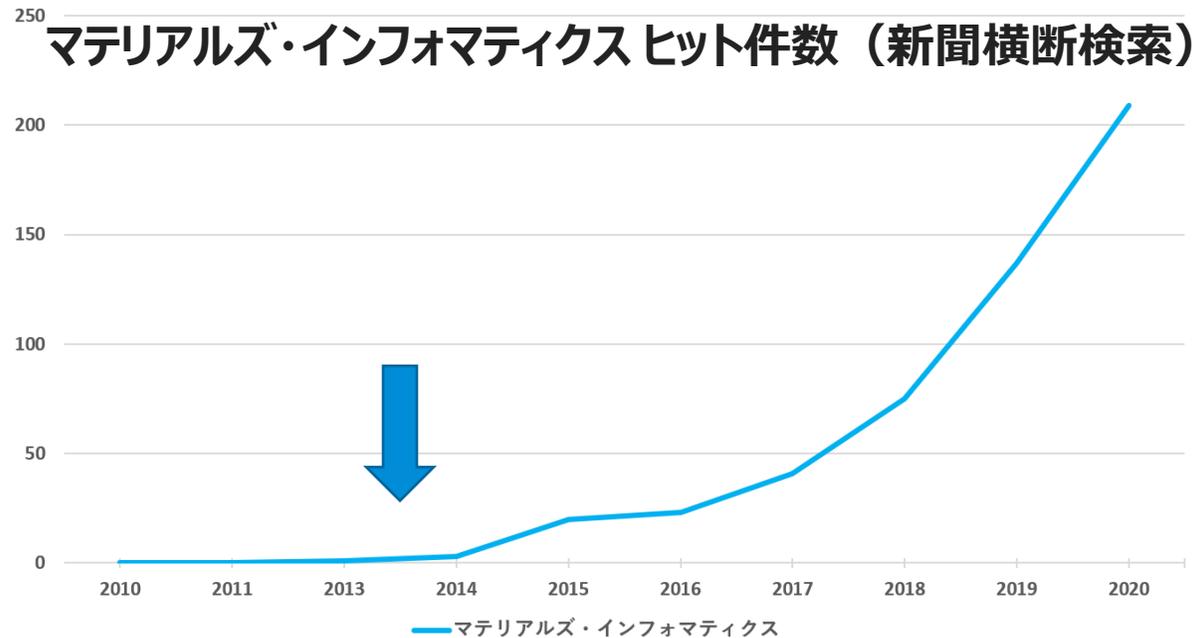
GAN

CNN (実用的)

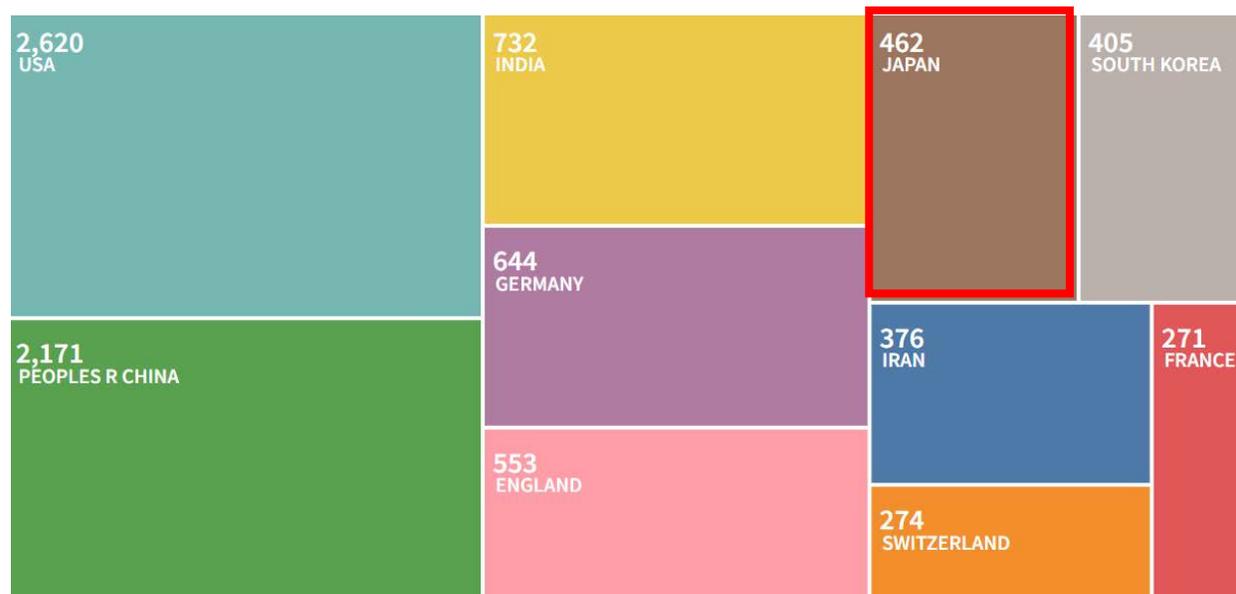
AlphaGo (強化学習)

VAE

Transformer
自動化・自律化の概念共有



出典：G-Searchデータベースサービス 2021年9月9日掲載

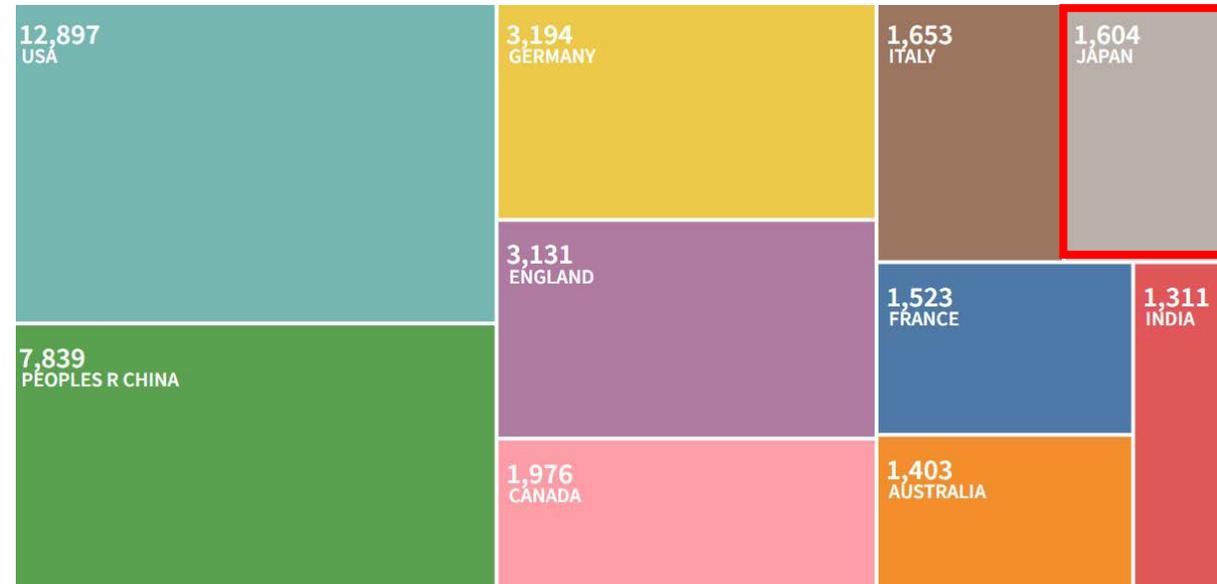
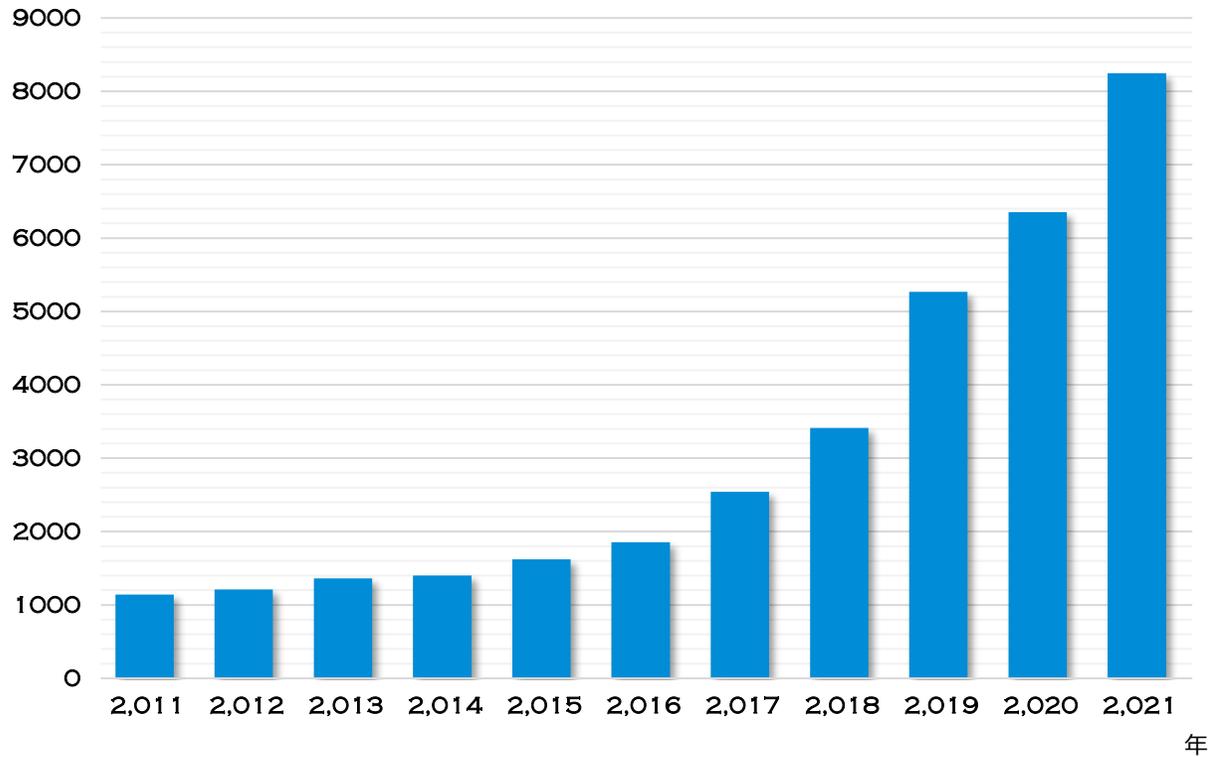


日本企業の取組み例

- 日立のLumadaで展開する「材料開発ソリューション」
- 三井化学は過去の有機材料の材料開発データを用いて、従来のMI技術と比べて高性能な新材料の開発に必要な実験の試行回数を4分の1に削減できることを確認
- 三菱ガス化学において、半導体材料において目標性能を満たす新素材探索の精度が約50%向上するほか、新素材探索に必要な実験時間が30～50%短縮すると確認
- NECは、予測の根拠を分かりやすいルール形で示せるAI技術「ルール発見型推論技術」を開発
- 製品不良要因分析における技術検証では、欠陥品に影響する要因を特定できるだけでなく、例えば「材料の温度が100度より高く、かつ設備の圧力が20hPaより高いとき、80%の確率で故障する」など、これまで専門家が想定していなかった要因が欠陥品発生に影響する可能性も見出すことができた。
- 長瀬産業は、IBMと共同開発した「TABRASA」というサービスを提供。
- 具体的な活用方法としては、まずAIに文献や実験データを読み込ませることで、「材料の化学構造と物性値とにどのような関連性があるか」を学習。ユーザーが求める材料の化学構造式を推測および提案。
- 横浜ゴムは、AIを活用したゴムの配合物性値予測システムを独自に開発し、タイヤ用ゴムの配合設計において実用を開始。
- この予測システムにより、膨大な仮想実験が可能となるため、開発のスピードアップやコスト削減、高性能な商品の開発に加え、経験の浅い技術者による配合設計が容易に
- トヨタ自動車は材料分析・データ解析クラウドサービス「WAVEBASE」を展開。
- 同サービスではスペクトルデータからAI技術などで特徴量を抽出し、主成分分析で性能に効く要素を特定。

【参考】 AI・機械学習を用いた論文の動向

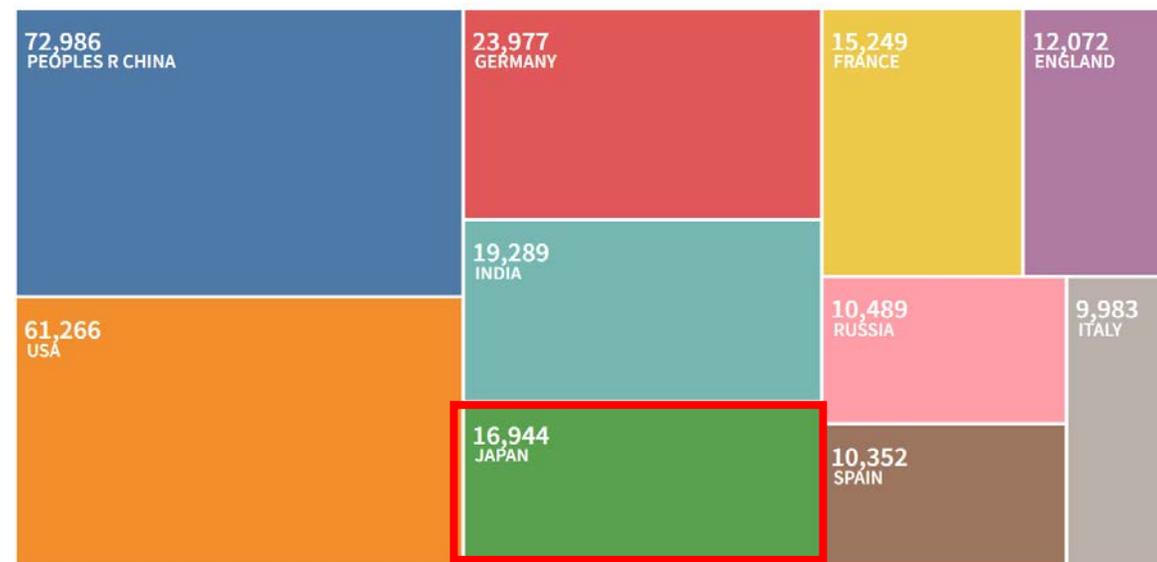
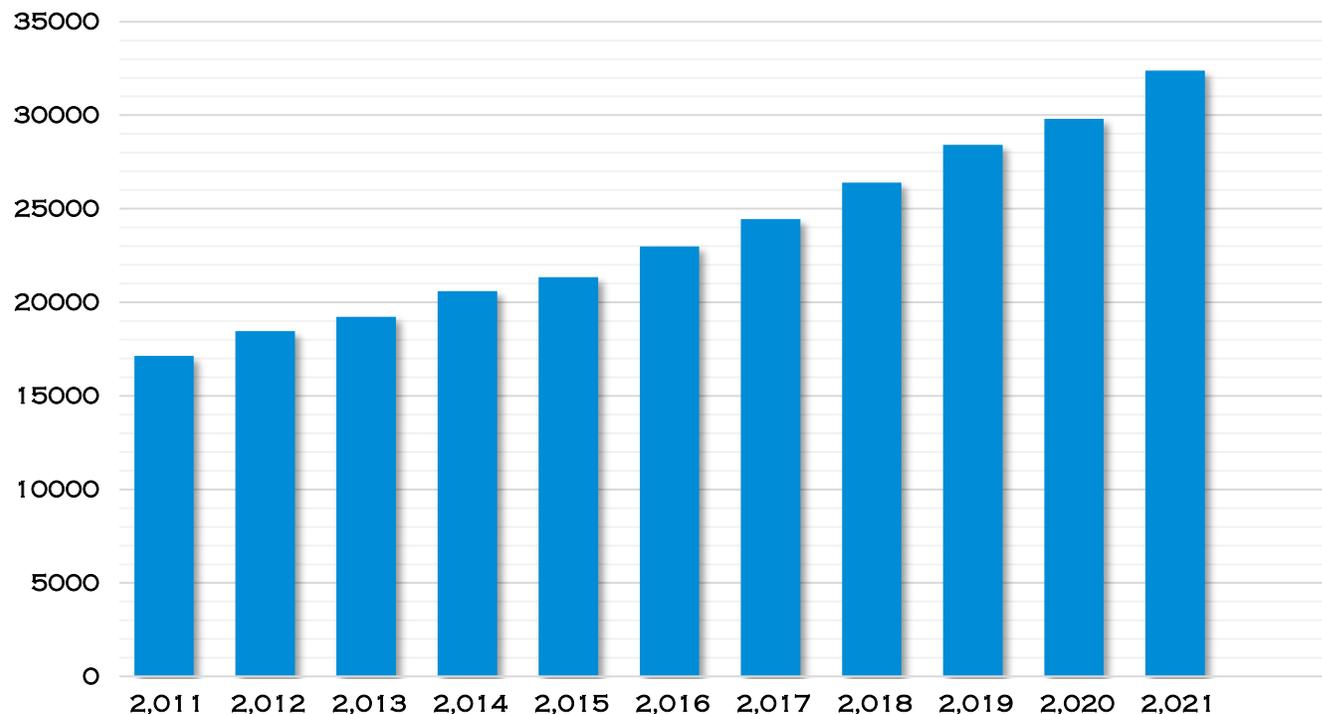
生命科学・医科学分野



Web of Scienceを元にCRDS作成

計算科学・計算化学を用いた論文の動向

化学・材料分野



年

Web of Scienceを元にCRDS作成

本日の話題提供の構成

1. AI・DXと科学技術イノベーションの動向
2. MIの発展の歴史
- 3. MIの研究動向（事例紹介）**
 - ① 予測／探索（計算シミュレーションの活用）
 - ② ディープラーニング（識別モデル、生成モデル、自然言語処理など）の活用
 - ③ 自動化・自律化（実験へのロボット、プログラム制御等の活用）
4. まとめと今後の展望

なぜマテリアルズ・インフォマティクス？

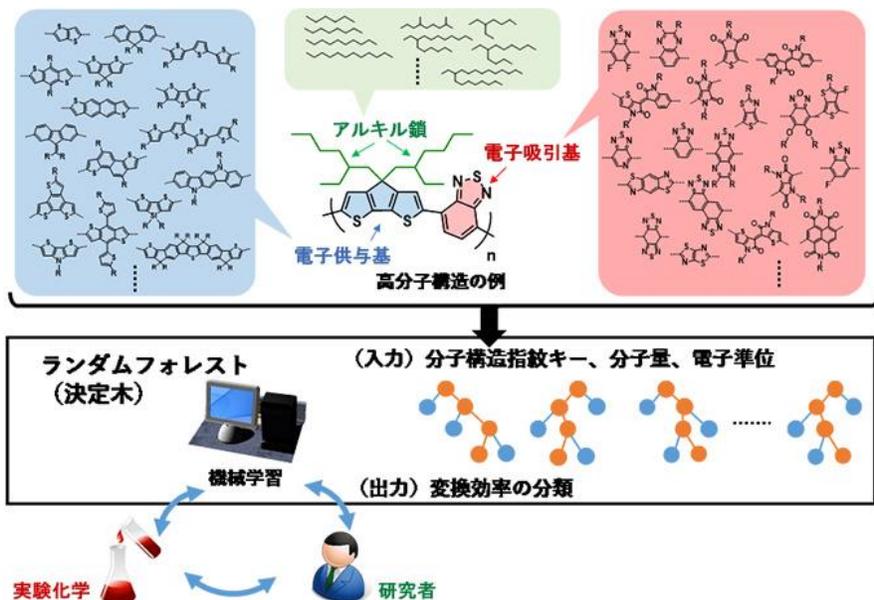
- 単純な元素組成、安定相の利用における技術的な限界
- 材料に対する**複数の機能、相反する特性の両立**などの高度な要求

機能材料	機能に対する要求	従来の材料例	現在開発中の材料例
蓄電池固体電解質	高速Liイオン伝導と広い電位窓の共存	LiPS	LiSnSiPS
蓄電池正極材料	高Li吸蔵・放出特性と安全性の両立	LiCoO	LiMnNiCoO
太陽電池	高い変換効率と長期信頼性の共存	Si、GaAs	CIGS、CH ₃ NH ₃ PbI ₃ (ペロブスカイト)
構造材料	軽量化と高強度、高強度・強靱性の両立	ハイツン、Al合金	CoCrFeMnNi (ハイエントロピー合金)
熱電変換材料	高電気伝導度と低熱伝導の両立	BiTe、PbTe	PbNaGeTe
磁石材料	高保磁力と高飽和磁化の両立	NdFeB (Dy添加)	NdLaCeFeB
ワイドギャップ半導体	高耐圧と高速動作 (高周波動作) の両立	SiC、GaN	α-Ga ₂ O ₃ (準安定相)
蛍光体	多様な発光波長と高輝度の共存	YAG	(Ca,Y)-α-SiAlON:Eu
触媒	高い触媒機能と耐熱性、低コストの共存	Pt、Rh	PdRu
水・ガス分離膜	高い物質選別性と高処理能力の両立	酢酸セルロース	金属有機構造体 (MOF)、ゼオライト
有機半導体	高移動度と塗布 (大面積化) の両立	ペンタセン (低分子)	PBTTT (高分子)

1. 予測／物質探索（有機材料）

高分子太陽電池、人工知能で性能予測（阪大）

- 1200個の高分子の化学構造と物性値のデータセットを集める
- 化学構造を1024個の指紋キー（数字の列）に変換し、加えて分子量や電子準位などの物性値を入力パラメータとし、光電変換効率のグループ（例えば4～5%のグループなど）を出力として与えるように機械学習を行う
- この分類器を使って新たな材料候補となる高分子の構造を抽出することに成功し、従来の計算化学では不可能であった溶解性を付与するアルキル鎖の選別も可能に

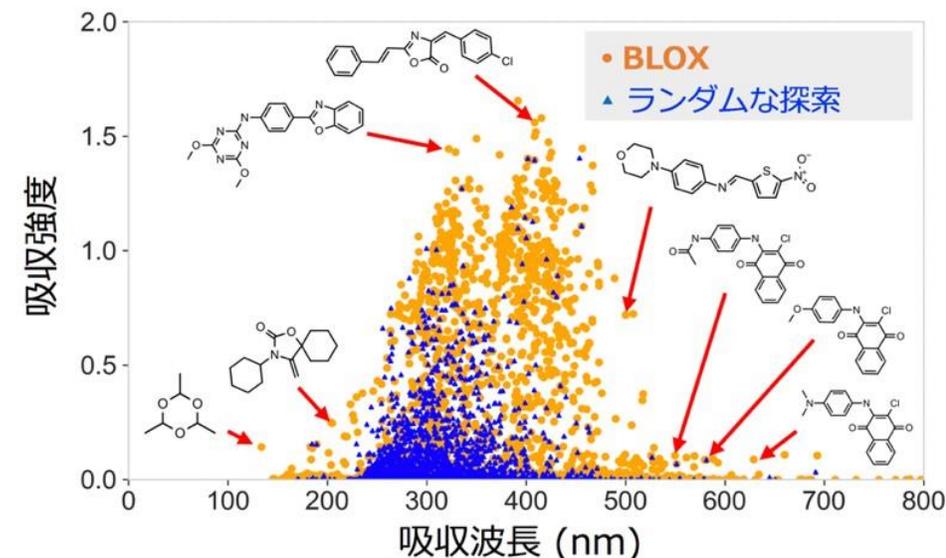


J. Phys. Chem. Lett. **2018**, 9, 10, 2639–2646

既存のDBやシミュレーションデータの利用

例外を発見するAIの開発～有機発光材料への応用～（NIMS等）

- 機械学習を組み合わせることで例外の度合いを数値化し、例外的な物質を効率的に発見するAIを開発し「BLOX」と名付けた。
- ZINCデータベースに含まれる10万個の分子と量子力学に基づいた分子シミュレーション技術と組み合わせた結果、例外的な光吸収特性を持つ有機化合物候補を多数発見した。そのうちの8個を実際の化合物で評価したところ、250ナノメートル以下や450nm以上の波長の光を強く吸収する例外的な特性を持つことを確認できた。



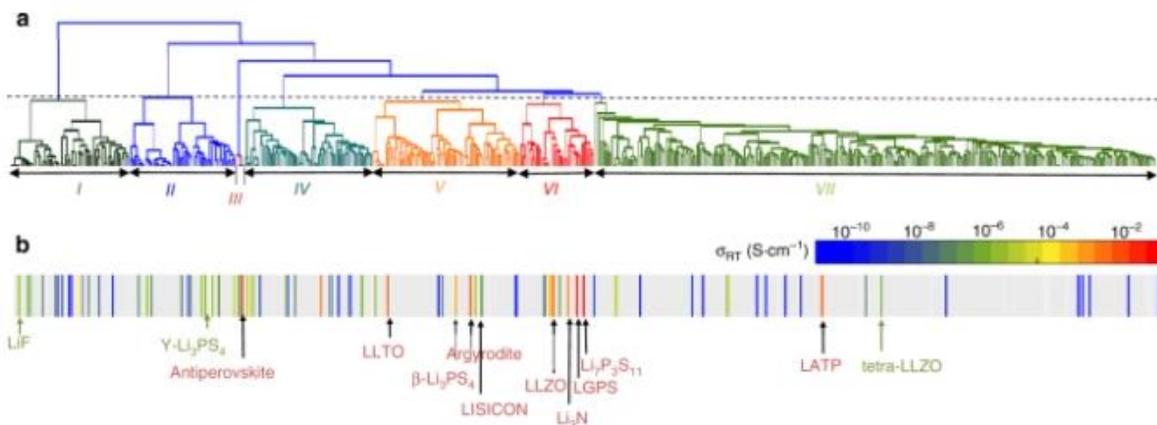
Chem. Sci., **2020**, 11, 5959–5968

予測／物質探索（無機材料）

既存のDBやシミュレーションデータの利用

固体リチウムイオン伝導体の教師なし発見 (米トヨタ研究所)

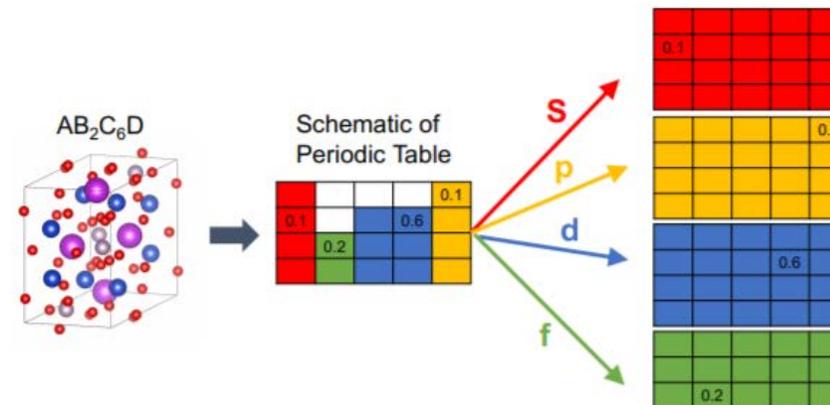
- 教師なし学習を適用して、無機結晶構造データベースからすべての既知のLi含有化合物をスクリーニングし、階層クラスタリングを経て、ab initio 分子動力学計算で高スループットスクリーニング
- 高速Li伝導性材料と貧弱なLi伝導性材料を区別することに成功し、室温導電率が 10^{-4} Scm^{-1} を超える固体Liイオン伝導体として16の新しい化合物を予測



Nature Communications volume 10, Article number: 5260 (2019)

超伝導体の予測技術（NICT、東大）

- 化合物を周期表情報（最外殻の価電子、軌道）のベクトルとして学習させることによる予測手法
- 物質の超伝導化を62%の精度で予測
- データベースに未登録の新素材超伝導体 CaBi_2 を発見



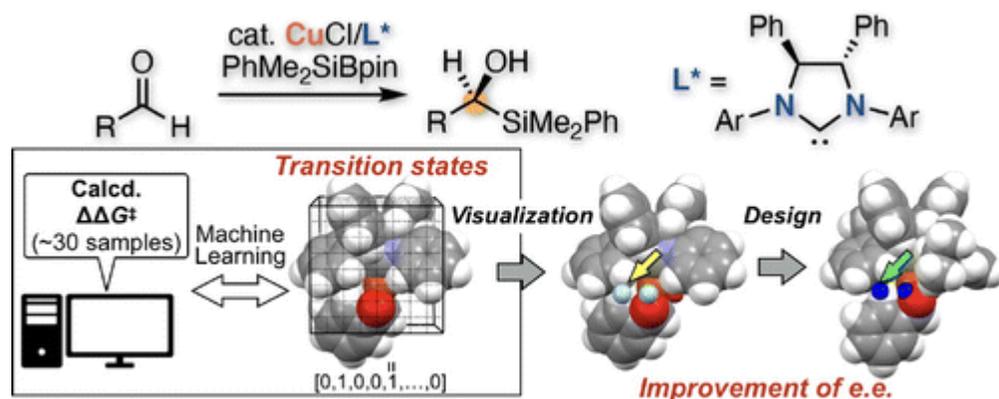
Main result	Training data	Test data
Prediction of superconductors from materials reported by Hosono et al. [2]	SuperCon and COD before 2010	Materials reported by Hosono et al.
Superconductor CaBi_2 found in candidate material list	SuperCon and COD in 2018	COD in 2018
FeSCs found from training data before discovery year	SuperCon and COD before 2008	FeSCs in Supercon in 2018

TABLE II: Summary of main results.

Phys. Rev. B 103, 014509 – Published 12 January 2021

計算機上で収集したデータの機械学習による不斉触媒設計 (金沢大学、理研)

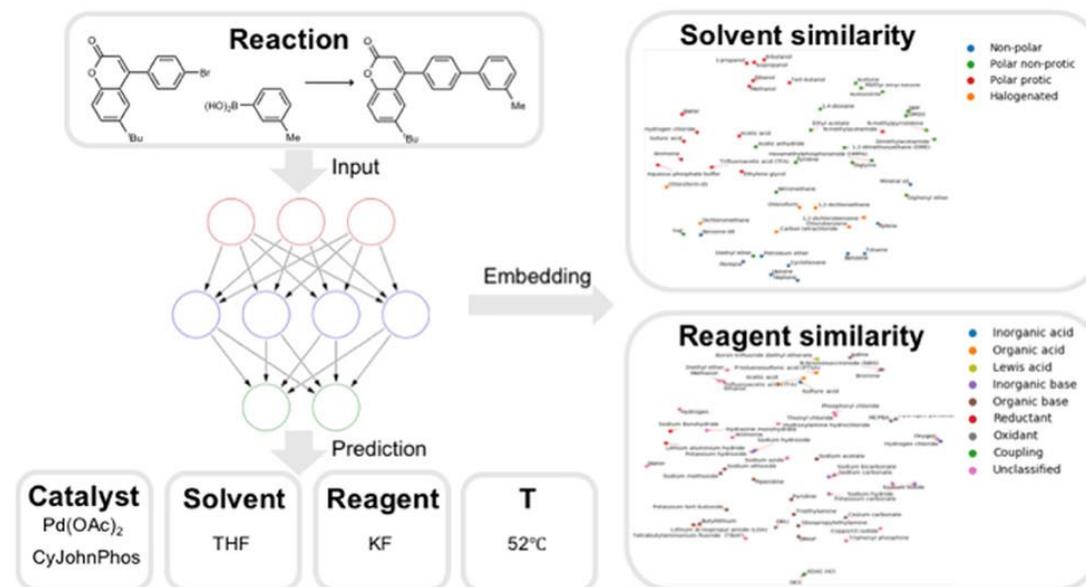
- モデルとした不斉触媒反応において、遷移状態計算により計算機上で集めたわずか30個のサンプルから、エナンチオ選択性が向上する「データ駆動型不斉触媒設計」に成功
- 選択性支配因子を抽出・可視化し、可視化した重要構造情報を基に研究者が設計した分子を用いて、再び遷移状態計算を行い、訓練データを追加する。最大73% eeを示す触媒を含む訓練データ18サンプルから始めて、サイクルを繰り返すことで、96% eeを示す不斉触媒および不斉配位子の設計に成功



Bull. Chem. Soc. Jpn. **2022**, 95, 271–277

反応条件を最適化を支援するDLシステム (MIT)

- 化学コンテキスト (触媒、溶媒、試薬)、および特定の有機反応に最も適した温度を予測するためのニューラルネットワークモデルを開発。
- Reaxysからの約1,000万の例でトレーニングされたモデルは、記録された触媒、溶媒、および試薬との密接な一致が69.6%の確率で上位10の予測内で見つかる条件を提案。



ACS Cent. Sci. **2018**, 4, 11, 1465–1476

- MI（化学・材料分野）が進んだ一つの理由として、第一原理（DFT）計算など計算機シミュレーションによって、構造と物性のデータセットを作成し、学習やテストに用いるという手法の存在が挙げられる。
- 機械学習技術としては、ベイズ推定やランダムフォレストなどを活用
- こういった手法はバイオ、生命科学ではほとんど見られない物質・材料固有の方法といえる。
- 第一原理計算はコストが高い。それを回避する手段としてディープラーニングを用いてDFTの計算コストを下げる研究も行われている。

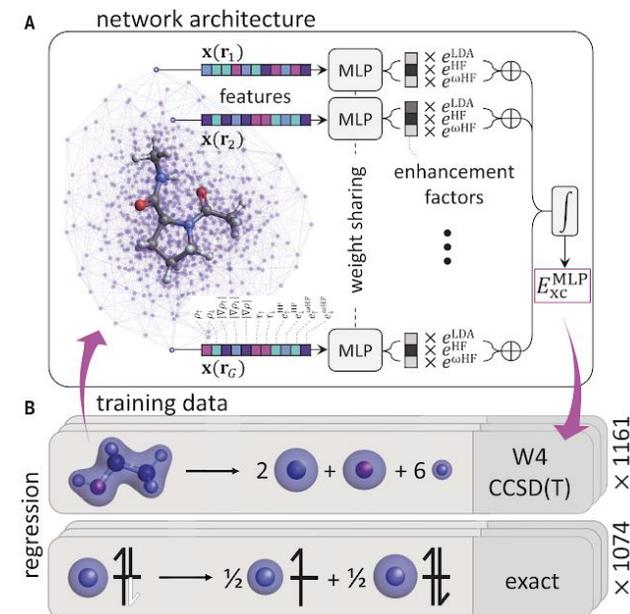
深層学習を駆使した汎用原子レベルシミュレーターを展開（プリファード・コンピューショナル・ケミストリー）

- 分子の物性を原子レベルでシミュレートする材料探索クラウドサービス「Matlantis」の提供を2021年に開始。
- これまで数時間～数日かかっていたシミュレーションを数秒単位で実施できる。
- 現在は55の元素をサポートしており、リチウムイオン電池の材料候補である LiFeSO_4F でのリチウム拡散、金属有機構造体の分子吸着、Cu-Au合金の秩序-無秩序転移、フィッシャー・トロプシュ触媒の探索で有効な性能を示した。

arXiv:2106.14583

有機分子内の電子の分布を予測することで分子の特性を示唆する機械学習モデルを開発（DeepMind）

- 最先端の汎関数と比較して改善された結果をもたらす新しいML交換相関の可能性を紹介
- 新しい汎関数を開発するときに、トレーニングセットに分数電荷とスピンを含めることの重要性を示す。分数電荷システムは自己相互作用エラーを減らすのに役立ち、ぶ分数スピンシステムは静的相関に関する情報を提供。
- この種の汎関数は、重要な多参照特性を持つ遷移金属含有システムでうまく機能する可能性がある。



Kirkpatrick, J. et al. Science 374, 1385–1389 (2021).

2. ディープラーニング ① 識別モデル

CNN(Convolutional Neural Network)

画像解析におけるAI：小さい画像に分割して特徴を抽出する

- **2012年**、「畳み込みニューラルネットワーク（CNN）」が画像認識で従来手法と比較して圧倒的精度を発揮（ILSVRC）。
- **2015年**、人間による画像認識のエラー率を下回る（ILSVRC）。

Classification (分類)

「その画像に何が写っているのか」

Detection (検出)

「その画像のどこに何があるのか」

Segmentation (セグメンテーション)

「個々のピクセルが何を意味しているのか（領域の識別）」

Semantic Segmentation

派生して、U-Net等セマンティック・セグメンテーションの手法が次々に登場

自動運転における物体認識



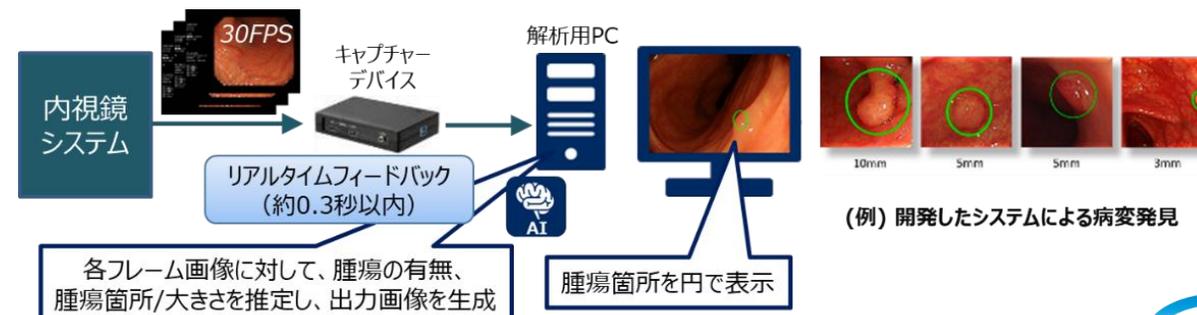
工場（生産ライン）等における材料や製品の欠陥抽出



AI

出典：オムロン社

内視鏡画像から腫瘍の存在診断、治療効果の判定

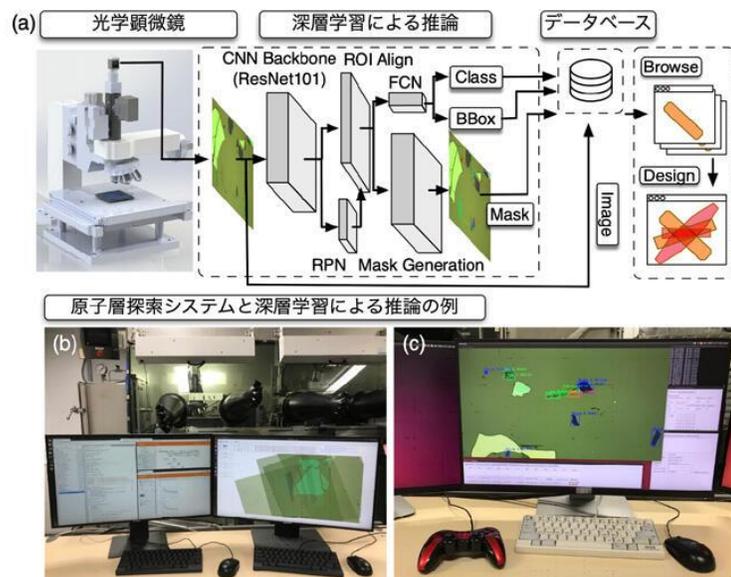


識別モデル（画像認識）

任意の画像認識

目的のシート状の原子層を、全自動で探索する 光学顕微鏡システム（東大）

- 東京大学の研究グループでは、シリコン基板の上に剥離された二次元結晶を光学顕微鏡により自動探索し、これらを任意の順番で自在積層するロボットシステムを2018年に開発。
- 加えて、**CNN**を搭載した光学顕微鏡により、熟練の研究者が長時間かけて行っていた認識作業を、ほぼリアルタイムで、90%以上の確率で再現すること、希少な単原子層を使った複合原子層の作製を可能にした。

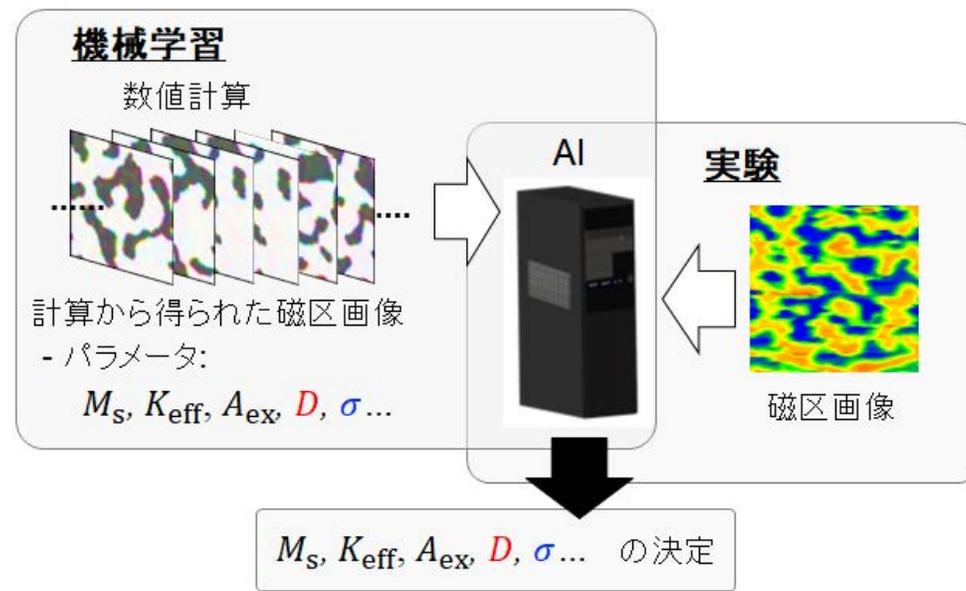


npj 2D Materials and Applications volume 4, Article number: 3 (2020)

画像と物性の相関

磁石の磁気パラメータの推定（東大など）

- 数値計算（コンピュータシミュレーション）によって得た磁区画像と、その時に利用した磁気パラメータを使って、**CNN**で学習。このAIに対し、実験的に計測された磁区画像を読み込ませ、磁気パラメータを推定
- 微細加工や電気測定なしに、評価が難しい複数の磁気パラメータを磁区画像から取得できることを実証



npj Computational Materials volume 7, Article number: 20 (2021)

ディープラーニング ②生成モデル

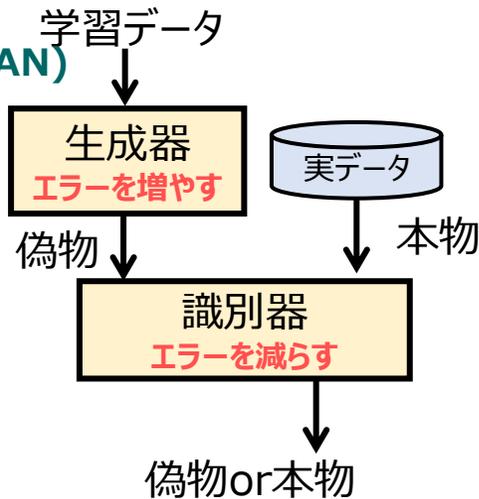
データの属するクラスを識別(例:CNN)

データの生成過程をモデル化

- 深層学習は、識別モデルだけでなく、ベイズの定理を用いた生成モデルにおいても著しく進展
- 特に「敵対的生成ネットワーク(GAN)」や「変分自己符号化器(VAE)」が注目され、生成系の応用が高品質化・拡大

敵対的生成ネットワーク (Generative Adversarial Networks: GAN)

- 本物と偽物を見分ける識別器と、それを騙す偽物を作る生成器を競わせる
- 本物と見分けのつかない偽物が作れる



有名人風の顔画像生成



<https://youtu.be/XOxxPcy5Gr4>

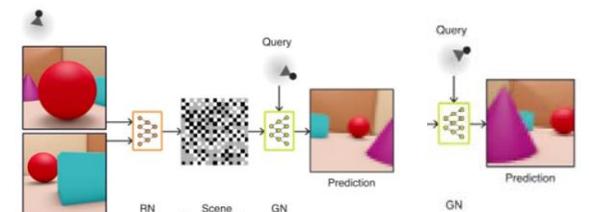
- 2014年、画像生成のためのアルゴリズム「敵対的生成ネットワーク (GAN)」発表。
- 2017年、ディープラーニングを用いた分子生成「VAE (Variational Auto-Encoder)」発表。

動作を別人にコピー



<https://youtu.be/PCBTZh41Ris>

見えない面も推測して3Dモデル生成 (Generative Query Network: GQN)



<https://youtu.be/RBJFngN33Qo>

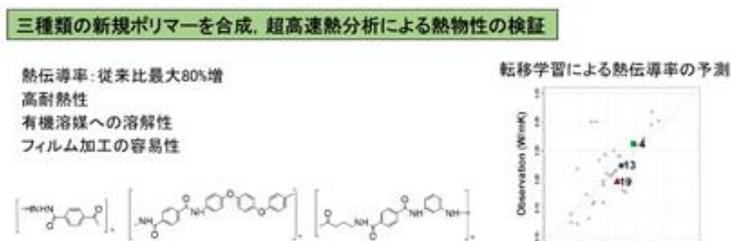
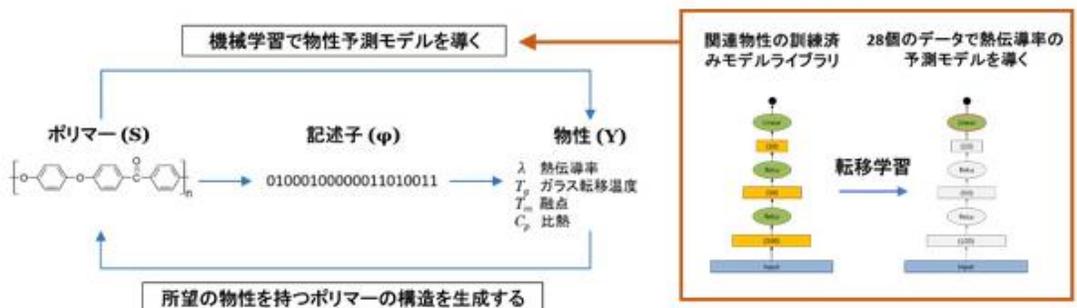
ACS Cent. Sci. 2018, 4, 2, 268–276

生成モデル（有機材料）

仮想（疑似）化合物を生成

高分子の熱伝導性の大幅な向上（NIMS、東工大）

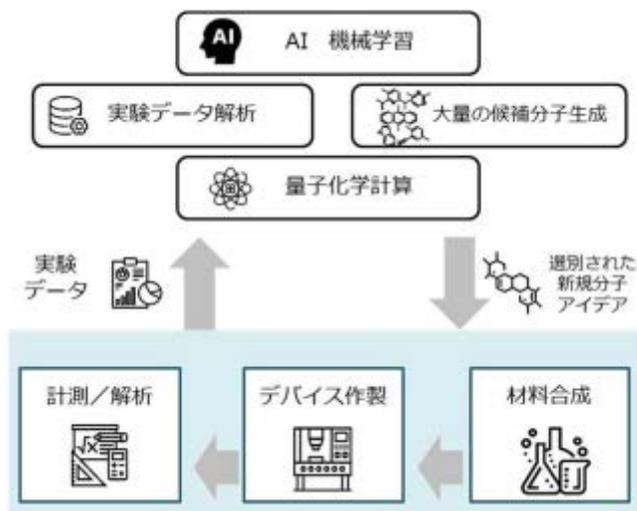
- ビッグデータの入手が可能なガラス転移温度等で機械学習のモデルを訓練し、モデルが獲得した「記憶」と少数の熱伝導率のデータを組み合わせることで、熱伝導率を高精度に予測できるモデルを導くことに成功（転移学習）
- このモデルを用いて高い熱伝導率をターゲットに1,000種類の仮想ライブラリを作製。その中から三種類の芳香族ポリアミドを合成し、熱伝導率0.41 W/mKに達する高分子を見出した（従来の高分子に比べて約80%の熱伝導率の向上）。



npj Computational Materials volume 5, Article number: 66 (2019)

次世代有機ELの効率化・長寿命化（Kyulux：九州大学発ベンチャー）

- 技術者が提案した材料を基にAIが材料設計ルールに従って候補材料を1万個以上生成。そのうち数百個の材料を量子化学計算し、結果を学習させて残りの材料の物性を予測する。予測結果から有望な材料を選定し、実際に合成して性能を確認する。さらに実験結果も機械学習して物性予測精度を高めている。
- 開発スピードを10倍以上に速められる。
- 赤色では目標の4倍の寿命を達成し、長寿命化が最も難しいといわれる青色においても、開発期間中に目標を達成し、100倍以上の寿命を実現



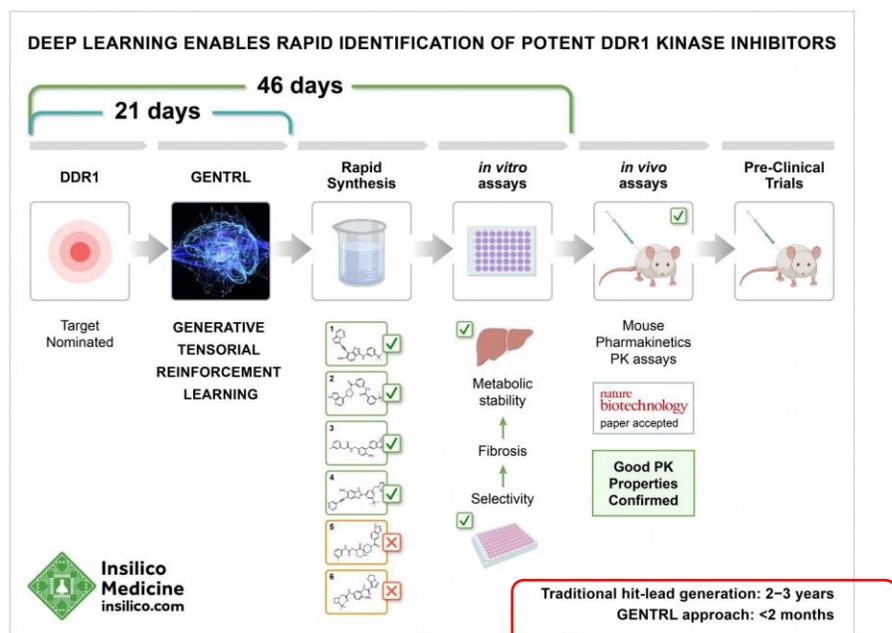
2021年11月プレス発表

生成モデルの活用は高コストな第一原理計算を回避するという意味も。

創薬への応用（生成モデルや強化学習などの活用）

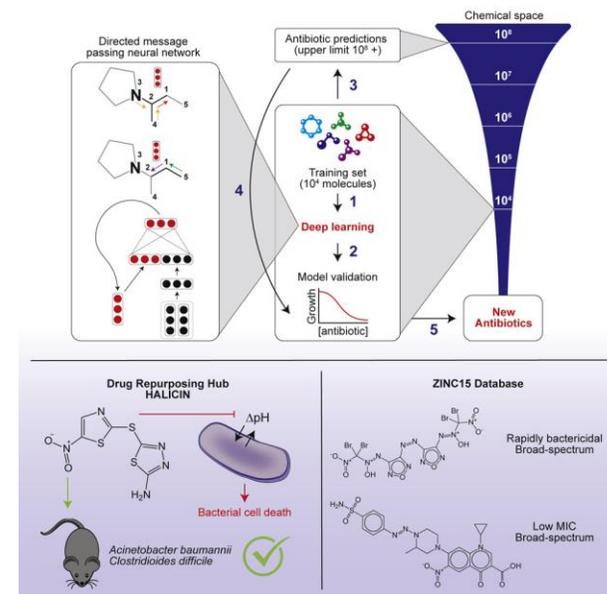
仮想（疑似）化合物を生成

- AI創薬ベンチャーのInsilico Medicine社（香港）では、新規の低分子設計のために、生成テンソル強化学習（GENTRL）を開発。合成の実現可能性、新規性、および生物活性を最適化することができる。
- これを使用して、線維化やその他の疾患に関するキナーゼターゲットであるディスコイジンドメイン受容体1（DDR1）の強力な阻害剤を20日間程で発見。



Nature Biotechnology volume 37, pages1038-1040(2019)

- MITでは、化合物のグラフ表現と抗菌活性を紐付けるニューラルネットワーク（Message Passing Neural Network）を用いて、大腸菌を殺すのに効果的であることが知られている約2500の化学構造で最初に訓練され、約6000の化合物ライブラリを用いてスクリーニングし、他の既知の抗生物質とは異なる構造を持つ潜在的なリード化合物を特定した。
- 「Halicin」と名付けられた抗生物質はこれまでのものとはまったく異なる作用メカニズムを持っており、臨床試験では数々の薬剤耐性菌を殺すことができた。



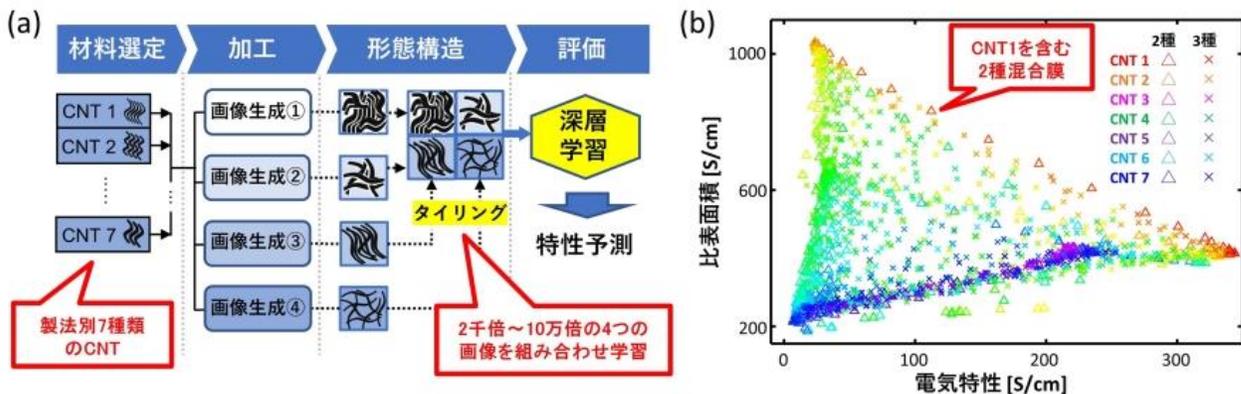
Cell Volume 180, Issue 4, 20 February 2020, Pages 688-702.e13

生成モデル（無機材料）

画像データ拡張（学習データとして利用）

AIが生成したCNT材料の構造画像から物性を予測（産総研等）

- 複雑な形態のCNT膜を、マクロな集合体からミクロな束状構造（バンドル構造）までいくつかの“階層”を持つ構造の走査型電子顕微鏡（SEM）画像を取得
- GANを用いて構造画像の学習および生成を行い、実際の実験と比べて98.8%もの時間を短縮し、材料物性の高精度な予測を実現



Communications Materials volume 2, Article number: 88 (2021)

仮想結晶構造を生成

GANやVAEを無機材料へ適用し、新たな結晶構造を探索する生成モデル開発についての論文も多数

(一例)

CrystalGAN: Learning to Discover Crystallographic Structures with Generative Adversarial Networks
arXiv:1810.11203

CCDCGAN
npj Computational Materials volume 7, Article number: 66 (2021)

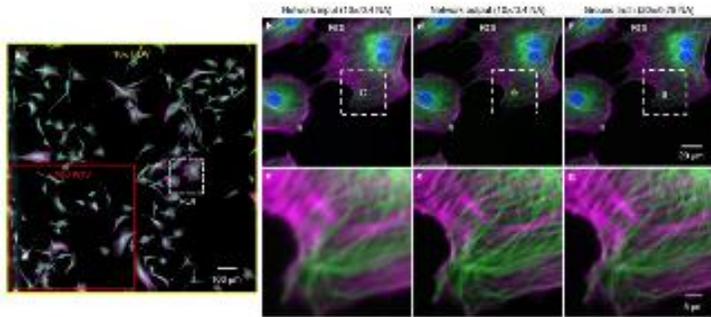
CubicGAN
Materials Science, 10.1007/s10853-021-06540-7, (2021).

Matter Volume 5, Issue 1, 5 January 2022, Pages 314-335

生成モデル 超解像化、3次元化、データ拡張等

GANによる光学顕微鏡画像の超解像化

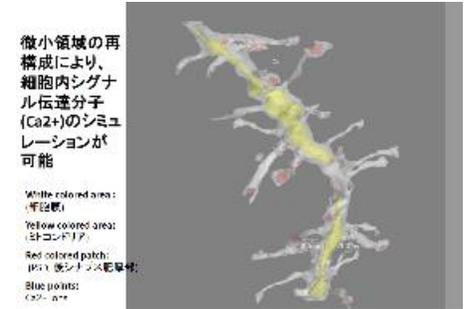
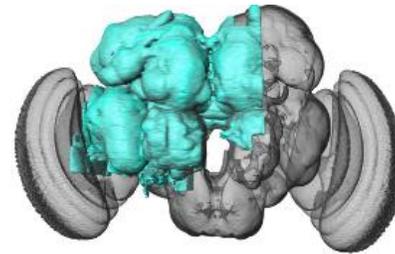
- 光学顕微鏡で高い分解能でサイズの大きいものを見ようとすると分解能と撮像に要する時間のトレードオフ関係がある。
- これまで1枚の撮像に分単位を要したが、秒単位を可能にし、ダイナミクスが30ナノメートル以下の分解能で見える。



Nature Methods volume 16, pages103–110(2019)

電子顕微鏡 (SEM連続断面観察法) を用いた細胞や臓器の3次元再構成

ショウジョウバエの脳にある2万5千個のニューロンとその間の2,000万のつながりを示す配線図が示された。



A Connectome of the Adult Drosophila Central Brain
bioRxiv January 21, 2020.

出典：石井信 (京都大学)

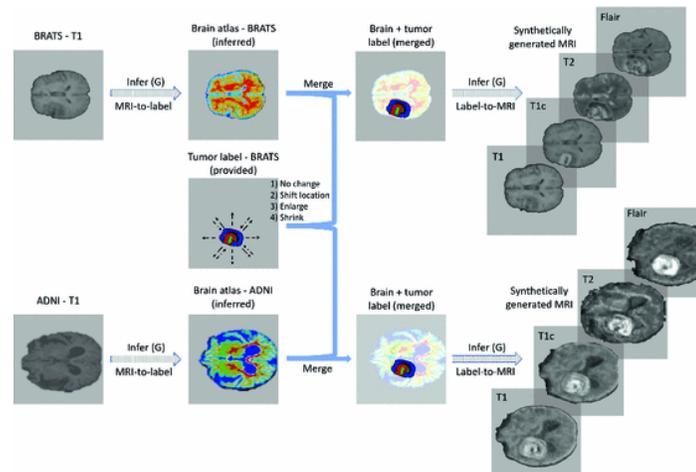
Scientific Reports volume 9,
Article number: 19413 (2019)

GANとCNNの両者を用いている。

GANによる医用画像 (学習用データ) の生成

- GANで生成された合成画像が学習用データに使われ、機械学習のアウトプット (識別器としての効率) が挙がる事例が現実に出始めている。

arXiv:1807.10225v2 [cs.CV] 13 Sep 2018



RNN、GNN、強化学習等

反応物から生成物を予測：化学反応予測 (IBM)

- 39万5496件の化学反応を含んだデータセットから未知の化学反応を予測
- 予測の正解率は、80%。

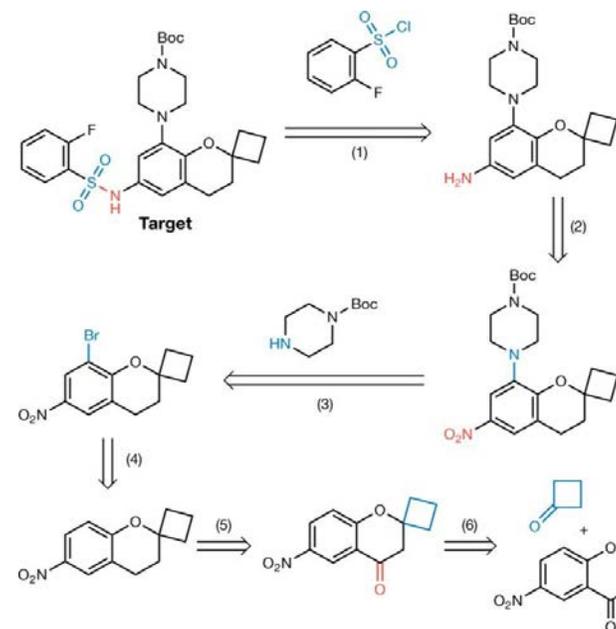
Chem. Sci., **2018**, 9, 6091-6098

反応予測アプリケーション IBM RXN for chemistry

特許と書籍から集めた
200万の反応を利用

生成物から反応物を予測する試み：逆合成予測 (ヴェストファーレン・ヴィルヘルム大学、BenevolentAI)

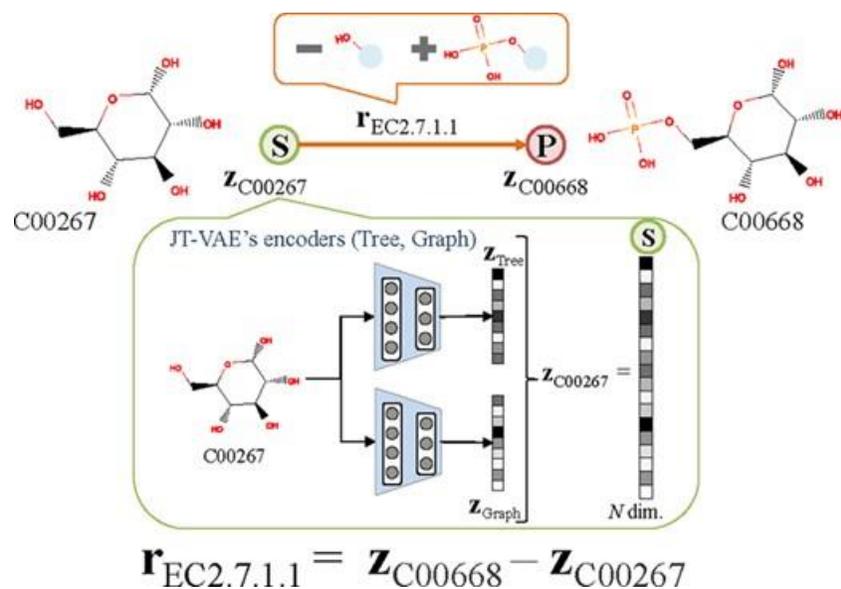
- AIによって提示された候補分子群をどうやって作るかが律速に。
- これまでに知られている一段階有機化学反応のほとんど全て約1240万をAIに学習させた。
- 2015年に報告された薬物候補合成の中間体の6段階のルートの例は、アルゴリズムによって5.4秒で発見された。



Nature 555, 604-610 (2018)

細胞内の代謝経路を新たに創り出す技術を開発 (日立)

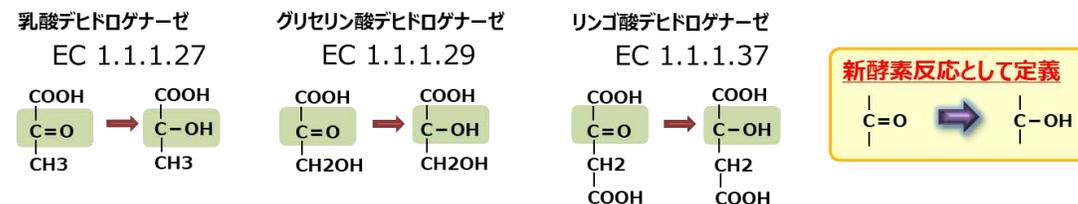
- 酵素反応前後の化合物の構造変化を予測するAI技術
- データに含まれない生化学反応、すなわち生物が実現し得る潜在的な代謝経路の存在を予測できる



BioProV : 人工代謝反応の設計ツール (理研)

- 目的の物質と反応経路の長さを入力すると合成のための候補代謝経路を提示

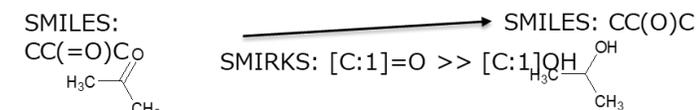
1. 約3,000の既知の酵素反応(EC number)を反応パターンに注目して400種類の反応に再分類化



2. 化合物と反応の定義・記述

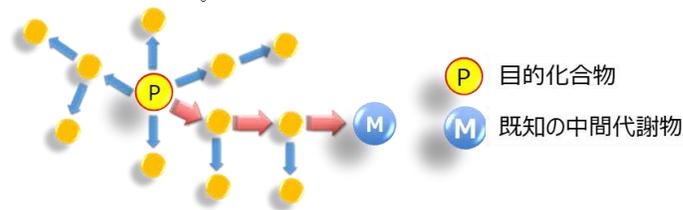
SMILES: 代謝化合物の記述
SMIRKS: 反応パターンの定義

化合物と反応の例



3. シミュレーションアルゴリズム

1. 目的化合物から反応パターンに沿って前駆体をランダムに生成する。
2. 既存の代謝化合物に到達するまで網羅的にシミュレーションを行う。



③ 自然言語処理

チャットボット、会話認識（文字起こし）、翻訳等に活用

- 2018年10月にGoogleの論文で発表された自然言語処理モデル「BERT」
- Bidirectional Encoder Representations from Transformers、「Transformerによる双方向のエンコード表現」と訳される。
- 文脈理解ができるように。

- イーロン・マスクが共同会長を務める、米国の非営利研究団体「OpenAI」が2020年に開発した言語モデル「GPT-3」
- 次に来る単語の予測を学習する。予測した単語を組み合わせて文章を出力する。学習は、大量の文章を用い、正解は与えられず、文脈にある単語との関係性について分析してパターンを学習していく。
- 自然言語と画像処理を組み合わせる新たな画像を生成する「DALL・E（ダリー）」が公表された（右図）。言葉を入力するだけで、それっぽいイラストや写真を自動生成できる。

- Microsoftは2021年10月、新たな**自然言語生成モデル**「Megatron-Turing Natural Language Generation（MT-NLG）」を紹介。
- 同モデルのパラメーター数は、既存の最多パラメーター数を持つ言語モデル「GPT-3」の約3倍となる約5300億個にもなり、補完や予測、読解、常識推論、自然言語推論、語義の曖昧性解消といったタスクの精度を飛躍的に高める。

テキストプロンプト
犬の散歩をするチュチュに大根の赤ちゃんのイラスト

A1で生成された画像



プロンプトを編集するか、他の画像を表示する➡

テキストプロンプト
アボカドの形をしたアームチェア[...]

A1で生成された画像



プロンプトを編集するか、他の画像を表示する➡

テキストプロンプト
「openai」という言葉が書かれた店頭[...]

A1で生成された画像



プロンプトを編集するか、他の画像を表示する➡

テキストと画像のプロンプト
上部の猫は下部のスケッチとまったく同じです

A1で生成された画像



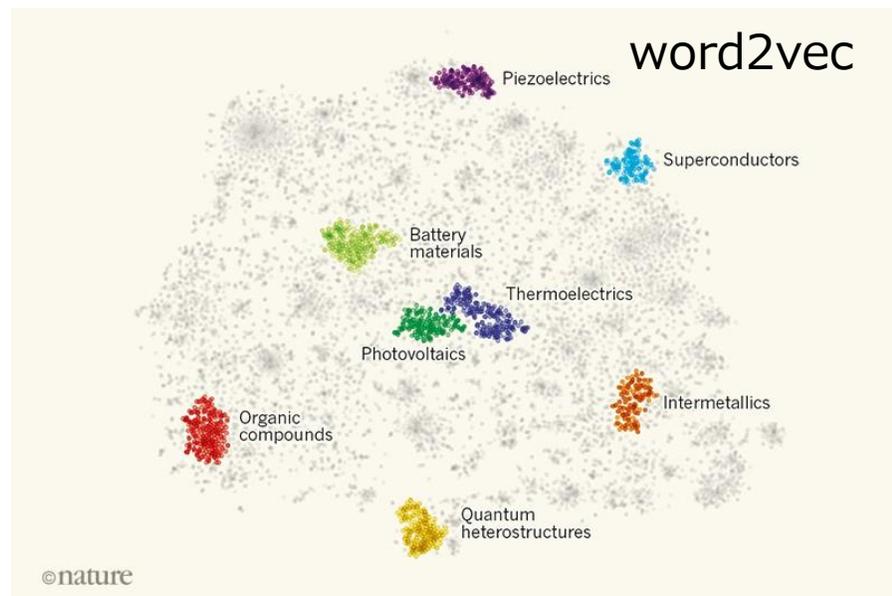
プロンプトを編集するか、他の画像を表示する➡

出典：OpenAI「DALL・E」

自然言語処理

教師なし学習により、材料科学の文献から暗黙知を取得（ローレンスバークレー）

- 330万の論文の要約のテキストを分析し、材料の名称を含む単語間の関係を特定。
- 個々の化合物を説明するために使用される単語間の意味関係に基づいて、データをクラスタ化。
- クラスタ（色付き）は、超伝導体、電池材料、有機化合物などの特定の種類の材料に対応
- 熱電特性があると報告されていないが、「熱電」という単語と高い意味関係を持っている材料を見つけた。



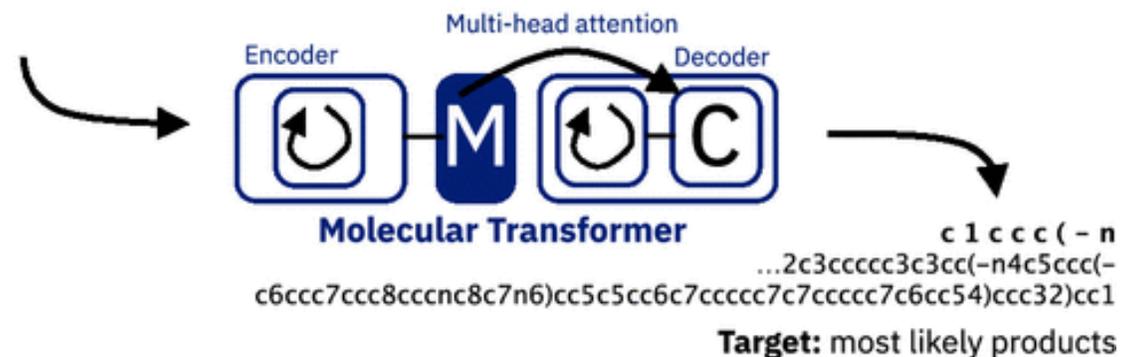
Nature. 2019 Jul;571(7763):95-98.

化学反応予測（IBM）

- 反応予測を、反応物、試薬、および生成物の簡略化された分子入カライントリシステム（SMILES）文字列（テキストベースの表現）間の相関。
- Transformerニューラルネットワークが、人間によるラベル付けなしで、生成物と反応物の間の原子マッピング情報を学習。
- 重要な原子マッピングを伴う非常に不均衡で化学的に複雑な反応に対しても、精度と速度の点で優れたパフォーマンスを示す。

Input: reactants-reagents (atom-wise tokenization)

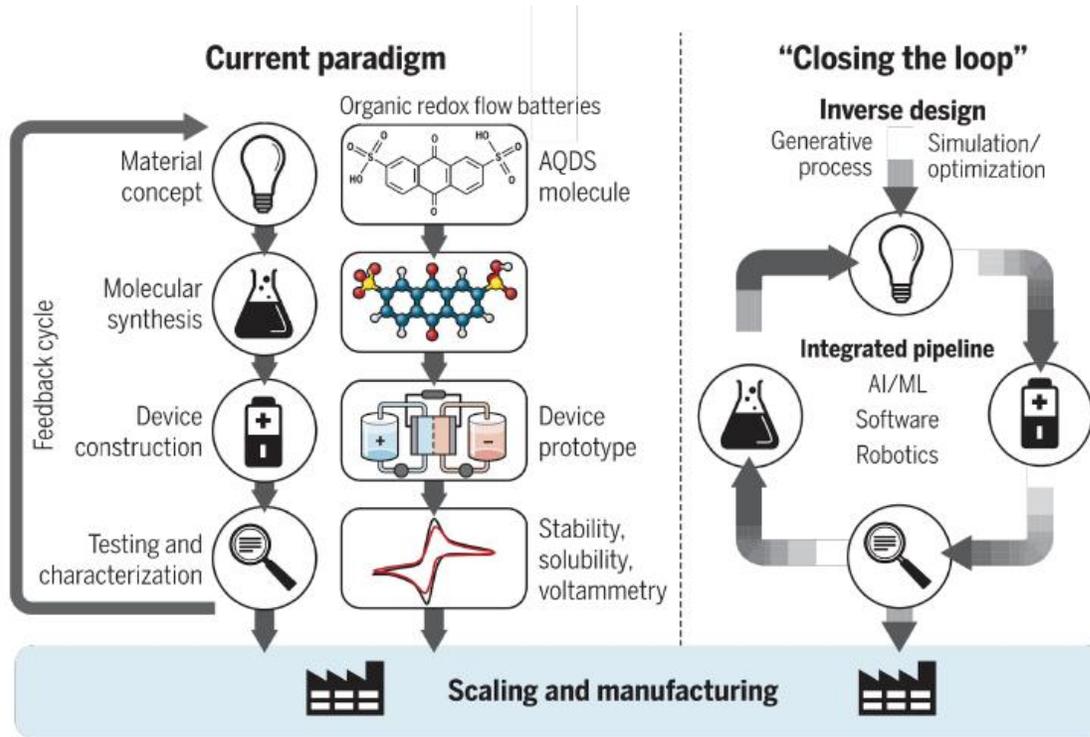
```
Brc1ccc2...c(c1)c1cc3c4ccccc4c4ccccc4c3cc1n2-c1ccc2c(c1)c1ccccc1n2-c1ccccc1.CCO.  
Cc1ccccc1.OB(O)c1ccc2ccc3cccn3c2n1.c1ccc([PH](c2ccccc2)(c2ccccc2)[Pd]([PH](c2ccccc2)  
(c2ccccc2)c2ccccc2)([PH](c2ccccc2)(c2ccccc2)c2ccccc2)[PH](c2ccccc2)(c2ccccc2)c2ccccc2)cc1
```



ACS Central Science 2019 5 (9), 1572-1583

3. (発見の) 自動化・自律化

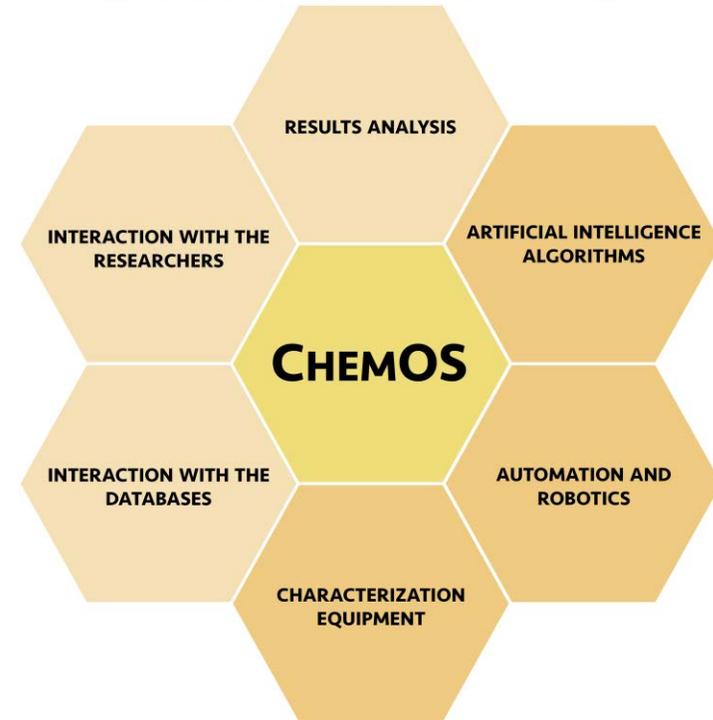
- 自律化に向かう方法論として、①「クローズドループ」の系を構築すること、②AI・機械学習、制御ソフトウェア（プログラム）、ロボティクスを用いてパイプラインを統合すること、を提案している。



Alán Aspuru-Guzik ブリティッシュコロンビア大学

SCIENCE • 27 Jul 2018 • Vol 361, Issue 6400 • pp. 360-365

適切なソフトウェアソリューション

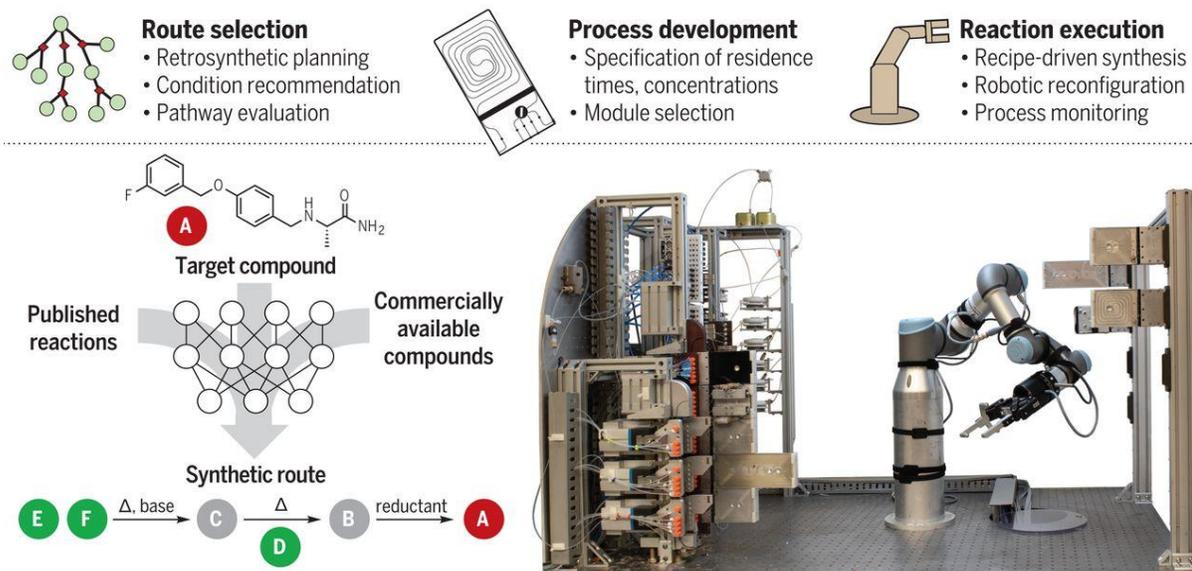


PLoS ONE 15 (4) : e0229862

- ①DBやシミュレーションからのデータあるいは②AI（生成モデル）によるデータ生成に加えて、③ハイスループット実験（実際にモノを作る、作ったモノを評価する）も系に入れる

再構成可能なフローケミストリープラットフォームによる多段階化学合成（MIT）

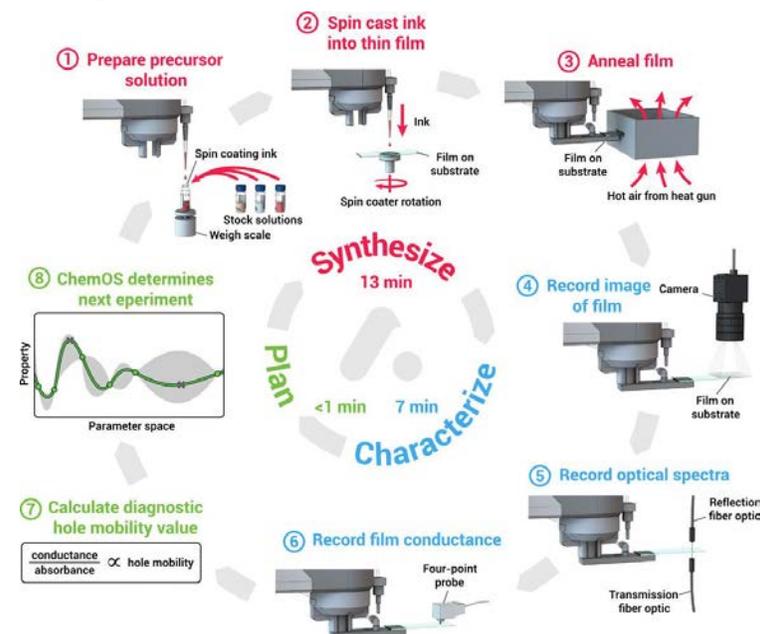
- 最初に、AIに導かれたソフトウェアが分子を合成するためのルートを提案
- 次に専門化学者がこのルートを確認して化学的なレシピに仕上げ
- 最後にハードウェアを自動的に組み立てて反応を実行するロボットプラットフォームにレシピを送信



SCIENCE • 9 Aug 2019 • Vol 365, Issue 6453

自動運転ロボットプラットフォーム（ブリティッシュコロンビア大学等）

- Adaと呼ばれるロボットプラットフォームによって、太陽電池に理想的な欠陥のない薄膜の作成に成功した。
- さまざまな溶液を混合し、フィルムにキャストし、熱処理やその他の処理ステップを実行し、フィルムの導電性をテストし、微細構造を評価し、記録した結果を用いてAIは各実験を解釈し、次に合成するものを決定する。
- 以前は9か月かかっていたものが、今では5日間で行える、としている。



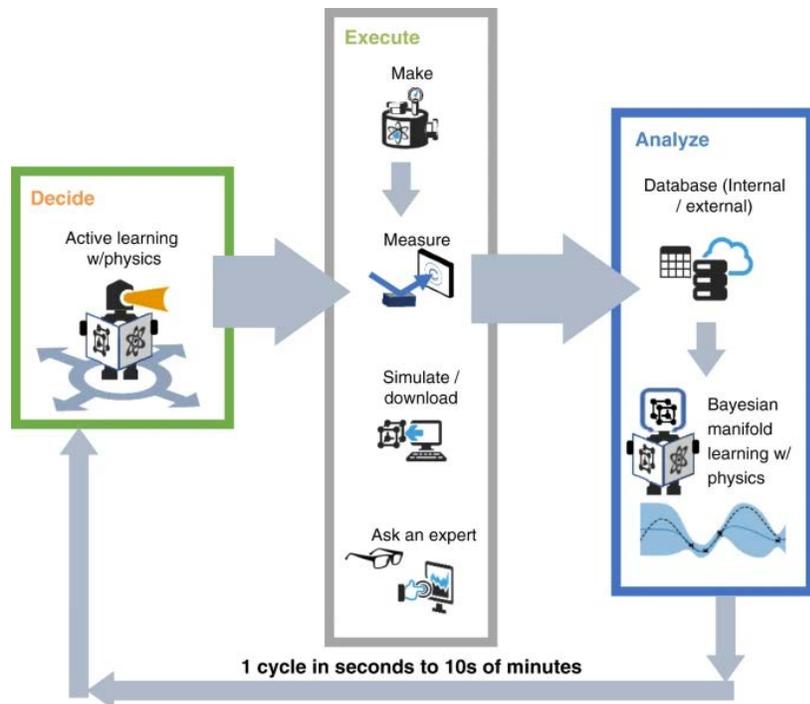
Science Advances 13 May 2020:Vol. 6, no. 20, eaaz8867

自動化

ハイスループット（コンビナトリアル）実験 + AI（クローズドループ方式）

ベイジアン能動学習とOn-the-flyクローズドループによる材料発見（メリーランド大学等）

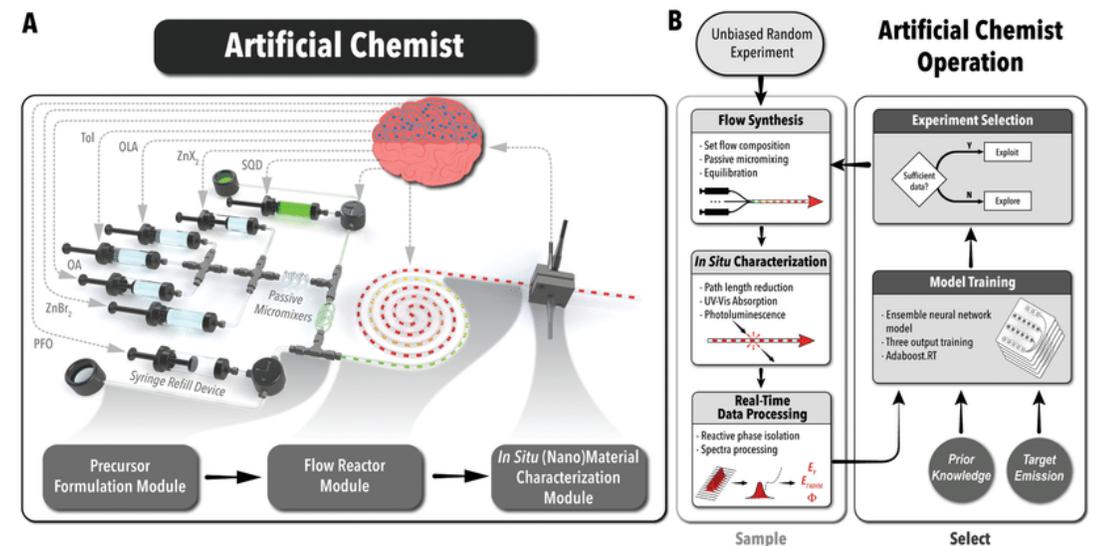
- シンクロトロンビームラインとラボでの組成スプレッドに対するX線回折測定実験のリアルタイム制御
- 自律サイクルは、組成スプレッド上の材料の組成データや計算された材料データを含むデータベースからのデータのロードから始まる。
- 新しいエピタキシャルナノコンポジット相変化メモリ材料が発見された。



Nature Communications volume 11, Article number: 5966 (2020)

人工化学者はオーダーメイドのペロブスカイト量子ドットを合成（ノースカロライナ州立大学等）

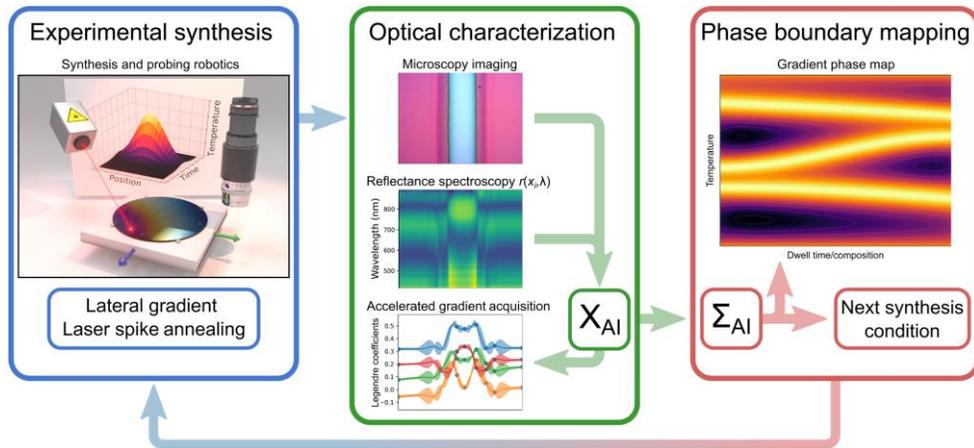
- 人工化学者が量子ドットを識別および生成できることを実証した。液相プロセス材料用に設計され、液体化学前駆体を使用して作成できる材料で機能する。具体的には、量子ドット、金属/金属酸化物ナノ粒子、有機金属フレームワーク（MOF）などの高価値材料が含まれる。
- これにより1日あたり500の量子ドット合成実験を実行できることを実証した。



July 2020 Advanced Materials 32(30):2001626

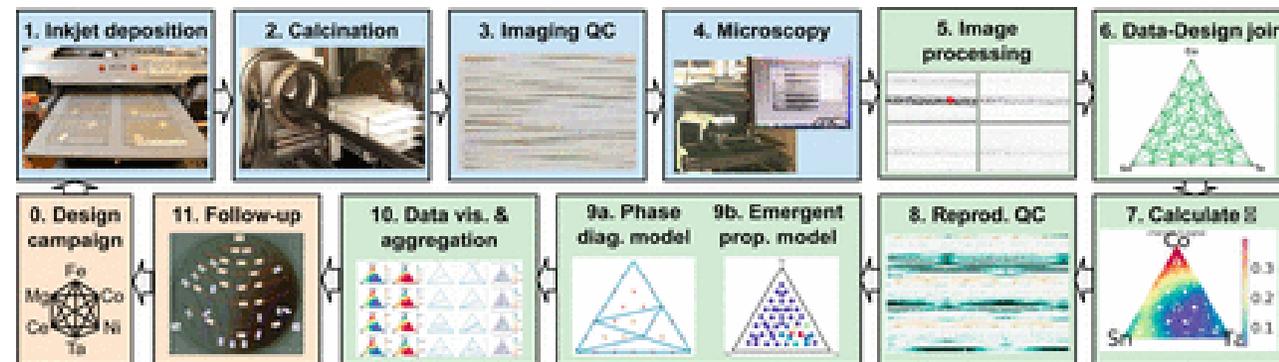
金属酸化物の準安定多形が形成される可能性が高い未踏の超高速アニーリング領域を探索（コーネル大学）

- 豊富な状態図を示すBi-O系を対象に準安定化合物の合成相境界を閉ループ方式で繰り返しマッピングする、完全に統合された自律的なフレームワークを提示
- 横方向勾配レーザースパイクアニーリングと光学的特性評価を使用したロボット材料合成



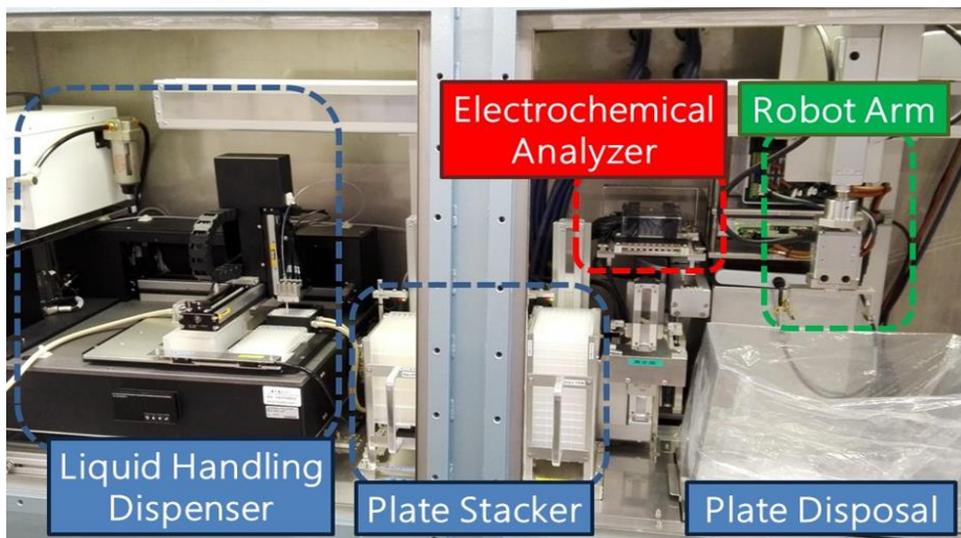
インクプリンターと機械学習を使用して新しい金属酸化物を発見（Google, カリフォルニア工科大学）

- 光学特性のハイスループット測定を使用して、単純な相混合では光学的傾向を説明できない組成を特定することにより、3価の金属酸化物組成空間の新規領域を探索
- Mg、Fe、Co、Ni、Cu、Y、In、Sn、Ce、およびTaについて、108の3カチオン酸化物システムから376,752の異なる組成を透明性、触媒活性、および強酸電解質での安定性からスクリーニング



ハイスループット電解液探索システム（NIMS）

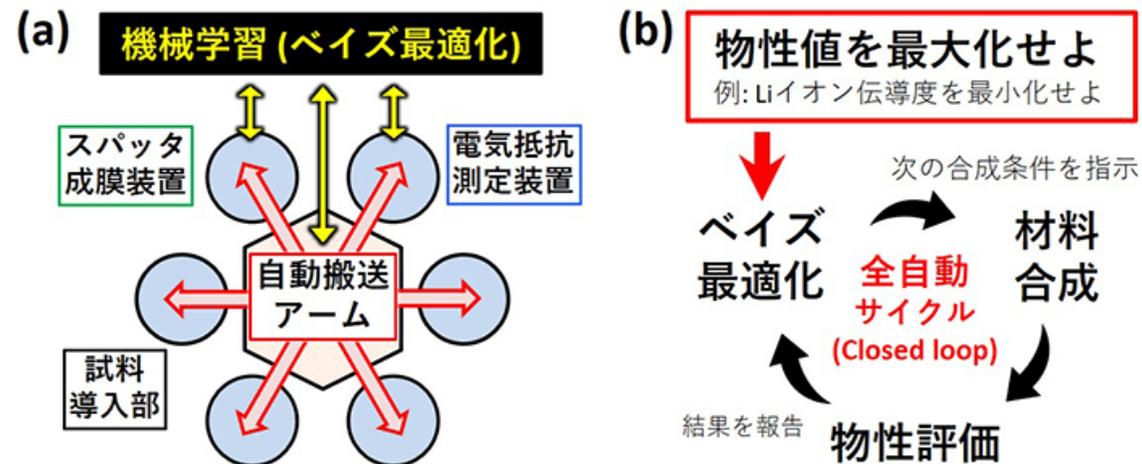
- リチウム空気電池
- 手動では1日10種類の電解液しか探索できなかったが、このシステムを使うことで1日1000種類を探索
- スピーディに大量にデータが集まるようになり、学習用データに活用できる。



Scientific Reports volume 9, Article number: 6211 (2019)

自律的に物質探索を進めるロボットシステムを開発（東工大）

- 機械学習と定常動作を繰り返す機械を組み合わせ、自律的に新規物質を探索するロボットシステムを開発
- 二酸化チタン薄膜の電気抵抗最小化に成功、従来の10倍の実験効率を達成



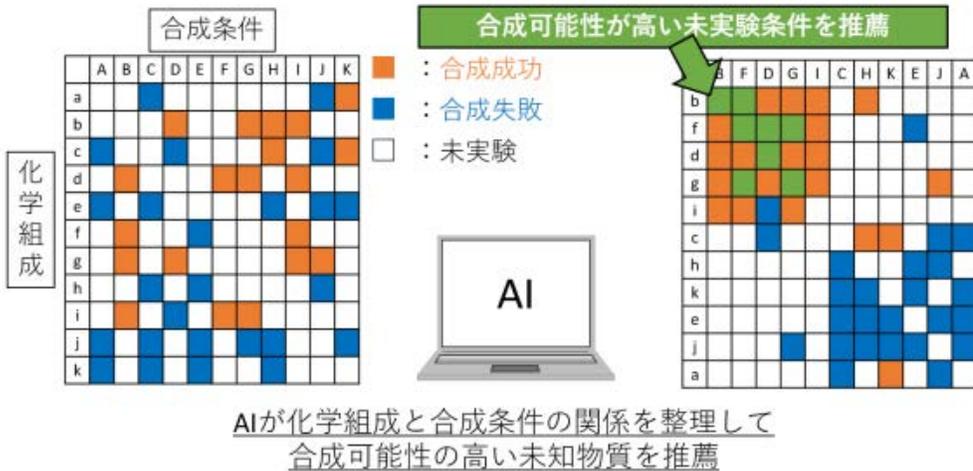
APL Materials 8, 111110 (2020)

自動化（日本）

ハイスループット（コンビナトリアル）実験 + AI

材料合成の推薦システム開発、ロボット協働実験で実際に合成（京大）

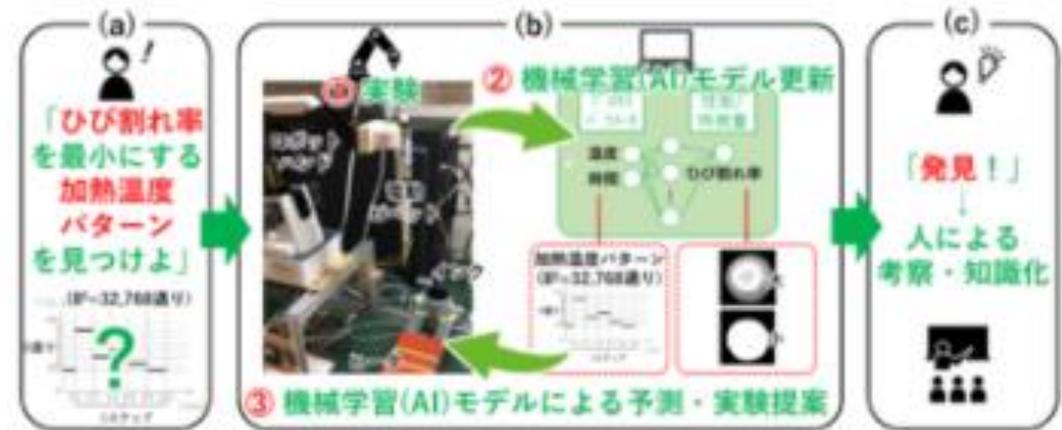
- 開発した合成条件推薦システムと 1,500 件規模の並列合成実験データに基づいて、数万件の未実験条件における合成可能性を評価し、上位 300 件の検証実験により 2 つの新規酸化物 $\text{La}_4\text{V}_2\text{O}_{11}$ 、 $\text{La}_7\text{Sb}_3\text{O}_{18}$ を発見



J Am Ceram Soc. 2022; 105: 853– 861.

ロボット実験×AIによる燃料電池のものづくり研究開発法の革新（東大）

- 粉体成膜プロセスインフォマティクスによりひび割れを最小とする加熱温度パターンについて3万候補から40回で新しい最適解の発見
- (a)人による問題設定。(b)実験に基づく粉体膜乾燥プロセスの最適化のループ、①実験②機械学習モデルの更新③機械学習モデルによる予測・実験提案（以下繰り返す）

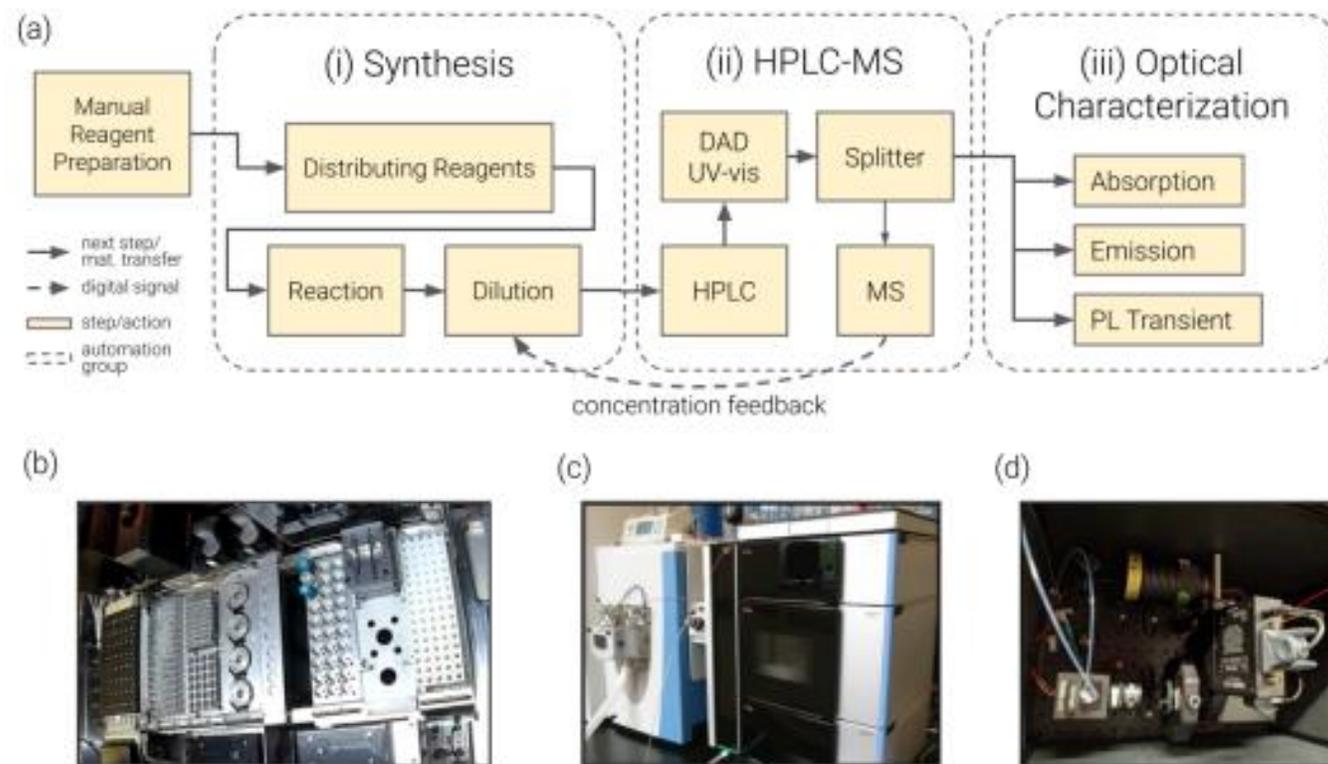


Scientific Reports volume 12, Article number: 1615 (2022)

- これまで挙げてきた事例のいずれも一部の自動化であり、「Human in the loop（人間の介入に大きく依存）」
- 自動化から自律化のボトルネックは？

有機半導体レーザー媒質分子を発見するための自動化されたプラットフォーム (トロント大学、九州大学)

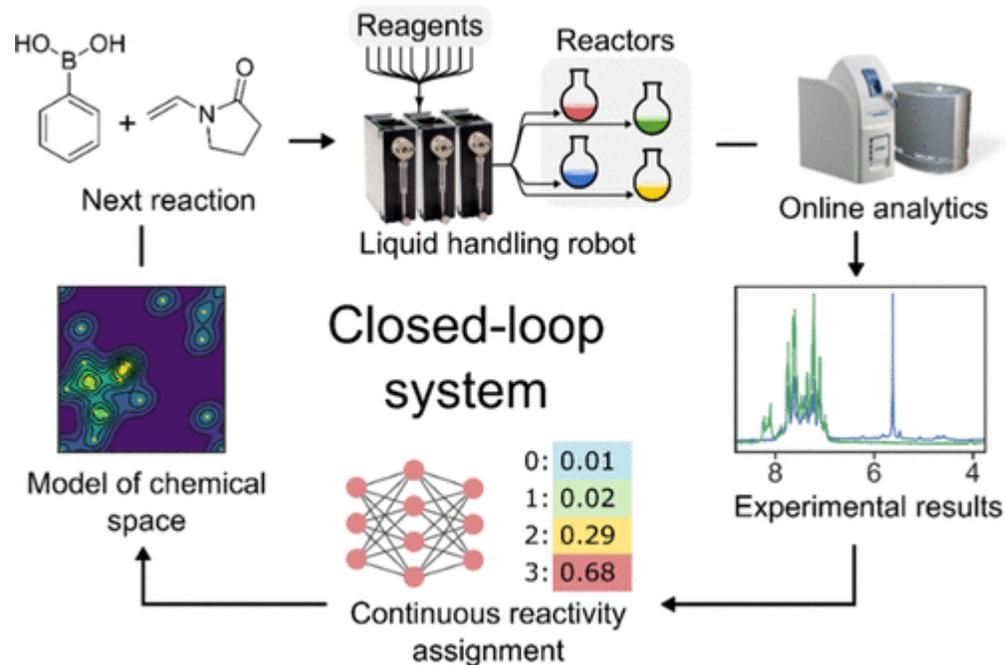
- 化学における自動化の取り組みのほとんどは、合成と化合物の同定に重点を置いており、ターゲット特性の特性評価はあまりされていない。
- 合成、識別、光学的特性評価の統合自動化された方法で実行
- このワークフローを使用して有機レーザーの候補をスクリーニングし、8つの潜在的な候補を発見
- 薄膜デバイスで4分子のレーザー発振しきい値をテストし、最先端のパフォーマンスを備えた2分子を発見



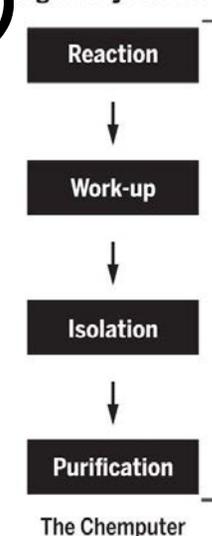
自律化の萌芽（化学研究環境のDX）

Chemputation（グラスゴー大学）

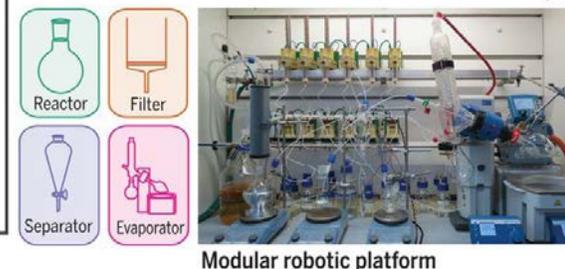
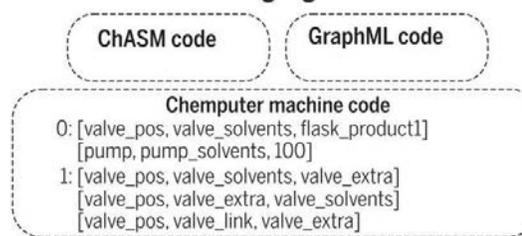
- モジュール化とプログラミング
- シミュレーション、AI（識別モデル、生成モデル、自然言語処理）など総動員
- 分子構造と反応性間の学習された一般的な関連に基づいて化学空間をナビゲートできるロボット化学発見システム



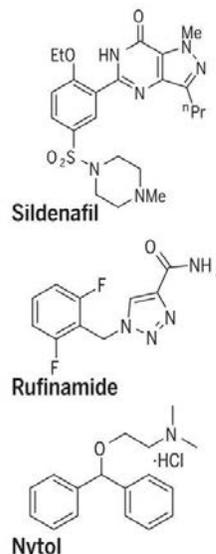
Abstraction of organic synthesis



Chemical programming language



Digital synthesis



Science. 2019 Jan 11;363(6423):eaav2211.

Modules are Plug and Play Fitting into the System

- Photochem
- Redox
- Bioreactor
- Solid
- Optimise
- HTS
- Microfluidic
- Chiral
- Catalyst discovery
- Cell interface
- Formulation
- NanoMaterial
- Polymer/Gel

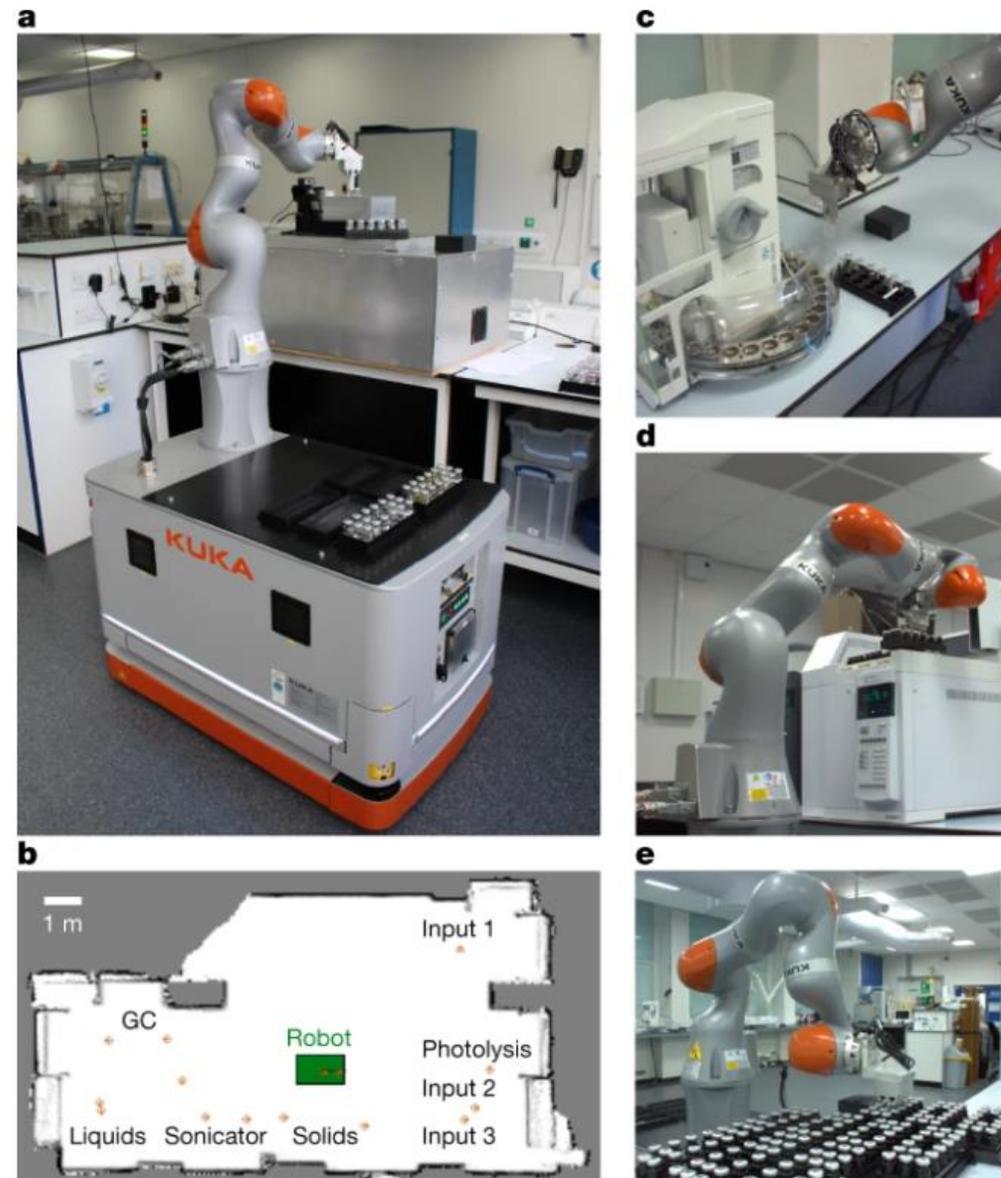


自律化の萌芽（自律移動ロボット）

自律運転ロボット化学者（リバプール大学）

- 「ロボット研究者」は研究室を自由に動き回り様々な設備を使用する。自律型ロボットは膨大な量の組み合わせを覚えることができる。ロボットの脳はベイジアンネットワークの計算アルゴリズムを用いており、入手したデータを処理し、推定したい事柄の確率を計算する。これにより膨大な実験数（9800以上のオプション）に適応し、過去の実験結果に基づき、どの実験を次に行うべきかを判断する。
- 水から水素を取り出すための新たな光触媒を発見するために使用された。ロボットは8日間で172時間働き、その間、688の実験を行い、自ら触媒を発見した。今回発見された触媒は既知の成分に比べ6倍の活性を持つ。

Nature volume 583, pages237–241 (2020)



自律化の萌芽（ラボオートメーション／クラウドラボ）

Chemspeed Technologies(1997)

ラボ用自動化装置「CHEMSPEED」は、60種以上の様々なツールを組み合わせることで自動で使い分けることにより、研究開発（特に化学合成）における一連の工程をオートメーション化する自動化システム

HighResBiosolutions(2004)

ラボラトリーオートメーションシステムおよびデバイスは、テクノロジーの変化に応じてオートメーションシステムをスケールアップおよび再構成する機能を提供

Strateos(2012)

ライフサイエンスの発見のためのリモートアクセスラボおよびラボ制御ソフトウェアの開発
データ、計算、自動化、およびハイスループットロボット工学によって推進される新しい知識を生み出す

Emerald Cloud Lab (2010)

メソッド設計、試料ロジスティクス、サンプル準備、機器操作、データ取得と分析、トラブルシューティング、廃棄物処理など、日常のラボ作業を処理する遠隔操作研究施設



ファブレス科学ラボ＋クラウドラボ＋ファウンドリベンチャー、
のように科学者は頭だけ使えば良い時代が来る？

【参考】微生物・細胞研究の自動化

Ginkgo Bioworks x Strateos

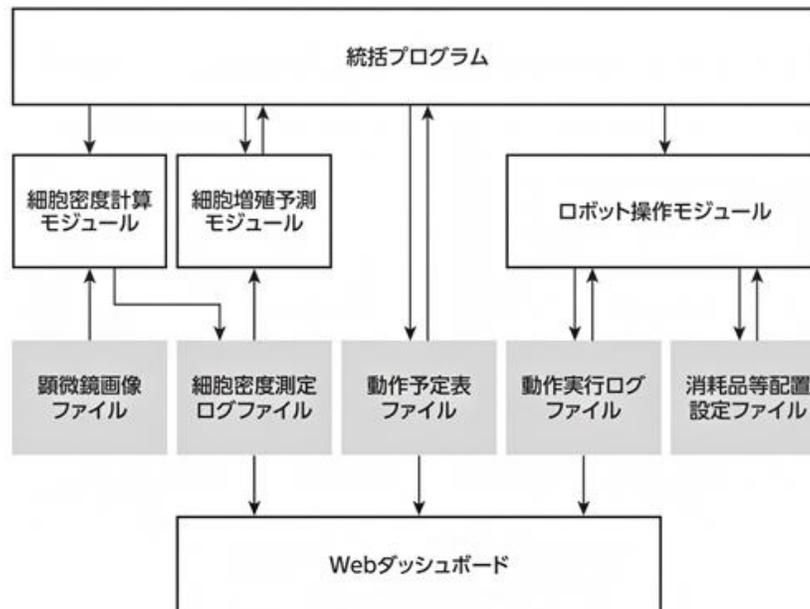
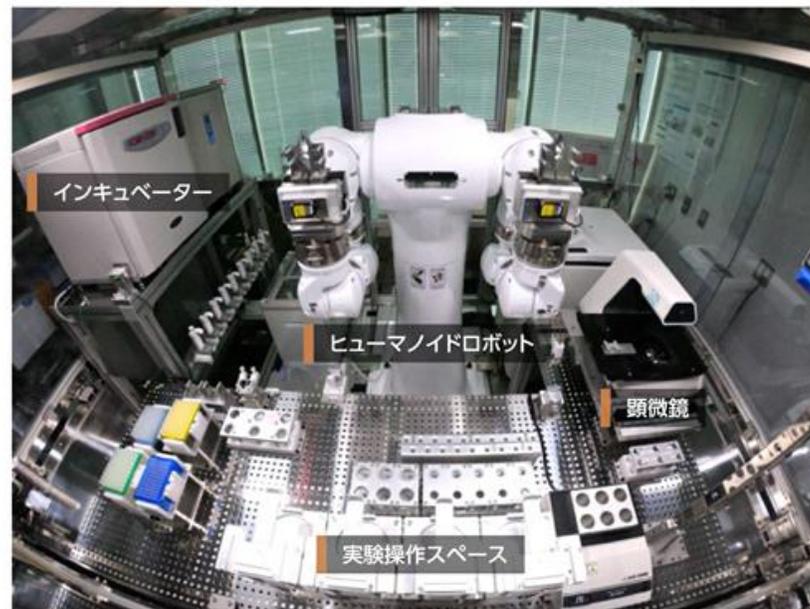
- Ginkgo社はボストンにあるファクトリー4か所で、自動化によるDBTLサイクルの高速化・高効率化を推進
- 2017年より両社はファクトリー自動化ソフト開発で協業
- DARPAのファンディングを受け、機械学習を活用した合成生物学のクラウド実験デザインのためのプラットフォームを共同開発
- Strateos社は、Eli Lilly社が設置したリモートアクセス可能な全自動クラウドラボのプラットフォームも開発

AstraZeneca x HighRes Biosolutions

- HighRes社が開発したロボットCoLabにより、フレキシブルかつ研究者との共存が可能な自動実験環境を、AZ社のケンブリッジ研究所に整備
- 1日に累計30万化合物のスクリーニングが可能
- HighRes社は、大手製薬企業その他、NIH、Broad Instituteにもシステムを提供

ロボットとAIによる自律細胞培養（理研）

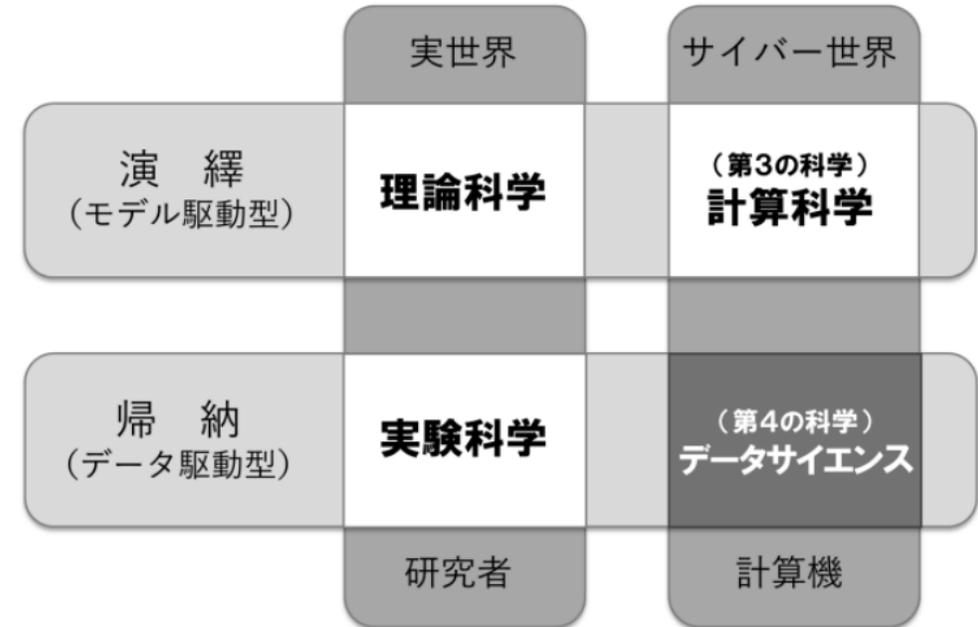
これまで人間が行ってきた基礎研究における細胞培養の動作・判断を、ロボットとAIに置き換えるシステムを開発



Autonomous科学

実験、理論、計算、データ科学の統合

- 探索空間をどのように広げるか（DB、シミュレーション、実験）によってプロセスは異なる。
- 特に実験プロセス（実空間での合成や評価）の導入がボトルネック
- 候補物のいくつかを実際に作成すること、探索空間を拡張するためにモノを作成してデータをとることの2点がありうる。
 1. 人による計画の指示（何をどうやって）
 2. バーチャル／リアルによる仮説の生成（候補の提示）
 3. 意思決定
 4. 合成プロセス（候補物の生成）
 5. 意思決定（次の工程の判断）
 6. 仮説の再生成
 7. 意思決定
 8. 以降これのサイクル



まとめと今後の展望

- 化学・材料のみならず、創薬や生物生産、ものづくり研究などにおいて研究開発の方法論が変遷してきている。
- 実験（ハイスループット、コンビナトリアル）科学、計算科学、データ科学の統合が起こっている（全般にDXが進んでいる）。

- 新しいAIを用いた新しいMI方法論は欧米発。手法の応用・適用は日本も追随。
- 現在はAI+自動化フェーズ、今後10年～かけて自律化（研究環境のDX）の実現へ
- 自動化と自律化ではハードルの高さや取り組み方が異なる。後者は世界で先行事例が数例、日本はほぼ手つかず
- あらゆるAIモデルが科学研究に応用され、今後も新しいAI技術が出てくる可能性は高く準備が必要
- 2010年前後にX-Infoの概念が提唱され約7年後に一定の研究の塊が立ち上がり。2018年頃にAutonomousの概念が提唱された。2025年頃に立ち上がる！？

- 材料探索のための物質・化合物空間は膨大
- 自動化・自律化への道は、利用可能な材料特性データの不足を解消するものでもある。
- 研究の上流工程（デザイン：探索空間の設定や探索方法）がより高いウエイトを占めることになるのではないか。
- 発見のプロセスのみならず、材料・デバイス開発（実用まで）の一連の階層プロセスのDX⇒プロセスインフォマティクスも必要。
- いずれにしても、自動化・自律化に向けては、化学・材料と情報・データ、工学の強力なパートナーシップと資金が必要であり、企業の方が存在感を発揮できる場面も多く出てくるのではないか。

ご静聴ありがとうございました。

本日の発表資料はCRDSのHP（下記URL）に掲載予定です。

<https://www.jst.go.jp/crds/sympo/20220325/index.html>

■ 俯瞰報告書はこちらから



<https://www.jst.go.jp/crds/report/by-report/02/index.html>

国内外の研究開発動向やイノベーションについて、ご関心のある方はお気軽にCRDSにコンタクトください。